FEATURAL RELATIONS IN THE BRAIN: THEORETICAL AND EXPERIMENTAL PERSPECTIVES ON GRAMMATICAL AGREEMENT,

2nd Edition

EDITED BY: Simona Mancini, Sendy Caffarra and Andrew Nevins

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88971-844-3
DOI 10.3389/978-2-88971-844-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FEATURAL RELATIONS IN THE BRAIN: THEORETICAL AND EXPERIMENTAL PERSPECTIVES ON GRAMMATICAL AGREEMENT,
2nd Edition

Topic Editors:
Simona Mancini, Basque Center on Cognition, Brain and Language, Spain
Sendy Caffarra, Stanford University, United States
Andrew Nevins, University College London, United Kingdom

Successful speaking and understanding hinges on the almost effortless capacity of speakers to decode and build dependencies among words in a sentence, based on covariance in some specific feature(s). Whenever two features covary, an agreement relation is established. Agreement is a widespread and varied phenomenon: its pervasiveness in some languages contrasts with its near absence in others, which poses a challenge for linguists and psycholinguists that attempt to explain the mechanics of its representation, processing and acquisition.

Agreement has been extensively investigated from a theoretical perspective, but also from the point of view of psycholinguistics and the cognitive neuroscience of language. Theoretical linguistics has provided an articulated system of structural representations and computations on which the establishment of agreement relations hinges, while psycholinguistics and cognitive neuroscience have aimed at unveiling the algorithms that underlie the use of these computations and their behavioral and neuro-physiological bases. The goal of this Research Topic is to draw together multiple and interdisciplinary work to highlight the state of the art in the study of agreement and propose new perspectives on this research topic.

Publisher's note: In this 2nd edition, the following article was added: Mancini S, Caffarra S and Nevins A (2021) Editorial: Featural Relations in the Brain: Theoretical and Experimental Perspectives on Grammatical Agreement. Front. Psychol. 12:754430. doi: 10.3389/fpsyg.2021.754430

Citation: Mancini, S., Caffarra, S., Nevins, A., eds. (2021). Featural Relations in the Brain: Theoretical and Experimental Perspectives on Grammatical Agreement, 2nd Edition. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-844-3

Table of Contents

frontiers in Psychology

EDITORIAL
published: 08 October 2021
doi: 10.3389/fpsyg.2021.754430

Editorial: Featural Relations in the Brain: Theoretical and Experimental Perspectives on Grammatical Agreement

Simona Mancini[1]*, *Sendy Caffarra*[1,2,3,4] and *Andrew Nevins*[5]

[1] Basque Center on Cognition, Brain and Language, San Sebastian, Spain, [2] Division of Developmental-Behavioral Pediatrics, Stanford University School of Medicine, Stanford, CA, United States, [3] Stanford University Graduate School of Education, Stanford, CA, United States, [4] Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy, [5] Division of Psychology and Language Sciences, University College London, London, United Kingdom

Keywords: syntax, agreement, sentence comprehension, sentence production, morphosyntactic features

Editorial on the Research Topic

Featural Relations in the Brain: Theoretical and Experimental Perspectives on Grammatical Agreement

Theoretical linguistics has provided an articulated system of structural representations and computations on which the establishment of agreement relations hinges, while psycholinguistics and neurolinguistics aim at unveiling the algorithms that underlie the use of these computations and their behavioral and neurophysiological bases. The goal of this special issue is to describe the state of the art in the theoretical and experimental study of agreement. Its 15 articles open a unique and privileged window onto a wide range of languages (from English and German to Romance languages like Italian, French, and Spanish, but also to less well-studied languages within psycholinguistics and neurolinguistics such as Georgian, Korean, Standard Modern Arabic, and South Slavic languages), through the lens of distinct features (person, gender, number, and tense), drawing evidence from a variety of experimental paradigms (e.g., offline elicitation tasks, self-paced reading, eye tracking, and event-related potentials) and diverse theoretically-grounded approaches.

Three main take-home messages emerge from the articles collected in this special issues. First, agreement does not a constitute a monolithic phenomenon: person, number, gender, and tense features have inherent structural and interpretive differences, which produce common but also feature-specific reflexes in comprehension and production. Second, the mechanisms that guide the parser in retrieving and encoding features during the building of an agreement relation obey distinct principles, depending on whether features are structurally accessible or not. Finally, agreement is not wholly circumscribed within syntax: its comprehension and production trigger the integration of information from distinct linguistic and non-linguistic domains. Let us see these points in more detail.

DISTINCT FEATURES, DISTINCT MECHANISMS? AGAINST FEATURES AS UNIFORM CONSTRUCTS

A landmark of numerous theoretical analyses within the generative framework is the idea that agreement features cannot be treated as a "bundle" under the same T head (Chomsky, 2014). The intrinsically different syntactic and interpretive properties that characterize e.g., person, number, tense, and gender agreement make it plausible to hypothesize the independent representation of

Edited and reviewed by:
Mila Vulchanova, Norwegian University of Science and Technology, Norway

***Correspondence:**
Simona Mancini
s.mancini@bcbl.eu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *06 August 2021*
Accepted: *06 September 2021*
Published: *08 October 2021*

Citation:
Mancini S, Caffarra S and Nevins A (2021) Editorial: Featural Relations in the Brain: Theoretical and Experimental Perspectives on Grammatical Agreement. Front. Psychol. 12:754430. doi: 10.3389/fpsyg.2021.754430

such features (Shlonsky, 2010; Sigurdsson, 2010; Rizzi and Cinque, 2016, among others). Hartmann and Heycock contribute to this research line by showing how person and number features can be structurally differentiated in several Germanic languages, such as Dutch, Faroese, German, and Icelandic.

A prolific strand of experimental research has been also devoted to investigating whether features' distinct representational properties have a processing reflex, of which the person-number dissociation hypothesis has been one of the main testing grounds (Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Mancini et al., 2011, 2017; Zawiszewski et al., 2016; Biondo et al., 2018 to name a few). Existing findings attributed qualitative differences in their processing to the different interpretive properties that characterize the two types of agreement: the link to discourse participant roles that is necessary to interpret 1st, 2nd, or 3rd person (a speaker, an addressee or a non-active participant), but not the singularity or the plurality of the individuals involved in the speech event. Interestingly, an alternative explanation is proposed here by den Dikken that centers on the distinct checking mechanisms in which the two features engage: both spec-head and Agree for number, while only spec-head for person.

While experimental studies overall agree that the response to agreement violations is stronger when person, rather than number is involved, they diverge on whether qualitative (Mancini et al., 2011, 2017; Biondo et al., 2018) or quantitative (Zawiszewski et al., 2016) differences emerge between these two features. Ackema and Neeleman's analysis proposes that the type of agreement controller and the distinct feature sets that pronouns and regular noun phrases (NPs) carry can reconcile these apparently contradicting results.

The literature on *gender* comprehension and production also corroborates the hypothesis that agreement features cannot be treated as uniform constructs, and that their granular properties do matter for comprehension and production. In particular, Wang and Schiller show that the strength of morphophonological representations determines whether speakers access gender information through a form-related route (as happens in Romance languages) or through a lexically-based route (as for example in German and Dutch). Moreover, speakers of different linguistic profiles (i.e., monolinguals and bilinguals) are sensitive to distributional differences between masculine and feminine classes, as Beatty-Martínez and Dussias's contribution reveals.

Other fine-grained aspects of agreement controllers and targets can play a crucial role in the establishment and comprehension of relations among words. Data from Spanish (Bañón and Rothman) and Georgian (Foley and Wagers) show that factors such as the morphological markedness of the subject and the canonicity of the verb form shape the parser's expectations, and thus its sensitivity to detecting errors and initiating reanalysis processes when anomalies are encountered (see also Tucker et al.'s contribution based on data from Modern Standard Arabic for a similar finding on how the morphological markedness of the verb impacts error detection on number and gender verbal agreement).

Thus, there are multiple fine-grained distinctions that can be made when we investigate agreement mechanisms. These specific differentiations can be based on the type of grammatical features involved, their distributional properties and their morphological markedness.

WHAT GUIDES THE ANALYSIS AND INTERPRETATION OF FEATURES DURING AGREEMENT PROCESSING?

An extremely productive line of theoretical and experimental research on agreement has focused on attraction, the phenomenon whereby the production of the correct number inflection on the verb can be disrupted by the presence of an intervening plural noun phrase, as in *The key to the cabinets were rusty* (Bock and Miller, 1991). In comprehension, attraction leads to illusions of grammaticality, i.e., to the acceptance of agreement anomalous sentences (Pearlmutter et al., 1999). Several psycholinguistic accounts exist that have attempted to explain the mechanisms behind agreement attraction in comprehension and production, among which the so-called retrieval accounts. Under this theoretical framework, attraction is an error of the memory system, whereby the cues of a certain head should be retrieved but the parser can select the wrong NP if there is a partial overlap in features. Using a variety of experimental paradigms, several papers in our special issue test retrieval accounts, reporting interesting findings across typologically different languages.

Parker and An suggest that in English, attraction depends on both retrieval and encoding mechanisms and that it is sensitive to both the semantic and syntactic properties of the attractor. Interestingly, Schlueter et al. show that the attractor can be erroneously interpreted as the thematic subject and that this is orthogonal to whether attraction happens. In their study in Korean, Kwon and Sturt show that misretrieval is more likely to occur if the distractor is nominative, rather than e.g., dative-marked, suggesting that at least in languages that overtly mark case, the grammatical role of a nominal element plays a more crucial role than e.g., mere proximity to the verb.

Are all linguistic features candidate cues that guide retrieval? Are all cues given similar weight? Biondo et al., address this question in an eye tracking study in English where they test readers' sensitivity to temporal concordance between an adverb and two verbs, a structurally accessible and a structurally inaccessible verb. They show that readers were sensitive to feature match between the adverb and a linearly distant but structurally accessible verb, while the evidence about the interference of a structurally inaccessible verb is not clear.

Tucker et al. show that inherent differences between features play a role also during the processing of attraction phenomena. Indeed, in Modern Standard Arabic subject-verb agreement, gender effects are larger and surface slightly later than number attraction effects, which calls for a revision of real-time models of agreement that posit the bundling of the two features in the computation of subject-verb agreement.

The emerging picture from all these results point to a diversity of agreement mechanisms, highlighting the differential impact of the grammatical role of nominal constituents and the structural

accessibility of grammatical features that are retrieved and encoded in the real time computation of agreement relations.

AGREEMENT BEYOND SYNTAX

Another important question that experimental and theoretical research on agreement aim to answer is whether its mechanisms and representations are circumscribed within syntax. Based on data from Serbo-Croatian, Mitić and Arsenijević suggest that the computation of agreement relations spans beyond purely syntactic boundaries and involves the interface between syntax and phonological form, in line with accounts that place some of the computation of agreement in the post-syntactic component (Bobaljik, 2008; Arregi and Nevins, 2012).

Further evidence for the impact of extra-syntactic factors in agreement processing comes from the analysis of online and offline patterns elicited by object cleft sentences. In their study on Italian cleft sentences, Chesi and Canal manipulate whether the subject and object NPs are 3rd person or 2nd person definite NPs in various combinations, in the attempt to elucidate the role played by the properties of different NPs and different persons. By collecting acceptability judgments and accuracy data from comprehension questions, as well as online reading times from eye tracking, Chesi and Canal show that sentence processing difficulty is not wholly driven by computing the syntactic analysis. Rather, there are aspects of the interpretation and discourse factors that play a major role.

Finally, Courteau et al. attempt to cast light onto how information across visual and linguistic domains impacts processing, thus shifting the focus on how agreement between the information contained in different domains is integrated. The results of their ERP study show that participants immediately detected number mismatches between pictures and acoustically-presented, grammatically correct linguistic material. These mismatches are processed in a way that is not fundamentally different from purely linguistic, within-sentence agreement violations, thus underlining the role of contextual information in the processing of agreement dependencies.

This set of papers thus highlight potential interactions between agreement mechanisms and non-linguistic domains, such as phonology, picture-based context, and discourse.

CONCLUSION

Agreement is a widespread and varied phenomenon: its pervasiveness in some languages contrasts with its near absence in others, which poses a challenge for linguists and psycholinguists that attempt to explain the mechanics of its representation and processing (Corbett, 2006). These inherent complexities notwithstanding, the 15 articles presented in this special issue clearly represent a step forward in the description of the architecture and mechanisms underlying this core linguistic function in terms of its representation and processing. The emerging picture from this collection of papers is that the mechanisms of grammatical agreement may be flexible, feature-specific, and in part non-strictly syntactic. We hope that the breadth of empirical contributions, novel methodological designs, and theoretical refinements presented herein pave the way for continued avenues of exploration of this pervasive aspect of natural language.

AUTHOR CONTRIBUTIONS

SM, SC, and AN equally contributed the development and running of the Research Topic, as well as to the writing of the editorial manuscript.

FUNDING

This work was supported by the Basque Government through (BERC 2018-2021 program) and by the Spanish State Research Agency (BCBL Severo Ochoa excellence accreditation SEV-2015-0490). SM acknowledges funding from the Spanish Ministry of Economy, Industry and Competitiveness (RYC-2017-22015), the Spanish Ministry of Science and Innovation (PID2020-113945RB-I00), and the Basque Government (PIBA_2020_1_0024). SC acknowledges funding from the Rita Levi Montalcini grant (Italian Ministry of Research).

ACKNOWLEDGMENTS

The authors wish to thank Diogo Almeida for insightful comments on an earlier version of this manuscript.

REFERENCES

Arregi, K., and Nevins, A. (2012). *Morphotactics: Basque Auxiliaries and the Structure of Spellout,* Vol. 86. New YorK, NY; London: Springer Science & Business Media.

Biondo, N., Vespignani, F., Rizzi, L., and Mancini, S. (2018). Widening agreement processing: a matter of time, features and distance. *Lang. Cogn. Neurosci.* 33, 890–911. doi: 10.1080/23273798.2018.1446542

Bobaljik, J. D. (2008). *Paradigms (optimal and otherwise): a case for Scepticism*. Available online at: http://www.unice.fr/scheer/egg/BLuka18/Bobalijk(2008).pdf

Bock, K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Chomsky, N. (2014). *The Minimalist Program*. Cambridge, MA: MIT Press.

Corbett, G. (2006). *Agreement*. Oxford: Oxford University Press.

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject–verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Mancini, S., Quiñones, I., Molinaro, N., Hernandez-Cabrera, J. A., and Carreiras, M. (2017). Disentangling meaning in the brain: left temporal involvement in agreement processing. *Cortex* 86, 140–155. doi: 10.1016/j.cortex.2016.11.008

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Memory Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Rizzi, L., and Cinque, G. (2016). Functional categories and syntactic theory. *Annu. Rev. Linguist.* 2, 139–163. doi: 10.1146/annurev-linguistics-011415-040827

Shlonsky, U. (2010). The cartographic enterprise in syntax. *Lang. Linguist. Compass* 4, 417–429. doi: 10.1111/j.1749-818X.2010.00202.x

Sigurdsson, H. Á. (2010). Language quarks. *Iberia Int. J. Theor. Linguist.* 1, 169–183. Available online at: http://www.siff.us.es/iberia/index.php/ij/article/view/6/2

Silva-Pereyra, J. F., and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029

Zawiszewski, A., Santesteban, M., and Laka, I. (2016). Phi-features reloaded: an event-related potential study on person and number agreement processing. *Appl. Psycholinguist.* 37, 601–626. doi: 10.1017/S014271641500017X

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 28 August 2018
doi: 10.3389/fpsyg.2018.01566

Not All Phrases Are Equally Attractive: Experimental Evidence for Selective Agreement Attraction Effects

Dan Parker and Adam An*

Linguistics Program, Department of English, College of William & Mary, Williamsburg, VA, United States

Research on memory retrieval during sentence comprehension suggests that similarity-based interference is mediated by the grammatical function of the distractor. For instance, Van Dyke and McElree (2011) observed interference during retrieval for subject-verb thematic binding when the distractor occurred as an oblique argument inside a prepositional phrase (PP), but not when it occurred as a core argument in direct object position. This contrast motivated the proposal that constituent encodings vary in the distinctiveness of their memory representations based on an argument hierarchy, which makes them differentially susceptible to interference. However, this hypothesis has not been explicitly tested. The present study uses an interference paradigm involving agreement attraction (e.g., Wagers et al., 2009) to test whether the argument status of the distractor determines susceptibility to interference. Results from two self-paced reading experiments show a clear contrast: agreement attraction is observed for oblique arguments (e.g., PP distractors), but attraction is nullified for core arguments (i.e., direct object and subject distractors). A follow-up experiment showed that this contrast cannot be reduced to the syntactic position of the distractor, favoring an account based on the semantic properties of the distractor. These findings support the proposal that interference is mediated by the argument status of the distractor and extend previous results by showing that the effect generalizes to a broader set of syntactic contexts and a wider range of syntactic dependencies. More generally, these results motivate a more nuanced account of real-time agreement processing that depends on both retrieval and encoding mechanisms.

Keywords: sentence comprehension, encoding, retrieval, agreement attraction, self-paced reading

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Sandra Villata,
University of Connecticut,
United States
Heidi Lorimor,
Bucknell University, United States

****Correspondence:***
Dan Parker
dparker@wm.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *04 May 2018*
Accepted: *06 August 2018*
Published: *28 August 2018*

Citation:
Parker D and An A (2018) Not All Phrases Are Equally Attractive: Experimental Evidence for Selective Agreement Attraction Effects. Front. Psychol. 9:1566. doi: 10.3389/fpsyg.2018.01566

INTRODUCTION

Sentence comprehension routinely relies on memory retrieval mechanisms to establish grammatical dependencies among the words and phrases in a sentence. For instance, to relate the verb *were* in (1) to its subject to establish subject-verb number agreement, memory retrieval mechanisms must access the encoding of the plural subject *girls* and ignore featurally similar information in non-target positions, such as the embedded plural noun *boys*.

(1) **The girls$_{PL}$** [that **the boys$_{PL}$** teased on the playground] **were** late for school.

Sometimes, featurally similar information in non-target positions intrudes on retrieval of the target, modulating acceptability and reading times. Such effects are commonly referred to as "similarity-based interference" (Gordon et al., 2001; Lewis and Vasishth, 2005; Lewis et al., 2006; Van Dyke and McElree, 2006; Van Dyke and Johns, 2012). The current study investigates the conditions under which such effects arise during retrieval for agreement processing.

Previous research on memory retrieval for dependency formation during real-time sentence comprehension has revealed a mixed profile of successes and failures with respect to interference effects. Some dependencies, like those involving subject-verb agreement, negative polarity item licensing, case licensing, and ellipsis, are highly susceptible to interference (Clifton et al., 1999; Pearlmutter et al., 1999; Vasishth et al., 2008; Wagers et al., 2009; Xiang et al., 2009, 2013; Martin et al., 2012; Dillon et al., 2013; Sloggett, 2013; Tanner et al., 2014; Lago et al., 2015; Tucker et al., 2015; Parker and Phillips, 2016). But other dependencies, like those involving reflexives, control, strong crossover binding, and bound variable pronouns, are more resistant to interference (Clifton et al., 1999; Dillon et al., 2013; Kush and Phillips, 2014; Kush et al., 2015, 2017), or require specific configurations for interference to obtain (Parker et al., 2015; Parker and Phillips, 2017).

The question of why different dependencies show different profiles with respect to interference remains unresolved (see Parker and Phillips, 2017, for discussion). However, many existing accounts agree that for the dependencies that do show interference, such as subject-verb agreement, interference reflects misretrieval of a feature-appropriate items from a structurally irrelevant position (Wagers et al., 2009; Dillon et al., 2013; Tanner et al., 2014; Lago et al., 2015; Tucker et al., 2015; Parker and Phillips, 2017; Tucker and Almeida, 2017). A key prediction of this retrieval-based account is that interference should generalize across a broad range of structural configurations, since the same error-prone retrieval mechanism should apply whenever a comprehender attempts agreement licensing (McElree, 2000; McElree et al., 2003; Lewis and Vasishth, 2005; Lewis et al., 2006).

However, recent research on retrieval for subject-verb thematic binding suggests that interference effects can also be modulated by the encoding mechanisms. For instance, Van Dyke and McElree tested sentences like those in (2). In both sentences, the critical verbs (*moaned* and *compromised*) require an animate subject, motivating the use of animacy as a retrieval cue for these dependencies (Van Dyke, 2007; Van Dyke and McElree, 2011). Despite similar retrieval requirements in (2a-b), Van Dyke and colleagues observed contrasting profiles: interference effects arose when a structurally-irrelevant animate distractor (in bold) occurred inside a prepositional phrase (PP), as in (2a), but not when it occurred as a direct object, as in (2b).

(2) (a) The pilot remembered that the lady who was sitting near ***the smelly man*** moaned about a friend.
(b) The attorney who the judge realized had rejected ***the witness*** in the case compromised.

This contrast is surprising because it is not predicted by existing retrieval accounts (McElree, 2000, 2006; McElree et al., 2003; Lewis and Vasishth, 2005). Existing accounts predict similar interference profiles for (2a-b), since the same interference-prone mechanism is assumed to apply whenever the comprehender attempts retrieval for thematic binding.

Van Dyke and McElree (2011) argued that the source of the contrast in (2) is the syntactic encoding. Specifically, they suggested that PPs and direct objects differ in the distinctiveness of their memory representations based on an argument hierarchy, making them differentially susceptible to interference. Many grammatical theories make a hierarchical distinction between core thematic arguments (e.g., subjects, direct objects), which play a prominent role in establishing the meaning of the sentence, and modifying oblique arguments, including PPs, which possess little discriminating syntactic information (e.g., PPs lack a theta role) and play a less prominent role in building meaning (Keenan and Comrie, 1977; Chomsky, 1981; Frazier and Clifton, 1996; Van Valin and LaPolla, 1997; Bresnan, 2001; Culicover and Jackendoff, 2005). Drawing on this distinction, Van Dyke and McElree (2011) hypothesized that the prominent grammatical function of core arguments makes the syntactic aspects of their memory encoding more distinctive, relative to oblique arguments, and hence easier to reject or accept based on their match to the syntactic retrieval cues. On this view, the distinctiveness of the syntactic features of the direct object in (2b) produces a salient mismatch with the subject retrieval cues of the verb, making them relatively easier to rule out. Conversely, less distinctive representations, like the oblique PP in (2a), are not salient enough to produce a strong mismatch with the syntactic retrieval cues, and hence are more likely to interfere, yielding the contrast observed in (2). Crucially, unlike previous accounts of interference that place the blame on the retrieval mechanisms, Van Dyke and McElree (2011) suggested that in the case of thematic binding, it is the encoding mechanisms that mediate interference.

Additional evidence of interference based on the thematic-semantic properties of the distractor encoding comes from Cunnings and Sturt (2018). Cunnings and Sturt manipulated sentence plausibility as a diagnostic of interference in sentences like (3). In (3), the critical verb *shattered* triggers a retrieval to recover its direct object. They manipulated whether the retrieval target, the direct object of the matrix verb, e.g., *the plate/letter*, was a plausible direct object of the critical verb, as well as the plausibility of a distractor embedded inside an intervening PP, e.g., *the cup/tie*.

(3) Sue remembered the plate/letter that the butler with **the cup/tie** accidently shattered today in the dining room.

Cunnings and Sturt observed a significant main effect of plausibility, such that implausible sentences were read more slowly than plausible sentences at the critical verb and spillover regions. They also found that this effect was modulated by the plausibility of the distractor, such that the plausibility effect was attenuated in sentences with a plausible distractor, e.g., *the cup*. These findings support Van Dyke and McElree's proposal that oblique arguments, such as PPs, trigger interference, and extend

their findings by showing that retrieval for thematic binding is sensitive to a broader range of thematic-semantic properties of the distractor encoding beyond animacy, e.g., [+shatterable].

A concern for the encoding hypothesis proposed by Van Dyke and McElree (2011) is that not all core arguments are equally resistant to interference. For instance, although they found that distractors in a direct object position resist interference during retrieval for thematic binding, they also found that distractors in a subject position reliably triggered interference, despite being a core argument. Van Dyke and McElree (2011) suggested that interference from subject distractors is expected because they match the syntactic cues from the verb, and it is only when a core argument mismatches the syntactic cues, as in the case of a direct object distractor, that they are precluded from retrieval, resulting in an effect they called 'syntactic gating.'

The finding that subject distractors trigger interference is also consistent with the recent proposal that the prominence of the distractor modulates interference (Cunnings and Felser, 2013; Engelmann et al., 2015; Patil et al., 2016). For instance, subjects are more prominent than direct objects in terms of their hierarchical position and discourse function, which makes them more salient in memory, and hence more likely to interfere at retrieval. On this view, argument status is but a single factor that determines susceptibility to interference.

The encoding hypothesis proposed by Van Dyke and McElree (2011) has important implications for our understanding of how we encode and navigate linguistic structures in memory. However, their proposal has never been explicitly tested, and the generality of the effects on which it is based remains unclear. Furthermore, the principle of argument status is based on both the syntactic and thematic-semantic properties of the constituent, and it remains unclear which of these properties is responsible for the observed contrast, making it difficult to distinguish the various accounts relating to argument status, cue-overlap (syntactic gating), and prominence. It is thus important to test whether the contrast observed in (2) generalizes to a broader set of structural environments and a wider range of linguistic dependencies to better understand what properties cause memory retrieval mechanisms to succeed and fail during sentence comprehension.

The Present Study

The present study uses interference effects in the comprehension of subject-verb agreement ('agreement attraction') to test Van Dyke and McElree's (2011) hypothesis that interference is mediated by the argument status of the distractor. Agreement attraction arises when a comprehender fails to notice that a plural-marked verb erroneously agrees with a distractor noun (termed an 'attractor') that is not its syntactic subject. It manifests as eased processing and boosted acceptability during agreement processing, relative to sentences that should be equally acceptable or unacceptable, resulting in an effect known as 'agreement attraction.' For instance, Wagers and colleagues used self-paced reading to examine the processing of grammatical and ungrammatical subject-verb agreement dependencies like those in (4). The sentence in (4b) is ungrammatical because the plural verb *were* does not agree in number with the head of its subject noun phrase (NP) *key*.

(4) (a) The key to *the cabinet(s)* unsurprisingly was rusty after years of disuse.
(b) *The key to *the cabinet(s)* unsurprisingly were rusty after years of disuse.

Wagers and colleagues found that in grammatical sentences like (4a), the number marking on the plural attractor *cabinets* did not impact acceptability or reading times after the verb. However, in ungrammatical sentences like (4b), the plural attractor *cabinets*, which matched the number of the verb *were*, boosted acceptability and facilitated reading times after the verb, relative to the ungrammatical condition with the singular noun *cabinet*. Wagers and colleagues argued that the facilitation observed in sentences like (4b) was due to incorrect retrieval of the plural attractor, which matches the plural retrieval cue at the verb. According to this account, encountering the plural verb *were* triggers a retrieval process to recover a constituent in memory that matches the cues [+subject] and [+plural]. In sentences that give rise to agreement attraction, like (4b), the target subject is encoded as [+subject] and [−plural], whereas the attractor is encoded as [−subject] and [+plural]. In this scenario, the retrieval processes triggered at the verb may retrieve the 'attractor' based on the partial match to the [+plural] cue, leading to the false impression that agreement is licensed (see also Dillon et al., 2013; Tanner et al., 2014; Lago et al., 2015; Tucker et al., 2015; Parker and Phillips, 2017; Tucker and Almeida, 2017).

Agreement attraction is not simply a case of proximity concord (Quirk et al., 1985) or local coherence (Tabor et al., 2004), as attraction is observed when the attractor does not intervene between the verb and its subject, as shown in (5).

(5) *The *runner(s)* who the driver see each morning always wave.

Agreement attraction provides an ideal test of Van Dyke and McElree's (2011) hypothesis that interference is mediated by the argument status of the distractor because susceptibility to attraction can be examined in a broad range of configurations, such as those with attractors in core and oblique argument positions. However, the vast majority of studies on agreement attraction have relied on a narrow range of configurations involving oblique PP attractors (see Hammerly et al., 2018, for a recent review), motivating further research. A small number of studies have reported evidence of attraction from constituents in core argument positions, such as matrix subjects like (5) (Clifton et al., 1999; Wagers et al., 2009) and direct objects embedded inside a relative clause (Dillon et al., 2013). But, there has not yet been a direct, side-by-side comparison of attractors in core argument and oblique positions for subject-verb agreement to evaluate Van Dyke and McElree's (2011) proposal. Furthermore, existing studies employed different experimental designs, items, and methodologies, making it difficult to compare interference profiles across configurations (see Jäger et al., 2017, for a Bayesian meta-analysis of attraction effects in comprehension). These issues are addressed in the current study.

Overview of Experiments

Three self-paced reading experiments were designed to test Van Dyke and McElree's (2011) hypothesis that interference is mediated by the argument status of the distractor using an agreement attraction paradigm. Specifically, we used the amount of attraction generated by core argument vs. oblique argument attractors to diagnose the distinctiveness of the respective encodings. Experiment 1 directly compared oblique (PP) argument attractors and core argument (direct object) attractors embedded inside a subject-modifying relative clause, as shown in (6), and Experiment 2 compared two types of core argument attractors (subject and direct object) in configurations like (7). To preview, attraction effects were observed for oblique attractors (PP attractors), but the effect was nullified for core argument attractors (subject and direct object attractors). These results are consistent with Van Dyke and McElree's (2011) proposal that interference is mediated by the argument status of the distractor, but challenge accounts that claim that subjects should produce more interference due to their prominence (cf. Engelmann et al., 2015).

(6) (a) **PP attractor**
*The waitress who sat near the girls unsurprisingly were unhappy . . .
(b) **Direct object attractor**
*The waitress who sat the girls unsurprisingly were unhappy . . .

(7) (a) **Direct object attractor**
*The celebrity who insulted the journalists certainly were upset . . .
(b) **Subject attractor**
*The celebrity who the journalists insulted certainly were upset . . .

Experiment 3 then tested core argument attractors in a syntactically oblique position (oblique agents), as shown in (8), to determine whether the lack of attraction for items in core argument positions is driven by their syntactic position or their thematic-semantic properties that jointly define their argument status.

(8) **Oblique agent attractor**
*The house that had been built by the workers sadly were falling . . .

Results showed that oblique agents resist attraction, which suggests that the lack of attraction for core arguments is not driven by the attractor's syntactic position, but rather its thematic-semantic properties. Taken together, the results of Experiments 1–3 provide converging evidence in favor of the proposal that interference effects are mediated by the argument status of the interfering item (Van Dyke and McElree, 2011), and motivate a more comprehensive account of agreement processing that must consider both encoding and retrieval mechanisms.

EXPERIMENT 1: DIRECT COMPARISON OF CORE VS. OBLIQUE ARGUMENTS

Experiment 1 directly compared PP and direct object attractors using self-paced reading to test Van Dyke and McElree's (2011) proposal that interference during retrieval for linguistic dependency formation is mediated by the argument status of the interfering item. According to their proposal, the encoding of oblique arguments, such as PPs, is less distinctive than that of core arguments like subjects and objects. On this view, the encoding of oblique arguments is not salient enough to trigger a mismatch to the syntactic cues at retrieval, making interfering items in oblique argument positions more likely to interfere at retrieval. If agreement attraction, as a specific kind of interference, is mediated by the argument status of the attractor, then we expect to find a substantially reduced or nullified attraction effect for sentences with a core argument direct object attractor, relative to sentences with an oblique argument PP attractor. However, if argument status does not mediate attraction, then we expect comparable attraction effects for PP and direct object attractors.

Based on previous studies of agreement attraction in comprehension (e.g., Wagers et al., 2009; Dillon et al., 2013), attraction is predicted to manifest as a reduced reading time disruption for ungrammatical sentences with a plural attractor, relative to ungrammatical counterparts with a singular attractor. By contrast, the absence of an attraction effect is predicted to appear as disrupted reading times for ungrammatical sentences, with no statistically significant difference in reading times between the ungrammatical sentences.

Participants

Participants were 60 native speakers of English who were recruited using Amazon's Mechanical Turk web service[1]. All participants in this and the following experiments provided informed consent and were screened for native speaker abilities. The screening probed knowledge of the constraints on English tense, modality, morphology, ellipsis, and syntactic islands. Participants were compensated $4.00. The experiment lasted approximately 25 min.

Materials

Experimental materials consisted of 48 sets of 8 items like those shown in **Table 1**. Three experimental factors were manipulated, including grammaticality (grammatical vs. ungrammatical), attractor number (singular vs. plural), and attractor argument status (direct object vs. PP). In all conditions, the target subject was modified by a subject relative clause that contained the attractor, followed by the main clause verb phrase, which consisted of the critical agreeing auxiliary verb and a 4–7 word spillover region. The target subject was always singular. The relative clause verb never overtly expressed agreement to prevent attraction before the critical region. Grammaticality was manipulated by varying the number feature of the critical agreeing verb (grammatical conditions = *was*, ungrammatical conditions = *were*). Attractor number was manipulated by

[1]https://aws.amazon.com/mturk

TABLE 1 | Sample set of items for Experiment 1.

Direct object attractor
Grammatical, PL attractor
The waitress who sat the girls unsurprisingly was unhappy about all the noise.
Grammatical, SG attractor
The waitress who sat the girl unsurprisingly was unhappy about all the noise.
Ungrammatical, PL attractor
The waitress who sat the girls unsurprisingly were unhappy about all the noise.
Ungrammatical, SG attractor
The waitress who sat the girl unsurprisingly were unhappy about all the noise.
PP attractor
Grammatical, PL attractor
The waitress who sat near the girls unsurprisingly was unhappy about all the noise.
Grammatical, SG attractor
The waitress who sat near the girl unsurprisingly was unhappy about all the noise.
Ungrammatical, PL attractor
The waitress who sat near the girls unsurprisingly were unhappy about all the noise.
Ungrammatical, SG attractor
The waitress who sat near the girl unsurprisingly were unhappy about all the noise.

SG, singular; PL, plural.

varying the number of the attractor, such that it appeared in either singular or plural form. Based on previous studies on agreement attraction in comprehension, such as Wagers et al. (2009) and Dillon et al. (2013), singular attractors were predicted to cause no attraction, whereas plural embedded attractors were potential sources of attraction, but only in the ungrammatical conditions, where the target subject and critical verb mismatched in number. Attractor argument status was manipulated by varying the position of the attractor, such that it appeared in either direct object or PP position immediately following the relative clause verb. Lexical items were chosen to create maximally similar sentences for direct object and PP attractor conditions. Crucially, the linear distance between the attractors and critical agreeing verbs was identical in each configuration to prevent biases due to differences in recency, decay, or passive memory dynamics unrelated to the processing of subject-verb agreement (e.g., Van Dyke and Lewis, 2003). The full set of experimental materials can be found in the **Supplementary Materials**.

The 48 target items were distributed across 8 lists in a Latin square design and combined with 96 grammatical filler sentences of similar length and complexity, such that each participant read a total of 144 sentences. All sentences were followed by a 'yes/no' comprehension question that addressed various parts of the sentence to prevent participants from developing superficial reading strategies that would allow them to answer the question without reading the entire sentence.

Procedure

The experiment was conducted using the online experiment platform Ibex (Drummond, 2018), which allows self-paced reading experiments to be deployed in a standard web browser. Sentences were initially masked by dashes, with white spaces and punctuation intact. Participants pushed the space bar to reveal each word. Presentation was non-cumulative, such that the previous word was replaced with dashes when the next word appeared. On-screen feedback was provided for incorrect answers to the comprehension questions. The order of presentation was randomized for each participant. To ensure that participants completed the task as directed, an instructional manipulation check was used (Oppenheimer et al., 2009). Instructional manipulation checks ensure that participants are completing the task as directed by asking them to ignore the standard response format and provide a confirmation that they have read the instructions.

Analysis

Only participants with at least 80% accuracy on comprehension questions were used in the analysis. Two participants were removed for performance below 80%. Four regions of interest were identified: the word immediately preceding the critical agreeing verb (pre-critical region), the agreeing verb (critical region), and the two words immediately following the verb (spillover regions 1 and 2, respectively). Based on previous studies that tested agreement attraction using self-paced reading, attraction effects were predicted to manifest starting at the regions immediately following the critical verb, e.g., spillover regions 1 and 2. Statistical analyses were carried out with linear mixed-effects models using the *lmerTest* package (Kuznetsova et al., 2014) in the *R* software environment (R Development Core Team, 2018). Analyses were carried out over the raw, untrimmed data, since recent research on attraction suggests that data transformations, such as those involving log-transformation or outlier removal (trimming), can obscure attraction effects (Staub, 2010; Lago et al., 2015; Tucker and Almeida, 2017; Villata et al., 2018).[2] Models were defined using orthogonal contrast coding to examine the effects of grammaticality, attractor number, and their interaction (grammaticality × attractor number) for each region of interest. Following Dillon et al. (2013), additional models were defined to focus on the effect of attraction (i.e., the amount of facilitation for ungrammatical sentences with a plural attractor relative to ungrammatical sentences with a singular attractor), labeled as 'attraction' in the coefficient tables, and the interaction of attraction with attractor argument status to determine whether PP and object attractors were differentially susceptible to attraction. All models were fit with a full variance-covariance matrix, i.e., a maximal random effects structure, with random intercepts and slopes for all fixed effect predictors by participants and items (Barr et al., 2013). If there was a convergence failure, or if the model converged but the correlation estimates were high, the random effects structure was simplified. A fixed effect was considered significant if its absolute *t*-value was greater than 2, which indicates that its 95% confidence interval did not include 0 (Gelman and Hill, 2007).

[2]For comparison, analyses using the log-transformed reading time values are reported in the **Supplementary Materials**. The results are comparable, and the presence/absence of attraction does not differ between analysis methods.

Results

Figure 1 shows the average word-by-word reading times for sentences with a PP attractor, and **Figure 2** shows the same for sentences with a direct object attractor. Mean reading times by condition at the regions of interest are provided in **Table 2**, and the results of the statistical analyses are reported in **Table 3**. Contrasting profiles were observed for prepositional and direct object attractors. In the PP attractor conditions, no effects were observed in the pre-critical or critical regions. As expected, the following spillover regions showed a main effect of grammaticality (Spillover 1 and 2), a main effect of attractor number (Spillover 1), a significant effect of attraction (Spillover 1 and 2), and a significant interaction between grammaticality and attractor number (Spillover 2). In these regions, ungrammatical sentences were read more slowly than grammatical sentences, but this processing disruption was nullified for ungrammatical sentences with a plural attractor, relative to ungrammatical sentences with a singular attractor. This pattern reflects the behavioral signature of agreement attraction, replicating previous results (e.g., Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015; Tucker et al., 2015; Parker and Phillips, 2017; Tucker and Almeida, 2017).

In the direct object attractor conditions, no effects were observed in the pre-critical or critical regions. The following spillover region showed a main effect of grammaticality (Spillover 1), carried by longer reading times in the ungrammatical sentences relative to grammatical sentences. In contrast to the PP attractor conditions, there was no evidence of attraction in any region, as reading times between ungrammatical conditions did not diverge.

The contrast between PP and direct object attractors with respect to attraction was supported by a significant interaction between attraction and attractor argument status, carried by the significant attraction effect for the PP attractor conditions.

Discussion

Experiment 1 directly compared PP and direct object attractors to test Van Dyke and McElree's (2011) proposal that interference effects are mediated by the argument status of the interfering item. This proposal claims that core arguments, such as direct objects, are encoded in memory more distinctly than oblique arguments, such as PPs, making them easier to reject when they mismatch the retrieval cues, and hence less likely to interfere at retrieval. Experiment 1 revealed that oblique arguments in PP position interfered during retrieval for agreement processing, yielding a clear agreement attraction effect, but core arguments in direct object position did not. These results are closely aligned with Van Dyke and McElree's (2011) proposal, and extend their findings by showing that the contrast between core and oblique arguments with respect to interference extends to a wider range of dependencies such as subject-verb agreement.

A concern with the results of Experiment 1 is that the critical interactions of grammaticality × attractor number and attraction × attractor argument status were observed two words after the critical verb. There are two reasons why we might see these effects appear after the critical word. First, recent work on the timing of agreement attraction effects suggests that attraction is an error-driven process that manifests in the late stages of agreement processing (Lago et al., 2015; Parker and Phillips, 2017). The observation of a late interaction is consistent

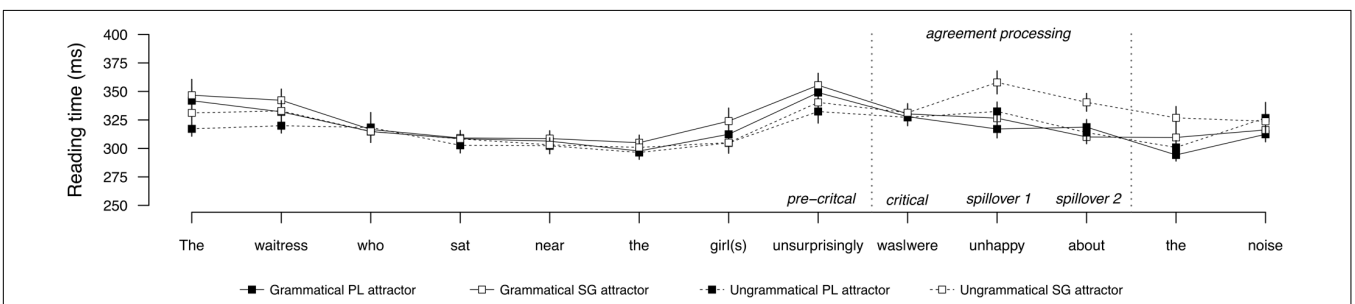


FIGURE 1 | Word-by-word reading times for the PP attractor conditions in Experiment 1. Error bars indicate standard error of the mean.

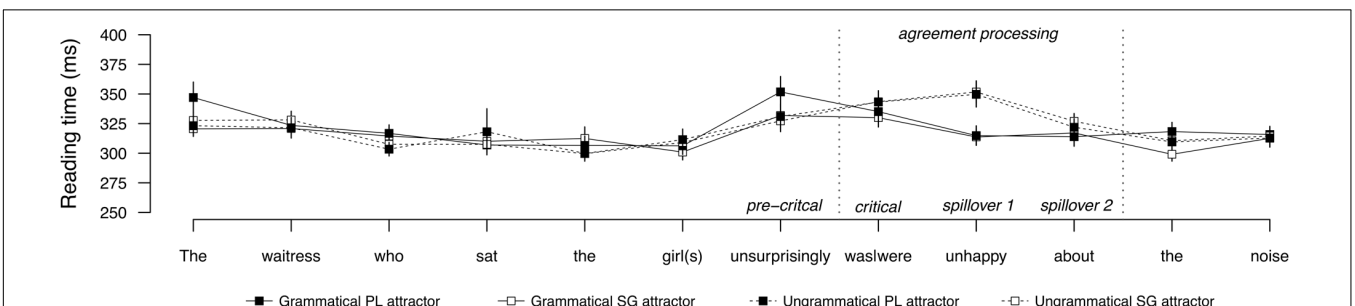


FIGURE 2 | Word-by-word reading times for the direct object attractor conditions in Experiment 1. Error bars indicate standard error of the mean.

TABLE 2 | Mean reading times (ms) by condition at the regions of interest for Experiment 1.

	Regions			
	Pre-critical	**Critical**	**Spillover 1**	**Spillover 2**
PP attractor				
Grammatical, PL attractor	349	327	317	318
Grammatical, SG attractor	355	330	326	310
Ungrammatical, PL attractor	332	327	332	313
Ungrammatical, SG attractor	340	331	357	340
Direct object attractor				
Grammatical, PL attractor	331	330	313	317
Grammatical, SG attractor	351	335	314	313
Ungrammatical, PL attractor	327	343	351	326
Ungrammatical, SG attractor	330	343	349	321

with this view. Second, observing an effect one or two regions downstream from the critical region is expected in self-paced reading tasks, since participants often adopt a fixed rhythm in advancing through the sentence (Witzel et al., 2012).

Another concern with Experiment 1 is that it failed to replicate attraction in configurations that have been shown to yield attraction in previous studies. For instance, Dillon et al. (2013) observed attraction when the attractor appeared as the direct object of a subject-modifying relative clause. This configuration is nearly identical to the direct object attraction condition tested in Experiment 1, which did not show attraction. One possibility is that the lack of attraction for direct object attractors in Experiment 1 is due to a lack of statistical power. There are three reasons why the current results are unlikely to reflect low power. First, we observed a positive attraction effect in maximally similar sentences involving PP attractors, which suggests that there was sufficient power to elicit attraction. Second, Experiment 1 had more power than previous studies that elicited attraction. For instance, Experiment 1 relied on 6 observations per condition, with 60 participants, yielding a total of 360 points for analysis. By comparison, Dillon et al. (2013) elicited attraction with less power (6 observations per condition, with 40 participants, for a total of 240 data points). Other studies that used self-paced reading to elicit attraction are similarly patterned. For example, Wagers et al. (2009) elicited attraction using self-paced reading in a design with exactly half the power of Experiment 1 in the current study. Thus, the lack of attraction under superficially similar conditions in the present study is unlikely to reflect an issue of statistical power. Third, we conducted a *post hoc* power analysis using the *simr* package in R over the final linear-mixed-effects model (including the main effects and their interaction) at the second spillover region, which showed the critical interaction between attraction and attractor position. According to this analysis, the observed power was at 74%, which suggests that lack of power is an unlikely cause for the contrast. However, since the standard recommendation is that the target power rate should be at least 80% (Cohen, 1962, 1988, 1992), the issue of power is addressed further in Experiment 2.

Another possibility is that the contrast between the current study and Dillon et al. (2013) reflects variability in the materials used by Dillon and colleagues. In their study, Dillon et al. (2013) reported the use of direct object attractors in their sample set of materials (see **Table 1** of their study). However, their full materials list shows that they used a combination of direct object and PP attractors exactly of the form tested in Experiment 1, with nearly 40% of their items using PP attractor configurations. It is possible that the attraction effects that they observed were driven by the PP conditions, in which case our studies pattern similarly with respect to attraction effects for prepositional and direct object attractors.

A third issue with Experiment 1 concerns the relationship between encoding accounts of interference (e.g., Van Dyke and McElree, 2011) and accounts of prominence and cue-matching (e.g., Engelmann et al., 2015). According to the encoding account, core arguments, like direct object attractors, are encoded in memory more distinctly than oblique arguments, like PP

TABLE 3 | Summary of statistical analyses for PP attractor conditions and direct object attractor conditions in Experiment 1.

	Regions											
	Pre-critical			**Critical**			**Spillover 1**			**Spillover 2**		
	$\hat{\beta}$	***SE***	***t***	$\hat{\beta}$	***SE***	***t***	$\hat{\beta}$	***SE***	***t***	$\hat{\beta}$	***SE***	***t***
PP attractor												
Grammaticality	−7.94	4.61	−1.72	0.09	2.87	0.03	**11.70**	**3.33**	**3.50**	**6.36**	**2.52**	**2.51**
Attractor number	3.64	4.49	0.81	1.64	2.87	0.57	**8.73**	**3.33**	**2.61**	4.53	2.52	1.79
Grammaticality × attractor number	0.38	4.49	0.08	0.44	2.87	0.15	4.00	3.33	1.21	**8.77**	**2.52**	**3.47**
Attraction	4.06	6.45	0.63	1.99	4.53	0.44	**12.72**	**4.99**	**2.54**	**13.27**	**4.47**	**2.96**
Direct object attractor												
Grammaticality	−6.35	4.53	−1.40	5.41	3.44	1.57	**18.23**	**3.99**	**4.56**	4.45	3.85	1.15
Attractor number	−5.96	4.53	−1.31	−1.25	3.44	−0.36	0.14	4.80	0.03	2.04	3.19	0.64
Grammaticality × attractor number	4.0	4.53	0.88	1.32	3.44	0.38	0.78	3.52	0.22	0.47	2.77	0.17
Attraction	−1.87	6.49	−0.28	0.11	5.53	0.02	0.94	7.42	0.12	2.53	3.89	0.65
Attraction × argument status	5.87	8.53	0.68	1.77	6.53	0.27	11.56	7.74	1.49	**10.73**	**5.26**	**2.04**

Significant coefficients ($|t| > 2$) are in bold.

attractors, and hence, are less likely to interfere. The strong view of this proposal would be that all core arguments, including direct objects and subjects, should resist interference, by virtue of their argument status, regardless of their syntactic position. However, recent accounts of prominence and cue-matching (e.g., Cunnings and Felser, 2013; Engelmann et al., 2015; Patil et al., 2016) predict divergent interference profiles for core arguments. Specifically, core arguments are predicted to trigger interference if they more closely match the retrieval cues from the verb, such as the subject cue, or are in a more prominent position in the sentence, such as in a subject position. As a result, subjects are predicted to be more likely to interfere at retrieval than items in less prominent positions, like direct objects, due to their heightened activation in memory (see Engelmann et al., 2015, for predictions from computational simulations). This possibility is tested in Experiment 2.

EXPERIMENT 2: DIRECT COMPARISON OF SUBJECT AND OBJECT ATTRACTORS

Experiment 1 showed that a core argument in direct object position did not trigger agreement attraction. It is possible that an attractor in subject position, despite its status as a core argument, might trigger attraction because it is highly accessible, both in terms of its match to the subject retrieval cue of the verb (Van Dyke and McElree, 2011) and its grammatical prominence (Engelmann et al., 2015). To test this hypothesis, Experiment 2 directly compared subject and direct object attractors in maximally similar configurations like those shown in (9) using self-paced reading.

(9) (a) **Direct object attractor**
*The celebrity who insulted the journalists certainly were upset . . .
(b) **Subject attractor**
*The celebrity who the journalists insulted certainly were upset . . .

If interference effects are mediated by the match to the retrieval cues or prominence, then we expect contrasting profiles for subject and object attractors, with stronger attraction effects predicted for subject attractors, since they provide a better match to the retrieval cues and are in a more prominent position. However, if interference effects are mediated by the argument status of the interfering item, as previously claimed (Van Dyke and McElree, 2011), then subject and object attractors should show similar profiles with respect to attraction because they share the same status as core arguments.

Participants

Participants were 120 native speakers of English who were recruited using Amazon's Mechanical Turk web service. This large sample size was chosen to increase statistical power (Vasishth and Nicenboim, 2016) to address the concern that the lack of attraction for core arguments in Experiment 1 was due to low power. Participants were compensated $4.00. The experiment lasted approximately 25 min.

Materials

Forty-eight item sets of the form shown in **Table 4** were constructed. The structure of the items followed the structure of the items used in Experiment 1, but held constant the argument status of the attractors, and instead manipulated their syntactic position. Attractors appeared either as the direct object of the relative clause verb, as in Experiment 1, or as the subject of the relative clause verb.

The 48 target items were distributed across 8 lists in a Latin square design and combined with the same 96 grammatical filler sentences from Experiment 1, such that each participant read a total of 144 sentences. All sentences were followed by a 'yes/no' comprehension question.

Procedure and Analysis

Experiment 2 used self-paced reading, following the same procedure and analysis methods used in Experiment 1. Three participants were removed for failing the instructional manipulation check, and an additional 14 participants were removed for performance below 80%, leaving a total of 103 participants for data analysis.

Results

Figure 3 shows the average word-by-word reading times for sentences with a direct object attractor, and **Figure 4** shows the same for sentences with a subject attractor. Mean reading times by condition at the regions of interest are provided in **Table 5**, and the results of the statistical analyses are reported in **Table 6**. No effects were observed in the pre-critical conditions for either subject or object attractor conditions. Both subject and object attractor conditions showed a main effect of grammaticality at

TABLE 4 | Sample set of items for Experiment 2.

Direct object attractor
Grammatical, PL attractor
The celebrity who insulted the journalists certainly was upset about the claims.
Grammatical, SG attractor
The celebrity who insulted the journalist certainly was upset about the claims.
Ungrammatical, PL attractor
The celebrity who insulted the journalists certainly were upset about the claims.
Ungrammatical, SG attractor
The celebrity who insulted the journalist certainly were upset about the claims.
Subject attractor
Grammatical, PL attractor
The celebrity who the journalists insulted certainly was upset about the claims.
Grammatical, SG attractor
The celebrity who the journalist insulted certainly was upset about the claims.
Ungrammatical, PL attractor
The celebrity who the journalists insulted certainly were upset about the claims.
Ungrammatical, SG attractor
The celebrity who the journalist insulted certainly were upset about the claims.

SG, singular; PL, plural.

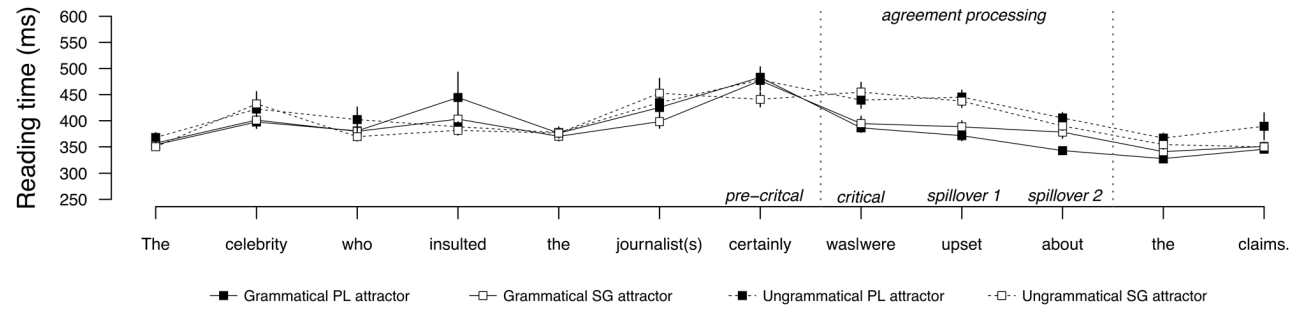


FIGURE 3 | Word-by-word reading times for the direct object attractor conditions in Experiment 2. Error bars indicate standard error of the mean.

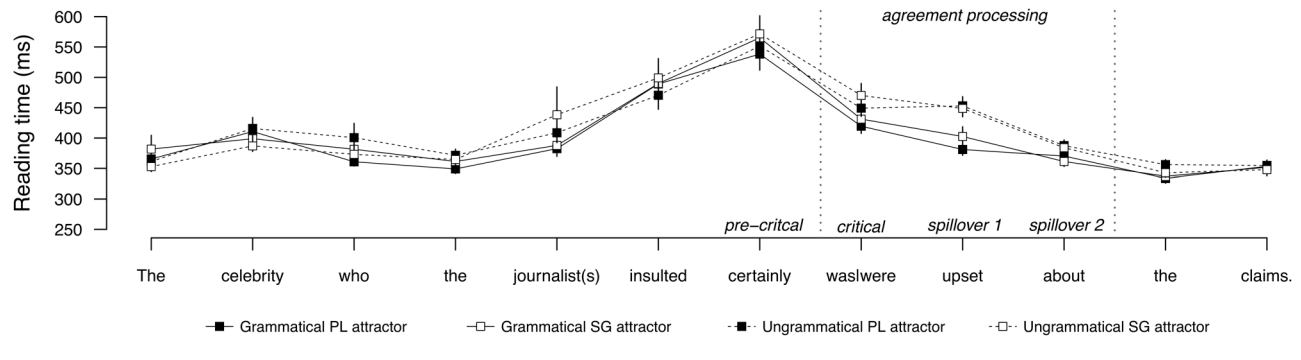


FIGURE 4 | Word-by-word reading times for the subject attractor conditions in Experiment 2. Error bars indicate standard error of the mean.

the critical verb region, which persisted to the second spillover region. There were no effects of attractor number or attraction in any region for subject and object attractor conditions. The second spillover region for object attractors showed a significant interaction between grammaticality and attractor number. Pairwise comparisons revealed that this interaction was carried by divergent reading times in the grammatical conditions, as grammatical sentences with a singular attractor were read more slowly than grammatical sentences with a plural attractor, relative to the ungrammatical conditions, which did not diverge. No other effects or interactions were observed.

Discussion

Experiment 2 compared subject and direct object attractors to test the hypothesis that a subject attractor should trigger attraction because it more closely matches the retrieval cues of the verb and is in a grammatically prominent position. Experiment 2 revealed two main findings. First, the prediction that subjects should trigger attraction was not supported by the reading time data from Experiment 2. Both subject and object attractor conditions showed a main effect of grammaticality, indicating that comprehenders were sensitive to the feature match between the verb and the target subject, but no evidence of attraction was found in any region from either subject or direct attractors. Second, Experiment 2 replicated the results of Experiment 1 by showing that direct objects resist attraction. This effect is notable given the high statistical power. This finding suggests that the lack of attraction for the direct object attractors in Experiment 1 cannot be reduced to low power. Taken together, the results of Experiments 1 and 2 are consistent with Van Dyke and McElree's (2011) proposal that interference is mediated by the argument status of the distractor, but challenge the recent proposal that subjects are more likely to interfere due to their prominence (cf. Engelmann et al., 2015).

A concern with Experiment 2 is the interaction between grammaticality and attractor number for the direct object attractors at the second spillover region. This effect was carried by divergent reading times in the grammatical conditions, as grammatical sentences with a singular attractor were read more slowly than grammatical sentences with a plural attractor, relative to the ungrammatical conditions, which did not diverge. This effect is unexpected under accounts that assume that attraction is an error-driven process that is triggered only when the verb form violates the number prediction made by the subject (Wagers et al., 2009; see also, Lago et al., 2015; Parker and Phillips, 2017). According to this account, retrieval is not engaged in the grammatical conditions because the prediction is satisfied. The alternative view is that retrieval always occurs at the verb, regardless of grammaticality. However, the fact that the same effect was not observed in the subject attractor conditions or in Experiment 1 suggests that this effect may reflect a Type I error.

Another concern is that Experiment 2 failed to replicate Van Dyke and McElree's (2011) syntactic gating effect, in which items in a subject position interfere at retrieval due to their match to the subject retrieval cues of the verb. There are two possibilities for why we might expect this difference. One

TABLE 5 | Mean reading times (ms) by condition at the regions of interest for Experiment 2.

	Regions			
	Pre-critical	**Critical**	**Spillover 1**	**Spillover 2**
Direct object attractor				
Grammatical, PL attractor	483	386	371	342
Grammatical, SG attractor	476	394	388	378
Ungrammatical, PL attractor	478	439	445	405
Ungrammatical, SG attractor	440	454	437	390
Subject attractor				
Grammatical, PL attractor	537	419	381	370
Grammatical, SG attractor	565	431	402	361
Ungrammatical, PL attractor	551	449	452	387
Ungrammatical, SG attractor	571	470	448	384

possibility is that we tested a different dependency. We tested subject-verb agreement, which is a morpho-syntactic feature-matching process, whereas Van Dyke and McElree (2011) tested thematic binding, which is an interpretive process that aids in establishing the meaning of the sentence. Both processes require retrieval of the local subject at the verb, but they have different grammatical functions. It is thus not unreasonable to assume that they might use different cues to guide retrieval based on their different grammatical requirements. For instance, agreement might rely more on morpho-syntactic cues like person and number, whereas thematic binding might rely more on thematic-semantic cues, like animacy. However, it remains unclear why retrieval mechanisms would use different cues to target the same position.

A more likely possibility is that the contrasting profiles for subject distractors reflect differences in feature similarity between the target and distractor NPs in memory. In the items tested by Van Dyke and McElree (2011), both the subject distractor and target overlapped substantially with the retrieval cues (both were animate subjects), which can reduce the distinctiveness of the target and increase the opportunity for interference at retrieval (Watkins and Watkins, 1975; Nairne, 1988, 1990; Anderson and Lebiere, 1998; Anderson et al., 2004; McElree, 2006). By contrast, the subject attractor and target in Experiment 2 of the current study were more distinct in feature content (plural vs. singular), increasing their distinctiveness at retrieval, reducing the chances of interference from cue-overlap. Crucially, this account is consistent with Van Dyke and McElree's (2011) general claim that interference is dependent on the *distinctiveness* of the information in memory. This account is also consistent with the recent proposal that interference depends on the degree to which the target and distractor match the retrieval cues (Parker and Phillips, 2017).

A more fundamental concern is that it is unclear why subject and object attractors differ from PP attractors with respect to interference. The results are consistent with Van Dyke and McElree's (2011) proposal that interference (measured here in terms of attraction) is mediated by the argument status of the interfering item. But an item's argument status is defined by both its syntactic and thematic-semantic properties. At this point, it is not clear which of these properties drives the contrasts observed in Experiments 1 and 2. This issue is addressed in Experiment 3.

EXPERIMENT 3

Experiments 1 and 2 revealed that PP attractors differ from subject and object attractors with respect to agreement attraction. However, the source of this contrast remains unclear. On the one hand, the contrast could reflect the thematic-semantic status of the attractor, as originally hypothesized (e.g., Van Dyke and McElree, 2011). On the other hand, the contrast could simply reflect the attractor's syntactic position. To distinguish these alternatives, Experiment 3 probed for attraction using core argument attractors that appeared in a PP position. Specifically, Experiment 3 tested configurations with an "oblique agent" attractor, where the attractor is a core thematic subject that appeared in a passive PP *by*-phrase (see **Table 5** for an example). If core arguments resist attraction by virtue of their thematic-semantic properties, then changes in their syntactic position should not impact their susceptibility to attraction. On this view, the oblique agent attractor should pattern with the core arguments from Experiments 1–2 (direct objects and subjects) by resisting attraction. However, if the contrast between core and oblique arguments is a consequence of their syntactic position, then the oblique agent should pattern with the PP attractor from Experiment 1 by triggering attraction.

Participants

Participants were 120 native speakers of English who were recruited using Amazon's Mechanical Turk web service. Participants were compensated $4.00. The experiment lasted approximately 25 min.

Materials

Twenty-four item sets of the form shown in **Table 7** were constructed. Two factors were manipulated, grammaticality and attractor number. Across all conditions, the target subject was modified by a passivized relative clause that contained the attractor in a prepositional *by*-phrase (oblique agent), followed by the main clause VP and spillover regions. The passivized relative clause verb never overtly expressed agreement to prevent spurious interference effects prior to the critical verb.

The 24 target items were distributed across 4 lists in a Latin square design and combined with the 48 grammatical filler sentences from Experiments 1–2, such that each participant read a total of 72 sentences. All sentences were followed by a 'yes/no' comprehension question.

Procedure and Analysis

Experiment 3 used self-paced reading, following the same procedure and analysis methods used in Experiments 1 and 2. Two participants were removed for failing the instructional manipulation check, and an additional 11 participants were removed for performance below 80%, leaving a total of 107 participants for data analysis.

TABLE 6 | Summary of statistical analyses for direct object attractor conditions and subject attractor conditions in Experiment 2.

	Regions											
	Pre-critical			Critical			Spillover 1			Spillover 2		
	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
Direct object attractor												
Grammaticality	−10.23	6.94	−1.47	**28.46**	**9.17**	**3.10**	**30.80**	**4.7**	**6.88**	**18.64**	**4.34**	**4.29**
Attractor number	−10.92	6.94	−1.57	5.48	7.80	0.70	2.50	4.48	0.55	5.027	3.92	1.28
Grammaticality × attractor number	−7.82	6.94	−1.12	1.51	7.93	0.19	−5.98	5.0	−1.17	**−12.65**	**4.16**	**−3.03**
Attraction	−18.86	12.71	−1.48	7.15	12.43	0.57	−3.95	6.89	−0.57	−7.66	5.20	−1.47
Subject attractor												
Grammaticality	1.45	12.03	0.12	**17.12**	**7.25**	**2.36**	**19.48**	**6.94**	**4.24**	**9.97**	**3.62**	**2.74**
Attractor number	15.43	12.05	1.28	2.46	7.25	0.33	4.34	5.74	0.75	−3.07	3.63	−0.84
Grammaticality × attractor number	−4.98	12.0	−0.41	8.73	7.25	1.20	−6.47	7.14	−0.90	1.37	3.76	0.36
Attraction	10.62	17.54	0.60	11.70	10.58	1.10	−2.15	8.33	−0.25	−1.78	5.25	−0.33
Attraction × argument status	28.98	18.14	1.59	4.97	16.12	0.30	1.82	10.93	0.16	5.96	7.34	0.81

Significant coefficients ($|t| > 2$) are in bold.

Results

Figure 5 shows the average word-by-word reading times for Experiment 3. Mean reading times by condition at the regions of interest are provided in **Table 8**, and the results of the statistical analyses are reported in **Table 9**. No effects were observed in the pre-critical or critical regions. A main effect of grammaticality was observed in spillover regions 1 and 2. There was no evidence of attraction in any region.

Discussion

The goal of Experiment 3 was to determine whether the contrast between PP vs. subject and direct object attractors observed in Experiments 1 and 2 reflects the thematic-semantic status of the attractor or its syntactic position. This was achieved by testing oblique agents, which are core thematic subjects that appear in an oblique PP position. Results showed that oblique agents resist attraction, patterning with the core arguments from Experiments 1 and 2. These results suggest that the modulation of the attraction effect observed across Experiments 1 and 2 cannot be reduced to the syntactic position of the attractor, or at least a PP position. Instead, the currents results favor the proposal that core arguments resist interference by virtue of their thematic-semantic properties (e.g., Van Dyke and McElree, 2011).

A concern with the results of Experiment 3 is that they appear to conflict with previous studies on attraction in production, which have shown that the syntactic position of the attractor modulates attraction. For instance, attractors that are syntactically similar to agreement controllers (e.g., they c-command the verb) lead to more attraction errors than those that only precede the verb (Franck et al., 2006, 2010, 2015). The current results are not incompatible with these findings, and we do not deny that syntactic position plays an important role in attraction, at least in production. Rather, the results of Experiment 3 suggest that the current contrast between core and oblique arguments with regards to attraction in comprehension cannot be reduced to syntactic position. An important goal for future research is to determine whether the current contrasts observed in comprehension extend to agreement production.

TABLE 7 | Sample set of items for Experiment 3.

Grammatical, PL attractor
The house that had been built by the workers sadly was falling into great disrepair.
Grammatical, SG attractor
The house that had been built by the worker sadly was falling into great disrepair.
Ungrammatical, PL attractor
The house that had been built by the workers sadly were falling into great disrepair.
Ungrammatical, SG attractor
The house that had been built by the worker sadly were falling into great disrepair.

SG, singular; PL, plural.

GENERAL DISCUSSION

Summary of Results

The goal of the current study was to test Van Dyke and McElree's (2011) hypothesis that interference effects are mediated by the argument status of the distractor. This hypothesis states that core arguments, such as subjects and objects, are encoded more distinctly in memory than oblique arguments, such as PP objects, because core arguments play a more prominent role in establishing the meaning of the sentence, making them easier to accept or reject as retrieval candidates. The current study tested this hypothesis with an interference paradigm involving agreement attraction in three self-paced reading experiments. Experiment 1 directly compared PP and direct object attractors, and Experiment 2 directly compared direct object and subject attractors. Results showed a clear contrast: attraction was observed for PP attractors, but not for direct object or subject attractors. Experiment 3 then tested whether this contrast is a consequence of the syntactic or thematic-semantic properties of the attractors by testing core thematic arguments embedded in a PP (oblique agents). Results showed that oblique agents resisted

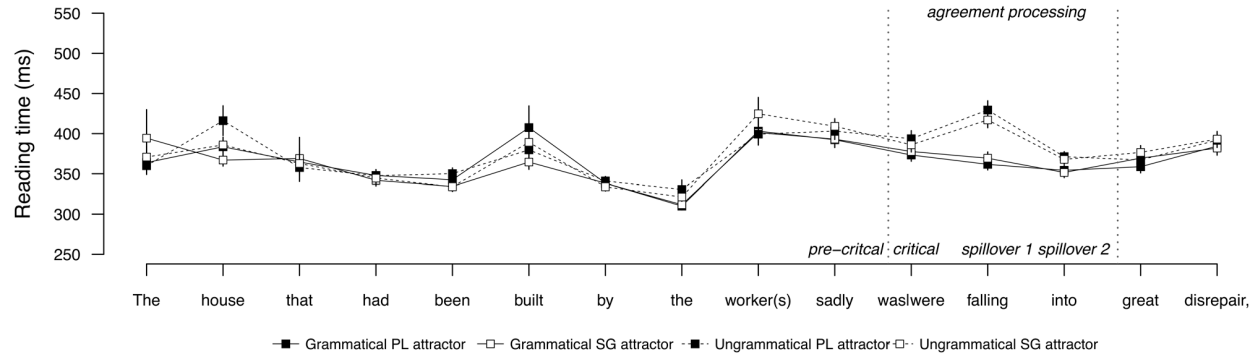


FIGURE 5 | Word-by-word reading times for Experiment 3. Error bars indicate standard error of the mean.

TABLE 8 | Mean reading times (ms) by condition at the regions of interest for Experiment 3.

	Regions			
	Pre-critical	**Critical**	**Spillover 1**	**Spillover 2**
Grammatical, PL attractor	392	373	361	354
Grammatical, SG attractor	393	377	369	351
Ungrammatical, PL attractor	403	393	429	371
Ungrammatical, SG attractor	409	386	416	367

attraction, patterning with the core arguments from Experiments 1 and 2. These results suggest that the contrast between core and oblique argument attractors is driven by their thematic-semantic properties, rather than their syntactic position.

Figure 6 provides a summary of the effects observed across each attractor position. This figure shows how PP attractors stand out, relative to subject, object, and oblique agent attractors, with respect to attraction effects. Taken together, the results of Experiments 1–3 provide converging evidence in favor of Van Dyke and McElree's (2011) proposal that interference is mediated by the argument status of the interfering item. They also extend previous results by showing that such effects generalize to a broader set of syntactic contexts and a wider range of syntactic dependencies, such as subject-verb agreement, and clarify that it is specifically the thematic-semantic properties of the argument that mediate interference.

Implications for Theories of Retrieval in Sentence Comprehension

The findings from the current study are unexpected under existing theories of memory retrieval in sentence comprehension, in the absence of a richer theory of memory representations and cues used in retrieval. Existing accounts, such as the prominent cue-based theory of memory retrieval, predict that interference effects for subject-verb agreement processing should generalize across syntactic contexts (Wagers et al., 2009; Dillon et al., 2013), based on the assumptions that the same interference-prone mechanism should apply whenever retrieval for agreement processing is required, and that interference is not mediated by the grammatical status of the attractor (McElree, 2000; McElree et al., 2003; Lewis and Vasishth, 2005). However, the current finding that agreement attraction is strongly modulated by the argument status of the attractor favors Van Dyke and McElree's (2011) proposal that interference is mediated by the encoding of the interfering item, motivating a more comprehensive account of agreement processing that depends on both retrieval and encoding mechanisms.

The current results also suggest that the relationship between argument status and interference is more tightly connected than previously assumed. For instance, Van Dyke and McElree (2011) found that core arguments in subject position trigger interference for thematic binding. However, Experiment 2 of the current study found that interference from subject distractors does not extend to subject-verb agreement. As suggested earlier, the positive effect found by Van Dyke and McElree (2011) may reflect a multiple match effect, where both the distractor and target overlap in feature content, reducing the distinctiveness of the target. Controlling for this difference, the generalization that emerges from these studies is that interference is dependent on the distinctiveness of the interfering item according to an argument hierarchy.

More broadly, the current results suggest that the memory architecture for language processing is more grammatically sophisticated than previously assumed. In particular, the current results, taken together with the findings reported in Van Dyke and McElree (2011), suggest that memory encoding mechanisms are attuned to fine-grained distinctions relating to the argument hierarchies described in the formal literature (Keenan and Comrie, 1977; Chomsky, 1981; Frazier and Clifton, 1996; Van Valin and LaPolla, 1997; Bresnan, 2001; Culicover and Jackendoff, 2005). These features of the grammar are often overlooked in many prominent models of sentence processing, including models that rely on superficial heuristics, "good enough" representations, local coherence, and other surface statistics (e.g., Townsend and Bever, 2001; Ferreira et al., 2002; Tabor et al., 2004; Ferreira and Patson, 2007; Karimi and Ferreira, 2016). Specifically, the results of the current study imply that interference effects are rooted in grammatical principles, e.g., an argument hierarchy, motivating a theory of sentence comprehension in which the parser and grammar are more

TABLE 9 | Summary of statistical analyses for Experiment 3.

	Regions											
	Pre-critical			Critical			Spillover 1			Spillover 2		
	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*	$\hat{\beta}$	*SE*	*t*
Grammaticality	6.73	4.44	1.51	7.15	3.99	1.79	**28.86**	**4.30**	**6.70**	**8.22**	**3.04**	**2.70**
Attractor number	−1.87	4.44	−0.41	0.85	3.99	0.21	1.10	4.30	0.25	1.45	3.04	0.47
Grammaticality × attractor number	1.41	4.44	0.31	−3.04	3.99	−0.76	−4.91	4.3	−1.14	−0.13	3.04	−0.04
Attraction	−16.20	14.84	−1.09	3.76	5.96	0.63	5.95	7.90	0.75	1.62	3.60	0.45

Significant coefficients (|t| > 2) are in bold.

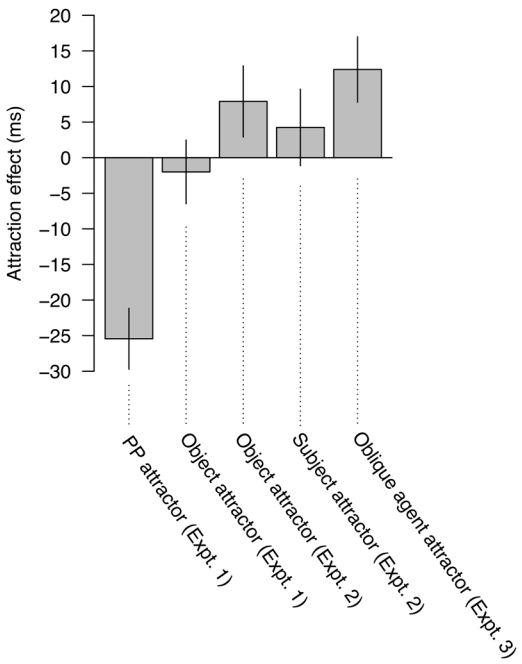


FIGURE 6 | Comparison of the profiles observed in Experiments 1–3 for PP, subject, object, and oblique arguments. The attraction effect for each attractor position was calculated by subtracting the RTs for the ungrammatical plural attractor condition from the ungrammatical singular attractor condition in the first spillover region.

closely aligned than previously assumed (e.g., Townsend and Bever, 2001; Ferreira et al., 2002; Ferreira and Patson, 2007).

Variability Across Studies

The current study showed that PP attractors trigger interference, but direct object and subject attractors do not. These results appear to be at odds with previous demonstrations of attraction that have used subject and object attractors. For instance, both Clifton et al. (1999) and Dillon et al. (2013) observed attraction from items that appeared in a direct object position, and Wagers et al. (2009) observed attraction from items in a matrix subject position. However, a closer examination of these contexts reveals critical differences that may explain why we see different profiles across studies.

For instance, a survey of the full materials list from Dillon et al. (2013) showed that a combination of both direct object and PP attractors was used in their study. It is possible that the attraction effects that they observed were triggered by the PP attractors, as shown in the current study. In Clifton et al. (1999) and Wagers et al. (2009), the attractors appeared in a subject or direct object position as the head of an object relative clause that contained the critical verb, e.g., *The musicians*$_{PL}$ *who the reviewer*$_{SG}$ *praise*$_{PL}$ . . . or *Lucine dislikes the people*$_{PL}$ *who the manager*$_{SG}$ *think*$_{PL}$ . . . In these configurations, the attractor must be retrieved at the verb anyway, independently of subject-verb agreement processing, to thematically bind the attractor as the object of the verb. It is possible that sensitivity to the number-matching attractor reflects the fact that multiple retrieval processes are triggered by the main agreeing verb, one of which targets the attractor. On this view, retrieval of the plural attractor as the object of the verb might give comprehenders the false impression that subject agreement is also licensed. No such effect is expected in the current study, as the critical verb always targeted the same item, namely the head noun of the main clause subject. A task for future research is to better understand how retrieval for agreement processing and thematic binding interact when they are triggered by the same verb.

Extensions to Other Dependencies

Van Dyke and McElree (2011) showed that core and oblique arguments differ with respect to interference during retrieval for thematic binding, and the current study extends those results by showing that the contrast generalizes to subject-verb agreement dependencies. These results raise the question of whether other dependencies should show similar effects. The evidence thus far is inconclusive, warranting further research.

One dependency that is ripe for investigation involves reflexive licensing. The leading consensus is that retrieval for reflexive licensing resists interference from all non-target items (see Dillon, 2014, for a review), except in specific configurations when the target subject provides a particularly poor match to the retrieval cues (Parker and Phillips, 2017). The majority of the existing studies on retrieval for reflexive licensing have tested attractors that appeared as core arguments (e.g., subjects and direct objects). To the best of our knowledge, there has only been one study that tested whether oblique arguments trigger interference for reflexives. Andrews et al. (2016) tested sentences

like *The motherly therapist(s) of the widow(s) eventually reassured themselves* . . . and found weak evidence of attraction from the number matching attractor *the widows* embedded in a PP. The findings from these studies provide some support for the current proposal that oblique argument attractors trigger interference, but core arguments do not. However, more systematic research is necessary to determine whether other dependencies pattern similarly. We leave investigation of this issue to future work.

CONCLUSION

The current study showed that oblique arguments triggered interference during retrieval for subject-verb agreement processing, but core arguments did not. These results were presented as evidence that retrieval interference is dependent on the distinctiveness of the items in memory according to an argument hierarchy. These effects might be a general property of verbal dependencies, including subject-verb agreement and thematic binding (e.g., Van Dyke and McElree, 2011). Taken together, these results shed new light on the principles that govern the accessibility of information in working memory, and show that interference effects are informative not only about retrieval mechanisms, but also about the nature of the encoding mechanisms.

ETHICS STATEMENT

The protocol was approved by the William & Mary Protection of Human Subjects Committee. All subjects gave written informed consent.

AUTHOR CONTRIBUTIONS

DP was responsible for the design and analysis of Experiments 1–2. DP and AA jointly designed and analyzed Experiment 3. The writing was done by DP, with contributions from AA.

FUNDING

This work was supported in part by a William & Mary Faculty Research Grant awarded to DP.

ACKNOWLEDGMENTS

We are grateful to Jeff Runner, Ian Cunnings, and Brian Dillon for their feedback and suggestions on our presentation of this work at the 2017 CUNY Conference. We would also like to thank the members of the William & Mary Computational & Experimental Linguistics Lab (CELL), Luiza Newlin-Lukowicz, and the two reviewers for their useful feedback on the material presented here.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01566/full#supplementary-material

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychol. Rev.* 111, 1036–1060. doi: 10.1037/0033-295X.111.4.1036

Anderson, J. R., and Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Andrews, C., Yacovone, A., Slogget, S., and Dillon, B. (2016). *Reflexives: we don't see the attraction. Paper Presented at the Poster at Architectures and Mechanisms of Language Process (AMLaP) 22*, Bilbao, Spain.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* 4:328. doi: 10.3389/fpsyg.2013.00328

Bresnan, J. (2001). *Lexical functional Syntax*. Oxford: Blackwell.

Chomsky, N. (1981). *Lectures on Government and Binding*. Berlin: Mounten de Gruyter.

Clifton, C. J., Frazier, L., and Deevy, P. (1999). Feature manipulation in sentence comprehension. *Riv. Ling.* 11, 11–39.

Cohen, J. (1962). The statistical power of abnormal social psychological research: a review. *J. Abnor. Soc. Psychol.* 65, 145–153. doi: 10.1037/h0045186

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

Culicover, P. W., and Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199271092.001.0001

Cunnings, I., and Felser, C. (2013). The role of working memory in the processing of reflexives. *Lang. Cognit. Process.* 28, 188–219. doi: 10.1080/01690965.2010.548391

Cunnings, I., and Sturt, P. (2018). Retrieval interference and semantic interpretation. *J. Mem. Lang.* 102, 16–27. doi: 10.1016/j.jml.2018.05.001

Dillon, B. (2014). Syntactic memory in comprehension of reflexive dependencies: an overview. *Lang. Ling. Compass* 8, 171–187. doi: 10.1111/lnc3.12075

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Drummond, A. (2018). *Ibex Farm*. Available at: http://spellout.net/ibexfarm

Engelmann, F., Jäger, L., and Vasishth, S. (2015). Cue confusion and distractor prominence can explain inconsistent interference effects. *Paper Presented at the 28th CUNY Conference on Human Sentence Processing*, University of Southern California, Los Angeles, CA.

Ferreira, F., Bailey, K. G. D., and Ferraro, V. (2002). Good-enough representations in language comprehension. *Curr. Direct. Psychol. Sci.* 11, 11–15. doi: 10.1111/1467-8721.00158

Ferreira, F., and Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Lang. Ling. Compass* 1, 71–83. doi: 10.1111/j.1749-818X.2007.00007.x

Franck, J., Colonna, S., and Rizzi, L. (2015). Task-dependency and structure-dependeny in number interference effects in sentence comprehension. *Front. Psychol* 6:349. doi: 10.3389/fpsyg.2015.00349

Franck, J., Lassi, G., Frauenfelder, U., and Rizzi, L. (2006). Agreement and movement: a syntactic analysis of attraction. *Cognition* 101, 173–216. doi: 10.1016/j.cognition.2005.10.003

Franck, J., Soare, G., Frauenfelder, U., and Rizzi, L. (2010). Object interference in subject-verb agreement: the role of intermediate traces of movement. *J. Mem. Lang.* 62, 166–182. doi: 10.1016/j.jml.2009.11.001

Frazier, L., and Clifton, C. J. (1996). *Construal*. Cambridge, MA: MIT Press.

Gelman, A., and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol. Learn. Mem. Cognit.* 26, 1411–1423. doi: 10.1037/0278-7393.27.6.1411

Hammerly, C., Staub, A., and Dillon, B. (2018). *The Grammatical Asymmetry in Agreement Attraction Reflects Response Bias: Experimental and Modeling Evidence*. Amherst, MA: University of Massachusetts Amherst.

Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.004

Karimi, H., and Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quart. J. Exp. Psychol.* 69, 1013–1040. doi: 10.1080/17470218.2015.1053951

Keenan, E. L., and Comrie, B. (1977). Noun phrase accessibility and universal grammar. *Ling. Inq.* 8, 63–99.

Kush, D., Lidz, J., and Phillips, C. (2015). Configuration-sensitive retrieval: resisting interference in processing bound variable pronouns. *J. Mem. Lang.* 82, 18–40. doi: 10.1016/j.jml.2015.02.003

Kush, D., Lidz, J., and Phillips, C. (2017). Looking forwards and backwards: the real-time processing of Strong and Weak Crossover. *Glossa* 2, 1–29. doi: 10.5334/gjgl.280

Kush, D., and Phillips, C. (2014). Local anaphor licensing in an SOV language: implications for retrieval strategies. *Front. Psychol.* 5:1252. doi: 10.3389/fpsyg.2014.01252

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2014). *lmerTest: Tests for Random and Diexed Effects for Linear Mixed Effect Models (Lmer Objects of lme4 package*. Available at: http://CRAN.R-project.org/package=lmerTest

Lago, S., Shalom, D., Sigman, M., Lau, E., and Phillips, C. (2015). Agreement attraction in Spanish comprehension. *J. Mem. Lang*. 82, 133–149. doi: 10.1016/j.jml.2015.02.002

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognit. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cognit. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

Martin, A. E., Nieuwland, M. S., and Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *NeuroImage* 59, 1859–1869. doi: 10.1016/j.neuroimage.2011.08.057

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *J. Psycholinguist. Res.* 29, 155–200. doi: 10.1023/A:1005184709695

McElree, B. (2006). "Accessing recent events," in *The Psychology of Learning and Motivation – Adcances in Research and Theory*, ed. B. H. Ross (San Diego, CA: Academic Press), 155–200.

McElree, B., Foraker, S., and Dyer, L. (2003). Memory structures that subserve sentence comprehension. *J. Mem. Lang.* 48, 67–91. doi: 10.1016/S0749-596X(02)00515-6

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Mem. Cognit.* 16, 343–352. doi: 10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Mem. Cognit.* 18, 251–269. doi: 10.3758/BF03213879

Oppenheimer, D. M., Meyvis, T., and Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* 45, 867–872. doi: 10.1016/j.jesp.2009.03.009

Parker, D., Lago, S., and Phillips, C. (2015). Interference in the processing of adjunct control. *Front. Psychol.* 6:1346. doi: 10.3389/fpsyg.2015.01346

Parker, D., and Phillips, C. (2016). Negative polarity illusions and the format of hierarchical encodings in memory. *Cognition* 157, 321–339. doi: 10.1016/j.cognition.2016.08.016

Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002

Patil, U., Vasishth, S., and Lewis, R. (2016). Retrieval interference in syntactic processing: the case of reflexive binding in english. *Front. Psychol.* 7:.329 doi: 10.3389/fpsyg.2016.00329

Pearlmutter, N., Garnsey, S., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. New York, NY: Longman.

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org

Sloggett, S. (2013). "Case licensing in processing: evidence from German," in *Poster at the 26th CUNY Conference on Human Sentence Processing*, University of South Carolina, Carolina, Columbia, SC.

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition* 114, 447–454. doi: 10.1016/j.cognition.2009.11.003

Tabor, W., Galantucci, B., and Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *J. Mem. Lang.* 50, 355–370. doi: 10.1016/j.jml.2004.01.001

Tanner, D., Nicol, J., and Brehm, L. (2014). The time-course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *J. Mem. Lang.* 76, 195–215. doi: 10.1016/j.jml.2014.07.003

Townsend, D. J., and Bever, T. G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*. Cambridge, MA: MIT Press.

Tucker, M. A., and Almeida, D. (2017). "The complex structure of agreement errors: evidence from distributional analyses of agreement attraction in arabic," in *Proceedings of the 47th Meeting of the North East Linguistics Society*, eds A. Lamont and K. Tetzloff (Amherst, MA: GLSA), 45–54. doi: 10.3389/fpsyg.2015.00347

Tucker, M. A., Idrissi, A., and Almeida, D. (2015). Representing number in the real-time processing of agreement: self-paced reading evidence from Arabic. *Front. Psychol.* 6:347. doi: 10.3389/fpsyg.2015.00347

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol. Learn. Mem. Cognit.* 33, 407–430. doi: 10.1002/lnc3.330

Van Dyke, J. A., and Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Lang. Ling. Compass* 6, 193–211. doi: 10.1016/S0749-596X(03)00081-0

Van Dyke, J. A., and Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *J. Mem. Lang.* 49, 285–316. doi: 10.1016/j.jml.2006.03.007

Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2011.05.002

Van Dyke, J. A., and McElree, B. (2011). Cue-dependent interference in comprehension. *J. Mem. Lang.* 65, 247–263. doi: 10.1017/CBO9781139166799

Van Valin, R. D., and LaPolla, R. J. (1997). *Syntax: Structure, Meaning, & Function*. Cambridge: Cambridge University Press. doi: 10.1080/03640210802066865

Vasishth, S., Brüssow, S., Lewis, R. L., and Drenhaus, H. (2008). Processing polarity: how the ungrammatical intrudes on the grammatical. *Cognit. Sci.* 32, 685–712. doi: 10.1111/lnc3.12201

Vasishth, S., and Nicenboim, B. (2016). Statistical methods for linguistic research: foundational ideas - part 1. *Lang. Ling. Compass* 10, 349–369.

Villata, S., Tabor, W., and Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: evidence from agreement. *Front. Psychol.* 9:2. doi: 10.3389/fpsyg.2018.00002

Wagers, M. W., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Watkins, O. C., and Watkins, M. J. (1975). Build-up of proactive inhibition as a cue overload effect. *J. Exp. Pscyhol. Hum. Learn. Mem.* 104, 442–452. doi: 10.1037/0278-7393.1.4.442

Witzel, N., Witzel, J., and Forster, K. (2012). Comparisons of online reading paradigms: eye tracking, moving-window, and maze. *J. Psycholinguist. Res.* 41, 421–438. doi: 10.1007/s10936-011-9179-x

Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain Lang.* 108, 40–55. doi: 10.1016/j.bandl.2008.10.002

Xiang, M., Grove, J., and Giannakidou, A. (2013). Dependency dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Front. Psychol.* 4:708. doi: 10.3389/fpsyg.2013.00708

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 14 January 2019
doi: 10.3389/fpsyg.2018.02754

A Review on Grammatical Gender Agreement in Speech Production

Man Wang[1]* *and Niels O. Schiller*[2,3]

[1] School of Foreign Languages, Qingdao University, Qingdao, China, [2] Leiden University Center for Linguistics, Leiden, Netherlands, [3] Leiden Institute for Brain and Cognition, Leiden, Netherlands

Grammatical gender agreement has been well addressed in language comprehension but less so in language production. The present article discusses the arguments derived from the most prominent language production models on the representation and processing of the grammatical gender of nouns in language production and then reviews recent empirical studies that provide some answers to these arguments.

Keywords: grammatical gender, agreement, lexico-syntactic feature, speech production, ERP

OPEN ACCESS

Edited by:
Sendy Caffarra,
Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Cristina Cacciari,
Università degli Studi di Modena e Reggio Emilia, Italy
Vera Kempe,
Abertay University, United Kingdom

***Correspondence:**
Man Wang
emilymanwang@163.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *12 October 2018*
Accepted: *20 December 2018*
Published: *14 January 2019*

Citation:
Wang M and Schiller NO (2019) A Review on Grammatical Gender Agreement in Speech Production. Front. Psychol. 9:2754. doi: 10.3389/fpsyg.2018.02754

INTRODUCTION

In order to successfully convey a message when speaking, speakers need to encode the to-be-produced speech in a grammatically correct way. Language systems differ in terms of whether or not grammatical gender is distinguished in the systems. Some language systems do not distinguish the grammatical gender of nouns, such as English and Chinese. Some other language systems (e.g., Romance languages, German, Dutch, and Russian, but also many non-Indo-European languages) distinguish nouns according to their grammatical gender (e.g., masculine versus feminine, common versus neutral). Very often, the grammatical gender of the nouns bears an opaque relation to the biological gender of its referent (i.e., the conceptual or natural gender; see Schiller and Caramazza, 2003; Schwichtenberg and Schiller, 2004).

Grammatical gender agreement is a crucial part of syntactic agreement within a noun phrase and within a sentence (e.g., in Spanish: 'La rosa es roja.' The$_{fem}$ rose$_{fem}$ is red$_{fem}$). It is stored in the mental lexicon as a lexico-syntactic feature of words (see Levelt et al., 1999a; Nickels et al., 2015). Unlike the feature 'number,' which always needs to be activated based on the concept (e.g., 'one cat' or 'two cats'; see Schiller and Caramazza, 2002) and requires the selection of the *–s* suffix in English for regular plural nouns (Nickels et al., 2015), 'gender' is an intrinsic feature of nouns (Corbett, 1991). Its activation has little to do with the concept and does not always have morphological or phonological consequences. For instance, in Romance languages such as Italian and Spanish, nouns' suffixes are morphologically and phonologically marked by the grammatical gender, although the gender-to-ending correspondence is not always transparent (see, e.g., Padovani et al., 2005).

Psycholinguistic models of language production have made distinctive assumptions about the representation and processing of grammatical gender in speech production. For instance, the WEAVER++ model distinguishes a conceptual stratum, a syntactic stratum and a word-form stratum (Levelt, 1992; Roelofs, 1992, 1993; Roelofs and Meyer, 1998; Levelt et al., 1999a,b) and words are linked to their syntactic features (i.e., grammatical gender, grammatical class, and number) at the syntactic stratum (see Levelt et al., 1999a; Nickels et al., 2015). This model distinguishes between the activation and selection of the syntactic features. Specifically, the grammatical gender is only selected when it is needed for production (Roelofs, 1992, 1993). The WEAVER++ model assumes the seriality of processing stages and a unidirectional link from a word to its syntactic features (see also, Jescheniak and Levelt, 1994). By contrast, although

constructed with the same layered architecture, the 'interactive' spreading-activation model (Dell, 1986, 1988, 1990; Dell and O'Seaghdha, 1991, 1992) assumes an interactive manner of activation flow. In other words, the links between layers are bi-directional. Alternatively, the 'Independent-Network' model (Caramazza, 1997; Caramazza and Miozzo, 1997) assumes three independent networks: the lexical-semantic network, the syntactic network and the phonological lexemes. In this latter model, the lexical-semantic network can directly activate the syntactic network and the phonological lexemes in parallel. Please note that both Dell's interactive model and the Independent-Network model reject the seriality and discreteness of activation flow and in principle allow the bypassing of the retrieval of grammatical gender to specify the phonological form of noun phrases when the grammatical gender of the nouns is not explicitly marked in their phonological forms (see Schriefers and Jescheniak, 1999 for a discussion).

There have been heated debates over the underlying mechanism of the selection of freestanding and bound gender-marking morphemes in speech production (see Jescheniak et al., 2014 for a thorough review). Jescheniak et al. (2014) reviewed empirical evidence and concluded that both gender-marked freestanding morphemes like determiners and bound morphemes like adjectival inflections are selected by competition at the phonological level in speech production (but see, Schiller and Costa, 2006). Compared to the review by Jescheniak et al. (2014), which focuses on the gender-marked morphemes, our review focuses on the activation and selection of the abstract gender features of the noun during speech production. Two major questions arise from the assumptions of the three most prominent language production models. The first one is whether or not grammatical gender is automatically activated and selected in speech production even when it is not explicitly needed for speech production. The second one is whether grammatical gender can be bypassed when the phonological form can be generated without knowing its gender. We will discuss empirical evidence on these arguments.

Empirical studies have collected evidence from speech errors as well as error-free speech. Studies that analyze speech errors give hints on the representation and processing of grammatical gender in speech production (see Schriefers and Jescheniak, 1999 for a thorough review). For example, German noun substitution errors show that the intended and intruded nouns were often of the same gender and this phenomenon occurs even without syntactic cues, consistent with a two-stage language production model (Marx, 1999). Evidence from Tip-of-the-Tongue (TOT) errors demonstrates that speakers can access to grammatical gender when no phonological cues are available, suggesting separate representations of lexico-syntactic features and phonological forms (Vigliocco et al., 1997; but see also Caramazza and Miozzo, 1997; Miozzo and Caramazza, 1997 in Italian; Gonzalez and Miralles, 1997 in Spanish; cf. Schriefers and Jescheniak, 1999, p. 589). Furthermore, studies of anomia and TOT states in Italian, Spanish, French and German where a noun is usually produced in a full NP (e.g., with a gender-marking determiner) show that patients have gender knowledge when they fail to name (e.g., Badecker et al., 1995; Vigliocco et al., 1996; Marx, 1999; cf. Friedmann and Biran, 2003). By contrast, Hebrew-speaking aphasic patients do not preserve grammatical gender in bare noun naming (Friedmann and Biran, 2003). However, for most speech error studies that investigate the grammatical gender representation and processing in language production, the results fail to give clear conclusions on lexical access in language production under error-free circumstances (for naturally occurring speech errors, see, e.g., Barbaud et al., 1982; Berg, 1992; Vigliocco et al., 1997; for experimentally elicited speech errors, see, Meyer and Bock, 1999; Vigliocco and Franck, 1999; see Schriefers and Jescheniak, 1999 for a detailed review on studies analyzing speech errors). Therefore, this article will focus on discussing studies that analyze error-free speech.

BEHAVIORAL STUDIES

Experimental studies have made use of the picture-word interference paradigm (PWI) to investigate the processing of syntactic features in speech production. The PWI paradigm (e.g., Glaser, 1992; see MacLeod, 1991 for a review) has been widely used to examine the language production process. Schriefers (1993) presented colored pictures to participants while a distractor word whose grammatical gender was either congruent or incongruent with that of the target picture was superimposed on the picture. Participants were asked to name the target pictures using noun phrases while ignoring the distractors. The experiment was conducted in Dutch with native Dutch speakers. In Dutch, there are two grammatical gender categories: neutral and common gender. The results of the study showed that participants were faster in naming the pictures when the grammatical gender of the distractor word (e.g., 'dak,' $\text{roof}_{\text{neuter}}$) was congruent with that of the target picture name (e.g., 'boek,' $\text{book}_{\text{neuter}}$) than an incongruent condition with a distractor (e.g., 'tafel,' $\text{table}_{\text{common}}$). This was also true with both article-adjective-noun (e.g., 'het groene boek,' the green book) and plain adjective-noun (e.g., 'groene boek,' green book) productions. The difference in naming latencies was called 'the gender congruency effect' and this effect was also observed in definite article-noun production (e.g., 'de tafel,' the table) in Dutch (van Berkum, 1997; La Heij et al., 1998; Schiller and Caramazza, 2003; Starreveld and La Heij, 2004; Schiller, 2013), in noun phrase naming in German (Schriefers and Teruel, 2000; Schiller and Caramazza, 2003), Chinese (Wang et al., 2006; Zhang and Liu, 2009), Konso (Tsegaye, 2017; Tsegaye et al., unpublished), Croatian (Costa et al., 2003), and Czech (Bordag and Pechmann, 2008).

Schriefers (1993) claimed that the target word's grammatical gender feature (e.g., neuter) and the distractor's gender feature (e.g., common) compete for selection when they are incongruent. The competition in the selection of the word's grammatical gender causes interference when producing the target noun phrase. This account has been called the *gender selection interference hypothesis* (GSIH; Schiller and Caramazza, 2003, 2006). This hypothesis assumes the selection of the grammatical gender. Another study by Schriefers and Teruel (1999) on French noun phrase production also showed the gender congruency effect even when the definite article and the

post-nominal adjective were identical for nouns of different grammatical genders (e.g., 'l'assiette jaune,' the$_{\text{fem or masc}}$ dish$_{\text{fem}}$ yellow$_{\text{fem or masc}}$) (but see Experiments 2 and 3 in Bordag and Pechmann, 2008). These findings, especially the latter one, suggest that the selection of grammatical gender cannot be bypassed, which runs against the interactive model and the Independent-Network model, both of which in principle allow such a bypass.

Nevertheless, conflicts have been found in later studies in various languages. The so-called gender congruency effect was not replicated in Italian definite article-noun phrase production (Miozzo and Caramazza, 1999; Cubelli et al., 2005) when the article is determined by both the grammatical gender (masculine versus feminine) and the phonological form (e.g., the onset) of the noun. Similar results were observed in other studies in Italian (Miozzo et al., 2002) and other Romance languages, such as Spanish, Catalan (Costa et al., 1999) and French (Alario and Caramazza, 2002; see Caramazza et al., 2001 for a review). Miozzo and Caramazza (1999) attributed the discrepancy to cross-linguistic differences in the selection of determiners. In Dutch, the selection of determiners depends on the noun's gender and number features, whereas the determiner selection in Italian also depends on the phonological form of the subsequent word. Furthermore, Schiller and Caramazza (2003) asked German and Dutch speakers to name pictures using "determiner and/or adjective" single or plural noun phrases. In German and Dutch, determiners are identical if the nouns are in plural forms. The so-called gender congruency effect was only obtained when to-be-named pictures were in singular forms, not in plural forms when the determiner was identical for all genders. The gender congruency effect was then interpreted as reflecting the competition in the selection of determiner forms, i.e., the *determiner selection interference hypothesis* (DSIH) (see also Schiller and Caramazza, 2006). These findings suggest that the selection of grammatical gender can be bypassed if its information is not necessary to determine the phonological form of the to-be-produced speech (see Jescheniak et al., 2014 for a detailed review over the selection of gender-marked morphemes in speech production). However, these results do not answer directly whether or not the grammatical gender feature is automatically activated when it does not have any phonological consequences.

Discrepancies were also observed in bare noun naming. No gender or determiner congruency effect was observed in bare noun naming in Dutch (La Heij et al., 1998; Starreveld and La Heij, 2004). In a Greek (L1) to German (L2) translation task, the gender congruency effect was only observed in noun phrases when the target utterance required gender agreement (Salamoura and Williams, 2007), although gender information in L2 is assumed to be computed anew during production rather than stored as a fixed feature in L1 (Bordag and Pechmann, 2007). By contrast, Cubelli et al. (2005) observed the grammatical gender interference effect in Italian bare noun production even when grammatical gender is not necessary for producing the target (but see also Finocchiaro et al., 2011). The gender congruency effect in bare noun naming was also found in Konso in a study by Tsegaye et al. (2013), Tsegaye (2017), Tsegaye et al. (unpublished) and in Czech where the congruency effect was shown with a comparable feature, i.e., declensional class (Bordag and Pechmann, 2009). Cubelli et al. (2005) concluded that the grammatical gender is selected even in bare noun production. The grammatical gender effect was observed both when the gender-to-ending correspondence is transparent (i.e., *-a* for feminine and *-o* for masculine) and when it is opaque (i.e., *-e* for either feminine or masculine). Paolieri et al. (2010, 2011) replicated this effect in both Italian and another Romance language, Spanish, which has an analogous gender system. Paolieri et al. (2011) extended the previous finding in that differential effects were observed when the morphological transparency of the ending vowel for gender varied. For instance, for the target word 'trattore' (tractor$_{masc}$), the gender congruency effect was stronger when the distractors had the same ending *-e* (e.g., 'peperone,' pepper$_{masc}$ vs. 'cicatrice,' scar$_{fem}$) in contrast to different endings (e.g., 'cappello,' hat$_{masc}$ vs. 'batteria,' drums$_{fem}$). Emerging evidence shows that in Romance languages such as Italian and Spanish, the selection of grammatical gender is not bypassed and the grammatical gender effect is related to the gender-to-ending transparency (Paolieri et al., 2011).

It seems that grammatical gender plays a crucial role in accessing the phonological form of the noun which may contribute to the selection of grammatical gender in bare noun production in Romance languages. Cubelli et al. (2005) proposed a Double Selection model, in which a word's lemma is linked to a semantic category node and a grammatical gender node in a two-layered structure. In spoken word production, both the lexico-semantic representation and the lexico-syntactic representation have to be selected prior to accessing the phonological form at the second layer. According to Cubelli et al. (2005), the discrepancy between the findings in Dutch and Italian bare noun productions is attributed to language-specific properties. The compulsory selection of grammatical gender is only present in languages with a complex morphological structure such as Italian, and can be bypassed in languages with a relatively simple morphological structure such as Dutch.

The Double Selection model proposed by Cubelli et al. (2005) is in line with the WEAVER++ model in that the grammatical gender information is accessed prior to the word's phonological form. Nevertheless, it disagrees with the WEAVER++ model by assuming a direct link between the semantic representation and the phonological representation. This, however, is in line with the prediction of the IN model and allows the bypass of grammatical gender selection in bare noun production as observed in Dutch. Furthermore, in contrast to the prediction of the IN model, the Double Selection model assumes the compulsory competition in the selection of grammatical gender as reflected by the grammatical gender effect in Italian bare noun production. This does not fully contradict the conjecture of the WEAVER++ model which assumes that the grammatical gender feature is activated but not selected if it is not needed for production (Roelofs, 1992, 1993) since the Double Selection model restricts the compulsory selection of the grammatical gender information to languages with a complex morphological structure. Nevertheless, whether the grammatical gender feature is automatically activated or not is still open to debate.

Unfortunately, the existing behavioral data cannot provide evidence for resolving this debate.

ELECTROPHYSIOLOGICAL STUDIES

In contrast with behavioral data, such as naming latencies which only reflect the outcome of the speech production process, electrophysiological data can provide fine-grained measurements of online processing of the speech production process (Luck, 2005). However, electrophysiological studies investigating the grammatical gender processing in language production are scarce. Van Turennout et al. (1998) measured the Lateralized Readiness Potentials (LRPs) in two versions of a combined forced choice task and go/no-go task and showed that the retrieval of grammatical gender feature precedes the retrieval of the phonological form information. Another study by Barber and Carreiras (2005) showed that grammatical gender disagreement elicited an N400 effect in (silent) sentence reading in Spanish.

In order to test whether lexico-syntactic features are activated and selected in bare noun production, Wang et al. (2018) investigated the Chinese language, where grammatical gender is not marked but nouns have a comparable lexico-syntactic feature, i.e., classifiers. It is compulsory to use a classifier between an article, a quantifier or another modifier and its associated noun (e.g., 'yi1 pi3 ma3,'[1] one *classifier-pi3* horse). Chinese classifiers bare a transparent semantic relationship to the noun but opaque in other cases (Tzeng et al., 1991) and are considered to have some functions of determiners in other languages (Cheng and Sybesma, 2005). Using the PWI paradigm, the authors asked participants to name the target picture in bare nouns with a distractor that was either classifier-congruent or -incongruent with that of a target picture. A stronger N400 effect was observed on the classifier-incongruent trials compared to the congruent trials (both semantically unrelated), suggesting the automatic activation of classifier feature in bare noun production. By contrast, no effect in naming latencies was observed between the classifier-congruent and -incongruent conditions, suggesting that the classifier feature is not selected in the process of bare noun production when it is not needed. The bypass of the selection of classifier feature is compatible with the hypothesis by Cubelli et al. (2005) given that Chinese has a very simple morphological structure. The findings are also compatible with the assumption by the WEAVER++ model that the lexico-syntactic feature is automatically activated but not selected in language production when it's not needed (Roelofs, 1992, 1993). Nevertheless, it is yet unclear to what extent these findings can be generalized to other language systems, especially those that distinguish the grammatical gender of nouns.

A few studies have investigated the processing of grammatical gender agreement in sentence comprehension. Molinaro et al. (2011) reviewed nine studies examining the neural correlates of either determiner-noun or noun-adjective gender mismatches. It has been observed that the N2pc component was modulated in a grammatical gender agreement task in Italian word pairs whose gender is transparently marked (Caffarra et al., 2013). The involvement of gender-to-ending is also shown in the investigation of language comprehension. Gender-to-ending transparency is shown to modulate grammatical gender effect in the gender categorization task. Specifically, the gender congruency effect was observed in morphologically complex words and even in pseudo-morphological words but not in nouns without morpheme-like parts (Meunier et al., 2008). In the following discussion on neural imaging evidence, it is suggested that language perception and production share a common neural network for grammatical gender processing (Heim et al., 2002; Miceli et al., 2002).

FUNCTIONAL MAGNETIC RESONANCE IMAGING (fMRI) STUDIES

Alongside the ongoing debate about whether grammatical gender is selected even when it is not needed for production investigated mainly with the behavioral measurements, researchers also investigated the neural correlates of grammatical gender retrieval using fMRI. Distinctive neural mechanisms seem to underlie the processing of the grammatical gender at different levels (Heim et al., 2002). While syntactic processing at the sentence level involves pronounced activation in the inferior part of Broca's area (e.g., Friederici et al., 2000; Indefrey et al., 2001), the selection of grammatical gender is correlated with the activation in the superior part of the Broca's area when participants were producing determiners (Heim et al., 2002; more specifically Brodmann's Area, (BA) 44, see Heim et al., 2009) or identifying the grammatical gender of a given word (Miceli et al., 2002). The superior part of Broca's area is found to be activated in both comprehension and production tasks, suggesting a common neural network for grammatical gender processing in language perception and production (Heim et al., 2002; Miceli et al., 2002).

Although the activation associated with accessing the grammatical gender information is located in the Broca's area – i.e., BA 44/45 – the focus of the activation varies depending on participants' processing strategies (Heim et al., 2005). Specifically, the direct access to gender information when performing the gender judgment features a network involving the inferior tip of BA 44. Alternatively, when participants adopt an indirect, form-related strategy, i.e., producing the definite determiner in order to judge the grammatical gender of the given word, they demonstrate a network of activation in BA 45/47, the superior part of BA 44 and the fronto-median wall (Heim et al., 2002, 2005, 2009; Miceli et al., 2002). The distinctive foci of networks were in line with a dual-route model for the retrieval of grammatical gender proposed by Gollan and Frost (2001) based on their behavioral study, with one route of direct grammatical gender access and the other being more form-based. Gollan and Frost (2001) also pointed out that the cross-linguistic variability in grammatical gender-marking may lead to variance in the speed and availability of the form-based route to grammatical gender. The influence of gender-marking regularity is confirmed by another fMRI study, showing activation in the

[1] As an example, "yi" indicates the phonetic notation of the lexical item, i.e., Pinyin of the word and the number 1 indicates the Lexical Tone 1.

left and right fronto-temporal areas (Padovani et al., 2005). By varying the gender-to-ending regularity of Italian words, the authors observed a complex activation network and suggested a lexically based route for words with "opaque" and "irregular" gender-to-ending correspondences and a form-based route for "transparent" words.

Emerging evidence suggests the importance of gender-to-ending regularity and the transparency of gender-marking. Furthermore, the distinctive routes of grammatical gender retrieval may result from the variability within and across languages in these two factors.

SUMMARY

The empirical studies discussed in the present article have investigated the representation and processing of grammatical gender or a similar lexico-syntactic feature in language production. It is generally agreed that grammatical gender is represented as a separate lexico-syntactic feature in the mental lexicon.

However, several issues still remain unsolved concerning the processing of grammatical gender in language production. Firstly, it seems that grammatical gender is not selected in bare noun production when it is not necessary for production in Dutch and Chinese but is selected in Italian and Konso (Tsegaye et al., 2013; Tsegaye, 2017; Tsegaye et al., unpublished) bare noun production. Further evidence is needed to confirm Cubelli et al. (2005)'s argument that the discrepancy is attributed to the complexity of morphological structure of the target language. Using another language other than Italian and Konso that has a complex morphological structure would illuminate this matter. Secondly, the study in Chinese provides evidence for the automatic activation of the lexico-syntactic feature, i.e., classifier, in bare noun production. To our knowledge, no direct evidence has been drawn to test whether it is the same with the grammatical gender feature. Thirdly, few studies have looked into the manner of activation flow between a word and its syntactic feature to determine when and how the lexico-syntactic feature is activated in language production.

Furthermore, it is still open to debate whether the selection of grammatical gender is bypassed in noun phrase production when the selection of grammatical gender does not have any phonological consequence. Nevertheless, emerging evidence has shown distinctive mechanisms underlying the selection of grammatical gender in Romance languages like Italian and Spanish, and Germanic languages like German and Dutch. For instance, the grammatical gender congruency effect in bare noun production was observed in Italian but not in German or Dutch; the determiner congruency effect was observed in German and Dutch but not in Romance languages (but see Schriefers and Teruel, 1999). fMRI studies also provide evidence for distinctive neural networks for the processing of grammatical gender and suggest that participants tend to adopt a more form-related route to access gender information in Romance languages where the gender-to-ending regularity modulates the gender effect. By contrast, participants tend to adopt a more lexically based route to access grammatical gender in Dutch and German where the noun's morpho-phonological form is generally not strongly marked by gender.

In sum, the present article reviewed recent empirical studies on the representation and processing of grammatical gender of nouns in language production. We may not have exhausted all relevant studies but the empirical evidence discussed above will provide reference in constructing the language production model.

AUTHOR CONTRIBUTIONS

Both authors have contributed equally to the manuscript and approved it for publication.

FUNDING

This research was supported by the Research Start-up Grant from Qingdao University.

ACKNOWLEDGMENTS

The authors thank the editor and the reviewers for the comments on the earlier versions of the manuscript.

REFERENCES

Alario, F.-X., and Caramazza, A. (2002). The production of determiners: evidence from French. *Cognition* 82, 179–223. doi: 10.1016/S0010-0277(01)00158-5

Badecker, W., Miozzo, M., and Zannuttini, R. (1995). The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. *Cognition* 57, 193–216. doi: 10.1016/0010-0277(95)00663-J

Barbaud, P. H., Ducharme, C. H., and Valois, D. (1982). D'un usage particulier du genre en canadien français: la feminisation des noms à initiate vocalique. *Can. J. Linguist.* 27, 103–133. doi: 10.1017/S0008413100023860

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Berg, T. (1992). Prelexical and postlexical features in language production. *Appl. Psycholinguist.* 13, 199–235. doi: 10.1017/S0142716400005567

Bordag, D., and Pechmann, T. (2007). Factors influencing L2 gender processing. *Biling. Lang. Cogn.* 10, 299–314. doi: 10.1017/S1366728907003082

Bordag, D., and Pechmann, T. (2008). Grammatical gender in speech production: evidence from Czech. *J. Psycholinguist. Res.* 37, 69–85. doi: 10.1007/s10936-007-9060-0

Bordag, D., and Pechmann, T. (2009). Externality, internality, and (in)dispensability of grammatical features in speech production: evidence from Czech declension and conjugation. *J. Exp. Psychol. Learn. Mem. Cogn.* 2, 446–465. doi: 10.1037/a0014874

Caffarra, S., Pesciarelli, F., and Cacciari, C. (2013). The interaction between language and visual spatial attention systems in grammatical gender processing. An N2pc study. *Cogn. Neurosci.* 4, 217–224. doi: 10.1080/17588928.2013.823392

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cogn. Neuropsychol.* 14, 177–208.

Caramazza, A., and Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: evidence from the 'tip-of-the-tongue' phenomenon. *Cognition* 64, 309–343. doi: 10.1016/S0010-0277(97)00031-0

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., and Alario, F.-X. (2001). "Across-linguistic investigation of determiner production," in *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, ed. E. Dupoux (Boston, MA: MIT Press), 209–226.

Cheng, L.-S., and Sybesma, R. (2005). "Classifiers in four varieties of Chinese," in *The Oxford Handbook of Comparative Syntax*, eds G. Cinque and R. S. Kayne (New York, NY: Oxford University Press), 259–292.

Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139166119

Costa, A., Kovacic, D., Fedorenko, E., and Caramazza, A. (2003). The gender congruency effect and the selection of freestanding and bound morphemes: evidence from Croatian. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1270–1282. doi: 10.1037/0278-7393.29.6.1270

Costa, A., Sebastián-Gallés, N., Miozzo, M., and Caramazza, A. (1999). The gender congruity effect: evidence from Spanish and Catalan. *Lang. Cogn. Process.* 14, 381–391. doi: 10.1080/016909699386275

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., and Job, R. (2005). Grammatical gender is selected in bare noun production: evidence from the picture word interference paradigm. *J. Mem. Lang.* 53, 42–59. doi: 10.1016/j.cortex.2018.05.014

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321. doi: 10.1037/0033-295X.93.3.283

Dell, G. S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *J. Mem. Lang.* 27, 124–142. doi: 10.1016/0749-596X(88)90070-8

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Lang. Cogn. Process.* 5, 313–349. doi: 10.1080/02643290802057311

Dell, G. S., and O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: a comment on Levelt et al. (1991). *Psychol. Rev.* 98, 604–614. doi: 10.1037/0033-295X.98.4.604

Dell, G. S., and O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition* 42, 287–314. doi: 10.1016/0010-0277(92)90046-K

Finocchiaro, C., Alario, F.-X., Schiller, N. O., Costa, A., Miozzo, M., and Caramazza, A. (2011). Gender congruency goes Europe: a cross-linguistic study of the gender congruency effect in Romance and Germanic languages. *Ital. J. Linguist.* 23, 161–198.

Friederici, A. D., Meyer, M., and von Cramon, D. Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang.* 74, 289–300. doi: 10.1006/brln.2000.2313

Friedmann, N., and Biran, M. (2003). When is gender accessed? A study of paraphasias in Hebrew anomia. *Cortex* 39, 441–463. doi: 10.1016/S0010-9452(08)70258-2

Glaser, W. R. (1992). Picture naming. *Cognition* 42, 61–105. doi: 10.1016/0010-0277(92)90040-O

Gollan, T. H., and Frost, R. (2001). Two routes to grammatical gender: evidence from Hebrew. *J. Psycholinguist. Res.* 30, 627–651. doi: 10.1023/A:1014235223566

Gonzalez, J., and Miralles, J. L. (1997). *La Informacion Sintatica Esta Disponible Durante un Estado de Punto de la Lengua [Syntactic Information is Available in a Tip-of-the-Tongue State]*. Castellon: Jaume I University.

Heim, S., Alter, K., and Friederici, A. D. (2005). A dual-route account for access to grammatical gender: evidence from functional MRI. *Anat. Embryol.* 210, 473–483. doi: 10.1007/s00429-005-0032-6

Heim, S., Friederici, A. D., Schiller, N. O., Rüschemeyer, S.-A., and Amunts, K. (2009). The determiner congruency effect in language production investigated with functional MRI. *Hum. Brain Mapp.* 30, 928–940. doi: 10.1002/hbm.20556

Heim, S., Opitz, B., and Friederici, A. D. (2002). Broca's area in the human brain is involved in the selection of grammatical gender for language production: evidence from event-related functional magnetic resonance image. *Neurosci. Lett.* 328, 101–104. doi: 10.1016/S0304-3940(02)00494-9

Indefrey, P., Brown, C. M., Hellwig, F., Amunts, K., Herzog, H., Seitz, R. J., et al. (2001). A neural correlate of syntactic encoding during speech production. *Proc. Natl. Acad. Sci. U.S.A.* 10, 5933–5936. doi: 10.1073/pnas.101118098

Jescheniak, J. D., and Levelt, W. J. (1994). Word frequency effects in speech production: retrieval of syntactic information and of phonological form. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 824–843. doi: 10.1037/0278-7393.20.4.824

Jescheniak, J. D., Schriefers, H., and Lemhöfer, K. (2014). Selection of freestanding and bound gender-marking morphemes in speech production: a review. *Lang. Cogn. Neurosci.* 29, 684–694. doi: 10.1080/01690965.2012.654645

La Heij, W., Mak, P., Sander, J., and Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychol. Res.* 61, 209–219. doi: 10.1007/s004260050026

Levelt, W. J. (1992). Accessing words in speech production: stages, processes and representations. *Cognition* 42, 1–22. doi: 10.1016/0010-0277(92)90038-J

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999a). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999b). Multiple perspectives on word production. *Behav. Brain Sci.* 22, 61–69. doi: 10.1017/S0140525X99451775

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychol. Bull.* 109, 163–203. doi: 10.1037/0033-2909.109.2.163

Marx, E. (1999). Gender processing in speech production: evidence from German speech errors. *J. Psycholinguist. Res.* 28, 601–621. doi: 10.1023/A:1023216927241

Meunier, F., Seigneuric, A., and Spinelli, E. (2008). The morpheme gender effect. *J. Mem. Lang.* 58, 88–99. doi: 10.1016/j.jml.2007.07.005

Meyer, A. S., and Bock, K. (1999). Representations and processes in the production of pronouns: some perspectives from dutch. *J. Mem. Lang.* 41, 281–301. doi: 10.1006/jmla.1999.2649

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., and Caramazza, A. (2002). The neural correlates of grammatical gender: an fMRI investigation. *J. Cogn. Neurosci.* 14, 618–628. doi: 10.1162/08989290260045855

Miozzo, M., and Caramazza, A. (1997). The retrieval of lexical-syntactic features in tip-of-the-tongue states. *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 1–14. doi: 10.1037/0278-7393.23.6.1410

Miozzo, M., and Caramazza, A. (1999). The selection of determiners in noun phrase production. *J. Exp. Psychol. Learn. Mem. Cogn.* 25, 907–922. doi: 10.1037/0278-7393.25.4.907

Miozzo, M., Costa, A., and Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: a matter of stimulus onset asynchrony? *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 388–391. doi: 10.1037/0278-7393.28.2.388

Molinaro, N., Barber, H. A., and Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Nickels, L., Biedermann, B., Fieder, N., and Schiller, N. O. (2015). The lexical-syntactic representation of number. *Lang. Cogn. Neurosci.* 30, 287–304. doi: 10.1080/23273798.2013.879191

Padovani, R., Calandra-Buonaura, G., Cacciari, C., Benuzzi, F., and Nichelli, P. (2005). Grammatical gender in the brain: evidence from an fMRI study on Italian. *Brain Res. Bull.* 65, 301–308. doi: 10.1016/j.brainresbull.2004.11.025

Paolieri, D., Lotto, L., Leoncini, D., Cubelli, R., and Job, R. (2011). Differential effects of grammatical gender and gender inflection in bare noun production. *Br. J. Psychol.* 102, 19–36. doi: 10.1348/000712610X496536

Paolieri, D., Lotto, L., Morales, L., Bajo, T., Cubelli, R., and Job, R. (2010). Grammatical gender processing in romance languages: evidence from bare noun production in Italian and Spanish. *Eur. J. Cogn. Psychol.* 22, 335–347. doi: 10.1080/09541440902916803

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition* 42, 107–142. doi: 10.1016/0010-0277(92)90041-F

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: retrieval of verbs. *Cognition* 47, 59–87. doi: 10.1016/0010-0277(93)90062-Z

Roelofs, A., and Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 922–939. doi: 10.1037/0278-7393.24.4.922

Salamoura, A., and Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: evidence from Greek and German. *Biling. Lang. Cogn.* 10, 257–275. doi: 10.1017/s1366728907003069

Schiller, N. O. (2013). "Psycholinguistic approaches to the investigation of grammatical gender in speech production: an overview and new data," in *The Expression of Gender*, ed. G. C. Corbett (New York, NY: Mouton de Gruyter), 161–190.

Schiller, N. O., and Caramazza, A. (2002). The selection of grammatical features in word production: the case of plural nouns in German. *Brain Lang.* 81, 342–357. doi: 10.1006/brln.2001.2529

Schiller, N. O., and Caramazza, A. (2003). Grammatical feature selection in noun phrase production: evidence from German and Dutch. *J. Mem. Lang.* 48, 169–194. doi: 10.1016/S0749-596X(02)00508-9

Schiller, N. O., and Caramazza, A. (2006). Grammatical gender selection and the representation of morphemes: the production of Dutch diminutives. *Lang. Cogn. Process.* 21, 945–973. doi: 10.1080/01690960600824344

Schiller, N. O., and Costa, A. (2006). Different selection principles of freestanding and bound morphemes in language production. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1201–1207. doi: 10.1037/0278-7393.32.5.1201

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 841–850. doi: 10.1037/0278-7393.19.4.841

Schriefers, H., and Jescheniak, J. D. (1999). Representation and processing of grammatical gender in language production: a review. *J. Psycholinguist. Res.* 28, 575–600. doi: 10.3389/fpsyg.2014.00589

Schriefers, H., and Teruel, E. (1999). "The production of Noun Phrases; a cross-linguistic comparison of French and German," in *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, Austin, TX.

Schriefers, H., and Teruel, E. (2000). Grammatical gender in noun phrase production: the gender interference effect in German. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 1368–1377. doi: 10.1037/0278-7393.26.6.1368

Schwichtenberg, B., and Schiller, N. O. (2004). Semantic gender assignment regularities in German. *Brain Lang.* 90, 326–337. doi: 10.1016/S0093-934X(03)00445-0

Starreveld, P., and La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Lang. Cogn. Process.* 19, 677–711. doi: 10.1080/01690960444000061

Tsegaye, M. T. (2017). *Plural Gender: Behavioral Evidence for Plural as a Value of Cushitic Gender with Reference to Konso*. Doctoral dissertation, Leiden University, Leiden.

Tsegaye, M. T., Mous, M., and Schiller, N. O. (2013). "Plural as a value of Cushitic gender: evidence from gender congruency experiments in Konso (Cushitic)," in *The Expression of Gender*, ed. G. C. Corbett (Berlin: Mouton de Gruyter), 191–214.

Tzeng, O. J., Chen, S., and Hung, D. L. (1991). The classifier problem in Chinese aphasia. *Brain Lang.* 41, 184–202. doi: 10.1016/0093-934X(91)90152-Q

van Berkum, J. J. (1997). Syntactic processes in speech production: the retrieval of grammatical gender. *Cognition* 64, 115–152. doi: 10.1016/S0010-0277(97)00026-7

Van Turennout, M., Hagoort, P., and Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science* 280, 572–574. doi: 10.1126/science.280.5363.572

Vigliocco, G., Antonini, T., and Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychol. Sci.* 8, 314–317. doi: 10.1111/j.1467-9280.1997.tb00444.x

Vigliocco, G., and Franck, J. (1999). When sex and syntax go hand in hand: gender agreement in language production. *J. Mem. Lang.* 40, 455–478. doi: 10.1006/jmla.1998.2624

Vigliocco, G., Garrett, M. P., and Martin, R. C. (1996). "The availability of syntactic information during anomia," in *Paper Presented at the 37th Meeting of the Psychonomic Society*, Chicago, IL.

Wang, L., Guo, J., Bi, Y., and Shu, H. (2006). Classifier congruency effect in the production of noun phrases. *Stud. Psychol. Behav.* 4, 34–38.

Wang, M., Chen, Y., and Schiller, N. O. (2018). Lexico-syntactic features are activated but not selected in bare noun production: electrophysiological evidence from overt picture naming. *Cortex* doi: 10.1016/j.cortex.2018.05.014 [Epub ahead of print].

Zhang, J.-J., and Liu, H.-Y. (2009). The lexical access of individual classifiers in language production and comprehension. *Acta Psychol. Sin.* 41, 580–593. doi: 10.3724/SP.J.1041.2009.00580

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 14 February 2019
doi: 10.3389/fpsyg.2019.00211

Processing Differences Between Person and Number: A Theoretical Interpretation

Peter Ackema[1] *and Ad Neeleman*[2]*

[1] The University of Edinburgh, Edinburgh, United Kingdom, [2] University College London, London, United Kingdom

The literature on processing of person and number agreement contains some apparently contradictory results. On the one hand, some ERP studies do not find a qualitative difference between person and number when an agreeing verb does not match the features of its subject, the controller of the agreement relation (Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2016). On the other hand, an ERP study reported in Mancini et al. (2011b) did find a qualitative difference between agreement violations in person and agreement violations in number, a result further corroborated by an fMRI study reported in Mancini et al. (2017). At the same time, there is also a trend on which the literature appears to agree: on the whole the response to agreement violations in person is stronger than the response to number agreement violations. In this paper we argue that the constellation of reported results can be accounted for by adopting a theory of person and number features that has the following two core properties: (i) pronouns are specified for both person and number, but regular NPs are specified for number only and do not carry any person specification; (ii) all of first, second and third person are characterized by one or more person features, whereas, in contrast, one of the numbers (singular) corresponds to the absence of number features.

Keywords: person, number, agreement, processing, features

OPEN ACCESS

Edited by:
Matthew Wagers,
University of California, Santa Cruz, United States

Reviewed by:
Brian Dillon,
University of Massachusetts Amherst, United States
Maziar Toosarvandani,
University of California, Santa Cruz, United States

***Correspondence:**
Ad Neeleman
a.neeleman@ucl.ac.uk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *10 October 2018*
Accepted: *22 January 2019*
Published: *14 February 2019*

Citation:
Ackema P and Neeleman A (2019) Processing Differences Between Person and Number: A Theoretical Interpretation. Front. Psychol. 10:211. doi: 10.3389/fpsyg.2019.00211

INTRODUCTION

The literature on processing of person and number agreement contains some apparently contradictory results. On the one hand, some ERP studies do not find a qualitative difference between person and number when an agreeing verb does not match the features of its subject, the controller of the agreement relation (see Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2016). On the other hand, an ERP study reported in Mancini et al. (2011a) did find a qualitative difference between agreement violations in person and agreement violations in number, a result further corroborated by an fMRI study reported in Mancini et al. (2017). There is nonetheless also a trend on which the literature appears to agree: on the whole the response to agreement violations in person is stronger than the response to number agreement violations.

The qualitative differences involve both the neuroanatomical and the electrophysiological level. Mancini et al. (2011a: 64) find that "while number agreement violations produced a left-anterior negativity followed by a P600 with a posterior distribution, the negativity elicited by person anomalies had a centro-posterior maximum and was followed by a P600 effect that was frontally distributed in the early phase and posteriorly distributed in the late phase." One conclusion from

Mancini et al.'s (2017) fMRI study is that "while the posterior portion of the (left middle temporal gyrus) is sensitive to both Person and Number Violations, the anterior portion of this region shows selective response for Person Violations" (p. 140).

In contrast, both Silva-Pereyra and Carreiras (2007) and Zawiszewski et al. (2016) explicitly note that they found no qualitative difference in the processing of person agreement violations versus number agreement violations. Both studies do find a quantitative effect. Zawiszewski et al. (2016) note that both person violations and person+number violations elicited larger P600 effects than number violations. Silva-Pereyra and Carreiras (2007) note that the P600 effect induced by a person+number violation is larger than either the effect of a person violation or the effect of a number violation; they did not find a significant difference between the latter two. The existence of a quantitative difference between person and number violations is further confirmed by the study by Mancini et al. (2017), who find a greater response for person compared to number in the region that is sensitive to both (the posterior portion of the left middle temporal gyrus, see above).

We will argue that the apparently contradictory findings can at least partly be understood in terms of the theory of phi-features developed in Ackema and Neeleman (2013, 2018). Two hypotheses play a crucial role in the account. First, pronouns are specified for both person and number, but regular Noun Phrases (which, following the theoretical literature, we will term R-expressions) are specified for number only and do not carry any person features. Second, all of first, second, and third person are characterized by one or more person features. By contrast, one of the numbers, namely singular, corresponds to the absence of number features. Only plurals (and other numbers, such as dual, trial, and paucal) carry one or more number features.

The first hypothesis bears on the contradictory findings described above, because some of the experiments use pronominal subjects as the controller of agreement, while others use R-expressions. The second hypothesis provides a handle on the quantitative difference between the effects of person versus number violations.

The paper is structured as follows. First, we will provide an outline of the theory of person and number features developed in our earlier work (see section "A Theory of Person and Number Features"). Then we will explain how this theory can inform a model of error detection and repair of agreement violations (see section "Detecting and Repairing Agreement Violations"). We will assess how this model fits the reported data in the Section "Accounting for Processing Differences Between Person and Number Violations." In the final section we mention a further possible test that could be used to assess our model.

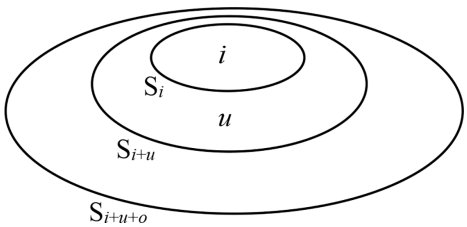
A THEORY OF PERSON AND NUMBER FEATURES

The Feature Make-Up of Pronouns

In Ackema and Neeleman (2013, 2018), we propose that there are two privative person features, dubbed PROX (for "proximate") and DIST (for "distal"). We interpret these features as functions, following insights in Harbour (2011, 2016). Both operate on an input set to deliver a subset as output.

The initial input set for the person system represents all potential referents in a given context (S_{i+u+o} in (1)). This set has a fixed structure. It contains a subset S_{i+u}, which itself contains a subset S_i. S_i has the speaker (*i*) as an obligatory member; its other members, if there are any, are associates of the speaker and/or further individuals identified as speaker. S_{i+u} has one addressee (*u*) as an obligatory member, in addition to all members of S_i; its other members, if there are any, are associates of the addressee and/or further individuals addressed by the speaker. S_{i+u+o} contains all members of S_{i+u}; its remaining members, if there are any, are neither associates of the speaker nor of the addressee.

(1)



We assume that the input set S_{i+u+o} is introduced by a category N_Π , which by definition forms the lexical core of a pronominal expression.

The feature PROX introduces a function that operates on an input set and discards its outermost "layer." Applied to S_{i+u+o} it delivers S_{i+u}. By contrast, DIST introduces a function that *selects* the outermost layer of its input set. Applied to S_{i+u+o} it delivers $S_{i+u+o} - S_{i+u}$.

This idea can be implemented as follows. Suppose that the various sets in (1) are ordered such that S_i is the predecessor of S_{i+u}, while S_{i+u} is the predecessor of S_{i+u+o} (we will abbreviate "predecessor" as Pred):

(2) a. $\text{Pred}(S_{i+u}) = S_i$

b. $\text{Pred}(S_{i+u+o}) = S_{i+u}$

If so, characterization of PROX and DIST is simple. The definitions in (3) have the desired effect that PROX discards, while DIST selects, those elements that are part of the outermost layer of the input set:

(3) a. $\text{PROX}(S) = \text{Pred}(S)$

b. $\text{DIST}(S) = S - \text{Pred}(S)$

We now consider how first, second and third person readings are derived, starting with the singular. The specification of the third person singular is straightforward: it should be DIST, as this feature will derive $S_{i+u+o} - S_{i+u}$, a set that excludes the speaker and any addressees.

A second person singular reading can be generated by applying both PROX and DIST. Notice that there is only one order of application that yields an interpretation. If PROX is applied first, S_{i+u} is selected, a set containing the speaker (and any of their associates) and individuals that the speaker addresses (and any

of their associates). Applying DIST to this set removes S_i, leaving only addressees (and any associates) as potential members – the required result [see (4)]. In the singular, this will yield a pronoun that refers to exactly one addressee.

(4) [[PRS PROX–DIST] N_Π]

= DIST(PROX(S_{i+u+o})) by definition

= DIST(S_{i+u}) by (3a)

= $S_{i+u} - S_i$ by (3b)

= S_u

The opposite order of function application is not coherent. DIST applied to S_{i+u+o} yields $S_{i+u+o} - S_{i+u}$ (a set that includes neither the speaker, nor any addressees). But this set is not layered [that is, Pred(S) is not defined for this set]. Therefore, PROX cannot apply to it.

Consider finally the first person. Notice that in the singular just applying PROX to S_{i+u+o} will not do. This is because the output it delivers, S_{i+u}, is a set with two obligatory members: the speaker and an addressee. Such a set obviously cannot be construed as singular[1]. Therefore, at least in the singular, a first person reading requires that PROX is applied to the output of PROX. As PROX discards the outermost layer of its input set, this will deliver S_i, a set whose only obligatory member is the speaker and which therefore permits a singular interpretation:

(5) [[PRS PROX–PROX] N_Π]

= PROX(PROX(S_{i+u+o})) by definition

= PROX(S_{i+u}) by (3a)

= S_i by (3a)

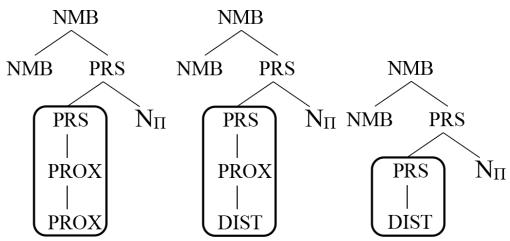
Note that in this person system all persons have one or more features[2]. This contrasts with the classical idea that third person corresponds to the absence of person information (see Benveniste, 1966). There are several previous theories in which the third person is characterized by a feature specification, among them Nevins (2007) and Harbour (2016). Evidence for this hypothesis is intricate, and cannot be reviewed here. It is based on a range of phenomena, including patterns of syncretism in pronominal and verbal agreement paradigms (see Harbour, 2016; Ackema and Neeleman, 2018), dissimilation phenomena in clitic clusters such as Spanish "spurious *se*" (Perlmutter, 1971; Grimshaw, 1997; Nevins, 2007) and person clashes in situations where double agreement has a single morphological reflex (Nevins, 2011; Ackema and Neeleman, 2018).

By contrast, in the number system, there is good evidence that, while there are features for numbers such as plural, dual, and trial, the singular corresponds to the absence of any number features. Evidence for the unmarked status of the singular includes Greenberg (1963: 94) observation that "there is no language in which the plural does not have some non-zero allomorph whereas there are languages in which the singular is expressed only by zero." Moreover, plural is both a target for morphological impoverishment rules and a context that triggers such rules. This behavior is typical of marked features (see Aalberse and Don, 2009, 2011; Nevins, 2011). Singular does not behave in the same way: it is neither a target nor a context for impoverishment.

We thus arrive at the following inventory of pronominal forms[3]:

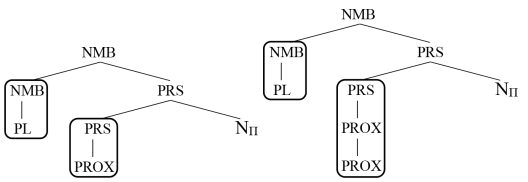
(6) ***a. 1st person*** ***b. 2nd person*** ***c. 3rd person***

Singular



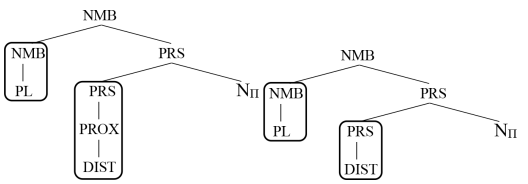
(7) ***a. 1st person inclusive*** ***b. 1st person exclusive***

Plural



c. 2nd person ***d. 3rd person***

Plural



Notice that in the system just outlined, the first person does not form a natural class with the third person to the exclusion of the second person. This fits well with the results of a large-scale study reported in Harbour (2016). Harbour looked at which systematic patterns of syncretism are attested cross-linguistically, where a systematic pattern of syncretism is one that is found in

[1]We assume that number in the context of pronouns is a feature that reflects the cardinality of its input set, rather than a feature (like person) that selects a subset from this input set (see Ackema and Neeleman, 2018: chapter 3 for more details). This means that there is no option to interpret a pronoun specified as [PROX] as singular by having it refer to just the *i* or just the *u* that are contained in the output of the person system (S_{i+u} in this case).

[2]In a system with privative features such as this, we may expect pronouns unmarked for person features. These arguably exist but correspond to impersonal rather than personal pronouns (an example of an impersonal pronoun is the English generic pronoun *one* as in *one can see the Eiffel Tower from here*); see for instance Egerland (2003); D'Alessandro (2007), Nevins (2007); Ackema and Neeleman (2018) for discussion.

[3]The difference between (7a) and (7b) corresponds to the distinction between first person inclusives and exclusives, a distinction that is morphologically marked in a number of languages. For our present purposes, this distinction is not relevant.

all paradigms of a given language. He reports that no language has a systematic syncretism for first and third person, whereas there are languages that have a systematic syncretism for first and second person, as well as languages that have a systematic syncretism for second and third person. On the assumption that the distribution of systematic syncretisms reflects the underlying distribution of features, this shows that no feature is shared uniquely by first and third person ("uniquely" meaning to the exclusion of second person).

R-Expressions Do Not Have Person

Even though third person pronouns have a person specification in the system outlined above, Ackema and Neeleman (2018) argue that R-expressions cannot carry person features[4]. They differ from pronouns in not being headed by N_Π . This means that they do not deliver S_{i+u+o} to any person features that the R-expression might contain, with the result that these features would be uninterpretable. The evidence that R-expressions do not carry any person information includes the following.

For a start, there are no first or second person R-expressions[5]. A first-person R-expression, for instance, would refer to the speaker and would obligatorily trigger first person agreement. It is certainly possible to use an R-expression to refer to the speaker or addressee (see for instance Collins and Postal, 2012 discussion of what they term "imposters"). However, this is never accompanied by obligatory first or second person agreement. Thus, the English examples in (8) are possible in certain registers, with the subject referring to the speaker. Nonetheless, these R-expressions cannot license first person agreement, let alone that they require it:

(8) a. The present author thinks/*think that this is not justifiable.

b. Yours Truly has/*have been awarded a Knighthood.

Further evidence that R-expressions like those in (8) are not specified as first person comes from the observation that in discourse they can easily be used ironically to refer to the addressee, as well as the speaker:

(9) A: Yours Truly has been awarded a Knighthood. (*Yours Truly* = speaker)

B: Well, then Yours Truly must be absolutely thrilled. (*Yours Truly* = addressee)

Crucially, the equivalent is not possible with pronouns, showing that these *are* specified for person. The following is impossible, for instance (no matter how ironic B's reply is intended to be):

(10) A: I have been awarded a Knighthood. (*I* = speaker)

B: #Well, then I must be absolutely delighted.(*I* = addressee)

Similar observations can be made for R-expressions that refer to the addressee.

Our proposal implies that R-expressions cannot carry a third person feature either. At first sight, this seems unlikely, given that R-expressions trigger what appears to be third person agreement. However, this is not a particularly compelling argument, because third person "agreement" also shows up in the absence of any possible controller for it: it can function as so-called default agreement. There are a number of languages in which finite clauses without a subject are allowed. In such clauses, the finite verb systematically shows up in its third person form[6]. While we cannot go into this here, it follows from the person system outlined above that a third person feature specification is the only one that need not be interpreted, and therefore the only one allowed on a verb in the absence of a nominal controller. If R-expressions indeed do not have person features, it follows that they should trigger default third person agreement.

There is evidence that R-expressions differ from third person pronouns. Their reference can contain speaker or addressee, as already illustrated in (8) and (9), and as corroborated by the examples in (11). In the latter examples, a first or second person pronoun refers back to an R-expression (underlining is used to indicate intended coreference). By contrast, a third person pronoun cannot be antecedent for a first or second person pronoun, as shown in (12). This follows if third person pronouns are specified as DIST, while R-expressions are not.

(11) a. Anyone who knows the Dutch realizes they no longer wear wooden shoes.

b. Anyone who knows the Dutch realizes we no longer wear wooden shoes.

c. Anyone who knows the Dutch realizes you no longer wear wooden shoes.

(12) a. Anyone who knows them realizes they no longer wear wooden shoes.

b. *Anyone who knows them realizes we no longer wear wooden shoes.

c. *Anyone who knows them realizes you no longer wear wooden shoes.

We conclude that R-expressions do not have person features that determine their reference. They never obligatorily trigger

[4]The question of whether R-expressions and third person pronouns differ in their specification for person is not often discussed in detail. In many proposals, the two categories are treated on a par in that both or neither are specified for person. There are a few exceptions that make the same cut that we take to be crucial; see in particular Sichel, 2000.

[5]Simona Mancini points out that Basque NPs that carry the proximate plural article *–ok* might be a counterexample to this generalization. We do not think so; Trask (2003: 122) notes that in the relevant varieties a phrase like *gizonok* variously means "we men," "you men," or "the men here." This is consistent with an analysis of *–ok* as a regular proximate marker, but not with an analysis as a person marker, as the person appears to vary across these translations. (For some discussion of expressions like *we men*, see Ackema and Neeleman, 2018: 155ff. and 294–295).

[6]A related observation can be made for English. There are finite clauses that have a subject, but not one that has phi-features. A case in point are clauses that have a sentential subject. In this context, too, the finite verb carries third person agreement (e.g., *That Mary wants to move to Ireland surprises/*surprise no one*).

first or second person agreement, and they can be co-referent with any pronoun[7].

DETECTING AND REPAIRING AGREEMENT VIOLATIONS

Given the theory outlined in the previous section, let us consider what might happen in processing when the input contains an agreement error. First, of course, the error must be detected. What this means is that the hearer/reader discovers that the features on the agreeing verb are not as expected given the feature specification of the subject. Second, a repair is carried out. We suggest that repair takes the form of deletion of a feature specification, insertion of a feature specification, or both. If both are required, this is a more costly operation than just deletion or insertion.

In principle, this repair can affect either the verb or the subject. There is evidence, at least in the realm of number, that there is variation in this regard. There is a preference to maintain the information on the most recently encountered element, but when repair of the preceding element is impossible it is the most recent element whose feature specification is changed (see Molinaro et al., 2008; Molinaro et al., 2011b). Repair of the subject is impossible, e.g., when the subject is a coordination, which cannot possibly be co-erced into a singular interpretation. For our purposes below, it is the nature of the repair that is crucial, not its location.

Let us first turn our attention to the detection of the agreement error. Here, we expect a qualitative difference between person and number in sentences in which the subject is an R-expression, but not in sentences in which the subject is a pronoun. Consider why. R-expressions are specified for number, but not for person. This implies that if the verb carries incorrect agreement, the type of error is qualitatively different for person and number. For number, the error is a clash: both subject and verb are specified for number, and the verb carries the wrong specification. For person, there is no clash, since the subject does not have person. Because of this, the verb should carry default agreement, which is identical to third person (see section "A Theory of Person and Number Features"). The error, therefore, is that the verb carries a non-default person specification instead[8]. Schematically, the difference between person errors and number errors with subjects that are R-expressions can be represented as follows, where the Greek letters are simply shorthand for a particular feature specification.

(13) NP [NMB: α]... V [NMB: β, PRS: γ]

Note that the specification for number is zero in the singular. This counts as a feature specification, because zero in the context of Num receives an interpretation ("n = 1"), distinct from the interpretation of a plural specification. Hence, there is a clash if subject and verb do not agree. However, given that R-expressions lack person altogether, there can be no clash with the person specification of the verb.

The situation is different when the subject is a pronoun, as pronouns have person as well as number. For both types of feature, then, the error will consist of a clash in feature specification:

(14) Pronoun [NMB: α, PRS: β]... V [NMB: γ, PRS: δ]

We are therefore led to expect that in the early stages of processing, when the detection of the agreement error takes place, person and number will behave alike with pronominal subjects, but will show a qualitative difference with subjects that are R-expressions.

Consider next what will happen at the repair stage. For number, there are two possible errors[9]. Either the subject is singular and the verb plural, or the other way around. Above we adopted the hypothesis that singular is a null feature specification. Hence, the two errors can be schematically represented as follows (both with pronouns and with R-expressions):

(15) a. NP/Pronoun [NMB: PL]... V [NMB: __)

b. NP/Pronoun [NMB: __]... V (NMB: PL)

In both cases, repair is a one-step process. It either consists of insertion of the specification Pl (if the unspecified element is repaired) or deletion of Pl (if the specified element is repaired).

For person, there are more possible errors, simply because there are more person specifications. However, what all these errors have in common is that repair cannot be a one-step process. All persons carry person features (see section "A Theory of Person and Number Features"), and therefore any change in the person specification of the repaired element must involve deletion of one person feature structure, followed by insertion of a different one. For example, if the subject is a first person pronoun, while the verb is third person, the situation is as follows:

(16) Pronoun [PRS: PROX-PROX]... V [PRS: DIST]

If it is the verb that is repaired, [DIST] will be deleted and [PROX-PROX] will be inserted. If it is the subject that is repaired, it is the other way around.

[7]We should note that this conclusion does not imply that there are no contexts in which R-expressions carry first or second person morphology. This is possible if the morphology in question is agreement morphology, and therefore not interpreted in the R-expression itself. In particular, when an R-expression has a predicative function, it may agree with a first or second person subject. This can be observed, e.g., in Classical Nahuatl (Launey, 2011) and Mohawk (Baker, 1996).

[8]In fact, in some pro drop languages plural R-expressions can be combined with a first or second person verb if they refer to a group of people containing the speaker or addressee (a phenomenon known as "unagreement," see Hurtado, 1985). Pronouns never permit this possibility. There is a variety of analyses available (e.g., Ackema and Neeleman, 2013; Höhn, 2016), but it should be clear that the hypothesis that R-expressions do not have person provides a good basis for an explanation. Unagreement is generally not possible in the singular, which is why the experiments reported below use either pronouns or singular R-expressions as subject. Unagreement gives rise to an ERP profile that is distinct from regular agreement as well as from agreement errors, see Mancini et al. (2011b).

[9]That is, possible errors in the context of the experiments discussed. There are more logically possible errors if we include numbers other than singular and plural (such as dual and trial). As far as we know, there is no experimental literature testing examples with other numbers yet, and we will put them to one side.

Thus, there is always a quantitative difference in the repair of person errors and number errors. The former is more costly, as it is a two-step process. Since repair obviously takes place after detection, this quantitative difference should present itself later in the process than the qualitative effects related to the detection of agreement errors in the context of R-expressions.

So far, we considered person and number errors separately, but of course the verb can carry a specification that is wrong for both person and number. If so, the repair process will be more costly still, as it must involve three steps: deletion or insertion of a number feature specification, deletion of a person specification, and insertion of a person specification. We may also expect differences in error detection, simply because a double error need not have the same effect as a single one, even disregarding the different nature of the person error with pronominals and R-expressions.

ACCOUNTING FOR PROCESSING DIFFERENCES BETWEEN PERSON AND NUMBER VIOLATIONS

The discussion in the previous section gives rise to the following expectations when the verb carries incorrect agreement:

(17) a. The subject is an R-expression: (i) In detection, person behaves differently from number; (ii) In repair, person errors are more costly than number errors.

b. The subject is a pronoun: (i) In detection, person behaves the same as number; (ii) In repair, person errors are more costly than number errors.

In order to evaluate whether these generalizations hold, we must know what the neurolinguistic footprints might be of detection of an agreement error and its repair.

There is a large amount of literature on the interpretation of different waveforms in ERP studies. Although there does not seem to be a clear consensus on this issue, there are certainly trends. To begin with, a negative waveform between 250 and 500 ms after stimulus onset seems to be associated with unexpected words in the input, including morphosyntactic violations. This negative deflection comes into two or three types. One is the N400. The N400 "is highly correlated (r = 0.9) with an offline measure of the eliciting word's expectancy" (Kutas and Federmeier, 2011: 624). This expectancy is often described in semantic terms, but there are indications that this may be too narrow, at least on the usual linguistic understanding of "semantic." For example, an N400 effect can be elicited by non-linguistic actions, such as cutting bread with a saw (Proverbio and Riva, 2009; Kutas and Federmeier, 2011). Furthermore, it can be elicited by words that are unexpected in form, but not in semantics. Thus, it is triggered by the form *an* for the indefinite article in English if, in the current context, the noun that is expected to follow starts with a consonant (DeLong et al., 2005). Indeed, for what counts as unexpected, simple word frequency appears relevant (Van Petten and Kutas, 1990). In short, although the N400 does not appear to index semantic anomaly exclusively, or even linguistic anomaly, in the context of language it seems correlated with detecting unexpected words in the input, including certain morpho-syntactic violations (Osterhout, 1997: 497; Tanner and Van Hell, 2014: 298).

There are other types of early negative deflection. In particular, Left Anterior Negativity (LAN) and Anterior Negativity (AN) are plausibly elicited specifically by morpho-syntactic errors, including verb agreement errors as well as case marking errors (Münte et al., 1993, 1997; Friederici et al., 1996; Gunter et al., 1997; Coulson et al., 1998; Friederici and Frisch, 2000).

The size of the LAN effect appears to be partly determined by the morpho-syntax of a language. In particular, it has been observed that it increases the more important agreement is for the parsing of grammatical dependencies (Friederici, 2011: 1381 and references cited there). Agreement is important for detecting the subject of a clause if the agreement paradigm is morphologically rich and therefore reliably indexes the subject's interpretation. Agreement is also important if the position of the subject in the clause is not fixed, so that word order does not provide a reliable clue as to what the subject is. Conversely, agreement is less important for detecting the subject if word order is strict, or if the morphological verbal agreement paradigm is poor (i.e., contains a lot of syncretisms). Nonetheless, there is some evidence that agreement violations induce a LAN effect also in languages with poorer agreement morphology and a relatively fixed word order, such as Dutch and English (see Osterhout and Mobley, 1995; Hagoort and Brown, 2000; see also Molinaro et al., 2011a for an overview).

Overall, it seems reasonable to correlate the detection of an unexpected agreement form of the verb in the input with a negative waveform in the relevant timeframe. This being said, we should acknowledge that there are some studies of agreement errors that do not find an early negativity effect (see Nevins et al., 2007). Our take on this is that there is a one-way implication: early negativity is in indication of error detection, but error detection is not guaranteed to produce early negativity[10].

A further clear trend in the literature involves the P600, a positive deflection starting around 500 ms after stimulus onset and lasting a few 100 ms. The P600 is said to be triggered by a range of linguistic anomalies or other difficulties, including those associated with syntactic processing. It is, e.g., triggered by complicated syntax (Kaan et al., 2000; Friederici et al., 2002), less preferred syntactic structure (Osterhout et al., 1994; Itzhak et al., 2010), and by syntactic garden-path effects, i.e., syntactic anomalies that result from misanalysis of an ambiguity rather than from ungrammaticality (Osterhout and Holcomb, 1992, 1993; Osterhout et al., 1994; Kaan and Swaab, 2003). It has also been observed with a variety of syntactic violations, see Gouvea et al. (2010: 150). In view of this, one may expect a P600 effect to index the repair of the morpho-syntactic structure that an agreement violation necessitates.

[10]If so, studies that do not find early negativity can also not identify potential differences in the detection of person and number agreement errors.

With the above in mind, let us consider whether the reported effects of agreement violations are in line with (17). One relevant ERP study is reported by Silva-Pereyra and Carreiras (2007). They tested Spanish sentences with pronominal subjects that contained agreement violations of the following three types:

(18) a. Pronoun_{1PL}... V_{1SG} (number disagreement)

b. Pronoun_{2SG}... V_{1SG} (person disagreement)

c. Pronoun_{2PL}... V_{1SG} (person and number disagreement)

Given the predictions in (17b), we expect there to be no qualitative differences between any of the examples where it concerns error detection. We expect the errors in (18b) and (18c) to give rise to a more costly repair than the one in (18a), as they involve person. In addition, we expect (18c) to be more costly in repair than (18b), as it involves a double violation.

Silva-Pereyra and Carreiras (2007) found that there were indeed no qualitative differences between person and number violations. They describe their findings as follows (where ND = number disagreement, PD = person disagreement, and NPD = disagreement for both person and number): "ND, PD, and NPD all elicited an anterior negativity (AN) and P600 pattern. An AN effect was only found in the NPD with a different topography from the classic LAN effect as it was lateralized to right and central sites. The P600 effect elicited by the NPD condition was larger than the agreement condition and that of ND and PD in the first window 500–700, while the three disagreement conditions elicited larger P600 amplitudes than the agreement condition in the second window 700–900" (p. 201).

These findings meet our expectations relatively well. The fact that no qualitative differences between person and number were found where it concerns early negativity is the result of pronominal subjects being used in the test sentences. The repair involved in the double violation condition is more costly than the repairs in either single violation condition. The only unexpected finding is the lack of a significant difference between number disagreement and person disagreement in the amplitude of the P600 effect.

Next, an ERP study by Zawiszewski et al. (2016) tested Basque sentences with agreement violations schematized in (19). In this study, too, the subject in all test sentences was a pronoun. The pronoun was always second person singular, while the agreement on the verb was varied to create number disagreement, person disagreement, or disagreement for both person and number.

(19) a. Pronoun_{2SG}... V_{2PL} (number disagreement)

b. Pronoun_{2SG}... V_{1SG} (person disagreement)

c. Pronoun_{2SG}... V_{1PL} (person and number disagreement)

On the basis of (17b), we again expect no qualitative differences in the effects triggered by the various violations. We do expect a quantitative difference in the repair stage of the process, where the violations that involve person should trigger a larger effect than the number violation. We also expect the double violation to give the largest effect in repair.

These expectations are largely met. Zawiszewski et al. (2016) found that, first, all violation types triggered an N400-P600 pattern. Second, person and person+number violations elicited larger P600 effects than number violations. To be more specific, with regards to the N400, "no differences were found between person and number violations or between person and person+number violations, while number violations elicited a larger negativity over left-posterior sites than person+number violations." With regards to the P600, "no differences were found between person and person+number violations, while both person and person+number violations elicited a larger P600 than number violations over posterior sites accompanied by a larger negativity over frontocentral sites" (p. 618). Zawiszewski et al. (2016):618) summarize their findings as follows: "Our results revealed qualitatively similar but quantitatively larger ERP signatures for person than for number violations."

This conclusion supports our main contention: if pronominal subjects are used, no qualitative differences between person and number violations are to be expected. Also, the hypothesis that violations involving person should always give rise to a more costly repair than violations only involving number, and hence to larger P600 effects, is confirmed. We do not have a specific account for the difference between the number violation and the double violation with respect to the amplitude of the N400, nor for the absence of a significant difference in the size of the P600 in the double violation condition and the person disagreement condition. (Note that the latter finding is the opposite of what Silva-Pereyra and Carreiras found).

A third relevant ERP study is the one by Mancini et al. (2011a). They tested Spanish sentences with R-expressions as subject. The test sentences were of the following types:

(20) a. NP_{SG}... V_{3PL} (number disagreement)

b. NP_{SG}... V_{2SG} (person disagreement)

Note that we have not labeled the NP subject as being third person, in line with our hypothesis that R-expressions do not have person. The structure in (20b) is therefore not actually a case of disagreement; rather, the verb does not show the expected default third person agreement that is selected when the controller does not have person features. Crucially, this is a different type of error than the one in (20a), where there is a clash between the number specification of the subject (namely ø) and the number specification of the verb; see section "A Theory of Person and Number Features" for more detailed discussion. Hence, as mentioned in (17a), we should find qualitative differences between the sentences with a person violation and the sentences with a number violation.

Mancini et al. indeed found that the parser is differentially sensitive to the two features. "While number agreement violations produced a left-anterior negativity followed by a P600 with a posterior distribution, the negativity elicited by person anomalies had a centro-posterior maximum and was

followed by a P600 effect that was frontally distributed in the early phase and posteriorly distributed in the late phase" (Mancini et al., 2011a: 64).

In addition, "both anomalies produce a P600 effect that has its maximum in posterior sites. Differences between number and person emerged in terms of the amplitude of this effect, which appears to be larger for the person mismatch" (Mancini et al., 2011a: 73). The latter observation confirms the second prediction in (17a), namely that repair of person violations is more costly than repair of number violations regardless of the nature of the subject.

A final relevant study, by Mancini et al. (2017), uses fMRI, rather than ERP, as its investigative technique. This implies that there is not enough temporal resolution to distinguish the detection and repair stages of the processing of sentences with agreement errors. However, fMRI can of course identify qualitative and quantitative differences in the parsing process, and so it does provide an opportunity to test the generalizations in (17) provided their temporal dimension, and hence the distinct reference to the detection and repair stages of the process, are removed. If we do this, the generalizations are as follows:

(21) a. The subject is an R-expression: Person behaves qualitatively differently from number and will have a quantitatively larger effect.

b. The subject is a pronoun: There are no qualitative differences between person and number, but person will have a quantitatively larger effect.

Mancini et al. tested Spanish sentences with an R-expression as subject, containing an agreement error of one of the following two types:

(22) a. NP_{SG}... V_{3PL} (number disagreement)

b. NP_{SG}... V_{2SG} (person disagreement)

Since the subjects are R-expressions, we expect the behavior in (21a). This is in line with what Mancini et al. found: "The direct contrast between Person and Number Violations permitted the uncovering of both quantitative and qualitative differences" (p. 147). More specifically, "A greater response for person compared to number was found in the left middle temporal gyrus (LMTG). However, critically, a posterior-to-anterior functional gradient emerged within this region. While the posterior portion of the LMTG was sensitive to both Person and Number Violations, the anterior portion of this region showed selective response for Person Violations" (p. 140).

To the best of our knowledge, there is no similar fMRI study that compares person and number agreement violations, but uses pronominal subjects instead. For now, then, the prediction in (21b) is left untested.

In sum, the studies that use pronominal subjects to explore agreement errors do not find qualitative differences between person and number, while the studies that use R-expressions as subject do find such differences. This is accounted for by a theory in which pronouns have a person specification, but R-expressions do not.

Studies differ in where they find significant quantitative differences between number violations, person violations and double violations. However, all significant differences that were found follow a hierarchy number < person < number+person. This is in line with the idea that repair of person violations is more costly than repair of number violations, though perhaps the difference in cost is relatively small.

A POSSIBLE FURTHER EXPERIMENT

Our proposal can be tested further, as one crucial data set currently remains unexplored. This involves third person pronouns. In traditional grammar, such pronouns are treated on a par with what we call R-expressions: both are third person. However, on the theory proposed here only the pronouns carry a third person specification; the R-expressions are personless. Hence, the prediction is that there should no contrastive behavior between third person pronouns and first and second person pronouns. The contrast should be between all pronouns on the one hand and R-expressions on the other hand. For example, an agreement violation between a third person singular pronoun and, say, a first person singular verb is of the type in (23a), and therefore involves a clash in person features. An agreement violation between a singular R-expression and a first person singular verb is, as discussed, of the type in (23b), which does not involve such a clash.

(23) a. Pronoun [NMB: __, PRS: DIST]... V [NMB: __, PRS: PROX-PROX]

b. NP [NMB: __]... V [NMB: __, PRS: PROX-PROX]

Relevant examples with third person pronominal subjects have not been tested, as far as we know. We expect that there are qualitative differences in error detection between the two conditions in (23).

CONCLUSION

We hope to have shown that theoretical accounts of agreement can be used to interpret experimental data, and that experimental data can be used to test theoretical accounts. In particular, we have argued that contrasts between the processing of person and number agreement violations may fall out from a specific theory of person and number features according to which (i) R-expressions do not have person, while pronouns do, and (ii) singular is the absence of a number feature, while all persons, including third person, have person features. To the extent that other theoretical accounts of person and number make different predictions, the experimental data can be said to confirm this theory.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Aalberse, S., and Don, J. (2009). Syncretism in Dutch dialects. *Morphology* 19, 3–14. doi: 10.1007/s11525-009-9132-y

Aalberse, S., and Don, J. (2011). Person and number syncretisms in Dutch. *Morphology* 21, 327–350. doi: 10.1007/s11525-010-9164-3

Ackema, P., and Neeleman, A. (2013). Subset controllers in agreement. *Morphology* 23, 291–323. doi: 10.1007/s11525-013-9218-4

Ackema, P., and Neeleman, A. (2018). *Features of Person: From the Inventory of Persons to their Morphological Realization*. Cambridge, MA: MIT Press.

Baker, M. (1996). *The Polysynthesis Parameter*. Oxford: Oxford University Press.

Benveniste, É (1966). *Problèmes de Linguistique Générale*. Paris: Gallimard.

Collins, C., and Postal, P. (2012). *Imposters: A Study of Pronominal Agreement*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262016889.001.0001

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi: 10.1080/016909698386582

D'Alessandro, R. (2007). *Impersonal Si-Constructions*. Berlin: Mouton de Gruyter. doi: 10.1515/9783110207514

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

Egerland, V. (2003). Impersonal pronouns in Scandinavian and Romance. *Work. Pap. Scand. Syntax* 71, 75–102.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D., and Frisch, S. (2000). Verb argument structure processing: the role of verb-specific and argument-specific information. *J. Mem. Lang.* 43, 476–507. doi: 10.1006/jmla.2000.2709

Friederici, A. D., Hahne, A., and Mecklinger, A. (1996). Temporal structure of syntactic processing: early and late event-related potential effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1219–1248. doi: 10.1037/0278-7393.22.5.1219

Friederici, A. D., Steinhauer, K., and Pfeifer, E. (2002). Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *PNAS* 99, 529–534. doi: 10.1073/pnas.012611199

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The linguistic processes underlying the P600. *Lang. Cogn. Process.* 25, 149–188. doi: 10.1080/01690960902965951

Greenberg, J. (1963). "Some universals of grammar with particular reference to the order of meaningful elements," in *Universals of Human Language*, ed. J. Greenberg (Cambridge, MA: MIT Press), 58–90.

Grimshaw, J. (1997). "The best clitic: constraint conflict in morphosyntax," in *Elements of Grammar*, ed. L. Haegeman (Dordrecht: Kluwer), 169–196.

Gunter, T. C., Stowe, L. A., and Mulder, G. (1997). When syntax meets semantics. *Psychophysiology* 34, 660–676. doi: 10.1111/j.1469-8986.1997.tb02142.x

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1

Harbour, D. (2011). Valence and atomic number. *Linguist. Inq.* 42, 561–594. doi: 10.1162/LING_a_00061

Harbour, D. (2016). *Impossible Persons*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262034739.001.0001

Höhn, G. (2016). Unagreement is an illusion: apparent person mismatches and nominal structure. *Nat. Lang. Linguist. Theory* 34, 543–592. doi: 10.1007/s11049-015-9311-y

Hurtado, A. (1985). "The Unagreement Hypothesis," in *Selected Papers from the Thirteenth Linguistic Symposium on Romance Languages*, eds L. D. King and C. A. Maley (Amsterdam: John Benjamins), 187–211. doi: 10.1075/cilt.36.12hur

Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., and Steinhauer, K. (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *Neuroreport* 21, 8–13. doi: 10.1097/WNR.0b013e328330251d

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Kaan, E., and Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110. doi: 10.1162/089892903321107855

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Launey, M. (2011). *An Introduction to Classical Nahuatl*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511778001

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011a). A person is not a number: discourse involvement in subject-verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011b). When persons disagree: an ERP study of unagreement in Spanish. *Psychophysiology* 48, 1361–1371. doi: 10.1111/j.1469-8986.2011.01212.x

Mancini, S., Quiñones, I., Molinaro, N., Hernandez-Cabrera, J. A., and Carreiras, M. (2017). Disentangling meaning in the brain: left temporal involvement in agreement processing. *Cortex* 86, 140–155. doi: 10.1016/j.cortex.2016.11.008

Molinaro, N., Barber, H. A., and Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Molinaro, N., Vespignani, F., Zamparelli, R., and Job, R. (2011b). Why "brother and sister" are not just "siblings": repair processes in agreement computation. *J. Mem. Lang.* 64, 211–232. doi: 10.1016/j.jml.2010.12.002

Molinaro, N., Kim, A., Vespignani, F., and Job, R. (2008). Anaphoric agreement violation: an ERP analysis of its interpretation. *Cognition* 106, 963–974. doi: 10.1016/j.cognition.2007.03.006

Münte, T. F., Heinze, H.-J., and Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *J. Cogn. Neurosci.* 5, 335–344. doi: 10.1162/jocn.1993.5.3.335

Münte, T. F., Matzke, M., and Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *J. Cogn. Neurosci.* 9, 318–329. doi: 10.1162/jocn.1997.9.3.318

Nevins, A. (2007). The representation of third person and its consequences for person-case effects. *Nat. Lang. Linguist. Theory* 25, 273–313. doi: 10.1007/s11049-006-9017-2

Nevins, A. (2011). Multiple agree with clitics: person complementarity vs. omnivorous number. *Nat. Lang. Linguist. Theory* 29, 939–971. doi: 10.1007/s11049-011-9150-4

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058

Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang.* 59, 494–522. doi: 10.1006/brln.1997.1793

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., and Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Lang. Cogn. Process.* 8, 413–438. doi: 10.1080/01690969308407584

Osterhout, L., Holcomb, P. J., and Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence for the application of verb information during parsing. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 786–803. doi: 10.1037/0278-7393.20.4.786

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Perlmutter, D. (1971). *Deep and Surface Structure Constraints in Syntax*. New York, NY: Holt, Rinehart, and Winston.

Proverbio, A. M., and Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neurosci. Lett.* 459, 142–146. doi: 10.1016/j.neulet.2009.05.012

Sichel, I. (2000). "Demonstrative pronouns, binding theory, and identity," in *Proceedings of the 24th Annual Penn Linguistics Colloquium* (Philadelphia, PA: University of Pennsylvania), 247–259.

Silva-Pereyra, J. F., and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029

Tanner, D., and Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301. doi: 10.1016/j.neuropsychologia.2014.02.002

Trask, L. (2003). "Morphology," in *A Grammar of Basque*, eds J. I. Hualde and J. Ortiz de Urbina (Berlin: Mouton de Gruyter), 113–362.

Van Petten, C., and Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Mem. Cogn.* 18, 380–393. doi: 10.3758/BF03197127

Zawiszewski, A., Santesteban, M., and Laka, I. (2016). Phi-features reloaded: an even-related potential study on person and number agreement processing. *Appl. Psycholinguist.* 37, 601–626. doi: 10.1017/S014271641500017X

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 05 April 2019
doi: 10.3389/fpsyg.2019.00751

Revisiting Masculine and Feminine Grammatical Gender in Spanish: Linguistic, Psycholinguistic, and Neurolinguistic Evidence

Anne L. Beatty-Martínez[1,2]* *and Paola E. Dussias*[1,2]

[1]*Center for Language Science, The Pennsylvania State University, University Park, PA, United States,* [2]*Department of Spanish, Italian and Portuguese, The Pennsylvania State University, University Park, PA, United States*

Research on grammatical gender processing has generally assumed that grammatical gender can be treated as a uniform construct, resulting in a body of literature in which different gender classes are collapsed into single analysis. The present work reviews linguistic, psycholinguistic, and neurolinguistic research on grammatical gender from different methodologies and across different profiles of Spanish speakers. Specifically, we examine distributional asymmetries between masculine and feminine grammatical gender, the resulting biases in gender assignment, and the consequences of these assignment strategies on gender expectancy and processing. We discuss the implications of the findings for the design of future gender processing studies and, more broadly, for our understanding of the potential differences in the processing reflexes of grammatical gender classes within and across languages.

Keywords: grammatical gender, gender assignment, language processing, language variation, Spanish

OPEN ACCESS

Edited by:
Sendy Caffarra,
Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Pascal Mark Gygax,
Université de Fribourg, Switzerland
Alan Garnham,
University of Sussex, United Kingdom

***Correspondence:**
Anne L. Beatty-Martínez
alb507@psu.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *05 February 2019*
Accepted: *18 March 2019*
Published: *05 April 2019*

Citation:
Beatty-Martínez AL and Dussias PE (2019) Revisiting Masculine and Feminine Grammatical Gender in Spanish: Linguistic, Psycholinguistic, and Neurolinguistic Evidence. Front. Psychol. 10:751. doi: 10.3389/fpsyg.2019.00751

INTRODUCTION

Linguistic factors have long been known to modulate word identification. Of relevance for the work presented here, studies examining grammatical gender provide evidence that information at one point in a sentence is used to anticipate other information downstream. Grammatical gender is a widespread feature in many of the world languages. Simply put, it refers to "classes of nouns reflected in the behavior of associated words" (Hockett, 1958, p. 231; see also Comrie, 1999). Linguists agree that a language is said to have a grammatical gender system if there is evidence for gender outside the nouns themselves. One such type of evidence is gender agreement (Corbett, 1991). Examples (1a) and (1b) from Spanish illustrate this:

(1)	(a)	La	televisión	es	roj**a**
		The$_{FEM}$	TV$_{FEM}$	is	red$_{FEM}$
		"The TV is red"			
	(b)	El	teleférico	es	roj**o**
		The$_{MASC}$	ski lift$_{MASC}$	is	red$_{MASC}$
		"The ski lift is red"			

In (1a), the form of the determiner is "*la*" and of the adjective is "*roja*" because "televisión" is a feminine noun. In other words, the determiner and the adjective agree in gender with the noun they accompany. In (1b), the determiner "*el*" and the adjective "*rojo*" agree with "teleférico" (a masculine noun).

A robust finding across languages with different gender systems (e.g., for Croatian, Costa et al., 2003; for French, Dahan et al., 2000; for German, Schmidt, 1986; for Italian, Bates et al., 1996; see Friederici and Jacobsen, 1999, for a review of early studies) is that when the gender of an article or adjective is congruent with that of the following noun, recognition of the noun is enhanced relative to a neutral baseline; when it is incongruent, recognition is delayed. This *gender congruency* effect has been reported in visual tasks (e.g., Jescheniak, 1999; Cubelli et al., 2005) and auditory tasks (e.g., Faussart et al., 1999; Dahan et al., 2000) and for languages with two genders (e.g., Barber and Carreiras, 2005) and more than two genders (e.g., van Berkum, 1996; Jacobsen, 1999). For instance, in Serbo-Croatian, lexical decision is faster for nouns preceded by adjective primes that match the nouns in gender than for those with mismatched preceding adjectives (Gurjanov et al., 1985). In addition, Cole and Segui (1994) reported that lexical decision is faster in French when primes are closed-class words (e.g., articles) relative to open-class words (e.g., adjectives), suggesting that the gender congruency effect changes as a function of word type. Results from Jakubowicz and Faussart (1998) have, in addition, shown that in a spoken lexical decision task, French adjectives phonetically marked for gender that intervened between an article and a noun (e.g., the adjective $petit_{MASC}$ [pəti] /$petite_{FEM}$ [pətit], as in "le/*la petit chien," the_{MASC}/*the_{FEM} $little_{MASC}$ dog_{MASC}) do not increase the magnitude of the gender congruency effect relative to an invariant adjective without gender marking (e.g., the adjective $pauvre_{MASC/FEM}$ [povʀ], as in "le/*la pauvre chien," the_{MASC}/*the_{FEM} poor dog_{MASC}). This is significant because it highlights the central role of articles in setting gender agreement features for the entire noun phrase (Jakubowicz and Faussart, 1998). For Spanish, the language under investigation in this review, Lew-Williams and Fernald (2007) showed that Spanish-speaking children and adults exploit gender information on articles to facilitate the processing of upcoming nouns. Using the *looking-while-listening* procedure, Lew-Williams and Fernald presented participants with two-picture visual scenes, in which objects either matched or differed in grammatical gender. Target items were embedded in fixed carrier phrases (e.g., "encuentra *el*/*la*," find the_{MASC}/the_{FEM}), and participants were instructed to find the named object. Results revealed that on different-gender trials, participants oriented their eyes toward target objects more quickly than on same-gender trials, yielding an anticipatory effect.

Importantly, studies reporting effects of prenominal gender marking on subsequent word identification have generally assumed that different gender classes (e.g., feminine and masculine in Spanish) modulate these effects with equal strength. Thus, with few exceptions (e.g., Gurjanov et al., 1985; Grosjean et al., 1994), studies have collapsed gender classes into a single analysis. Despite this general practice, in the work presented here, we discuss evidence from linguistic, psycholinguistic, and neurolinguistic studies, suggesting that grammatical gender classes may differentially contribute to the identification of nouns. Central to this proposal is the assumption that individuals of all language backgrounds are equipped with the ability to develop sensitivity to distributional information in language (Clayards et al., 2008; Gennari and MacDonald, 2009; Beatty-Martínez and Dussias, 2018). Our starting point is that words form relations along phonetic dimensions which contribute toward the creation of exemplar clusters. Categories are formed by placing exemplars in a conceptual space either closer to or further from each other depending upon the degree of dissimilarity of the members of a class (i.e., schematicity; Clausner and Croft, 1997). In the following sections, we provide evidence for this claim by examining distributional asymmetries between masculine and feminine gender in Spanish.

ON THE DIFFERENTIAL BEHAVIOR OF MASCULINE AND FEMININE GENDER IN SPANISH

Evidence From Monolingual Speakers

In Spanish, masculine has an unmarked or default status that sharply distinguishes it from feminine. One piece of evidence comes from loanwords, which are overwhelmingly assigned masculine gender. In a study by De la Cruz Cabanillas et al. (2007), 82% of the gendered loanwords in their corpus were masculine. In addition, masculine gender is also used in Spanish to refer to groups of individuals that include at least one male. As such, the noun phrase "los padres de Ana" (the_{MASC} fathers of Ana) can refer to Ana's father and mother; "mis hijos" (my sons) can include daughters but not vice-versa; and "los estudiantes" (the_{MASC} students) can refer to groups of students in which all but one person are male.[1] The unmarked status of Spanish masculine gender is further highlighted by agreement phenomena. When prepositions, conjunctions, and other non-gender marked words are used as nouns, they take masculine prenominals (e.g., reemplaza *este* "aunque" por *un* "sin embargo", replace $this_{MASC}$ "still" for a_{MASC} "nevertheless") and masculine determiners are used in nominalizations (e.g., "*el* fumar mata," the_{MASC} smoking kills). A study by Eddington and Hualde (2008) presented intriguing evidence showing that native speakers of Spanish make errors when assigning gender to certain Spanish feminine nouns. In Spanish, the phonological pattern most typically associated with feminine gender is the presence of a final /a/ phoneme, illustrated in nouns such as "*casa*" (house), "*mesa*" (table), "*arpa*" (harp), and "*águila*" (eagle). Endings for masculine nouns include the vowels *-o* and *-e*, as well as

[1]Although not the focus of the current view, there is also some evidence of grammatical gender asymmetries when referring to human beings (or animate beings in general). More specifically, some have argued that the generic use of masculine forms in gendered languages may lead to biased representations of gender during language processing (Spanish: Carreiras et al., 1996; see Gabriel and Gygax, 2016, for a detailed discussion on this issue).

a number of consonants (e.g., *-l* ["*caracol,*" $\text{snail}_{\text{MASC}}$], *-n* ["*tren*," $\text{train}_{\text{MASC}}$], *-j* ["*reloj*," $\text{watch}_{\text{MASC}}$]), reflecting the fact that Spanish masculine phonological endings are less restricted. Feminine nouns, however, have an additional complicating rule. When the onset of a Spanish feminine noun is a stressed /a/, singular definite determiners ("*la*," the_{FEM}) and determiners ending in /-una/ ("*una*," a_{FEM}; "*alguna*," some_{FEM}; "*ninguna*," none_{FEM}) must carry masculine gender if they immediately precede the noun.[2] The reason appears to be a phonetic infelicity involving word-final /a/ immediately followed by stressed word-initial /a/. This is shown in the examples (2a) and (2b) below:

(2) (a) *una* costos**a** arp**a**
aFEM $\text{expensive}_{\text{FEM}}$ harp_{FEM}
"an expensive harp"

(b) *un* arp**a** costos**a**
a_{MASC} harp_{fem} $\text{expensive}_{\text{fem}}$
"an expensive harp"

What Eddington and Hualde (2008) found is that this variation produces confusion in native speakers, which results in the (incorrect) use of masculine prenominal modifiers appearing to the left of these nouns and feminine post-nominal modifiers appearing to the right:

(3) (a) Echa *todo el* agu**a** fr**í**a en el barreño
pour all_{MASC} the_{MASC} $\text{water}_{\text{FEM}}$ cold_{FEM} in the basin
"pour all the cold water in the basin"

instead of

(b) Echa *toda el* agu***a*** fr***í a*** en el barreño
pour all_{FEM} the_{MASC} $\text{water}_{\text{FEM}}$ cold_{FEM} in the basin
"pour all the cold water in the basin"
(Eddington and Hualde, 2008, p. 4)

Psycholinguistic evidence also highlights the unmarked status of Spanish masculine gender. Domínguez et al. (1999) found that for masculine and feminine words closely matched in frequency, mean reaction times during a lexical decision task were shorter for the masculine than the feminine forms. Another source of linguistic evidence comes from studies on Spanish gender acquisition. Pérez-Pereira (1991) observed that monolingual Spanish-speaking children made use of a noun's phonological shape (i.e., whether nouns ended in *-a* or *-o*) when assigning gender to determiners. However, Pérez-Pereira also observed that children were more likely to assign masculine gender to nouns with irregular (i.e., ambiguous) phonological cues, suggesting a masculine default strategy in gender assignment (Harris, 1991). One question raised by these results is whether the preference for masculine gender stems from distributional frequency differences in language input to children. Smith et al. (2003) examined a corpus of child-directed speech and developed a connectionist model of gender assignment to mirror the type frequency patterns to which a child is exposed over time. Analysis of the corpus revealed an equal number of masculine and feminine nouns. However, upon closer inspection, distributional frequency differences between regular (i.e., nouns ending in *-a* or *-o*) and irregular nouns emerged: "while regular feminine nouns were slightly more frequent than regular masculine nouns, irregular masculine nouns outnumbered irregular feminine nouns by roughly 2 to 1" (Smith et al., 2003, p. 306). The model, which was incrementally trained on this input, produced a similar bias toward masculine gender when tested on novel words, suggesting that the frequency distribution, particularly the interaction between gender and word form ambiguity, plays a direct role in gender assignment.

A potential limitation of the Smith et al. (2003) study is that it did not examine the role of phonological factors beyond the word-final phoneme in determining gender assignment. Contrary to previous claims in the literature (Harris, 1985; Roca, 1989), the correspondence between the gender of a noun and its phonological shape is not fortuitous. Eddington (2002) used an exemplar-based model to determine the gender of a noun based on its phonological shape. The database for the simulation included a list of highly frequent nouns in Spanish taken from Juilland and Chang-Rodríguez's (1964) frequency estimates. Each noun was encoded to include its phonemic makeup (e.g., the word's final phoneme) and the syllabic structure of the penultimate and final syllables. When the penultimate rhyme and final syllable variables were included in the model, the algorithm successfully assigned gender to 95% of nouns. To determine whether native speakers were able to exploit the same systematic correspondences as the model, Eddington tested a group of monolingual Spanish-speaking adults on a gender assignment task using novel words with ambiguous endings (i.e., final phonemes other than *-a* and *-o*). The results produced a clear bias toward masculine gender assignment, replicating previous findings. Notably, an assessment of success and error rates for each of the variables confirmed a high degree of association between the model and native speakers' intuitions.

Altogether, the Eddington (2002) results suggest that speakers establish and make use of phonological factors besides word-final phonemes to assign grammatical gender. Eddington suggests that the structure of the nouns themselves provides an explanation for speakers' bias toward masculine due to a markedness asymmetry between the two genders. In a marked/unmarked relation, the marked member of the opposition (i.e., feminine gender) has a densely clustered category, settling on a tighter range of variance. The unmarked category (i.e., masculine gender), on the other hand, covers a wider range of configurations (Greenberg, 1966). "[W]hat this means for gender is that a random throw of the dart onto a map of nouns organized according to phonological similarities, has a much higher probability of landing in a neighborhood of masculine nouns, even if they do not dominate feminine nouns numerically"

[2]While exceptions exist (e.g., '*el* d*í*a', the_{MASC} day_{MASC}; '*l*a mano', the_{FEM} hand_{FEM}), *–a* and *-o* endings been shown to be highly correlated with masculine and feminine gender respectively (Bull, 1965; Harris, 1991; Eddington, 2002; Clegg, 2010). We refer to *-a* and *-o* endings as predictor variables for gender assignment rather than gender morphemes. These variables are probabilistic, some being more reliable than others (see Table 1 in Eddington, 2002, for a list of other phonemic variables interpreted as relevant for gender assignment).

(Eddington, 2002, p. 66). We return to the role of morphological markedness on gender processing in the section devoted to electrophysiological evidence.

Evidence From Bilingual Speakers

The evidence presented above raises the question of whether Spanish masculine and feminine articles differentially affect the time course of noun processing. One potential disadvantage of the current monolingual work is that most studies have employed offline grammaticality judgments or speech elicitation experiments with novel words out of context, which are artificial tasks. In this respect, bilingualism can be used as a tool to examine questions that are sometimes not easily studied with monolingual populations. We adopt a broad definition of bilingualism to include speakers who actively use two or more languages, regardless of whether those languages were acquired in early childhood or later in life. In this section, we will review gender assignment strategies in bilingual speakers with a special emphasis on codeswitching[3], the alternation between languages within and between utterances in bilingual discourse. Like monolinguals, bilingual speakers of Spanish and another language have been shown to have a similar preference to assign masculine gender to determiners for loanwords (Smead, 2000; Aaron, 2015), with the exception of established loanwords that are strongly morphologically integrated in Spanish (e.g., "la troca," the truck; Clegg and Waltermire, 2009). However, a characteristic of many bilingual communities of the Spanish-speaking world is to routinely switch between Spanish and another language when speaking to other bilinguals. We propose that codeswitching provides a special testbed for the study of distributional asymmetries in gender assignment while circumventing some of the obstacles outlined above (Myers-Scotton and Jake, 2015). Specifically, codeswitched noun phrases (NPs) are abundant in Spanish-English codeswitched speech (Timm, 1975; Pfaff, 1979; Poplack, 1980). Because "mixed" NPs (i.e., NPs that appear in two languages) are highly frequent in the everyday speech of some bilingual populations, they provide a valuable alternative for examining gender assignment strategies as a means to reveal the underlying mechanisms that are responsible for asymmetrical distributions. How so? Because when bilinguals codeswitch, they make opportunistic decisions about how to integrate the two linguistic systems on the fly (Green and Wei, 2014). Their production choices provide, in turn, a window on speakers' prior linguistic experience (Beatty-Martínez et al., 2018a). For example, corpus studies on Spanish-English codeswitching have noted that bilinguals are more likely to produce mixed NPs with Spanish determiners and English nouns (e.g., "*el* dog," the_{SPAN} dog_{ENG}) over mixed NPs with the opposite configuration (e.g., "the *perro*," the_{ENG} dog_{SPAN}; Jake et al., 2002; Valdés Kroff, 2016; Beatty-Martínez et al., 2018a; Królikowska et al., 2019; cf. Blokzijl et al., 2017). Similarly, many studies have reported a masculine tendency in the assignment of grammatical gender for Spanish-English mixed NPs similar to the sentences in (4a) and (4b) below (Montes-Alcalá and Lapidus Shin, 2011; Valdés Kroff, 2016; cf. Liceras et al., 2008). What makes this observation particularly interesting is that many English nouns in mixed NPs have a clear Spanish translation equivalent, so the opportunity to examine how these switches are integrated in spontaneous conversation sheds light on the asymmetrical relationship between masculine and feminine by revealing which linguistic mechanisms are at play in a way that is otherwise obscured in monolingual speech.

(4) (a) La señora colocó *un* **knife** next to every plate
The woman placed a_{MASC} $\text{knife}_{\text{MASC}}$

(b) La señora colocó *un* **spoon** next to every plate
The woman placed a_{MASC} $\text{spoon}_{\text{FEM}}$

Current work in our research group is aimed at determining the extent to which codeswitching patterns are community-specific or generalizable across different speech communities of the Spanish-speaking world. To explore this issue, we have designed a conversational paradigm to obtain spontaneous speech samples of bilingual speakers (Beatty-Martínez and Dussias, 2017; Beatty-Martínez et al., 2018a). In the task, participants are assigned the role of director and are instructed to communicate to a matcher addressee how to arrange a series of images printed on a map. To maximize ecological validity, no language restrictions are imposed; that is, participants are free to use whichever language they choose. The project resulted in four comparable corpora of over 100 Spanish-English bilingual young adults from four linguistically distinct interactional contexts (San Juan (PR), El Paso (TX), State College (PA), and Granada (Spain)). Based on these data, Królikowska et al. (2019) asked whether all groups showed the attested preference for masculine determiners before switching to an English noun, regardless of the gender of the translation equivalent.

Figure 1 illustrates an asymmetric relation between masculine and feminine grammatical gender assignment across all four groups. For bilinguals in San Juan and State College, the data show an overwhelming preference for masculine determiners, regardless of the grammatical gender of the Spanish translation equivalent. Moreover, while bilinguals in Granada and El Paso also exhibited higher rates of masculine determiners overall, they also produced higher rates of feminine determiners than the other two groups. Specifically, masculine and feminine determiners were produced at similar rates for nouns with feminine translation equivalents (e.g., "*la* spoon," the_{FEM} $\text{spoon}_{\text{FEM}}$).

Although more work is needed to unpack these results, one possible explanation for the variability between these four contexts is that bilinguals from these communities exhibit different rates of codeswitching overall. **Figure 2** depicts rates of unilingual (e.g., English: "the dog"; or Spanish: "el perro") and mixed (e.g., "el dog") NPs across the four testing locations. Bilinguals from San Juan had the highest rate of codeswitching at almost 24%, while bilinguals from Granada had the lowest at 2%. Therefore, one possibility is that the more the bilinguals engage in codeswitching, the greater the tendency to assign the default masculine gender to mixed NPs. This is an important

[3]For ease of exposition, we adopt a broad definition of codeswitching to include single word and multiword constituents.

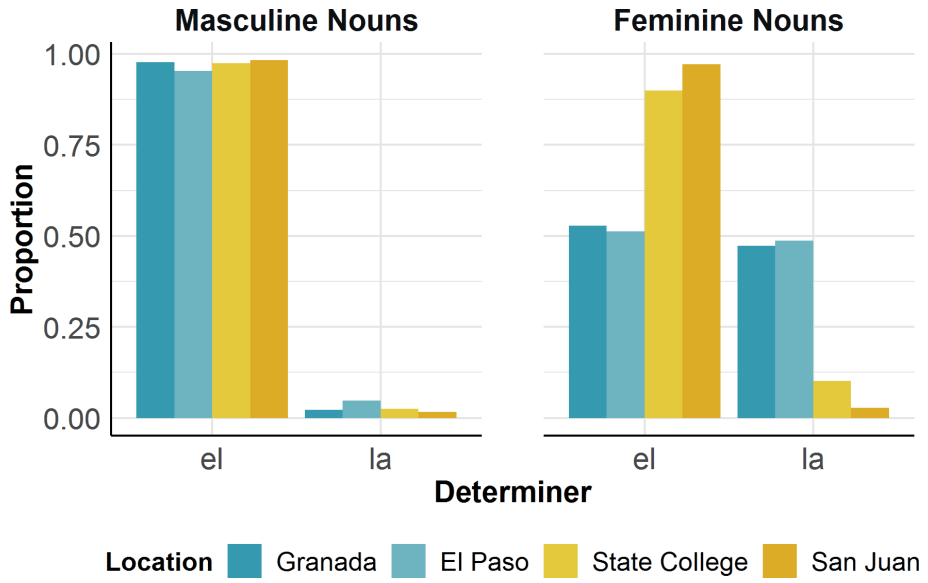


FIGURE 1 | Distribution of mixed NPs across four bilingual communities in Królikowska et al. (2019).

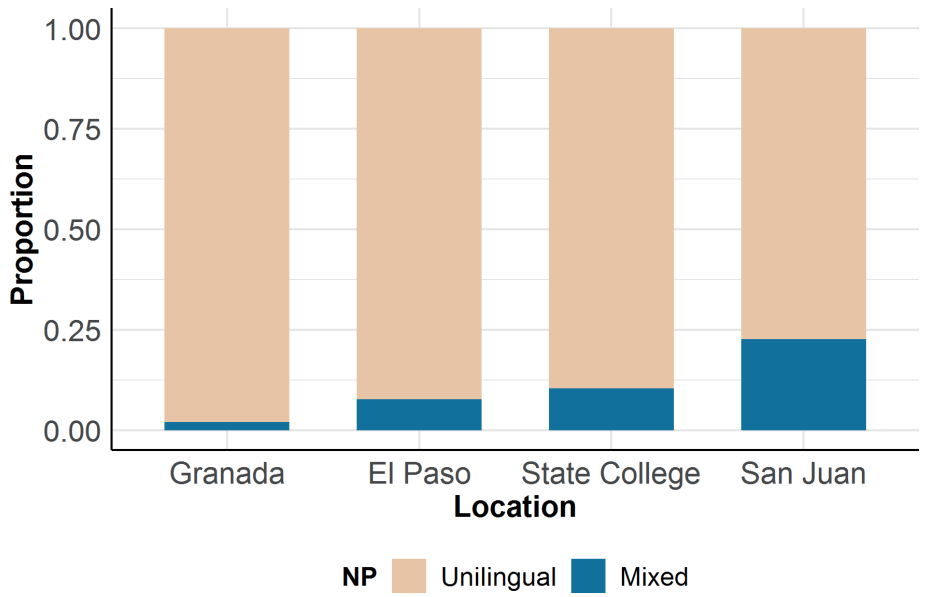


FIGURE 2 | Rates of expression of unilingual and mixed NPs across four bilingual communities in Królikowska et al. (2019).

observation that supports previous claims that codeswitching preferences reflect community norms and are therefore not necessarily generalizable across bilingual populations, even when examining the same language pair (Poplack, 1988; Aaron, 2015; Beatty-Martínez et al., 2018a).

Because most English words differ from typical Spanish words with respect to their phonological shape (Clegg, 2010; Butt and Benjamin, 2013), it is difficult to determine whether the masculine default strategy is, at least to some degree, driven by phonological factors (Poplack et al., 1982; DuBord, 2004; Montes-Alcalá and Lapidus Shin, 2011). Below, we consider two recent studies that examined how the phonological shape of nouns from different source languages (i.e., Basque and Purepecha) can influence the choices speakers make in terms of the choice of gender assignment.

Parafita Couto et al. (2015) examined grammatical gender assignment strategies of Spanish-Basque NPs in naturalistic speech and auditory judgement data. Basque differs from Spanish and English in its morphological behavior and NP word order. In Basque, the definite determiner *-a* appears suffixed to the noun (e.g. "sagarr-*a*," the apple) which is coincidentally homophonous with the regular feminine endings in Spanish

(e.g., "la manzana"). The naturalistic data indicated a preference for the feminine determiner when it was congruent with the Basque phonological ending *-a*, providing converging evidence for the role of a noun's phonological shape in gender assignment.

In a similar study, Bellamy et al. (2018) examined gender assignment in Spanish-Purepecha mixed NPs using a production task and an online acceptability judgement task. Like Basque, Purepecha has bound suffixes terminating in *-a* that coincides with phonological cues to feminine gender assignment in Spanish. In the production task, participants overwhelmingly preferred to use masculine determiners, irrespective of the noun ending or Spanish translation equivalent. In the acceptability judgement task, participants also preferred masculine assignment except in cases where nouns ended in *-a*. Bellamy et al. interpreted this result to indicate that orthography can lead speakers to re-interpret the *-a* ending suffix, a marker of feminine gender. Furthermore, the discrepant findings of these tasks provide evidence that the modality of the task can influence gender agreement strategies in Spanish speakers. Taken together, these studies highlight how preferences in gender agreement are susceptible to both cross-language effects and the type of task. In the next section, we consider how bilingual language experience can lead to the same adaptive consequences in predictive processing.

IMPLICATIONS FOR LANGUAGE PROCESSING

Eye-Tracking Evidence

We discussed earlier how the study of codeswitching provides a unique lens through which the differential status of masculine and feminine gender in Spanish can be examined. The distributional patterns outlined in the "Evidence From Bilingual Speakers" section on the use of grammatical gender in Spanish-English mixed noun phrases raise the question of whether the asymmetries observed in Spanish-English mixed NPs has consequences for the comprehension system, as would be predicted by experience-based models of language processing (e.g., MacDonald, 2013; Dell and Chang, 2014). Initial results indicate that they do. In a series of eye-tracking experiments, Valdés Kroff et al. (2016) capitalized on competitor (Allopenna et al., 1998) and anticipatory (Lew-Williams and Fernald, 2007) effects reported in studies of spoken language processing using the visual world paradigm (Tanenhaus et al., 1995) to examine whether the overwhelming preference for the Spanish masculine article in codeswitched noun phrases had any consequences for the comprehension system. Target items in the codeswitching condition were made up of a Spanish preamble ("Encuentra *el*/*la*," find the$_{MASC}$/the$_{FEM}$) followed by an English target noun, yielding mixed NPs such as "*Encuentra el candy*." To provide a test of the hypothesis that speakers exploit feminine but not masculine cues on determiners to anticipate upcoming nouns, they incorporated an additional manipulation. The mixed NPs contained pairs of items that were phonological competitors in English. For example, *candy* and *candle* overlap phonologically in the first syllable [kæn], but critically their Spanish translations differ in grammatical gender; *candy* is English for dulce$_{MASC}$ or caramelo$_{MASC}$ and *candle* is English for vela$_{FEM}$. Because in mixed NPs, the pattern from corpus studies suggests that the definite article *el* surfaces with English nouns whose Spanish translations are both masculine and feminine, the prediction was that the gender information encoded in the article would not facilitate the processing of sentences such as "Encuentra *el* candy." Instead, the presence of phonological competitors should evince a competitor effect, and this is precisely what they found. When a masculine article was heard in the presence of the picture pair *candle-candy*, the results showed a clear competitor effect, suggesting that the masculine article *el* was not informative when bilinguals were asked to select a noun. In other words, it functioned as a default article in Spanish-English codeswitching. When a feminine article was heard in the presence of the same two pictures (i.e., "Encuentra *la* candle"), the results showed a different pattern. Participants failed to display an anticipatory effect and instead experienced an extended delay in processing for target items that did not match in grammatical gender (e.g., *la* candy) likely reflecting the rarity of this type of mixed NP in Spanish-English codeswitching.

Electrophysiological Evidence

Thus far, we have argued that the distributional asymmetry between masculine and feminine gender reflects underlying differences in the representation of the two genders. In this section, we turn to electrophysiological studies of grammatical gender to examine possible differences in processing and representation for masculine and feminine nouns in unilingual and mixed NPs. In contrast to behavioral measures, which reflect the cumulative outcome of several processes, the event-related potentials (ERPs) technique can provide high temporal resolution indices at different stages of processing, which is reflected in modulations of distinguishable components. Importantly, ERPs have been found to be modulated by different linguistic processes, including morphological markedness (Deutsch and Bentin, 2001; Kaan, 2002; Alemán Bañón and Rothman, 2016), making this technique particularly suitable to uncover potential differences in the processing of masculine and feminine grammatical gender.

ERPs have been widely employed to investigate the time course of noun phrase grammatical gender processing in both monolingual (Wicha et al., 2004; Barber and Carreiras, 2005; Caffarra and Barber, 2015) and bilingual (Caffarra et al., 2017a) speakers. The general finding is that grammatical gender violations in Spanish elicit a biphasic pattern, consisting of a Left Anterior Negativity (LAN) around 300 ms after stimulus onset and a subsequent P600 after 500 ms.[4] The LAN effect has been suggested to reflect initial processes for detection of a

[4]We note that the LAN is not consistently observed across studies examining morphosyntactic violations (e.g., Wicha et al., 2004; Alemán Bañón et al., 2012). Recent reports have questioned whether this effect even exists (Tanner, 2015), although others disagree (Molinaro et al., 2015; Caffarra et al., 2017b).

morphosyntactic violation (Osterhout, 1997). The P600 effect has been linked to processes of reanalysis and repair of syntactic anomalies (Osterhout and Holcomb, 1992; Friederici et al., 1996; Kaan et al., 2000).

Caffarra and Barber (2015) investigated whether distributional gender cues conveyed by Spanish noun endings (i.e., *-a* for feminine and *-o* for masculine) can influence gender processing in native Spanish speakers. Nouns with regular endings elicited a greater sustained negativity around 200 ms after the stimulus onset suggesting that Spanish speakers are sensitive to noun endings (see Halberstadt et al., 2018 for related findings with second language speakers of Spanish using eye-tracking methodology). Notwithstanding, a LAN-P600 biphasic pattern was similarly reported for gender violations for both regular and irregular nouns. Based on these findings, the authors concluded that grammatical gender agreement processes rely mostly on the representation of gender, regardless of distributional gender cues conveyed by noun endings. Using the same paradigm, Caffarra et al. (2017a) replicated these results with Spanish-Basque bilinguals but observed that participants who reported using Spanish more regularly were able to detect violations for irregular nouns earlier and more easily than those who were Basque dominant. These results highlight the role of regular correspondence between the word form and a specific gender class and, more broadly, of experience that users have with language in category learning and representation. At the same time, these findings also suggest that lexical representations may become more entrenched with greater language experience, resulting in more efficient processing.

A few studies have investigated gender agreement processes of masculine and feminine genders separately using ERPs. Alemán Bañón and Rothman (2016) examined the brain's sensitivity to noun-adjective agreement violations during online sentence comprehension. ERPs were time-locked to adjectives appearing predicatively in relative clauses. In their design, half of the items were masculine and the other half were feminine. They found that both types of gender agreement violations yielded robust P600 effects albeit earlier for feminine-marked adjectives. Alemán Bañón and Rothman interpreted the difference in latency as evidence that violations realized on marked predicates are easier to detect and thus revised more quickly, consistent with previous work on syntactic processes of diagnosis and repair (e.g., Friederici, 1998; Kaan, 2002). Notwithstanding, the processing of noun-adjective agreement has been shown to differ from the processing of gender assignment with nouns (Dewaele and Véronique, 2001; Barber and Carreiras, 2003; Kupisch et al., 2013), and while adjectives and nouns have overlapping cues to gender, there are differences in marking consistency between the two elements. It follows that a manipulation of gender agreement ultimately addresses a different question than the one we ask here: If the attested distributional asymmetries in gender assignment reflect differences intrinsic to the structure of nouns (e.g., Eddington, 2002) and speakers have been shown to attend to and make use of these cues in production, what consequences do these adjustments have for lexical processing and representation?

To our knowledge, only two studies have compared gender processes in nouns as a function of their gender in Spanish. Beatty-Martínez and Dussias (2017) examined gender processing in mixed NPs for bilinguals differing in codeswitching experience (i.e., codeswitchers and non-codeswitchers). In their design, the gender of the target noun (i.e., the gender of its translation equivalent in Spanish; e.g., masculine: "knife," cuchillo$_{MASC}$ or feminine: "spoon," cuchara$_{FEM}$) was manipulated such that it either agreed in gender with the preceding determiner (congruent condition: "*el* knife," the$_{MASC}$ knife$_{MASC}$) or not (incongruent condition: "*la* knife," the$_{FEM}$ knife$_{MASC}$). For codeswitchers, masculine targets in incongruent mixed NPs (e.g., "*la* knife") were more difficult to integrate relative to masculine targets in congruent mixed NPs (e.g., "*el* knife"; **Figure 3A**). Importantly, incongruent mixed NPs with masculine determiners (e.g., "*el* spoon") did not result in processing difficulties (**Figure 3B**). The authors interpreted this result as evidence for bilinguals' sensitivity to distributional codeswitching patterns (i.e., incongruent mixed NPs with feminine determiners are rarely attested in naturalistic codeswitching; Valdés Kroff, 2016; Beatty-Martínez et al., 2018a). Non-codeswitchers, on the other hand, only showed sensitivity to agreement violations for mixed NPs involving feminine translation equivalents: incongruent mixed NPs (e.g., "*el* spoon") elicited a P600 effect (**Figure 3D**). While the absence of the P600 in incongruent mixed NPs involving masculine translation equivalents (e.g., "*la* knife"; **Figure 3C**) is likely due to substantial variability in participants' responses (McLaughlin et al., 2010; Qi et al., 2017), these differences in themselves are likely indications of the differential representation of masculine and feminine gender.

An alternative explanation proposed in the Caffarra et al. (2017a) study is that knowledge and usage of a second language may influence the strength of gender lexical representation, and that therefore, bilinguals may not rely on gender features in the same way as native speakers. We would like to take this proposal a step further and assume that variability in grammatical gender processing exists even among monolinguals processing their native language (see Tanner et al., 2014, for a discussion on "native-like" processing). We consider a recent study whose findings may provide insights into this issue. Beatty-Martínez et al. (2018b) examined the electrophysiological correlates of masculine and feminine gender violations in native monolingual Spanish speakers. Specifically, ERPs were recorded while participants read sentences in Spanish that were either well-formed or contained grammatical gender violations. Half of the target nouns were masculine (e.g., "cuchillo," knife) and half were feminine (e.g., "cuchara," spoon) in gender. When collapsed across gender, the gender violation showed the classical LAN-P600 biphasic pattern. However, splitting the data by noun gender revealed different ERP patterns to masculine and feminine gender. Responses to masculine grammatical gender violations had far greater variability and showed a reduced P600 (**Figure 4A**). This is consistent with previous studies showing reduced sensitivity to morphological violations involving unmarked elements (Deutsch and Bentin, 2001; Kaan, 2002; Alemán Bañón and Rothman, 2016). As illustrated in **Figure 4B**, feminine gender violations elicited a more robust P600 response that was modulated by

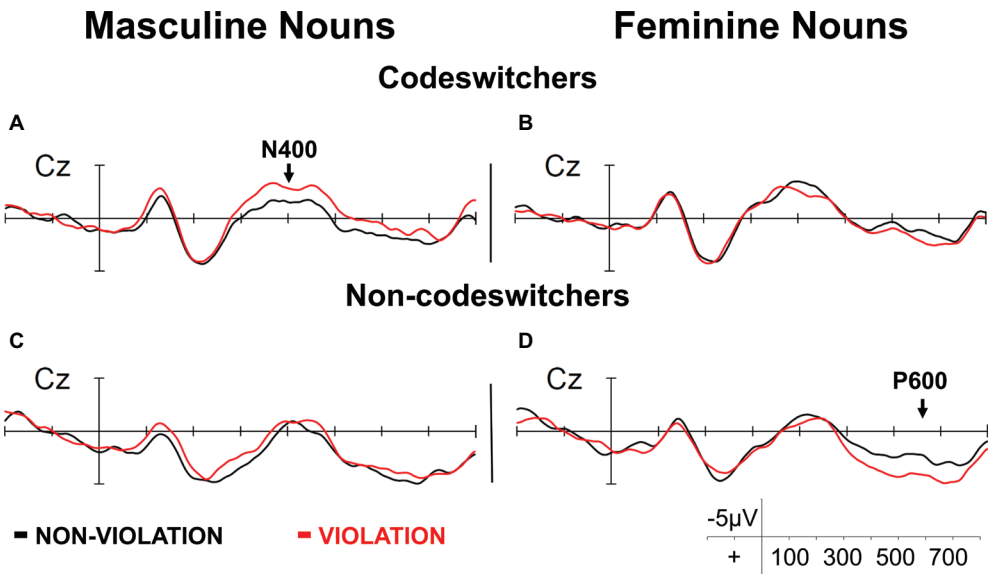


FIGURE 3 | ERPs time-locked to the onset of masculine **(A,C)** and feminine **(B,D)** nouns for codeswitchers **(A,B)** and non-codeswitchers **(C,D)** at the electrode site Cz. Figure adapted from Beatty-Martínez and Dussias, 2017, Copyright (2017), with permission from Elsevier.

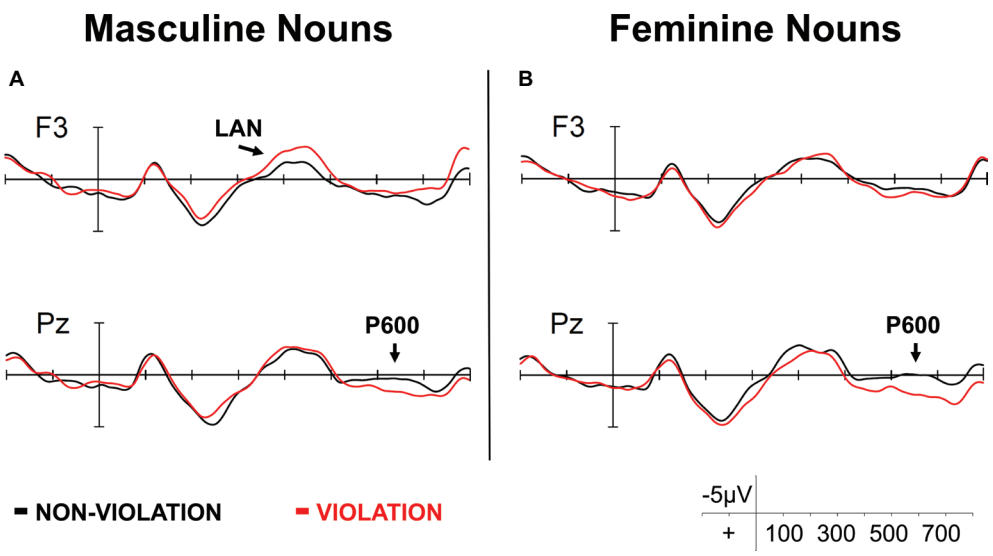


FIGURE 4 | ERPs time-locked to the onset of masculine **(A)** and feminine **(B)** nouns at F3 and Pz electrode sites adapted from Beatty-Martínez et al. (2018b).

vocabulary knowledge: individuals with higher Spanish vocabulary were more sensitive to grammatical gender violations with feminine nouns. We interpret this finding to suggest that as vocabulary increases, so does the strength of the representation of noun clusters, supporting the more general idea that experience with language affects the structure of categories and has an impact on cognitive representations (e.g., Bybee, 2010). Together, the results in this section provide support for a differential representation between masculine and feminine gender by demonstrating that variability in gender processing exists even among groups traditionally assumed to be homogenous.

CONCLUSION

The main objective of this paper was to examine distributional asymmetries between masculine and feminine gender, the resulting biases in gender assignment, and the consequences of these assignment strategies on gender expectancy and processing. While the available evidence is not conclusive, a striking feature that emerges from this review is an underlying difference in the representation and processing of masculine and feminine gender in Spanish. What does this difference mean for our understanding of grammatical gender?

The processing results reported here, together with the acquisition data, suggest that assumptions made in past processing literature, which have treated different gender classes similarly, is unwarranted. Grammatical gender has been extensively studied in a wide variety of disciplines, yet there is often little crosstalk between different fields of study. Within the second language processing literature for example, grammatical gender has served as the benchmark of native-like attainment, with some studies reporting differential sensitivity in the second language and others arguing against such differences. The evidence presented here contributes to this debate through a consideration of distributional factors in explaining differences in grammatical gender processing.

While distributional asymmetries are not necessarily language specific, we caution against generalizing the specific biases arising in Spanish across other gendered languages for several reasons. First, languages differ with respect to how gender classes are distributed. While masculine and feminine gender are distributed approximately equally in Spanish (Bull, 1965), other languages with a binary gender system have a less balanced distribution (e.g., about 3:1 ratio for masculine and neuter nouns in Dutch; van Berkum, 1996). Gendered languages also differ in the degree to which gender assignment can be made in terms of phonological shape or morphological composition. For example, historical sound change in French turned regular feminine endings to schwas (e.g., "fenestra → fenêtre," window), resulting in greater phonic ambiguity in the endings of masculine and feminine nouns (Nelson, 2005). Moving forward, we suggest that more interdisciplinary studies are needed to exploit the consequences of distributional regularities on language processing. More broadly, processing research must proceed from a distinct set of assumptions regarding the status of grammatical gender, adopting an approach in which gender is not viewed as a single concept but rather recognized as a complex and granular phenomenon, whose processing reflexes may exhibit surprising asymmetries.

AUTHOR CONTRIBUTIONS

Both authors have contributed equally to the manuscript and approved it for publication.

FUNDING

The writing of this paper was supported in part by A Ford Foundation Dissertation Fellowship to AB-M and by NSF grants BCS-1535124 and OISE-1545900 to PD.

ACKNOWLEDGMENTS

We are grateful to Rena Torres Cacoullos and Christian A. Navarro-Torres for helpful comments and discussions during the preparation of this manuscript. We would also like to thank the editor and the reviewers for the comments on the earlier versions of the manuscript.

REFERENCES

Aaron, J. E. (2015). Lone English-origin nouns in Spanish: the precedence of community norms. *Int. J. Biling.* 19, 459–480. doi: 10.1177/1367006913516021

Alemán Bañón, J., Fiorentino, R., and Gabriele, A. (2012). The processing of number and gender agreement in Spanish: an event-related potential investigation of the effects of structural distance. *Brain Res.* 1456, 49–63. doi: 10.1016/j.brainres.2012.03.057

Alemán Bañón, J., and Rothman, J. (2016). The role of morphological markedness in the processing of number and gender agreement in Spanish: an event-related potential investigation. *Lang. Cogn. Neurosci.* 31, 1273–1298. doi: 10.1080/23273798.2016.1218032

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439. doi: 10.1006/jmla.1997.2558

Barber, H., and Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: an ERP study. *Cortex* 39, 465–482. doi: 10.1016/S0010-9452(08)70259-4

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Bates, E., Devescovi, A., Hernandez, A., and Pizzamiglio, L. (1996). Gender priming in Italian. *Percept. Psychophys.* 58, 992–1004. doi: 10.3758/BF03206827

Beatty-Martínez, A. L., Bruni, M., Navarro-Torres, C. A., Togato, G., and Dussias, P. E. (2018a). Linguistic and experiential factors reveal fundamental variation in grammatical gender processing. Poster Presented at the 59th Annual Meeting of the Psychonomic Society. New Orleans, LA.

Beatty-Martínez, A. L., and Dussias, P. E. (2017). Bilingual experience shapes language processing: evidence from codeswitching. *J. Mem. Lang.* 95, 173–189. doi: 10.1016/j.jml.2017.04.002

Beatty-Martínez, A. L., and Dussias, P. E. (2018). Tuning to languages: experience-based approaches to the language science of bilingualism. *Ling. Van.* 4, 1–9. doi: 10.1515/lingvan-2017-0034

Beatty-Martínez, A. L., Valdés-Kroff, J. R., and Dussias, P. E. (2018b). From the field to the lab: a converging methods approach to the study of codeswitching. *Language* 3:19. doi: 10.3390/languages3020019

Bellamy, K., Parafita Couto, M. C., and Stadthagen-Gonzalez, H. (2018). Investigating gender assignment strategies in mixed Purepecha–Spanish nominal constructions. *Language* 3:28. doi: 10.3390/languages3030028

Blokzijl, J., Deuchar, M., and Parafita Couto, M. C. (2017). Determiner asymmetry in mixed nominal constructions: the role of grammatical factors in data from Miami and Nicaragua. *Language* 2:20. doi: 10.3390/languages2040020

Bull, W. (1965). *Spanish for teachers: Applied linguistics*. (New York, NY: Ronald Press Co).

Butt, J. B., and Benjamin, C. (2013). *A new reference grammar of modern Spanish*. (London, UK: Routledge).

Bybee, J. (2010). *Language, usage and cognition*. (New York, NY: Cambridge University Press).

Caffarra, S., and Barber, H. A. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Res.* 1605, 83–92. doi: 10.1016/j.brainres.2015.02.018

Caffarra, S., Barber, H., Molinaro, N., and Carreiras, M. (2017a). When the end matters: influence of gender cues during agreement computation in bilinguals. *Lang. Cogn. Neurosci.* 32, 1069–1085. doi: 10.1080/23273798.2017.1283426

Caffarra, S., Mendoza, M., and Davidson, D. (2017b). Is the LAN effect in morphosyntactic processing an ERP artifact? *bioRxiv*:218594. doi: 10.1101/218594

Carreiras, M., Oakhill, J., and Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: evidence from English and Spanish. *Q. J. Exp. Psychol.* 49A, 639–663. doi: 10.1080/713755647

Clausner, T., and Croft, W. (1997). The productivity and schematicity of metaphor. *Cogn. Sci.* 21, 247–282. doi: 10.1207/s15516709cog2103_1

Clayards, M., Tanenhaus, M., Aslin, R., and Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809. doi: 10.1016/j.cognition.2008.04.004

Clegg, J. H. (2010). Native Spanish speaker intuition in noun gender assignment. *Language Design: Journal of Theoretical and Experimental Linguistics* 12, 5–18.

Clegg, J., and Waltermire, M. (2009). Gender assignment to English-origin nouns in the Spanish of the Southwestern United States. *Sw. J. Ling.* 28, 1–17.

Cole, P., and Segui, J. (1994). Grammatical incongruency and vocabulary types. *Mem. Cogn.* 22, 387–394. doi: 10.3758/BF03200865

Comrie, B. (1999). Grammatical gender systems: a linguist's assessment. *J. Psycholinguist. Res.* 28, 457–466. doi: 10.1023/A:1023212225540

Corbett, G. (1991). *Gender*. (Cambridge, England: Cambridge University Press).

Costa, A., Kovacic, D., Fedorenko, E., and Caramazza, A. (2003). The gender congruency effect and the selection of freestanding and bound morphemes: evidence from Croatian sumption that the contextual effects observed with this paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1270–1282. doi: 10.1037/0278-7393.29.6.1270

Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., and Job, R. (2005). Grammatical gender is selected in bare noun production: evidence from the picture-word interference paradigm. *J. Mem. Lang.* 53, 42–59. doi: 10.1016/j.jml.2005.02.007

Dahan, D., Swingley, D., Tanenhaus, M. K., and Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in French. *J. Mem. Lang.* 42, 465–480. doi: 10.1006/jmla.1999.2688

De la Cruz Cabanillas, I., Martínez, C. T., Prados, M. D., and Redondo, E. C. (2007). English loanwords in Spanish computer language. *J. ESP* 26, 52–78. doi: 10.1016/j.esp.2005.06.002

Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. B* 369:20120394. doi: 10.1098/rstb.2012.03

Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224. doi: 10.1006/jmla.2000.2768

Dewaele, J.-M., and Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: a cross-sectional study. *Biling. Lang. Cognit.* 4, 275–297. doi: 10.1017/S136672890100044X

Domínguez, A., Cuetos, F., and Segui, J. (1999). The processing of grammatical gender and number in Spanish. *J. Psycholinguist.* 28, 485–498. doi: 10.1023/A:1023216326448

DuBord, E. M. (2004). Gender assignment to English words in the Spanish of Southern Arizona. *Divergencias: Revista de estudios lingüísticos y literarios* 2, 27–39.

Eddington, D. (2002). Spanish gender assignment in an analogical framework. *J. Quant. Linguist.* 9, 49–75. doi: 10.1076/jqul.9.1.49.8482

Eddington, D., and Hualde, J. I. (2008). El abundante agua fría: hermaphroditic Spanish nouns. *Stud. Hispanic Lusoph. Linguist.* 1, 5–32. doi: 10.1515/shll-2008-1004

Faussart, C., Jakubowicz, C., and Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Int. J. Lexicogr.* 11, 75–101.

Friederici, A. D. (1998). "Diagnosis and reanalysis: two processing aspects the brain may differentiate" in *Reanalysis in Sentence Processing: Studies in Theoretical Psycholinguistics*. eds. J. D. Fodor and F. Ferreira (Dordrech: Kluwer), 177–200. doi: 10.1007/978-94-015-9070-9_6

Friederici, A. D., Hahne, A., and Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1219–1248. doi: 10.1037/0278-7393.22.5.1219

Friederici, A. D., and Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *J. Psycholinguist. Res.* 28, 467–484. doi: 10.1023/A:1023264209610

Gabriel, U., and Gygax, P. (2016). "Gender and linguistic sexism" in *Language as social action: Advances in intergroup communication*. eds. H. Giles and A. Maass (New York, NY, US: Peter Lang Publishing), 177–192.

Gennari, S. P., and MacDonald, M. C. (2009). Linking production and comprehension processes: the case of relative clauses. *Cognition* 111, 1–23. doi: 10.1016/j.cognition.2008.12.006

Green, D. W., and Wei, L. (2014). A control process model of code-switching. *Lang. Cogn. Neurosci.* 29, 499–511. doi: 10.1080/23273798.2014.882515

Greenberg, J. H. (1966). "Language universals" in *Current Trends in Linguistics*. ed. T. A. Sebeok (The Hague: Mouton), 60–112.

Grosjean, F., Dommergues, J. Y., Cornu, E., Guillelmon, D., and Besson, C. (1994). The gender-marking effect in spoken word recognition. *Percept. Psychophys.* 56, 590–598. doi: 10.3758/BF03206954

Gurjanov, M., Lukatela, G., Lukatela, K., Savić, M., and Turvey, M. T. (1985). Grammatical priming of inflected nouns by the gender of possessive adjectives. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 692–701. doi: 10.1037/0278-7393.11.1-4.692

Halberstadt, L., Valdés Kroff, J. R., and Dussias, P. E. (2018). Grammatical gender processing in L2 speakers of Spanish: the role of cognate status and gender transparency. *JSLS* 1, 5–30. doi: 10.1075/jsls.17023.hal

Harris, J. (1985). "Spanish word markers" in *Current issues in Hispanic phonology and morphology*. ed. F. H. Nuessel Jr. (Bloomington, IN: Indiana University Linguistics Club), 34–54.

Harris, J. (1991). The exponence of gender in Spanish. *Linguist Inquiry.* 22, 27–62.

Hockett, C. F. (1958). *A course in modern linguistics*. (New York, NY: Macmillan).

Jacobsen, T. (1999). Effects of grammatical gender on picture and word naming: evidence from German. *J. Psycholinguist. Res.* 28, 499–514. doi: 10.1023/A:1023268310519

Jake, J. L., Myers-Scotton, C., and Gross, S. (2002). Making a minimalist approach to codeswitching work: adding the matrix language. *Biling. Lang. Cognit.* 5, 69–91. doi: 10.1017/S1366728902000147

Jakubowicz, C., and Faussart, C. (1998). Gender agreement in the processing of spoken French. *J. Psycholinguist. Res.* 27, 597–617. doi: 10.1023/A:1023297620824

Jescheniak, J. D. (1999). Gender priming in picture naming: modality and baseline effects. *J. Psycholinguist. Res.* 28, 729–737. doi: 10.1023/A:1023229329967

Juilland, A., and Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. (The Hague, NL: Mouton).

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *J. Psycholinguist. Res.* 31, 165–193. doi: 10.1023/A:1014978917769

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Królikowska, M., Bierings, E., Beatty-Martínez, A., Navarro-Torres, C., Dussias, P., and Parafita Couto, M. C. (2019). "Gender-assignment strategies within the bilingual determiner phrase: four Spanish-English communities examined" in *Poster Presented at the 3rd Conference on Bilingualism in the Hispanic and Lusophone World (BHL).* (Leiden: NL).

Kupisch, T., Akpinar, D., and Stöhr, A. (2013). Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguist. Approaches Biling.* 3, 150–179. doi: 10.1075/lab.3.2.02kup

Lew-Williams, C., and Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychol. Sci.* 18, 193–198. doi: 10.1111/j.1467-9280.2007.01871.x

Liceras, J. M., Fernández Fuertes, R., Perales, S., Pérez-Tattam, R., and Spradlin, K. T. (2008). Gender and gender agreement in bilingual native and non-native grammars: a view from child and adult functional-lexical mixings. *Lingua* 118, 827–851. doi: 10.1016/j.lingua.2007.05.006

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Front. Psychol.* 4:226. doi: 10.3389/fpsyg.2013.00226

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., et al (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Lang. Learn.* 60, 123–150. doi: 10.1111/j.1467-9922.2010.00604.x

Molinaro, N., Barber, H. A., Caffarra, S., and Carreiras, M. (2015). On the left anterior negativity (LAN): the case of morphosyntactic agreement. *Cortex* 66, 156–159. doi: 10.1016/j.cortex.2014.06.009

Montes-Alcalá, C., and Lapidus Shin, N. (2011). Las keys versus el key: feminine gender assignment in mixed-language texts. *Stud. Inform. Control.* 8, 119–143. doi: 10.1075/sic.8.1.06mon

Myers-Scotton, C., and Jake, J. L. (2015). "Cross-language asymmetries in code-switching patterns: implications for bilingual language production" in *The*

Cambridge handbook of bilingual processing. ed. J. Schwieter (Cambridge: Cambridge University Press), 416–458.

Nelson, D. (2005). French gender assignment revisited. *Word* 56, 19–38. doi: 10.1080/00437956.2005.11432551

Osterhout, L. (1997). On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain Lang.* 59, 494–522. doi: 10.1006/brln.1997.1793

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., and Oyharçabal, B. (2015). Gender conflict resolution in Spanish–Basque mixed DPs. *Biling. Lang. Cognit.* 18, 304–323. doi: 10.1017/S136672891400011X

Pérez-Pereira, M. (1991). The acquisition of gender: what Spanish children tell us. *J. Child Lang.* 18, 571–590. doi: 10.1017/S0305000900011259

Pfaff, C. (1979). Constraints on language mixing: intrasentential code-switching and borrowing in Spanish/English. *Language* 55, 291–318. doi: 10.2307/412586

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of codeswitching. *Linguistics* 18, 581–618. doi: 10.1515/ling.1980.18.7-8.581

Poplack, S. (1988). "Contrasting patterns of code-switching in two communities" in *Codeswitching: Anthropological and Sociolinguistic Perspectives*. ed. M. Heller. (Berlin: Mouton), 215–244.

Poplack, S., Pousada, A., and Sankoff, D. (1982). Competing influences on gender assignment: variable process, Stable Outcome. *Lingua* 57, 1–28. doi: 10.1016/0024-3841(82)90068-7

Qi, Z., Beach, S. D., Finn, A. S., Minas, J., Goetz, C., Chan, B., et al (2017). Native-language N400 and P600 predict dissociable language-learning abilities in adults. *Neuropsychologia* 98, 177–191. doi: 10.1016/j.neuropsychologia.2016.10.005

Roca, I. M. (1989). The organisation of grammatical gender. *Trans. Phil. Soc.* 87, 1–32. doi: 10.1111/j.1467-968X.1989.tb00617.x

Schmidt, R. (1986). Was weiß der Artikel vom Hauptwort? Ein Beitrag zur Verarbeitung syntaktischer Beziehungen beim Lesen. *Z. Exp. Angew. Psychol.* 33, 150–163.

Smead, R. N. (2000). On the assignment of gender to Chicano anglicisms: processes and results. *Biling. Rev.* 25, 277–297.

Smith, P., Nix, A., Davey, N., López Ornat, S., and Messer, D. (2003). A connectionist account of Spanish determiner production. *J. Child Lang.* 30, 305–331. doi: 10.1017/S0305000903005622

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi: 10.1126/science.7777863

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: a commentary on "Grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex* 66, 149–155. doi: 10.1016/j.cortex.2014.04.007

Tanner, D., Inoue, K., and Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Biling. Lang. Cognit.* 17, 277–293. doi: 10.1017/S1366728913000370

Timm, L. (1975). Spanish-English code-mixing: el porqué y el how-not-to. *RPH* 28, 473–482.

Valdés Kroff, J. R. (2016). "Mixed NPs in Spanish-English bilingual speech: using a corpus-based approach to inform models of sentence processing" in *Spanish-English Codeswitching in the Caribbean and the US*. eds. R. E. Guzzardo, C. M. Mazak, and M. C. Parafita (Amsterdam, NL: John Benjamins), 281–300.

Valdés Kroff, J. R., Dussias, P. E., Gerfen, C., Perrotti, L., and Bajo, M. T. (2016). Experience with code-switching modulates the use of grammatical gender during sentence processing. *Linguist. Approaches Biling.* 7, 163–198. doi: 10.1075/lab.15010.va

van Berkum, J. J. A. (1996). The psycholinguistics of grammatical gender: studies in language comprehension and production. doctoral dissertation. (Nijmegen (NL): Max Planck Institute for Psycholinguistics).

Wicha, N. Y. Y., Moreno, E. M., and Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* 16, 1272–1288. doi: 10.1162/0898929041920487

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 April 2019
doi: 10.3389/fpsyg.2019.00746

Being a *Participant* Matters: Event-Related Potentials Show That Markedness Modulates Person Agreement in Spanish

José Alemán Bañón[1]* *and Jason Rothman*[2,3]

[1] *Centre for Research on Bilingualism, Department of Swedish and Multilingualism, Stockholm University, Stockholm, Sweden,* [2] *Department of Languages and Linguistics, Faculty of Humanities, Social Sciences, and Education, The Arctic University of Norway, Tromsø, Norway,* [3] *Centro de Ciencia Cognitiva, Universidad Nebrija, Madrid, Spain*

The present study uses event-related potentials to examine subject–verb person agreement in Spanish, with a focus on how markedness with respect to the speech participant status of the subject modulates processing. Morphological theory proposes a markedness distinction between first and second person, on the one hand, and third person on the other. The claim is that both the first and second persons are *participants* in the speech act, since they play the *speaker* and *addressee* roles, respectively. In contrast, third person refers to whomever is neither the *speaker* nor the *addressee* (i.e., it is unmarked for person). We manipulated speech participant by probing person agreement with both first-person singular subjects (e.g., *yo*. . .*lloro* "I. . .cry-$_{1ST\,PERSON-SG}$") and third-person singular ones (e.g., *la viuda*. . .*llora* "the widow. . .cry-$_{3RD\,PERSON-SG}$"). We also manipulated agreement by crossing first-person singular subjects with third-person singular verbs (e.g., *yo*. . .**llora* "I. . .cry-$_{3RD\,PERSON-SG}$") and vice versa (e.g., *la viuda*. . .**lloro* "the widow. . .cry-$_{1ST\,PERSON-SG}$"). Results from 28 native speakers of Spanish revealed robust positivities for both types of person violations, relative to their grammatical counterparts between 500 and 1000 ms, an effect that shows a central-posterior distribution, with a right hemisphere bias. This positivity is consistent with the P600, a component associated with a number of morphosyntactic operations (and reanalysis processes more generally). No negativities emerged before the P600 (between 250 and 450 ms), although both error types yielded an anterior negativity in the P600 time window, an effect that has been argued to reflect the memory costs associated with keeping the errors in working memory to provide a sentence-final judgment. Crucially, person violations with a marked subject (e.g., *yo*. . .**llora* "I. . .cry-$_{3RD\,PERSON-SG}$") yielded a larger P600 than the opposite error type between 700 and 900 ms. This effect is consistent with the possibility that, upon encountering a subject with marked features, feature activation allows the parser to generate a stronger prediction regarding the upcoming verb. The larger P600 for person violations with a marked subject might index the reanalysis process that the parser initiates when there is a conflict between a highly expected verbal form (i.e., more so than in the conditions with an unmarked subject) and the form that is actually encountered.

Keywords: ERP, P600, late anterior negativity, markedness, person agreement, prediction, Spanish

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Nicoletta Biondo,
Basque Center on Cognition, Brain and Language, Spain
Itamar Kastner,
Humboldt-Universität zu Berlin, Germany

***Correspondence:**
José Alemán Bañón
jose.aleman.banon@biling.su.se

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *12 October 2018*
Accepted: *18 March 2019*
Published: *24 April 2019*

Citation:
Alemán Bañón J and Rothman J (2019) Being a Participant Matters: Event-Related Potentials Show That Markedness Modulates Person Agreement in Spanish. Front. Psychol. 10:746. doi: 10.3389/fpsyg.2019.00746

INTRODUCTION

The present study uses event-related potentials (ERPs) to investigate the processing of subject–verb person agreement in Spanish. An example of how person information is encoded in the Spanish verb is provided in (1). As can be seen, the form of the verb *entrenar* "to train," which is inflected in the simple present for singular subjects, varies systematically depending on whether the subject is the speaker (*yo*, first-person singular), the addressee (*tú*, second-person singular), or someone else (*el atleta* "the athlete").

(1) a. Yo entren**o**.
I train-1ST PERSON-SG
b. Tú entren**as**.
You-SG train-2ND PERSON-SG
c. El atleta entren**a**.
The athlete train-3RD PERSON-SG

A number of theoretical proposals have drawn a distinction between first and second person on the one hand, and third person, on the other (e.g., Jakobson, 1971; Harris, 1996; Harley and Ritter, 2002; McGinnis, 2005; Bianchi, 2006). The idea is that both first and second person are *participants* in the speech act, since they play the *speaker* and *addressee* roles, respectively. Third person, in contrast, is not a speech participant and merely refers to someone who is neither the speaker nor the addressee. This distinction bears directly upon the concept of markedness, the observation that different feature values carry differential weight (e.g., Battistella, 1990; Bonet, 1995; Corbett, 2000; Cowper, 2005). The claim is that third person, not being a speech participant, is unmarked relative to first and second person (e.g., Harley and Ritter, 2002; Bianchi, 2006; Wechsler, 2011). Our study investigates if and how markedness with respect to the speech participant status of the subject modulates person agreement resolution online. We do this by comparing sentences with a first-person singular subject (speaker role) to sentences with a third-person singular subject (default person).

An influential proposal formalizing this markedness distinction between first/second and third person is Harley and Ritter (2002). Harley and Ritter (2002) offer a feature geometry analysis for person (and number) where features, such as *participant*, are privative rather than binary. For the person feature, this means that only first and second person have the status of true grammatical persons. In contrast, third person carries no person specification at all (see also Benveniste, 1971; Kayne, 2000; McGinnis, 2005; Adger and Harbour, 2006; Wechsler, 2011). Contrastive proposals treat third person as a true grammatical person, one that is specified as "non-participant." This is, for example, what Nevins (2007, 2011) argues for third-person pronouns (but not for lexical determiner phrases "DP," which he assumes carry no person specification). Crucially, despite these differences with respect to third-person pronouns, there is consensus that only the first and second persons are participants in the speech act. In fact, Bianchi (2006, p. 2026) suggests that this distinction might be universal.

This conceptual distinction between first/second and third person is consistent with typological data showing (a) that third person often distributes differently from first and second person crosslinguistically and (b) that third person is morphologically unmarked. For example, Forchheimer (1953) points out that some languages have specific pronouns for the first and second persons, but not the third (i.e., demonstratives are used instead, as in Halh Mongolian or Telugu; see Harley and Ritter, 2002). In addition, in some languages, first and second person show overt agreement, but third person does not. This is indeed what Harris (1996) argues for Spanish (i.e., that there is only first- and second-person verbal morphology). Finally, third-person pronouns are more likely to show gender distinctions than first- or second-person pronouns. Since the third person is not a speech participant, its referent in the speech act is independent from the discourse and, thus, more likely to show distinctions that are also independent from the discourse, such as gender[1]. We see this in Spanish, where gender distinctions only emerge in the third-person pronoun[2].

(2) a. yo
1ST PERSON-SG
b. tú
2ND PERSON-SG
c. él/ella/ello
3RD PERSON-SG-MASC/FEM/NEUT

An interesting question that arises is whether these markedness distinctions impact the establishment of person dependencies online. In the psycholinguistic literature, a self-paced reading study by Carminati (2005) provides psycholinguistic validity for the differential treatment of first and second person on the one hand, and third person on the other. Carminati examined bi-clausal sentences in Italian where she manipulated the type of cue that served to disambiguate a null pronoun toward its antecedent (e.g., *Quando Maria ha litigato con me, ero*... "when Maria quarreled with me," *pro* was-1ST PERSON – SG). The logic behind this paradigm is that, in Italian, null pronouns show a strong preference toward the subject position (i.e., *Maria*). Carminati found that having to establish co-reference between a null pronoun and a non-preferred antecedent (i.e., the object, the underlined first-person pronoun *me*) carried a smaller penalty (in terms of reaction time) when the disambiguating verb was inflected for first or second person, relative to third person (e.g., *Quando ho litigato con Maria, era*... "when quarreled-1ST PERSON – SG with Maria," *pro* was-3RD PERSON – SG). In contrast, no differences emerged between the first and second persons. This suggests that first- and second-person cues are stronger than third-person

[1]See Harley and Ritter (2002) for further discussion. See Forchheimer (1953) for a more elaborate list of differences between first/second and third person.

[2]The Spanish plural personal pronouns *nosotros/nosotras* "we-MASC/FEM" and *vosotros/vosotras* "you-PL – MASC/FEM" might seem to contradict this observation. However, as Harley and Ritter (2002) point out, these are bimorphemic pronouns, where the actual person morphemes (*nos, vos*) show no gender distinction. They only show person and number specification. Likewise, the morpheme *otros/otras* shows number and gender specification, but not person.

cues, consistent with the possibility that they carry greater cognitive weight.

Outside the domain of person agreement, the literature on *agreement attraction* has provided additional evidence for the psycholinguistic validity of markedness, in this case for number and, to a lesser extent, gender (attraction is argued not to be possible for person; e.g., Den Dikken, 2011; Nevins, 2011). In attraction, a finite verb agrees in number with a noun other than its controller subject, one that is structurally inaccessible, as in *The key to the cabinets *are*...* (production: Bock and Miller, 1991; Antón-Méndez et al., 2002; comprehension: Pearlmutter et al., 1999; Wagers et al., 2009; Dillon et al., 2013; Acuña Fariña et al., 2014; Lago et al., 2015). Importantly, attraction tends to occur when the attractor noun (i.e., *cabinets*) is plural (i.e., marked for number). Singular nouns (i.e., unmarked for number) rarely attract. Thus, both Carminati's study (2005) and the literature on attraction provide interesting evidence that markedness impacts the processing of person and number dependencies, at least in contexts that involve more than one trigger noun (whether or not they are licensed as controllers). In the present study, we examine whether markedness differences with respect to the speech participant status of the subject (speaker vs. default person) modulate person agreement in simpler sentences with an unambiguous subject.

One possibility is that the marked status of the subject will allow the parser to compute agreement as a top-down mechanism (e.g., Nevins et al., 2007; Wagers and Phillips, 2014). A number of proposals assume that agreement is a predictive procedure (e.g., Gibson, 1998, 2000; Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015; but see for example, Nicol et al., 1997; Pearlmutter et al., 1999), but little is known as to the role of markedness in predictive processing. Nevins et al. (2007) proposed that, for subject–verb agreement, feature activation at the subject might allow the parser to generate a stronger prediction regarding the form of the upcoming verb (Wagers and McElree, unpublished also posit that the parser can conclude more from the presence than the absence of a feature). This is a possibility that we evaluate in the present study. Herein, we use ERPs, brain responses which are time-locked to stimuli of interest and which provide high temporal resolution.

ERP LITERATURE ON AGREEMENT

The ERP literature on agreement (as a general phenomenon) has mainly focused on the P600, a positive-going wave that typically emerges between 500 and 900 ms in central-posterior electrodes (see Molinaro et al., 2011a for a review). The functional significance of the P600 is still debated. It was initially interpreted as an index of difficulty at the level of the syntax (reanalysis, repair, integration), as it was found for morphosyntactic anomalies (e.g., Hagoort et al., 1993; Osterhout and Mobley, 1995; Friederici et al., 1996), garden-path sentences (e.g., Osterhout and Holcomb, 1992), and grammatical but complex sentences that require the integration of displaced elements (e.g., Kaan et al., 2000). Some have also argued that the P600 encompasses two separate phases, which are sensitive to different factors and show different topography (e.g., Hagoort and Brown, 2000). This proposal has received interest in the agreement literature, where the late phase of the P600 (∼700–900 ms, argued to be sensitive to repair mechanisms) has been found to be modulated by feature distinctions. For example, Barber and Carreiras (2005) found it to be larger for gender than number in Spanish, and Mancini et al. (2011a) found it to be larger for person than number in Spanish (but see Alemán Bañón et al., 2012; Chow et al., 2018a).

The finding that certain types of semantic anomalies (e.g., Kolk et al., 2003; Kuperberg et al., 2003, 2006; Kim and Osterhout, 2005) and non-linguistic stimuli (Patel et al., 1998) sometimes also yield a P600 has prompted alternative proposals where the P600 is viewed as an index of reanalysis in general, as opposed to core morphosyntactic processing (see Tanner et al., 2017). For example, van de Meerendonk et al. (2010) argue that the P600 reflects the reanalysis process triggered by a strong conflict between a highly expected linguistic element (e.g., a word, a morpheme) and the encountered input, thus assuming that the P600 is sensitive to the violation of top–down expectations. Other proposals argue that the P600 reflects (non-exclusively morphosyntactic) combinatorial processing (Kuperberg, 2007) or well-formedness checking (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008). We do not elaborate on these proposals here, since the purpose of our study is not to tease them apart (see also Brouwer et al., 2012; Van Petten and Luka, 2012). What is important for the purposes of the present study is that the P600 consistently emerges for agreement errors across languages, agreement types (e.g., person, number, gender), and syntactic contexts (e.g., subject–verb, determiner–noun, noun–adjective, etc.) (see Table 1 in Molinaro et al., 2011a, p. 910).

The same is not true of a negativity that sometimes precedes the P600 between ∼300 and 500 ms. In some studies, this negativity shows an anterior distribution, sometimes with a left hemisphere bias. In others, it is more broadly distributed, spanning over central-posterior areas. This topographical variability has generated much debate regarding the identity of this component. Some refer to it as a Left Anterior Negativity (LAN), a component argued to index automatic morphosyntactic processing (e.g., Friederici et al., 1996; Friederici, 2002; De Vincenzi et al., 2003; Barber and Carreiras, 2005; Molinaro et al., 2008; Mancini et al., 2011a; Caffarra and Barber, 2015) or the working memory costs associated with the processing of long-distance dependencies (e.g., Kluender and Kutas, 1993; Fiebach et al., 2002; see a review in Molinaro et al., 2011a). In the agreement literature, it has been argued that the Left Anterior Negativity is more likely to emerge when the dependency is local (e.g., determiner–noun), the agreement cues are overt, and the reference site is hemisphere-neutral (e.g., Molinaro et al., 2011a,b).

Other researchers have argued that the LAN is reminiscent of the N400 (e.g., Service et al., 2007; Guajardo and Wicha, 2014; Tanner and van Hell, 2014; but see Molinaro et al., 2015), a component related to lexical retrieval and semantic integration (see Lau et al., 2008 for a review). Recent work by Caffarra et al. (2019), however, suggests that the LAN can characterize agreement progressing independently of the N400 (at least, for

determiner–noun gender errors in Spanish). Yet, others have argued that agreement violations yield either a LAN or an N400, depending on the levels of representation (e.g., morphosyntax, discourse) that are disrupted by the error (e.g., Mancini et al., 2011a). Importantly, in many studies on agreement, this negativity is simply absent (e.g., Nevins et al., 2007; Frenck-Mestre et al., 2008; Hammer et al., 2008), even for local agreement errors in languages with rich morphosyntax (e.g., Wicha et al., 2004; Alemán Bañón et al., 2012, 2014). Herein, we will focus mainly on the P600, which is the most consistent ERP signature of agreement, although we will also investigate the LAN. In the next section, we review how these components have informed our understanding of how person dependencies are established in real-time comprehension.

ERP LITERATURE ON PERSON AGREEMENT

A number of studies have used ERP to investigate agreement, but only a few have manipulated person dependencies. Silva-Pereyra and Carreiras (2007) found robust P600 effects for single person violations in Spanish between 700 and 900 ms (e.g., *yo entiendo/*entiendes* "I-$_{1ST\ PERSON-SG}$ understand-$_{1ST\ PERSON-SG}$/*understand-$_{2ND\ PERSON-SG}$). This positivity emerged earlier (500–700 ms) for combined person + number violations. In addition, only combined violations showed an anterior negativity (300–450 ms), which was not left-lateralized. Rossi et al. (2005) also reported this biphasic pattern (LAN-P600) for single person violations in German, although both components emerged later in Rossi et al.'s study.

Nevins et al. (2007) examined subject–verb agreement in Hindi with a design that includes both single (number, gender) and combined errors (number + gender, person + gender). Crucially for the purposes of the present study, they examined whether agreement is computed as a bottom-up or top-down (i.e., predictive) mechanism. In the latter case, Nevins et al. hypothesized that combined violations would yield a larger P600 than single errors, since the distance between the predicted and encountered forms increases as a function of the number of features violated. Their results showed equally robust P600 effects for single number, single gender, and combined number + gender violations (not preceded by a LAN). Combined person + gender errors yielded an earlier and larger P600 than all other error types, but a follow-up study suggested that this was due to person being orthographically more marked/salient in the Devanagari script. Thus, these results are inconclusive as to whether agreement checking takes place top-down. However, Nevins et al. suggest that this might have been due to their using subjects with a default status (i.e., third person, singular, masculine), which might have failed to activate the relevant features. We address this question in our study, by specifically manipulating the markedness of the subject with respect to the person feature (first vs. third person).

In another study looking at Spanish, Mancini et al. (2011a) found that person violations (e.g., *el cocinero* *cocinaste... "the cook-$_{3RD\ PERSON-SG}$ cooked-$_{2ND\ PERSON-SG}$") yielded an N400-P600 biphasic pattern, relative to control sentences (e.g., *los cocineros cocinaron*... "the cook-$_{3RD\ PERSON-PL}$ cooked-$_{3RD\ PERSON-PL}$"), whereas number violations (e.g., *el cocinero* *cocinaron*... "the cook-$_{3RD\ PERSON-SG}$ cooked-$_{3RD\ PERSON-PL}$") elicited a LAN-P600 biphasic pattern. In addition, the early phase of the P600 (500–800 ms) was broader, and the late phase (800–1000 ms) larger, for person relative to number errors. The authors argue that the qualitative differences between person (N400) and number (LAN) reflect the different interpretative procedures associated with each feature. Their claim is that only person violations disrupt the process of building a discourse representation, since the parser cannot assign a speech role (speaker, addressee) to the subject (see Tanner and van Hell, 2014 for an alternative proposal regarding N400 effects for agreement errors).

These qualitative differences between person and number were not replicated by Zawiszewski et al. (2016). The authors compared the effects of person, number, and person + number violations in Basque (e.g., *zuk...utzi duzu/*dut/*duzue/*dugu* "you-$_{2ND\ PERSON-SG}$ left have-$_{2ND\ PERSON-SG}$/*left have-$_{1ST\ PERSON-SG}$/*left have-$_{2ND\ PERSON-PL}$/*left have-$_{1ST\ PERSON-PL}$) and found an N400-P600 biphasic pattern (and a late frontal negativity) for all error types. Interestingly, the P600 was larger in the two conditions with a person mismatch, which the authors interpret as evidence that person is more salient than number, although they cannot rule out that this was due to orthographic differences between the critical words (e.g., Nevins et al., 2007). The N400 effect for person (and number) violations is accounted for by the fact that the Basque verb also instantiates object agreement, which requires the parser to check thematic relations (upon encountering a disagreeing verb).

To our knowledge, the only study that has manipulated markedness in an examination of person agreement is Mancini et al. (2018). The authors probed two types of person dependencies in Basque that differed with respect to the speech participant status of the subject (first-person plural: marked vs. third-person plural: unmarked). Their design encompassed errors where a first-person plural subject mismatched a third-person plural verb (*japoniarr-ok...ikasi dugu/*dute* "Japanese-$_{1ST\ PERSON-PL}$ learned have-$_{1ST\ PERSON-PL}$/*learned have-$_{3RD\ PERSON-PL}$") and errors where a third-person plural subject mismatched a first-person plural verb (*japoniarr-ek...ikasi dute/*dugu* "Japanese-$_{3RD\ PERSON-PL}$ learned have-$_{3RD\ PERSON-PL}$/*learned have-$_{1ST\ PERSON-PL}$"). The authors hypothesized that the latter error type would yield a qualitatively different P600, because the marked person features of the verb (first-person) could extend to the unmarked subject (third-person) and "rescue" the violation. In fact, such a mismatch is ungrammatical in Basque, but not in languages like Bulgarian, Modern Greek, Swahili, or Spanish (example from Spanish: *los investigadores somos tenaces* "the researchers-$_{3RD\ PERSON-PL}$ are-$_{1ST\ PERSON-PL}$ tenacious"), a phenomenon known as *unagreement* (e.g., Hurtado, 1985; Höhn, 2016). Both error types yielded an N400, but only "first-person plural subject + third-person plural verb" errors showed a P600. The authors argue in favor of their hypothesis, although they cannot rule out the possibility that participants treated violations on

first-person plural verbs as grammatical unagreement (they accepted them at a rate of 42% in the judgment task, and ERPs were calculated without excluding incorrectly judged trials), especially as they were highly proficient bilingual speakers of Spanish. This would be consistent with Torrego and Laka's claim (2015) that unagreement is grammatical in Basque, although it is subject to individual differences. Importantly, previous work by Mancini et al. (2011b) showed a qualitatively similar processing profile (N400, no P600) for unagreement sentences in Spanish. Thus, although Mancini et al.'s results (2018) are interesting, the evidence that outright violations with unmarked subjects are salvageable requires further exploration (see Mancini et al., 2018 for counterarguments).

Importantly, Mancini et al.'s results (2018) show that markedness does modulate person agreement. Whether "third-person plural subject + first-person plural verb" combinations yielded no P600 effect because (1) the unmarked status of the subject makes an outright person violation less disruptive (potentially due to the participants' bilingualism with Spanish, a language that clearly allows this) or (2) because the Basque grammar itself simply allows it (e.g., Torrego and Laka, 2015), what is important is that the speech participant status of the subject affects person agreement resolution. Thus, Mancini et al.'s study (2018) adds to a small ERP literature showing that markedness modulates agreement processing (e.g., Deutsch and Bentin, 2001; Kaan, 2002). Outside the realm of person agreement, a previous study from our own lab (Alemán Bañón and Rothman, 2016) was the first to investigate how markedness affects the processing of noun–adjective number and gender agreement (in Spanish). In that study, we examined markedness by manipulating the number/gender of the trigger nouns and their agreeing adjectives (e.g., *una catedral que parecía inmensa* "a cathedral-$_{FEM-SG}$ that looked huge-$_{FEM-SG}$"). Following Nevins et al. (2007), one of our hypotheses was that the parser might be more likely to engage in predictive processing when the controller noun carried marked features (gender: feminine; number: plural), due to feature activation. In that case, our prediction was that errors of the kind "marked noun + unmarked adjective" might result in a larger P600 than the opposite error type, given that a prediction would be generated but unmet. Instead, we found that violations realized on marked adjectives (the opposite error type) yielded an earlier P600 for both number and gender. In addition, the P600 was larger for number errors realized on plural adjectives (e.g., Deutsch and Bentin, 2001; Kaan, 2002). Although our results provide evidence that markedness modulates agreement, they do not provide evidence that markedness triggers predictive processing. One possibility, however, is that the syntactic frame where we examined agreement was not sufficiently constraining to allow for the generation of strong predictions. That is, although an adjective carrying agreement features was likely to appear after the structure "Noun that looked/seemed. . .," other continuations were possible (e.g., *una catedral que parecía desafiar la gravedad* "a cathedral-$_{FEM-SG}$ that seemed to defy gravity"). However, the same is not true of subject–verb agreement, where the presence of a subject allows for the strong prediction that a verb will appear further down the line. We address this question in the present study.

THE PRESENT STUDY: RESEARCH QUESTIONS AND PREDICTIONS

The present study examines the processing of two types of person dependencies in Spanish. Crucially, the study is among the first to investigate how the online resolution of person agreement is impacted by markedness. Samples of the structure where we manipulated markedness (and agreement) can be seen in (3-6). The agreement relation of interest is that between the subject and the verb (underlined). Our design examines markedness by manipulating the speech participant status of the subject, such that half of the sentences had a first-person subject (marked for person: speaker role; see 3 and 4) and the other half, a third-person subject (unmarked for person; see 5 and 6). Agreement was manipulated by crossing each subject type with a verb showing the opposite person inflection. Unlike Mancini et al. (2018), we only used singular subjects and, thus, both types of person violations had an unambiguously ungrammatical status in Spanish (i.e., singular unagreement is not licensed in Spanish; see Torrego, 1996).

(3) Yo	a menudo acaricio	a	los caballos.
I-$_{1ST\ PERSON-SG}$	often pet-$_{1ST\ PERSON-SG}$	$_{CASE}$	the horses
(4) Yo	a menudo acelero		en la autopista.
I-$_{1ST\ PERSON-SG}$	often speed up-$_{1ST\ PERSON-SG}$		on the highway.
(5) El cartero	a menudo acaricia	a	los gatos.
the postman-$_{3RD\ PERSON-SG}$	often pet-$_{3RD\ PERSON-SG}$	$_{CASE}$	the cats
(6) El conductor	a menudo acelera		en la carretera.
the driver-$_{3RD\ PERSON-SG}$	often speed up-$_{3RD\ PERSON-SG}$		on the road.

As a first step, we will examine which ERP components are associated with violations of person agreement. Based on the previous literature, our prediction is that both types of person violations will yield a P600, which is a reliable finding across studies (Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2016; Mancini et al., 2011a, 2018). Predictions regarding negative effects (LAN, N400) preceding the P600 are less straightforward, since these effects only emerged in the studies by Mancini et al. (2011a, 2018) and Zawiszewski et al. (2016) (and Rossi et al., 2005 found a LAN). In addition, Zawiszewski et al. (2016) interpret the N400 as evidence that person violations compromise thematic role assignment, given that the Basque verb also instantiates object agreement, an operation that does not apply to Spanish.

Our main research question concerns how markedness will impact person agreement resolution. We evaluate two possible scenarios. First, "third-person subject + first-person verb" violations could yield an earlier and larger P600 relative to "first-person subject + third-person verb errors." This is because first-person verbs are marked relative to third-person ones (e.g., Harris, 1996). This would be consistent with what we found in Alemán Bañón and Rothman (2016) and would constitute further evidence that the parser can more easily detect

violations realized on marked elements or that these are more disruptive (e.g., Friederici et al., 2001; Kaan, 2002; Nevins et al., 2007). Alternatively, if Nevins et al.'s (2007) proposal that the parser is more likely to engage in predictive processing when the subject carries marked features is on the right track, it is possible that violations of the type "first-person subject + third-person verb" (hereinafter "marked subject violations") will yield a larger P600 than "third-person subject + first-person verb" errors (hereinafter "unmarked subject violations"). It is also possible that the positivity will span over frontal areas, given recent proposals linking frontal positivities to prediction disconfirmation (e.g., DeLong et al., 2011; see Van Petten and Luka, 2012 for a review). This is because the marked status of the first-person subject (i.e., speaker) would activate the person feature, allowing the parser to generate a prediction regarding the specification of the upcoming verb. The same is not true of lexical subjects such as *el conductor* "the driver," which do not carry a person feature (e.g., Bianchi, 2006)[3]. To sum up, Alemán Bañón and Rothman's (2016) proposal predicts that the verb's markedness (as in 7) will impact processing at the violating verb, whereas Nevins et al.'s proposal predicts that it is the subject's markedness (as in 8) that will impact processing at the verb.

(7) la viuda *lloro
the widow-UNMARKED cry-MARKED

(8) yo *llora
I-MARKED cry-UNMARKED

MATERIALS AND METHODS

Before the testing began, the study was reviewed by the relevant research ethics committee at the University of Reading and received clearance (project number: 2014-031-JAB). All participants provided their informed written consent to take part in the study.

Participants

The participants include 28 native speakers of Spanish (16 females; age range: 18–38; mean age: 27). Data from 27 of these participants (from a different study) were reported in Alemán Bañón and Rothman (2016). All participants indicated being right-handed, and this was confirmed via the Edinburgh Handedness Inventory (Oldfield, 1971). In addition, they all reported having no history of cognitive or neurological damage/diseases. They all spoke one or more foreign languages (mainly English) to varying levels of proficiency, and four of them identified themselves as speakers of another one of Spain's co-official languages (Catalan, Galician) or Spanish Sign Language. They all received financial compensation for their time.

[3]Nevins et al. (2007) approached this question by comparing single to double violations, which did not differ from one another, possibly due to the use of subjects with default agreement features. If subject markedness determines, at least to some extent, whether agreement processing is predictive, then differences should emerge when comparing two types of single violations that differ with respect to subject markedness, as in the present study.

Materials

The materials comprise 160 single-clause sentences assigned to one of the four conditions in **Table 1**. All sentences follow the structure: subject + temporal adverb *a menudo* "often" + verb in the simple present + continuation (i.e., direct object or prepositional phrase). Half of the sentences (see conditions 1–2 in **Table 1**) include a lexical DP subject (e.g., *el cazador* "the hunter"), which corresponds to the default person (third person). In the grammatical version (condition 1), the verb is in the third-person singular. In the ungrammatical version (condition 2), the verb is incorrectly inflected as first-person singular, which is marked for person. In the other 80 sentences (conditions 3–4), the subject is the first-person singular pronoun *yo* (marked person: speaker). In the correct version (condition 3), the verb carries first-person singular inflection. In the ungrammatical version (condition 4), the verb shows third-person singular features and is, therefore, incorrectly underspecified for person. We chose the first as opposed to the second person as the marked subject for two reasons. First, only the first person allowed us to match the target verbs for length (e.g., *lloro* "cry-1ST PERSON–SG" vs. *llora* "cry-3RD PERSON–SG"; compare to *lloras* "cry-2ND PERSON–SG"). Second, there is substantial variability with respect to the use of the second person across varieties of Spanish, even within European Spanish (e.g., Green, 1988).

In sum, markedness was manipulated via the speech participant status of the subject and its corresponding verb (*el cazador. . .caza* "the hunter-3RD PERSON–SG hunt-3RD PERSON–SG," *yo. . .cazo* "I-1ST PERSON–SG hunt-1ST PERSON–SG") and agreement

TABLE 1 | Sample of the materials, including the conditions examining person agreement with third-person singular subjects (grammatical, ungrammatical), the conditions examining person agreement with first-person singular subjects (grammatical, ungrammatical), and the fillers.

3rd person singular subject
Grammatical
1. *El cazador a menudo acampa en la montaña.*
The hunter-3RD PERSON–SG often camp-3RD PERSON–SG in the mountain
Unmarked-subject violation
2. *El cazador a menudo *acampo en la montaña.*
The hunter-3RD PERSON–SG often camp-1ST PERSON–SG in the mountain
1st person singular subject
Grammatical
3. *Yo a menudo canto en la ducha.*
I-1ST PERSON–SG often sing-1ST PERSON–SG in the shower
Marked-subject violation
4. *Yo a menudo *canta en la ducha.*
I-1ST PERSON–SG often sing-3RD PERSON–SG in the shower
Fillers
Nosotros somos muy comprensivos y ellos tambièn.
We-1ST PERSON–PL are very understanding and they-3RD PERSON–PL too
Ellas son más puntuales que tú.
They-3RD PERSON–PL are more punctual than you-2ND PERSON–SG

was manipulated by pairing up first-person subjects with third-person verbs, and third-person subjects with first-person verbs. The adverb *a menudo* "often" intervened between the subject and verb in order to create some linear distance between the agreeing elements. We reasoned that this might give the parser a better opportunity to engage in predictive processing, since additional time is available for prediction generation (e.g., Chow et al., 2016, 2018b). Thus, if subject–verb agreement is ever predictive, we thought that this would be an appropriate set-up to explore such a possibility.

For the conditions with third-person subjects, we used lexical subjects (as opposed to third-person singular pronouns) for two reasons. First, it allowed us to diversify the stimuli as much as possible. Most importantly, as discussed in Section "Introduction," there is disagreement in the literature regarding whether third-person pronouns carry any person specification (e.g., Harley and Ritter, 2002 argue that they do not; Nevins, 2007 argues the reverse). In contrast, there seems to be agreement that lexical DPs are underspecified for person (e.g., Den Dikken, 2011; Nevins, 2011). Since the same could not be done in the conditions with first-person subjects, the fillers were designed so as to mitigate the salience of the first-person singular pronoun *yo*, which participants saw in 80 sentences. Therefore, the fillers involved 40 instances of the second-person singular pronoun *tú* "you," 40 instances of the first-person plural pronouns *nosotros/nosotras* "we-$_{\text{MASC/FEM}}$," and 80 instances of the third-person plural pronouns *ellos/ellas* "they-$_{\text{MASC/FEM}}$"). All materials are provided in **Supplementary File 1**.

Each inflected verb (e.g., *llora* "cry-$_{\text{3RD PERSON–SG}}$," *lloro* "cry-$_{\text{1ST PERSON–SG}}$") was used twice, once with a third-person singular subject and once with a first-person singular subject (e.g., *La viuda a menudo llora/*lloro en la iglesia* "the widow often cry-$_{\text{3RD PERSON–SG}}$/*cry-$_{\text{1ST PERSON–SG}}$ in church"; *Yo a menudo lloro/*llora en las películas* "I often cry-$_{\text{1ST PERSON–SG}}$/*cry-$_{\text{3RD PERSON–SG}}$ at the movies"). This was done to ensure that all properties associated with a given verb (e.g., meaning, argument structure, lexical aspect, etc.) would be held constant across the two markedness conditions. With the exception of the subject, all sentences across the two markedness conditions were therefore identical up to the critical verb. Since the testing took place in two separate sessions, we distributed the materials in such a way that participants would only see one token of each verb per session.

Since the verbs were the same across markedness conditions, they were controlled with respect to number of characters [mean length of verbs inflected as third-person singular: 6.56; mean length of verbs inflected as first-person singular: 6.57; $t(79) = 0.445$, $p = 0.658$]. Mean length was, however, not exactly the same, due to five verbs showing certain conjugational or orthographic idiosyncrasies (e.g., *conduce* "drive-$_{\text{3RD PERSON–SG}}$" vs. *conduzco* "drive-$_{\text{1ST PERSON–SG}}$"; *sigue* "follow-$_{\text{3RD PERSON–SG}}$" vs. *sigo* "follow-$_{\text{1ST PERSON–SG}}$"). It was not possible to match the critical verbs with respect to frequency of use. We calculated the log frequency of each form with the EsPal database (Duchon et al., 2013), and found that third-person singular forms were significantly more frequent than first-person singular ones. This is unsurprising, given that default forms (i.e., third-person singular) have a wider syntactic distribution. Notice that a similar issue arose in Mancini et al.'s (2011a) study and that information about frequency is not provided in most other ERP studies on person agreement (e.g., Rossi et al., 2005; Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Zawiszewski et al., 2016)[4]. Finally, the position of the critical verb was always mid-sentence, and it was similar across markedness conditions (conditions 1–2: word #5; conditions 3–4: word #4).

These materials were intermixed with 240 sentences (160 ungrammatical) from a separate study that examines noun–adjective number and gender agreement, but does not manipulate subject–verb agreement (reported in Alemán Bañón and Rothman, 2016). All 80 fillers were grammatical, which brought the ratio of grammatical to ungrammatical sentences to 1/1. A sample of each filler type is provided in **Table 1**.

Procedure

The testing was divided into two 3-hour sessions (e.g., O'Rourke and Van Petten, 2011; Alemán Bañón et al., 2012). Each EEG recording included 240 sentences (with an equal number of items per condition, including the fillers) and took approximately 1 h. Participants read the sentences quietly. The sentences were presented one word at a time, in random order. After each sentence, participants provided a grammaticality judgment, similar to previous ERP studies on person agreement (e.g., Rossi et al., 2005; Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Mancini et al., 2011a, 2018; Zawiszewski et al., 2016). Participants received instructions to favor accuracy over speed while judging the sentences, to avoid blinks and muscle movements while reading them, and to rest their eyes between trials. At the beginning of each session, participants completed an eight-trial practice set (four ungrammatical) so that they would become acquainted with the task. None of the practice trials involved agreement errors or nouns/verbs from the experimental stimuli. Participants received feedback for the first three practice trials. The experiment began right after. Each session comprised six 40-sentence blocks, separated by five short breaks. Sentence presentation was carried out in *Paradigm*, by Perception Research Systems Inc. (Tagliaferri, 2005).

Each trial began with a fixation cross, which remained in the center of the screen for 500 ms. Then, the presentation of the sentence began, one word at a time, using the Rapid Serial Visual Presentation method. Each word remained on the screen for 450 ms, followed by a 300 ms pause (e.g., Alemán Bañón et al., 2012; see Molinaro et al., 2011a). Upon presentation of the last word (marked with a period), there was a 1000 ms pause. Right after, participants saw the prompts for the Grammaticality Judgment Task (GJT), the words *Bien* "good" and *Mal* "bad" for grammatical and ungrammatical sentences, respectively. The prompts remained visible until participants provided a response, which they did with their left hand (middle and index fingers, respectively). After the behavioral response, we added an inter-trial interval ranging between 500 and 1000 ms, pseudo-randomly varied at 50 ms increments.

[4]Mancini et al. (2018) circumvented this issue by looking at auxiliary verbs in Basque.

EEG Recording and Analysis

The EEG was recorded with the Brain Vision Recorder software (Brain Products, GmbH, Germany) from 64 sintered Ag/AgCl electrodes mounted in an elastic cap (Easycap, Brain Products, GmbH, Germany). The placement of the electrodes followed the 10% system (midline: FPz, Fz, Cz, CPz, Pz, POz, Oz; hemispheres: FP1/2, AF3/4, AF7/8, F1/2, F3/4, F5/6, F7/8, FC1/2, FC3/4, FC5/6, FT7/8, FT9/10, C1/2, C3/4, C5/6, T7/8, CP1/2, CP3/4, CP5/6, TP7/8, TP9/10, P1/2, P3/4, P5/6, P7/8, PO3/4, PO7/8, O1/2). Electrode AFz served as the ground electrode and FCz as the online reference. The recordings were then re-referenced offline to the average of near-mastoid electrodes (TP7/8). Electrodes FP1/2, located above the eye-brows, were used to monitor blinks. Electrode IO was placed on the outer canthus of the right eye to capture horizontal eye movements. Electrode impedances were kept below 10 kΩ for all electrodes. The recordings were amplified by a BrainAmp MR Plus amplifier (Brain Products, GmbH, Germany) with a bandpass filter of 0.016–200 Hz, and digitized at a sampling rate of 1 kHz.

We analyzed the EEG data with the Brain Vision Analyzer 2.0 software (Brain Products, GmbH, Germany). After re-referencing the EEG, it was segmented into epochs relative to the critical verb. Epochs started 300 ms before the critical verb (i.e., the pre-stimulus baseline) and ended 1200 ms post-onset. Trials with blinks, horizontal eye movements, excessive alpha waves, or excessive muscle movement were manually rejected before analysis (based on visual inspection). We also discarded trials associated with incorrect responses in the GJT. This resulted in approximately 10% of data loss. After cleaning the data, the mean number of trials per condition ranged between 33 and 37 out of 40 (Condition 1: 37; Condition 2: 33; Condition 3: 36; Condition 4: 36), and this difference was significant, $F(2.01,54.31) = 11.049$, $p < 0.01$. Follow-up tests showed that the number of artifact-free trials in Condition 2 was lower than in all other conditions [Condition 2 vs. Condition 1: $F(1,27) = 18.973$, $p < 0.001$, $q^* = 0.008$; Condition 2 vs. Condition 3: $F(1,27) = 14.415$, $p = 0.001$, $q^* = 0.017$; Condition 2 vs. Condition 4: $F(1,27) = 11.758$, $p < 0.01$, $q^* = 0.025$], which did not differ from one another. Although this is not ideal, it should not be problematic for mean amplitude analyses (as opposed to peak analyses, which we did not conduct). As explained by Luck (2014, supplement, chapter 8, pp. 4–5), when measuring mean amplitudes, different numbers of trials per condition will not yield a spurious effect and should not be considered a confound. Following artifact rejection, data were baseline-corrected relative to the pre-stimulus baseline and averaged per condition and per subject. Finally, we applied a 30-Hz low-pass filter to the waveforms.

Event-related potentials were then quantified as mean amplitudes in two time windows: 250–450 ms, which corresponds to the LAN/N400, and 500–1000 ms, which corresponds to the P600. Both time-windows are consistent with previous reports on agreement. Importantly, they are the same time windows that we examined in Alemán Bañón and Rothman (2016). Thus, both time windows are the best estimates of where effects of agreement/markedness should emerge. For statistical analysis, we also used the same nine regions of interest (ROI) as in Alemán Bañón and Rothman (2016). Each ROI was calculated by averaging across the mean amplitudes of all electrodes in the region (left anterior: F1, F3, F5, FC1, FC3, FC5; right anterior: F2, F4, F6, FC2, FC4, FC6; left medial: C1, C3, C5, CP1, CP3, CP5; right medial: C2, C4, C6, CP2, CP4, CP6; left posterior: P1, P3, P5, P7, PO3, PO7; right posterior: P2, P4, P6, P8, PO4, PO8; midline anterior: Fz, FCz; midline medial: Cz, CPz; midline posterior: Pz, POz). The resulting values were then submitted to a repeated-measures ANOVA with Markedness (first-person singular subject, third-person singular subject), Agreement (grammatical, ungrammatical), Anterior–Posterior (anterior, medial, posterior), and Hemisphere (left, right) as the repeated factors. Since the hemisphere and midline regions comprise different numbers of electrodes, they were analyzed separately. For the analyses on the midline regions, Anterior–Posterior was the only topographical factor in the model. The Geisser and Greenhouse correction was applied in cases where sphericity could not be assumed. In such cases, we report corrected degrees of freedom (Field, 2005). A false discovery rate correction (Benjamini and Hochberg, 1995) was applied to all follow-up tests, to avoid an inflated Type I error. For all follow-up tests, we provide both the raw p-value and the adjusted significance level (q^*), that is, the significance level below which we consider effects significant.

RESULTS

All relevant data are provided in **Supplementary File 2**.

Behavioral

Table 2 provides the percentage of accurate responses in the GJT for each of the four experimental conditions (together with standard deviations). D-prime scores are also provided in the rightmost column. As can be seen, accuracy was generally very high (above 90% across the board), although participants were less accurate rejecting "unmarked subject violations." A repeated-measures ANOVA with Markedness (first-person, third-person singular subject) and Agreement (grammatical, ungrammatical) as the repeated factors revealed a main effect of Markedness, $F(1,27) = 9.051$, $p < 0.01$, a main effect of Agreement, $F(1,27) = 10.731$, $p < 0.01$, and a Markedness by Agreement interaction, $F(1,27) = 10.662$, $p < 0.01$. Follow-up tests to the interaction revealed that the main effect of Agreement was only significant in the conditions with third-person singular subjects,

TABLE 2 | Mean accuracy rates in the Grammaticality Judgment Task for the conditions examining person agreement with first-person singular subjects (i.e., marked subjects) vs. third-person singular subjects (i.e., unmarked subjects) ($N = 28$).

	Grammatical	Violation	D-prime score
Marked-subject	98 (2)	98 (4)	4.1 (0.4)
Unmarked-subject	98 (2)	92 (9)	3.6 (0.7)

Standard deviations are provided between parentheses.

$F(1,27) = 13.316$, $p = 0.001$, $q^* = 0.025$, driven by the fact that participants were less accurate rejecting ungrammatical sentences than accepting grammatical ones.

ERP Effects

Figure 1 plots ERPs for all four experimental conditions in the six ROIs computed for analysis. As can be seen, approximately 500 ms after presentation of the critical verb, both types of person violations yielded a positivity relative to their grammatical counterparts. In both cases, the positivity shows a central-posterior distribution and a slight right hemisphere bias, consistent with the P600 (e.g., Barber and Carreiras, 2005). In addition, the positivity does not go back to baseline before the end of the epoch (at 1200 ms). The positivity appears more robust for "marked subject violations," as it almost completely engulfs the positivity for the opposite error type, especially between 700 and 900 ms. This is also visible in **Figure 2**, which plots the magnitude of the violation effects for both types of person dependencies in four time windows of interest.

Also at approximately 700 ms, both types of person violations become more negative than grammatical sentences in the left anterior region, an effect that also remains visible until the end of the epoch (see **Figures 1**, **2**). This late left anterior negativity also appears larger for "marked subject violations." Preceding the P600, no evidence for a LAN or an N400 is apparent in **Figures 1** or **2** for either type of person violation (e.g., Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007). The following statistical analyses were conducted in the 250–450 ms time window (i.e., LAN effects should emerge in left anterior; N400 effects should emerge primarily in central-parietal regions) and the 500–1000 ms time window (i.e., P600 effects should emerge in central-posterior regions, possibly spanning over frontal regions for "marked subject violations").

Time Window Between 250 and 450 ms (LAN/N400)

Results of the omnibus ANOVA for the 250–450 ms time window are provided in **Table 3**. As can be seen, the ANOVA revealed two relevant interactions, Agreement by Hemisphere by Anterior–Posterior and Markedness by Agreement by Anterior–Posterior. To follow up on the former, we examined the main effect of Agreement within each of the six relevant ROIs, but no significant effects emerged. To evaluate the second interaction, we examined the Markedness by Agreement interaction, which is directly relevant to our discussion, at each level of Anterior–Posterior. The Markedness by Agreement interaction was significant in the Anterior and Posterior regions, but only before correcting for Type I error [posterior: $F(1,27) = 6.435$, $p = 0.0172$, $q^* = 0.017$; anterior: $F(1,27) = 4.491$, $p = 0.043$, $q^* = 0.033$]. In the posterior area, the interaction was driven by the fact that "unmarked subject violations" tended to be more negative than grammatical sentences, possibly signaling an N400 effect. However, this effect, which is too small to be visible in the waveforms, was only marginal, even before correcting for Type I error, $F(1,27) = 3.149$, $p = 0.087$, $q^* = 0.008$. In contrast, "marked subject violations" tended to be more positive than their grammatical counterparts (possibly signaling the onset of the P600), a comparison that also failed to reach significance. In the anterior area, the interaction was driven by the fact that "unmarked subject violations" were more positive than their grammatical counterparts, while the opposite error type yielded more negative waveforms than correct sentences. None of these comparisons reached significance either.

As shown in **Table 3**, the omnibus ANOVA revealed that the Markedness by Agreement by Anterior–Posterior interaction was also significant in the midline. Follow-up tests to this interaction yielded a similar pattern of effects to the hemispheres. That is, the Markedness by Agreement interaction was marginal in midline anterior, but only before correcting for Type I error, $F(1,27) = 4.035$, $p = 0.055$, $q^* = 0.017$. This interaction was driven by the fact that "unmarked subject violations" were more positive than grammatical sentences, while "marked subject violations" yielded a negativity relative to grammatical sentences. Only the negativity found for "marked subject violations" was significant, but only before adjusting the p-values, $F(1,27) = 5.069$, $p = 0.033$, $q^* = 0.008$. Visual inspection of the waveforms shows that this is the beginning of the late anterior negativity, which becomes robust in the subsequent time window.

To summarize, our analyses in the 250–450 ms time window revealed no reliable LAN or N400 effects for either type of person violation, as is clear from **Figure 2** (250–450 ms time window). What we see is a trend toward an earlier onset of the late anterior negativity for "marked subject violations." Additional analyses were conducted in the 300–500 ms time window (e.g., Silva-Pereyra and Carreiras, 2007; Mancini et al., 2011a), which revealed a similar pattern. Thus, we do not report them here.

Time Window Between 500 and 1000 ms (P600)

Table 3 summarizes the results of the omnibus ANOVA in the 500–1000 ms time window. As can be seen, the ANOVA revealed a main effect of Agreement, which was qualified by an interaction with Hemisphere and an interaction with Anterior–Posterior. In addition, the Agreement by Hemisphere by Anterior–Posterior interaction was significant. To follow up on the three-way interaction, we first examined the main effect of Agreement within each of the six relevant ROIs. The main effect of Agreement was significant in right posterior, $F(1,27) = 47.476$, $p < 0.001$, $q^* = 0.006$; left posterior, $F(1,27) = 29.587$, $p < 0.001$, $q^* = 0.012$; and right medial, $F(1,27) = 22.144$, $p < 0.001$, $q^* = 0.019$. In addition, it was marginal in left medial before correcting for Type I error, $F(1,27) = 3.748$, $p = 0.063$, $q^* = 0.037$. In all cases, person violations overall yielded more positive waveforms than grammatical sentences, consistent with the P600. The main effect of Agreement was also significant in left anterior, $F(1,27) = 16.206$, $p < 0.001$, $q^* = 0.025$, but here violations yielded more negative waveforms than grammatical sentences.

At least two factors seem to contribute to this three-way interaction. First, the positivity appears larger in the right hemisphere, as **Figures 1**, **2** clearly show. This was confirmed by the fact that, when comparing the main effect of Agreement in right posterior and left posterior, the Agreement by Hemisphere interaction was significant, $F(1,27) = 8.54$, $p < 0.01$, $q^* = 0.031$, and driven by the positivity being larger in right posterior. However, when comparing the main effect of Agreement in right posterior and right medial, the Agreement by Anterior–Posterior interaction was not significant. The second factor that seems to

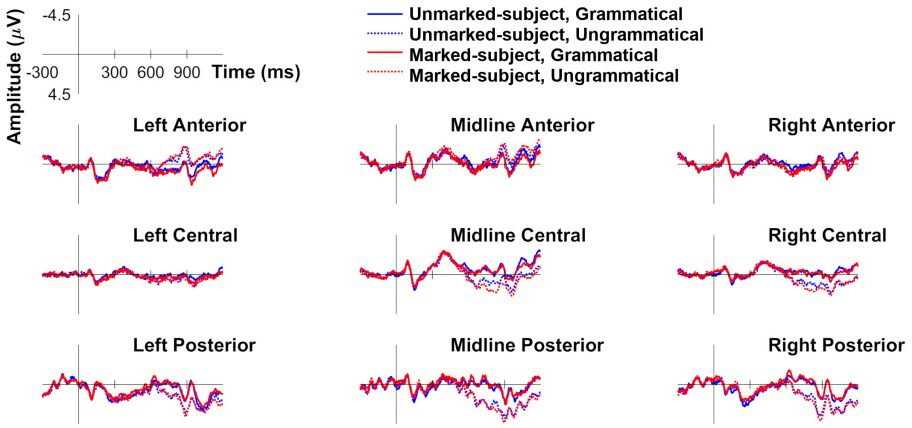


FIGURE 1 | Grand average ERP waveforms for the conditions examining person agreement with unmarked (third person) and marked (first person) subjects: unmarked-subject grammatical, unmarked-subject ungrammatical, marked-subject grammatical, marked-subject ungrammatical.

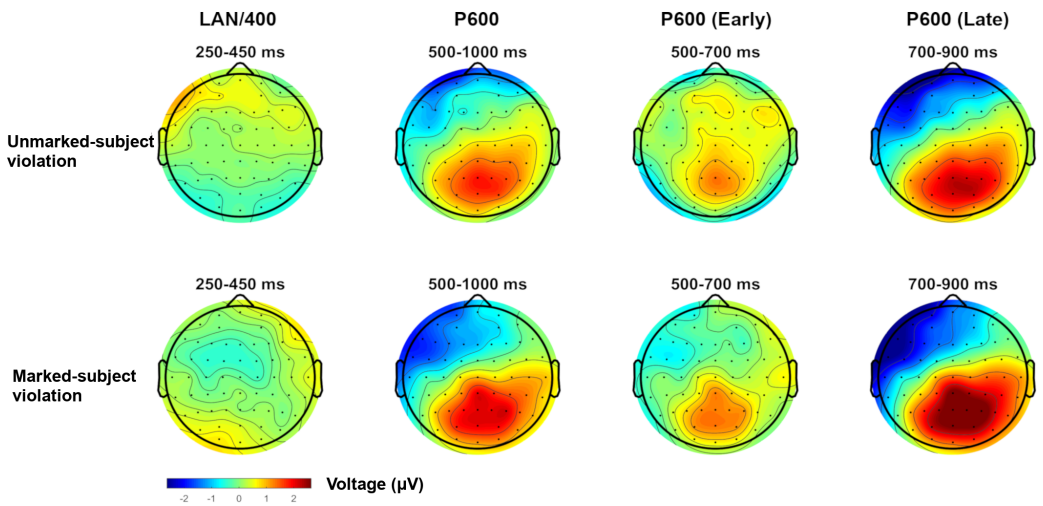


FIGURE 2 | Topographic plots for the two types of person violations (unmarked-subject violation, marked-subject violation) in the 250–450, 500–1000, 500–700, and 700–900 ms time windows. Plots were computed by subtracting the grammatical sentence from the violation condition.

contribute to the interaction is the fact that an effect of different polarity (i.e., a negativity) emerged for violations in left anterior.

The omnibus ANOVA also revealed a significant Markedness by Agreement by Anterior–Posterior interaction (see **Table 3**). Since Markedness and Agreement are the two relevant linguistic factors in our study, we followed up on this interaction by examining the Markedness by Agreement interaction at each level of Anterior–Posterior. The interaction was only significant in the anterior portion of the scalp, $F(1,27) = 6.568$, $p = 0.016$, $q^* = 0.02$, driven by the fact that "marked subject violations" were more negative than their grammatical counterparts, $F(1,27) = 9.581$, $p = 0.005$, $q^* = 0.01$. However, no effects emerged for the opposite type of person error. The larger late left anterior negativity for "marked subject violations" is clearly visible in **Figure 2** (500–1000 ms time window).

In the midline, the effects were qualitatively similar to the hemispheres (see **Table 3**). The Markedness by Agreement by Anterior–Posterior interaction was marginal ($p = 0.051$), and it was driven by the fact that Markedness and Agreement only interacted in midline anterior, but only before correcting for Type I error, $F(1,27) = 3.44$, $p = 0.075$, $q^* = 0.017$. Similar to the hemispheres, this interaction was driven by the fact that "marked subject violations" were more negative than grammatical sentences (before adjusting the p-values), $F(1,27) = 6.533$, $p = 0.017$, $q^* = 0.008$, while the reverse error type yielded no effects.

Finally, follow-up tests to the Agreement by Anterior–Posterior interaction (see **Table 3**) revealed main effects of Agreement in midline posterior, $F(1,27) = 50.31$, $p < 0.001$, $q^* = 0.017$, and midline medial, $F(1,27) = 23.059$, $p < 0.001$, $q^* = 0.033$, driven by person violations being more positive than grammatical sentences.

To summarize, our analyses in the 500–1000 ms time window revealed robust P600 effects for both types of person violations in

central-posterior areas of the scalp, with a slight right-hemisphere bias. The larger P600 effect that can be seen for "marked subject violations" relative to the reverse error type was, however, not statistically supported in this time window. In the same time window as the P600, person violations also showed an anterior negativity, mainly in left anterior but also present in midline anterior. This negativity is driven by "marked subject violations," as confirmed by the Markedness by Agreement interaction.

Time Window Between 700 and 900 ms (Late Phase of the P600)

To further explore the P600 magnitude difference between the two types of person violations, we conducted additional analyses in the 700–900 ms time window, corresponding to the late phase of the P600 (e.g., Barber and Carreiras, 2005; Silva-Pereyra and Carreiras, 2007). This is when both types of person violations seem to differ the most, as can be seen in **Figures 1**, **2**. We created an additional ROI including the electrodes from all four regions where the P600 was significant: right medial, right posterior, midline medial, and midline posterior. This approach allows us to compare the two types of person violations in all ROIs where we know the P600 emerged, without directly comparing regions with different numbers of electrodes (hemisphere regions: six electrodes; midline regions: two electrodes). A repeated-measures ANOVA with Markedness and Agreement as the repeated factors revealed a significant main effect of Agreement, $F(1,27) = 48.455$, $p < 0.001$, and a significant Markedness by Agreement interaction, $F(1,27) = 4.508$, $p < 0.05$. The interaction was driven by the fact that "marked subject violations" yielded a larger positivity (relative to their grammatical counterparts) than the reverse type of person error.

Additional analyses were conducted in the 500–700 ms time window, which confirmed that the larger P600 for "marked subject violations" was restricted to the 700–900 ms time window (see the topographical plot for the 500–700 ms time window in **Figure 2**). These analyses only revealed a significant main effect of Agreement, $F(1,27) = 12.478$, $p < 0.01$.

DISCUSSION

The present study used ERP to investigate subject–verb person agreement in Spanish, with a focus on how markedness differences with respect to the speech participant status of the subject influence agreement resolution at the verb. We manipulated markedness by probing both third-person singular lexical subjects, such as *la viuda* "the widow," and subjects consisting of the first-person singular pronoun *yo* "I." Crucially, while first person is marked (i.e., it plays the *speaker* role in the speech act), third person functions as a default, since it plays neither the *speaker* nor the *addressee* role. Our design also manipulated agreement, by crossing third-person singular subjects with first-person singular verbs and vice versa. We hypothesized that person violations might yield an earlier and larger P600 when realized on a marked verb (*la viuda*. . .**lloro* "the widow-$_{\text{3RD PERSON–SG}}$ cry-$_{\text{1ST PERSON–SG}}$") relative to an unmarked one (*yo*. . .**llora* "I-$_{\text{1ST PERSON–SG}}$ cry-$_{\text{3RD PERSON–SG}}$"). This is because violations have been argued to be more disruptive when they are realized on marked items (e.g., Deutsch and Bentin, 2001; Kaan, 2002; Nevins et al., 2007). In addition, this would be in line with what we found for noun–adjective number and gender agreement in Spanish

TABLE 3 | Results of the omnibus ANOVA in the 250–450 and 500–1000 ms time windows.

	250–450 ms	500–1000 ms
Lateral regions: Effects		
Markedness × Agreement × Anterior × Hemisphere	$F(1.31,35.47) = 2.553$	$F(1.56,42) = 0.03$
Agreement × Anterior × Hemisphere	$F(1.31,35.5) = 4.212$*	$F(1.34,36.07) = 4.68$*
Markedness × Anterior × Hemisphere	$F(2,54) = 1.016$	$F(2,54) = 0.294$
Markedness × Agreement × Hemisphere	$F(1,27) = 0.016$	$F(1,27) = 2.491$
Agreement × Hemisphere	$F(1,27) = 0.351$	$F(1,27) = 15.028$***
Markedness × Hemisphere	$F(1,27) = 1.244$	$F(1,27) = 0.729$
Markedness × Agreement × Anterior	$F(1.26,34.13) = 6.53$**	$F(1.42,38.31) = 5.405$*
Agreement × Anterior	$F(1.19,32.12) = 0.033$	$F(1.12,30.25) = 24.035$***
Markedness × Anterior	$F(1.4,37.75) = 1.183$	$F(1.22,33) = 0.638$
Markedness × Agreement	$F(1,27) = 0.003$	$F(1,27) = 0.102$
Agreement	$F(1,27) = 0.013$	$F(1,27) = 13.75$***
Markedness	$F(1,27) = 0.399$	$F(1,27) = 0.436$
Midline regions: Effects		
Markedness × Agreement × Anterior	$F(1.42,38.44) = 3.434$*	$F(2,54) = 3.153$^
Agreement × Anterior	$F(1.33,36.03) = 0.452$	$F(1.33,35.9) = 38.588$***
Markedness × Anterior	$F(2,54) = 0.9$	$F(2,54) = 1.142$
Markedness × Agreement	$F(1,27) = 0.09$	$F(1,27) = 0.143$
Agreement	$F(1,27) = 0.071$	$F(1,27) = 18.414$***
Markedness	$F(1,27) = 1.35$	$F(1,27) = 0.36$

*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ^$p \leq 0.1$. Where applicable, degrees of freedom were adjusted using the Greenhouse-Geisser correction.*

with the same participants (Alemán Bañón and Rothman, 2016). Alternatively, we evaluated the possibility that the marked status of the first-person subject would allow the parser to generate a stronger prediction regarding the upcoming verb due to feature activation (e.g., Nevins et al., 2007). If such is the case, we predicted that violations with a first-person singular subject (*yo. . .***llora* "I-$_{\text{1ST PERSON–SG}}$ cry-$_{\text{3RD PERSON–SG}}$") would show a larger (or more broadly distributed) P600 than violations with unmarked subjects (*la viuda. . .***lloro* "the widow-$_{\text{3RD PERSON–SG}}$ cry-$_{\text{1ST PERSON–SG}}$").

Our results revealed that both types of person violations elicited a robust positivity relative to grammatical sentences between 500 and 1000 ms, consistent with the P600, a component that is sensitive to a number of morphosyntactic operations, including agreement (e.g., Hagoort et al., 1993; Osterhout and Mobley, 1995; Nevins et al., 2007; Mancini et al., 2011a). Subsequent analyses revealed that this effect was larger for "marked subject violations," relative to the opposite error type between 700 and 900 ms. Our results did not reveal any reliable negativities preceding the P600 for either type of person violation (e.g., Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; cf. Mancini et al., 2011a, 2018; Zawiszewski et al., 2016). However, an anterior negativity did emerge in the P600 time window [similar to Alemán Bañón and Rothman's study (2016) for both number and gender errors], which was also impacted by markedness, as it was larger for "marked subject violations." We discuss these effects below.

Effects of Agreement

The P600 effects for both types of person violations are consistent with a large literature on agreement processing (e.g., Osterhout and Mobley, 1995; Barber and Carreiras, 2005; Alemán Bañón et al., 2012), including all previous studies on person agreement (Rossi et al., 2005; Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007; Mancini et al., 2011a, 2018; Zawiszewski et al., 2016). As previously discussed, the functional significance of the P600 is still a matter of debate. Initial proposals viewed the P600 as an index of syntactic reanalysis and repair (or syntactic difficulty, more generally) (e.g., Osterhout and Holcomb, 1992; Hagoort et al., 1993). Subsequent ones have posited that the P600 reflects reanalysis processes in general (i.e., not exclusively morphosyntactic) (e.g., Kuperberg, 2007; Bornkessel-Schlesewsky and Schlesewsky, 2008), or conflict monitoring (van de Meerendonk et al., 2010). Our results do not adjudicate between these proposals (nor was it the purpose of the study), but they are consistent with them. That is, the P600 effects for person violations here might reflect the reprocessing costs associated with trying to reconcile conflicting information (i.e., morphosyntactic and discourse information) in light of top-down expectations.

The lack of an N400 effect for both types of person violations deserves some discussion. An N400 effect was reported by Mancini et al. (2011a) for person violations in Spanish, and for person errors in Basque by both Zawiszewski et al. (2016) and Mancini et al. (2018). Mancini et al. (2011a) interpret this effect as evidence that person violations disrupt the assignment of a discourse role to the subject, due to the failure to map morphosyntactic and discourse information (i.e., person inflection on the verb + speech participant role). We agree that this is indeed possible, but we remain skeptical about how generalizable this account is, since our results did not reveal N400 effects for either type of person error (consistent with Nevins et al., 2007 and Silva-Pereyra and Carreiras, 2007).

Finally, person violations in the present study also elicited a late anterior negativity in the same time window where the P600 emerged. This effect has been reported in previous studies on agreement that required participants to provide a sentence-final judgment (e.g., Sabourin and Stowe, 2004; Gillon-Dowens et al., 2010; Alemán Bañón et al., 2012; Alemán Bañón and Rothman, 2016; Zawiszewski et al., 2016). One position in the literature is that this late negativity reflects the cost of keeping the ungrammaticalities in working memory until the end of the sentence. This interpretation is consistent with our results. It also explains why the negativity was less robust for "unmarked subject violations" relative to the opposite error type, as participants were less accurate rejecting the former in the GJT (92 vs. 98% accuracy, respectively). One possibility is that the parser can better maintain the feature specification of the subject in the focus of attention when the subject is marked, which would explain why our participants were more accurate rejecting "marked subject violations" in the GJT (e.g., Wagers and McElree, unpublished). Another possibility is that, Spanish being a null-subject language, the salience of an overt personal pronoun facilitated the detection of the ungrammaticalities at the verb. We come back to this possibility below.

In Alemán Bañón and Rothman (2016), we hinted that this late anterior negativity might be a phase reversal of the P600 (Nunez and Srinivasan, 2006), since both effects showed similar latency, but the reverse scalp distribution (right posterior vs. left anterior). The same is true of the late anterior negativity in the present study (see **Figure 2**). That both components were impacted by markedness in a similar way makes us wonder the extent to which these two components are independent from one another (although Osterhout and Hagoort, 1999 point out that two different ERPs can be impacted by the same factor). We, therefore, remain cautious in interpreting this effect.

Effects of Markedness

Our results revealed that person agreement violations realized at the verb yielded P600 effects of different magnitude in the 700–900 ms time window, depending on the speech participant status of the subject. More specifically, violations with a first-person singular subject, which corresponds to the *speaker* role, yielded a larger positivity than errors with a third-person (lexical) subject, which is underspecified for person (e.g., Harley and Ritter, 2002). This pattern of results is consistent with the proposal that, upon encountering a subject with marked features, feature activation allows the parser to generate a stronger prediction regarding the upcoming verb (e.g., Nevins et al., 2007). We do not argue that the larger P600 reflects prediction disconfirmation itself, since the effect was not frontally distributed (e.g., DeLong et al., 2011; Van Petten and Luka, 2012) (see **Figure 2**). The larger P600 for person violations with a marked subject might index the reanalysis process that

the parser initiates when there is a conflict between a highly expected verbal form (i.e., more so than in the conditions with an unmarked subject) and the form that was actually encountered (e.g., van de Meerendonk et al., 2010).

These results are not consistent with our previous investigation on the role of markedness in the processing of noun–adjective number and gender agreement in Spanish, which involved the same participants (Alemán Bañón and Rothman, 2016). In that study, we found that violations realized on marked adjectives (plural for number; feminine for gender) yielded earlier and, in the case of number, larger P600 effects than violations realized on unmarked adjectives. Here, we found the reverse. It is possible that differences between the target structures where we examined agreement in each study explain this discrepancy. As we discussed above, the configuration where we examined noun–adjective agreement (e.g., *una catedral que parecía inmensa* "a cathedral-$_{FEM-SG}$ that looked huge-$_{FEM-SG}$") might not have been sufficiently constraining to allow the parser to generate strong predictions regarding upcoming adjectives, since other continuations were possible (e.g., *una catedral que parecía desafiar la gravedad* "a cathedral that seemed to defy gravity"). In fact, adjective phrases are always optional, although some structures might make adjectives more predictable (e.g., *una fruta muy jugosa* "a fruit-$_{FEM-SG}$ very juicy-$_{FEM-SG}$," where the adverb *muy* "very" makes it very likely that an adjective will follow; see Alemán Bañón et al., 2012). The same is not true of subject–verb agreement, where the presence of a subject DP allows for the strong prediction that a verb phrase (VP), headed by a verb, will appear in order to satisfy the phrase structure rule for sentence building (e.g., Chomsky, 1957, 1995). It is, therefore, possible that markedness influences agreement processing in different ways at different stages, depending on the nature of the computation itself (see Dillon et al., 2013, who suggested agreement attraction in comprehension to be sensitive to the predictability of the dependency).

The results of the present study differ from those by Mancini et al. (2018) in a number of ways, although there are certain similarities. Unlike Mancini et al. (2018), our results did not reveal reliable N400 effects for either type of person violation, although this is consistent with previous studies (e.g., Nevins et al., 2007; Silva-Pereyra and Carreiras, 2007). With respect to the P600, the present study found that "marked subject violations" yielded a larger P600 than the reverse configuration. A similar asymmetry between violations with marked vs. unmarked plural subjects emerged in Mancini et al.'s study (2018), except that, in their study, only violations with first-person plural subjects yielded a P600. Recall, however, that the ungrammatical status of "third-person plural subject + first-person plural verb" errors in Mancini et al.'s study (2018) was uncertain, given that participants accepted them at a rate of 42% in the judgment task, consistent with theoretical accounts of person agreement in Basque (e.g., Torrego and Laka, 2015; see Mancini et al., 2018 for counterarguments). In addition, the authors did not discard incorrectly judged trials from analysis. The same was not true of our study, where "unmarked subject violations" were unambiguously ungrammatical. In fact, our participants only accepted them at a rate of 8% (and we discarded incorrectly judged trials from analysis). This might explain, partly, why a P600 did not emerge for errors with unmarked subjects in Mancini et al.'s study.

Mancini et al. (2018) interpret their results as evidence that, when the subject carries no person specification (i.e., third-person plural), encountering a verb with first-person plural features (i.e., marked for person) allows the parser to extend the verb's person specification to the subject. The authors point out that such a process only applies to plural subjects, which include more than one entity. For example, first-person plural includes the *speaker* + associates, and second-person plural includes the *addressee* + associates. In contrast, singular subjects are atomic entities that can only take their canonical speech role. What this means is that Mancini et al.'s proposal cannot explain our findings, since we found an asymmetry in the same direction as they did, but for person errors with singular subjects that differed with respect to markedness. However, Mancini et al.'s results can be explained in terms of an interplay between markedness and top-down expectations. That is, it is possible that the marked status of the first/second-person plural suffix *–ok*, relative to the third-person plural suffix *–ek*, allowed the parser to generate a stronger prediction regarding the upcoming verb. Future studies should explore this possibility, for example, by looking at person dependencies with plural subjects in non null-subject languages, where "third-person plural subject + first-person plural verb" configurations are more categorically disallowed (Höhn, 2016)[5].

We must point out, however, that first- and third-person subjects in our study differed with respect to more than just feature specification. While the first-person conditions involved a personal pronoun, the third-person conditions involved referential DPs, and the reader might rightfully wonder how this could have affected our results. Recall that we opted for lexical DPs (as opposed to third-person pronouns) because there is consensus in the literature that they carry no person specification. Therefore, only first-person subjects should have allowed for prediction generation with respect to person morphology at the verb[6].

One possibility is that sentences with first-person subjects were more salient than sentences with referential subjects because Spanish licenses *pro* drop and personal pronouns are often null. While this is indeed possible, we point out that overt pronouns are syntactically licensed and pragmatically appropriate as subjects in Spanish. Null pronouns are preferred

[5]Recall that, in the behavioral literature, Carminati (2005) found that disambiguating verbs inflected for first person carried a smaller reaction-time penalty than verbs inflected for third person. If our proposal is on the right track, it is possible that such an effect arose because of the unmarked nature of the matrix subject, which did not allow for strong predictions regarding upcoming verbs.

[6]Lexical DPs might have activated other features (e.g., *la viuda* "the widow" is [+feminine]), but these features were not manipulated at the verb (i.e., the Spanish verb only encodes number, which was held constant, and person). In addition, the use of third-person pronouns would not have mitigated this issue, since they also encode other features, such as gender or animacy (*él/ella/ello* "he/she/it"). In fact, this relates to markedness asymmetries, as we discussed in Section "Introduction."

as subjects if their referent can be inferred from context (topic maintenance), whereas overt pronouns tend to be used when there is a discourse switch to another referent (topic shift) (e.g., Lubbers-Quesada and Blackwell, 2009) or for contrastive focus (e.g., Rothman, 2009). This division of labor clearly emerges in cases of anaphora resolution such as *the man pushed the boy when he/Ø...* Here, null pronouns have been found to prefer subjects (*the man*) (e.g., Alonso-Ovalle et al., 2002; Carminati, 2005; Filiaci et al., 2014) and overt pronouns, objects (*the boy*) (e.g., Alonso-Ovalle et al., 2002; cf. Filiaci et al., 2014). Our materials, however, did not require anaphoric resolution. In fact, since each sentence was presented with no prior context (one that would determine topic maintenance or shift), the use of an overt pronoun does not seem overtly salient. In addition, we are skeptical that the use of third-person singular pronouns would have ameliorated this issue (even beyond theoretical considerations). Such a strategy would have made third-person pronouns more salient, because of their lower proportion in the language overall. For example, Morales (1997) shows that the proportion of overt first-person singular pronouns in European Spanish (our participants' variety) is 28%, compared to 8% for third-person pronouns (see similar results in Duarte and Soares da Silva, 2016).

Another possibility is that the parser might have extracted feature information more easily from personal pronouns than lexical DPs, which encode lexical information that can slow down processing. While we cannot rule out this possibility, we point out that the adverb *a menudo* "often" intervened between the subject and the verb. Thus, since we used a 750 ms stimulus onset asynchrony, participants had 1800 ms to extract person information from the subject before encountering the verb (*la viuda a menudo VERB*). This time interval should have allowed participants to generate predictions (e.g., Chow et al., 2018b). Alternatively, the semantic features of lexical DPs might have impacted processing at the verb, either by allowing the parser to predict the type of event encoded by the verb, or by allowing combinatorial processing with the verb's semantic features (even if the verb itself was not predicted). While this is also possible, we point out that the verb was held constant in the grammatical and ungrammatical conditions (*la viuda...llora/*lloro*). Thus, this should not have impacted the violation effect. We examined the possibility that lexical DPs might have allowed the parser to predict the event described by the verb by calculating the cloze probability of the target verbs in these conditions. The results of this cloze test (N = 33) show that mean cloze probability (across items) was very low (mean = 0.03; SD: 0.1), and that only one item had a cloze probability over 0.67, which corresponds to high probability (e.g., Block and Baldwin, 2010). Thus, the target verbs in the conditions with DP subjects were, overall, not predictable. Future research should investigate how markedness modulates person agreement while controlling for these differences, for example, by introducing the two subjects in a previous context, in order to reduce the salience of *yo* in the sentence where agreement is manipulated, and by using demonstratives in lieu of lexical DPs or third-person pronouns (see 9).

(9)	El atleta	y yo vamos al	gimnasio.
	The athlete	and I go	to-the gym
	a.	Yo	entreno/*entrena...
		I	train-1ST PERSON−SG/train-3RD PERSON−SG
	b.	Éste	entrena/*entreno...
		This	train-3RD PERSON−SG/train-1ST PERSON−SG

Two additional issues, however, might seem to undermine our claims. First, the mean number of trials for "unmarked subject violations" (Condition 2) was significantly lower than in the other three conditions. Thus, one could easily argue that the smaller P600 for person errors with an unmarked subject could be accounted for by signal-to-noise ratio differences across the conditions being compared. We can provide two counterarguments, one methodological and one theoretical. First, as discussed above, Luck (2014) points out that differences with respect to the mean number of trials per condition may affect analyses based on peak amplitudes, which we did not conduct, but not comparisons based on mean amplitudes, which are the basis for our conclusions. We therefore assume that the P600 size differences between the two error types are not epiphenomenal. Notice also that, albeit significant, the numerical differences in number of items across conditions were rather small (Condition 1: 37; Condition 2: 33; Condition 3: 36; Condition 4: 36) and the mean number of good items per condition was well above 30 across the board. Our second argument is that we only retained for analysis artifact-free trials that the participants had correctly judged in the GJT (unlike Mancini et al., 2018). As discussed in Section "Results," participants were least accurate rejecting "unmarked subject violations" (Condition 2). Thus, the fact that Condition 2 encompassed fewer trials than the other conditions is not independent from how markedness impacts person agreement resolution online, which is our main research question.

The second issue concerns differences in lexical frequency between the critical verbs. Recall that first-person verbs were significantly less frequent than third-person ones. How could this have affected our results? There is evidence in the literature that lexical frequency is inversely related to the amplitude of the N400 (e.g., Neville et al., 1992; Kutas et al., 2006), a component associated with lexical access and retrieval. That is, less frequent words tend to show a larger N400. One possibility is that violations realized on first-person verbs (*la viuda...*lloro* "the widow-3RD PERSON−SG cry-1ST PERSON−SG") yielded more negative effects than their grammatical counterparts (*la viuda...llora* "the widow-3RD PERSON−SG cry-3RD PERSON−SG") in the N400 time-window, due to the fact that the verb was less frequent in the violation condition. In turn, this might have attenuated the following P600. Moreover, the reverse could have happened in the conditions with a marked subject. That is, violations on third-person singular verbs (*yo...*llora* "I-1ST PERSON−SG cry-3RD PERSON−SG") might have elicited a smaller N400 relative to their grammatical counterparts (*yo...lloro* "I-1ST PERSON−SG cry-1ST PERSON−SG"), due to the fact that the verb was more frequent in the ungrammatical condition. In turn, this might have amplified the size of the subsequent P600. In fact, the results reported for the N400 time window are compatible

with this scenario. Those analyses revealed a trend toward an N400 for "unmarked subject violations," and a trend toward a positivity for "marked subject violations." Crucially, however, the effects of markedness in our study emerged between 700 and 900 ms, in the late phase of the P600. If differences in the N400 time window (caused by differences in lexical frequency between the critical verbs) were responsible for the difference in P600 size across markedness conditions, those differences should have been largest in the early phase of the P600, right after the N400 (500–700 ms), which was not the case. To rule out this possibility, we recalculated effects in the 700–900 ms time window by using the N400 time window as a baseline (we used both the 250–450 and the 300–500 ms time windows) (e.g., Hagoort, 2003; Wicha et al., 2004; Martín-Loeches et al., 2006). These analyses revealed a similar pattern of results as with a pre-stimulus baseline. That is, the P600 was larger for marked subject violations, relative to violations with a third-person subject[7]. Thus, we can safely assume that the markedness effects that we found in the P600 time window are, at least to some extent, independent of baseline differences.

CONCLUSION

The data reported in the present study showed that subject–verb person agreement resolution in Spanish is impacted by the speech participant status of the subject. More specifically, we found that person violations where the subject is the *speaker* (i.e., first person, marked for person) yielded a larger P600 between 700 and 900 ms than violations where the subject is not a speech participant (i.e., third person, the default person). We interpreted these findings as evidence that, upon encountering a marked element (i.e., the subject), feature activation allows the parser to generate a stronger prediction regarding the form of the upcoming verb (e.g., Nevins et al., 2007). When this prediction is not met, the result is a larger P600 relative to cases when no feature information is available at the subject.

[7] We provide the relevant results of the analyses using a 300–500 ms baseline. In the midline, we found a Markedness by Agreement interaction, $F(1,27) = 4.571$, $p < 0.05$, driven by "marked subject violations" yielding a larger P600 than the reverse error type. In the hemispheres, we found a marginal Markedness by Agreement by Hemisphere interaction, $F(1,27) = 3.983$, $p = 0.056$, and a marginal Markedness by Agreement by Anterior–Posterior by Hemisphere interaction, $F(1.32,35.79) = 2.785$, $p = 0.09$. Follow-ups showed that the Markedness by Agreement interaction was significant in the right hemisphere, $F(1,27) = 4.278$, $p < 0.05$, driven by the fact that "marked subject violations" yielded a larger P600 than the reverse error type.

AUTHOR CONTRIBUTIONS

JA conceptualized the study, designed the materials, collected the data with help from research assistants, conducted the analyses, and wrote the original draft. JR supervised the study, and contributed to the original draft.

ACKNOWLEDGMENTS

We thank the reviewers for their valuable suggestions, Jon Andoni Duñabeitia for helping us with data collection, and Vincent DeLuca and Dave Miller for their help with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00746/full#supplementary-material

REFERENCES

Acuña Fariña, J. C., Meseguer, E., and Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua* 143, 108–128. doi: 10.1016/j.lingua.2014.01.013

Adger, D., and Harbour, D. (2006). The syntax and syncretisms of the person case constraint. *Syntax* 10, 2–37. doi: 10.1111/j.1467-9612.2007.00095.x

Alemán Bañón, J., Fiorentino, R., and Gabriele, A. (2012). The processing of number and gender agreement in Spanish: an event-related potential investigation of the effects of structural distance. *Brain Research* 1456, 49–63. doi: 10.1016/j.brainres.2012.03.057

Alemán Bañón, J., Fiorentino, R., and Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: an event-related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Lang. Res.* 30, 275–306. doi: 10.1177/0267658313515671

Alemán Bañón, J., and Rothman, J. (2016). The processing of number and gender agreement in Spanish: an event-related potential investigation of the effects of structural distance. *Lang. Cogn. Neurosci.* 31, 1273–1298. doi: 10.1016/j.brainres.2012.03.057

Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., and Clifton, C. (2002). Null vs. Overt pronouns and the topic-focus articulation in Spanish. *Riv. Linguist.* 14, 151–169.

Antón-Méndez, I., Nicol, J. L., and Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax* 5, 1–25. doi: 10.1111/1467-9612.00045

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Battistella, E. L. (1990). *Markedness: The Evaluative Superstructure of Language*. Albany: SUNY Press.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Benveniste, E. (1971). *Problems in General Linguistics*. Coral Gables, FL: University of Miami Press.

Bianchi, V. (2006). On the syntax of personal arguments. *Lingua* 116, 2023–2067. doi: 10.1016/j.lingua.2005.05.002

Block, C., and Baldwin, C. (2010). Cloze probability and completion norms for 498 sentences: behavioral and neural validation using event-related potentials. *Behav. Res. Methods* 42, 665–670. doi: 10.3758/BRM.42.3.665

Bock, K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Bonet, E. (1995). Feature structure of romance clitics. *Nat. Lang. Linguist. Theory* 13, 607–647. doi: 10.1007/BF00992853

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Res. Rev.* 59, 55–73. doi: 10.1016/j.brainresrev.2008.05.003

Brouwer, H., Fitz, H., and Hoeks, J. C. J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Res.* 1446, 127–143. doi: 10.1016/j.brainres.2012.01.055

Caffarra, S., and Barber, H. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence Reading. *Brain Res.* 1605, 83–92. doi: 10.1016/j.brainres.2015.02.018

Caffarra, S., Mendoza, M., and Davidson, D. (2019). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain Lang.* 191, 9–16. doi: 10.1016/j.bandl.2019.01.003

Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy (Person > Number > Gender) and implications for linguistic theory. *Lingua* 115, 259–285. doi: 10.1016/j.lingua.2003.10.006

Chomsky, N. (1957). *Syntactic Structures*. The Hague/Paris: Mouton.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chow, W. Y., Lau, E., Wang, S., and Phillips, C. (2018a). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Lang. Cogn. Neurosci.* 33, 803–828. doi: 10.1080/23273798.2018.1427878

Chow, W. Y., Nevins, A., and Carreiras, M. (2018b). Effects of subject-case marking on agreement processing: ERP evidence from Basque. *Cortex* 99, 319–329. doi: 10.1016/j.cortex.2017.12.009

Chow, W. Y., Smith, C., Lau, E., and Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Lang. Cogn. Neurosci.* 31, 577–596. doi: 10.1080/23273798.2015.1066832

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139164344

Cowper, E. (2005). A note on number. *Linguist. Inq.* 36, 441–455. doi: 10.1162/0024389054396935

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain Lang.* 85, 280–296. doi: 10.1016/S0093-934X(03)00055-5

DeLong, K. A., Urbach, T. P., Groppe, D. M., and Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology* 48, 1203–1207. doi: 10.1111/j.1469-8986.2011.01199.x

Den Dikken, M. (2011). Phi-feature inflection and agreement: an introduction. *Nat. Lang. Linguist. Theory* 29, 857–874. doi: 10.1007/s11049-011-9156-y

Deutsch, A., and Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224. doi: 10.1006/jmla.2000.2768

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Duarte, M. E., and Soares da Silva, H. (2016). "Microparametric variation in Spanish and Portuguese: the null subject parameter and the role of the verb inflectional paradigm," in *The Morphosyntax of Spanish and Portuguese in Latin America*, eds M. A. Kato and F. Ordóñez (Oxford: Oxford Scholarship Online), 1–26. doi: 10.1093/acprof:oso/9780190465889.003.0001

Duchon, A., Perea, M., Sebastián Gallés, N., Martí, A., and Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behav. Res. Methods* 45, 1246–1258. doi: 10.3758/s13428-013-0326-1

Fiebach, C. J., Schlesewsky, M., and Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German wh-questions. *J. Mem. Lang.* 47, 250–272. doi: 10.1016/S0749-596X(02)00004-9

Field, A. (2005). *Discovering Statistics Using SPSS*, 2nd Edn. London: SAGE Publications.

Filiaci, F., Sorace, A., and Carreiras, M. (2014). Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Lang. Cogn. Neurosci.* 29, 825–843. doi: 10.1080/01690965.2013.801502

Forchheimer, P. (1953). *The Category of Person in Language*. Berlin: Walter de Gruyter. doi: 10.1515/9783111562704

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., and Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: evidence from ERPs. *Acta Psychol.* 128, 528–536. doi: 10.1016/j.actpsy.2007.12.007

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D., Hahne, A., and Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *J. Exp. Psychol.* 22, 1219–1248. doi: 10.1037/0278-7393.22.5.1219

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., and Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cogn. Brain Res.* 11, 305–323. doi: 10.1016/S0926-6410(00)00065-3

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). "Dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, eds A. Marantz, Y. Miyashita, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.

Gillon-Dowens, M., Vergara, M., Barber, H., and Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *J. Cogn. Neurosci.* 22, 1870–1887. doi: 10.1162/jocn.2009.2134

Green, J. (1988). "Spanish," in *The Romance Languages*, eds M. Harris and N. Vincent (London: Routledge), 79–130.

Guajardo, L. F., and Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component: event-related potentials to gender-marked post-nominal adjectives. *Neuroimage* 91, 262–272. doi: 10.1016/j.neuroimage.2013.09.077

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/089892903322370807

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1

Hagoort, P., Brown, C. M., and Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Lang. Cogn. Process.* 8, 439–483. doi: 10.1080/01690969308407585

Hammer, A., Jansma, B. M., Lamers, M., and Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Res.* 1230, 177–191. doi: 10.1016/j.brainres.2008.07.004

Harley, H., and Ritter, E. (2002). Person and number in pronouns: a feature-geometric analysis. *Language* 78, 482–526. doi: 10.1353/lan.2002.0158

Harris, J. (1996). "The morphology of Spanish clitics," in *Evolution and Revolution in Linguistic Theory: Essays in honor of Carlos Otero*, ed. H. Campos (Washington, DC: Georgetown University Press), 168–197.

Höhn, G. F. (2016). Unagreement is an illusion: apparent person mismatches and nominal structure. *Nat. Lang. Linguist. Theory* 34, 543–592. doi: 10.1007/s11049-015-9311-y

Hurtado, A. (1985). "The unagreement hypothesis," in *Selected papers from the Thirteenth Linguistic Symposium on Romance Languages*, eds L. King and C. Maley (Amsterdam: John Benjamins), 187–211. doi: 10.1075/cilt.36.12hur

Jakobson, R. (1971). *Shifters, Verbal Categories, and the Russian verb. Selected Writings: Word and Language*, Vol. 22. The Hague: Mouton, 130–147. doi: 10.1515/9783110873269.130

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject-verb dependencies: an ERP study. *Journal of Psycholinguistic Research* 31, 165–193. doi: 10.1023/A:1014978917769

Kaan, E., Harris, A., Gibson, E., and Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Lang. Cogn. Process.* 15, 159–201. doi: 10.1080/016909600386084

Kayne, R. S. (2000). *Parameters and Universals*. New York, NY: Oxford University Press.

Kim, A., and Osterhout, L. (2005). The independence of combinatory semantic processing: evidence from event-related potentials. *J. Mem. Lang.* 52, 205–225. doi: 10.1016/j.jml.2004.10.002

Kluender, R., and Kutas, M. (1993). Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *J. Cogn. Neurosci.* 5, 196–214. doi: 10.1162/jocn.1993.5.2.196

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., and Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain Lang.* 85, 1–36. doi: 10.1016/S0093-934X(02)00548-5

Kuperberg, G., Caplan, D., Sitnikova, T., Eddy, M., and Holcomb, P. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Lang. Cogn. Process.* 21, 489–530. doi: 10.1080/01690960500094279

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Res.* 1146, 23–49. doi: 10.1016/j.brainres.2006.12.063

Kuperberg, G. R., Sitnikova, T., Caplan, D., and Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cogn. Brain Res.* 17, 117–129. doi: 10.1016/S0926-6410(03)00086-7

Kutas, M., Van Petten, C. K., and Kluender, R. (2006). "Psycholinguistics electrified II (1994-2005)," in *Handbook of Psycholinguistics*, eds M. Traxler and M. A. Gernsbacher (New York, NY: Elsevier), 659–724.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., and Phillips, C. (2015). Agreement attraction in Spanish comprehension. *J. Mem. Lang.* 82, 133–149. doi: 10.1016/j.jml.2015.02.002

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Lubbers-Quesada, M., and Blackwell, S. E. (2009). "The L2 Acquisition of Null and Overt Spanish Subject Pronouns: A Pragmatic Approach," in *Proceedings of the 11th Hispanic Linguistics Symposium*, eds J. Collentine, M. Garcia, B. Lafford and F. M. Martín (Somerville, MA: Cascadilla Proceedings Project), 117–130.

Luck, S. (2014). *An Introduction to the Event-Related Potential Technique*, 2nd Edn. Cambridge, MA: The MIT Press.

Mancini, S., Massol, S., Duñabeitia, J. A., Carreiras, M., and Molinaro, N. (2018). Agreement and illusion of disagreement: an ERP study on Basque. *Cortex* doi: 10.1016/j.cortex.2018.08.036 [Epub ahead of print].

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011a). A person is not a number: discourse involvement in subject-verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011b). When persons disagree: an ERP study of Unagreement in Spanish. *Psychophysiology* 48, 1361–1371. doi: 10.1111/j.1469-8986.2011.01212.x

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., and Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: a brain potential study of noun-adjective agreement in Spanish. *Brain Res.* 1093, 178–189. doi: 10.1016/j.brainres.2006.03.094

McGinnis, M. (2005). On markedness asymmetries in person and number. *Language* 81, 699–718. doi: 10.1353/lan.2005.0141

Molinaro, N., Barber, H. A., Caffarra, S., and Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex* 66, 156–159. doi: 10.1016/j.cortex.2014.06.009

Molinaro, N., Barber, H. A., and Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Molinaro, N., Vespignani, F., and Job, R. (2008). A deeper reanalysis of a superficial feature: an ERP study on agreement violations. *Brain Res.* 1228, 161–176. doi: 10.1016/j.brainres.2008.06.064

Molinaro, N., Vespignani, F., Zamparelli, R., and Job, R. (2011b). Why brother and sister are not just siblings: repair processes in agreement computation. *J. Mem. Lang.* 64, 211–232. doi: 10.1016/j.jml.2010.12.002

Morales, A. (1997). La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico. *Hispania* 80, 153–165. doi: 10.2307/345995

Neville, H. J., Mills, D. L., and Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cereb. Cortex* 2, 244–258. doi: 10.1093/cercor/2.3.244

Nevins, A. (2007). The representation of third person and its consequences for person-case effects. *Nat. Lang. Linguist. Theory* 25, 273–313. doi: 10.1007/s11049-006-9017-2

Nevins, A. (2011). Multiple agree with clitics: person complementarity vs. omnivorous number. *Nat. Lang. Linguist. Theory* 29, 939–971. doi: 10.1007/s1049-011-9150-4

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94. doi: 10.1016/j.brainres.2007.05.058

Nicol, J. L., Forster, K. I., and Veres, C. (1997). Subject-verb agreement processes in comprehension. *J. Mem. Lang.* 36, 569–587. doi: 10.1006/jmla.1996.2497

Nunez, P. L., and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*, 2nd Edn. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195050387.001.0001

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

O'Rourke, P. L., and Van Petten, C. (2011). Morphological agreement at a distance: dissociation between early and late components of the event-related brain potential. *Brain Res.* 1392, 62–79. doi: 10.1016/j.brainres.2011.03.071

Osterhout, L., and Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Lang. Cogn. Process.* 14, 1–14. doi: 10.1080/016909699386356

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Rossi, S., Gugler, M. F., Hahne, A., and Friederici, A. D. (2005). When word category information encounters morphosyntax: an ERP study. *Neurosci. Lett.* 384, 228–233. doi: 10.1016/j.neulet.2005.04.077

Rothman, J. (2009). Pragmatic deficits with syntactic consequences: l2 pronominal subjects and the syntax-pragmatics interface. *J. Pragmatics* 41, 951–973. doi: 10.1016/j.pragma.2008.07.007

Sabourin, L., and Stowe, L. A. (2004). Memory effects in syntactic ERP tasks. *Brain Cogn.* 55, 392–395. doi: 10.1016/j.bandc.2004.02.056

Service, E., Helenius, P., Maury, S., and Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *J. Cogn. Neurosci.* 19, 1193–1205. doi: 10.1162/jocn.2007.19.7.1193

Silva-Pereyra, J. F., and Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Res.* 1185, 201–211. doi: 10.1016/j.brainres.2007.09.029

Tagliaferri, B. (2005). *Paradigm. Perception Research Systems*. Available at: www.perceptionresearchsystems.com

Tanner, D., Grey, S., and van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology* 54, 248–259. doi: 10.1111/psyp.12788

Tanner, D., and van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301. doi: 10.1016/j.neuropsychologia.2014.02.002

Torrego, E. (1996). On quantifier float in control clauses. *Linguist. Inq.* 27, 111–126.

Torrego, E., and Laka, I. (2015). "The syntax of φ-features: agreement with plural DPs in Basque and Spanish," in *Ibon Sarasola, Gorazarre. Homenatge, homenaje*, eds B. Fernandez and P. Salaburu (Bilbao: UPV/EHU), 633–646.

van de Meerendonk, N., Kolk, H. H., Vissers, C. T., and Chwilla, D. J. (2010). Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. *J. Cogn. Neurosci.* 22, 67–82. doi: 10.1162/jocn.2008.21170

Van Petten, C., and Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *Int. J. Psychophysiol.* 83, 176–190. doi: 10.1016/j.ijpsycho.2011.09.015

Wagers, M. W., Lau, E., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Wagers, M. W., and Phillips, C. (2014). Going the distance: memory and control processes in active dependency construction. *Q. J. Exp. Psychol.* 67, 1274–1304. doi: 10.1080/17470218.2013.858363

Wechsler, S. (2011). Mixed agreement, the person feature, and the index/concord distinction. *Nat. Lang. Linguist. Theory* 29, 999–1031. doi: 10.1007/s11049-011-9149-x

Wicha, N. Y., Moreno, E. M., and Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy and gender agreement in Spanish sentence reading. *J. Cogn. Neurosci.* 16, 1272–1288. doi: 10.1162/0898929041920487

Zawiszewski, A., Santesteban, M., and Laka, I. (2016). Phi-features reloaded: an event-related potential study on person and number agreement processing. *Appl. Psycholinguist.* 37, 601–626. doi: 10.1017/S014271641500017X

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 May 2019
doi: 10.3389/fpsyg.2019.00942

Plural Conjuncts and Syncretism Facilitate Gender Agreement in Serbo-Croatian: Experimental Evidence

Ivana Mitić[1] *and Boban Arsenijević*[2]*

[1] Department of Serbian Language, University of Niš, Niš, Serbia, [2] Institute for Slavic Languages, University of Graz, Graz, Austria

The literature on agreement in South Slavic generalizes that conjunct agreement in gender is only possible when all conjuncts are plural (e.g., Bošković, 2009). Marušič et al. (2015) and Arsenijević and Mitić (2016a,b) attest a significant level of patterns contradicting this claim in elicited production experiments. They weaken the earlier generalization to a facilitating role of plural number for conjunct agreement in gender. However, the stimuli in the two respective experiments involve syncretism between the members of conjunction. The syncretism removes the possibility – at Phonological Form at least – that by agreeing with one conjunct, the verb disagrees with the other. It is hence expected to result in a similar surface effect as the facilitation by plurals, which makes it a potential confound variable. We report and discuss the results of an experiment aimed to test both the effect of syncretism and the reality of the facilitating effects of plural number. The results of the experiment yield positive answers to both questions: syncretism is a facilitating factor, but plural number nevertheless has its facilitating effect too – as confirmed by the stimuli without syncretism. Since syncretism is a phenomenon in which phonological information plays a central role, our findings support syntactic models of agreement which extend to the interface with phonology. Moreover, our results reveal a double similarity of conjunct agreement with agreement attraction, in both showing a (stronger) attraction effect of plural number compared to singular, and in being sensitive to syncretism (cf. Badecker and Kuminiak, 2007; Malko and Slioussar, 2013; i.e., Bader and Meng, 2002; Hartsuiker et al., 2003; Slioussar, 2018).

Keywords: agreement, syncretism, gender, number, Serbo-Croatian

OPEN ACCESS

Edited by:
Simona Mancini,
Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Zeljko Boskovic,
University of Connecticut, United States
Cristiano Chesi,
Istituto Universitario di Studi Superiori di Pavia (IUSS), Italy

***Correspondence:**
Boban Arsenijević
b.arsenijevic@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *19 November 2018*
Accepted: *08 April 2019*
Published: *07 May 2019*

Citation:
Mitić I and Arsenijević B (2019) Plural Conjuncts and Syncretism Facilitate Gender Agreement in Serbo-Croatian: Experimental Evidence. Front. Psychol. 10:942. doi: 10.3389/fpsyg.2019.00942

INTRODUCTION

Relevance of the Research

Grammatical agreement is a hallmark property of human language. Agreement in person, gender, and/or number of features between the subject and the verb is one of its prototypical instantiations. Consider the person and number agreement in the English example in (1).

(1) John smoke-s. vs. John and Bill smoke.

The properties of agreement, especially in conflicting situations, where different (sources of) information can be identified for the same feature, present a highly informative window into the nature of the features and their representation and processing in the brain.

One such conflicting context emerges when the subject consists of two or more conjoined nominal expressions with different number or gender features. What feature does the verb display in such contexts? Does it agree with one of the conjuncts (yielding what is referred to as conjunct agreement), and with which one, or does it display some other (default) feature? What ending should the verb display in (2)?

(2) Flaše i ogledala *Serbo-Croatian*
bottle.FPl and mirror.NPl
su izbačen-?.
AuxPl thrown.out-?[1]
"The bottles and the mirrors have been thrown out."

Sometimes, the conjuncts within the subject have different values of number and gender, but these combinations have phonologically identical exponents (a phenomenon known as syncretism) – leading to an even more complex situation. If the ending on the verb in (3) were *-a*, would it stand for FSg, NPl, or would it be underspecified between them?

(3) Flaša i ogledala *Serbo-Croatian*
bottle.FSg and mirror.NPl
je/su izbačen-a.
AuxSg/Pl thrown.out-?
"The bottles and the mirrors have been thrown out."

Here, the suffix *-a* stands in one case for the combination FSg, and in the other for NPl. In neither of the two occurences is it possible to identify the individual realizations of number and gender: the two features have a so-called fused realization.

The present research looks into this type of construction: verbs agreeing with a conjunction of two nouns with different number and gender features characterized by a fused syncretic realization, and informs two questions about syntactic features:

I. Does a fused morphological realization of two features by one simplex affix, in this case number and gender, imply that they are also computed as a bundle, or are they rather separately computed features bound by certain dependency relations?
II. Does syncretism in the morphological realization of combinations of different values of a set of features affect their processing in agreement?

Both these questions have theoretical linguistic as well as psycholinguistic relevance. In theoretical linguistics, they have been investigated for a wide range of languages, from Arabic (e.g., Aoun et al., 1994), to Hindi (e.g., Bhatt and Walkow, 2013), and to Slavic (e.g., Bošković, 2009), with a rich body of literature discussing the theoretical consequences of these facts (McCloskey, 1986; Munn, 1999; Doron, 2000; Citko, 2004, among many others).

In psycholinguistics, the question of the bundled vs. independent representation of number and gender has been investigated a.o. in Vigliocco et al. (1996), De Vincenzi (1999), De Vincenzi and Di Domenico (1999), Faussart et al. (1999), Igoa et al. (1999), Hinojosa et al. (2003), Barber and Carreiras (2005), Carminati (2005), Nevins et al. (2007), and Fuchs et al. (2015). Syncretism has been observed to play a role in agreement attraction – a process whereby the target of agreement displays the features of an unexpected expression referred to as the attractor. Typically, this is a nominal expression which intervenes in the linear order between the grammatical controller (by default, the subject) and the target (*the verb*). Consider example (4), where instead of the singular feature of the subject (*the box*), the verb receives the plural feature of the attractor (*the books*).

(4) The box with the books are in the basement.

The more features an expression shares with the controller, the more likely it is to act as an attractor. Syncretism between the controller and the attractor is one such similarity: it has been observed that having an ending syncretic with the ending of the grammatical controller of agreement increases the chances an expression will attract agreement (Bader and Meng, 2002; Hartsuiker et al., 2003; Slioussar, 2018). Moreover, plural number has been shown to be a stronger attractor than singular (Badecker and Kuminiak, 2007; Malko and Slioussar, 2013) – which makes for another parallel with the attractive power of the plural number on conjuncts in competition with singular.

Mixed Agreement in Gender and Number: Empirical Facts and Theoretical Relevance

Both the traditional and formal literature on conjunct agreement in Serbo-Croatian (henceforth SC), from Maretić (1899) to Bošković (2009), draw the empirical generalization that agreement in gender with a single conjunct obtains only when all conjuncts are plural (Pl).[2] In other cases – whether with all singular (Sg) conjuncts, or with a combination of Sg and Pl – mixed gender conjunction triggers default agreement (MPl). The empirical picture as reported is illustrated in (5).

(5) a. Flaše i ogledala su *SC*
bottle.FPl and mirror.NPl AuxPl
izbačen-e/izbačen-a/izbačen-i.
thrown.out-FPl/-NPl/-MPl[3]

[1]The following abbreviations are used in the paper: Aux for *auxiliary*, ConjP for the *conjunction phrase*, F for *feminine gender*, FCA for *first conjunct agreement*, GenP for *gender phrase*, LCA for *last conjunct agreement*, DEF for *default agreement*, M for *masculine gender*, MIX for *mixed agreement*, N for *neuter gender*, N^0 and n^0 for the *nominal categorial head*, NP for *noun phrase*, NumP for *grammatical number phrase*, Pl for *plural*, Refl for *reflexive*, SC for *Serbo-Croatian*, and Sg for *singular*.

[2]Babić (1998) and Bojović (2003) provide a number of exceptions, but most of their examples involve special kinds of conjunction – that clearly involving ellipsis, that where all the conjuncts after the first conjunct are its appositives, or that where conjunction has a disjunctive interpretation. It is worth noting that the different investigations used different methodologies, and relied on different formats and types of data. Babić (1998) and Bojović (2003) mostly have corpus examples with conjoined subjects of different forms, Bošković (2009) has own examples of the form "one NP and all NPs," while Marušič et al. (2007, 2015), Willer-Gold et al. (2016, 2018) as well as the present paper observe conjoined bare nouns from the experimental perspective.

"The bottles and the mirrors have been thrown out."

b. Flaša i ogledalo su
bottle.FSg and mirror.NSg AuxPl
*izbačen-e/*izbačen-a/izbačen-i.
thrown.out-FPl/-NPl/-MPl
"The bottle and the mirror have been thrown out."

c. Flaša i ogledala su
bottle.FSg and mirror.NPl AuxPl
*izbačen-e/*izbačen-a/izbačen-i.
thrown.out-FPl/-NPl/-MPl
"The bottle and the mirrors have been thrown out."

Arsenijević and Mitić (2016a,b) present experimental evidence that this is not entirely correct, and that with all Sg conjuncts – gender agreement in SC may still target a single conjunct. They report a significant level of production, as well as an only partial degradation of acceptability of sentences like (5b) when the verb agrees in gender with the first or with the last conjunct (henceforth First Conjunct Agreement, shorter FCA, and Last Conjunct Agreement, shorter LCA), suggesting that (6) is a more accurate empirical report than (5b).

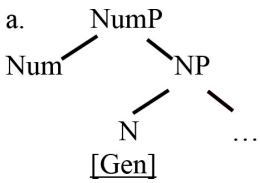
(6) Flaša i ogledalo su *SC*
bottle.FSg and mirror.NSg AuxPl
(?)izbačen-e/(?)izbačen-a/izbačen-i.
thrown_out-FPl/-NPl/-MPl
"The bottle and the mirror have been thrown out."

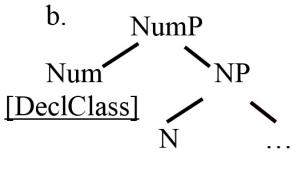
With FCA or LCA in gender, examples of this type manifest mixed agreement: agreement where gender has a single conjunct as a control, while number takes plural – either as the value of the entire conjunction, or as the semantically default value, but crucially a value that is not represented on any of the conjuncts.[3] This pattern has been observed also on combinations of conjuncts of different number (Sg and Pl) in Slovenian (Marušič et al., 2015: 25–26), a language with very similar behavior to SC when it comes to conjunct agreement. Their study is also the first study in South Slavic conjunct agreement that examines the behavior of doubly mixed conjunctions: those where the conjunct share neither the value for gender, nor for number (in particular, the combinations of neuter singular and feminine plural, and of neuter plural and feminine singular were examined: NSg&FPl, FPl&NSg, NPl&FSg, FSg&NPl).

Theoretical Modeling of Number, Gender, and Agreement in These Two Features

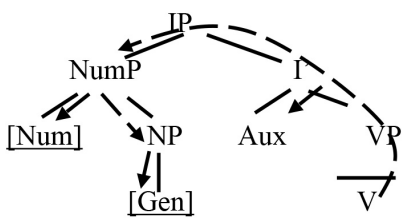
The investigation reported and discussed in the present paper targets the empirical issues of the effect of syncretism on agreement and of the attracting power of the plural number for agreement in gender. It has consequences for the question whether number and gender are represented as one feature-bundle or separately, and whether they enter agreement together or apart. It also has consequences for the question of whether agreement extends to the syntax–phonology interface. Further than that, it does not directly bear on any particular analysis or theoretical model of the representation of gender and of the operation of agreement. But in the interest of a better understanding of the phenomena discussed, and their theoretical relevance, we briefly present a somewhat simplified model of gender and number representation and agreement.[4]

At least since Ritter (1993), models have been entertained in which gender and number are syntactically represented separately, in two different projections within the nominal domain. Ritter argues that in languages where the grammatically relevant feature is gender itself, as in Hebrew, it figures as a feature of the nominal lexical category head with a derivational value (it derives a noun from another word or from the root), as in (7a), while in those where the relevant nominal property is rather the declension class (or the "word marker," as she calls it), as in Romance, this property is represented as a feature on number, in NumP, as in (7b).

(7) a.


b.


Both views lend themselves well to analyses arguing for an attraction effect of number regarding gender agreement. Assuming that the verb searches (probes) the local structural domain for number and gender features, obeying certain structural restrictions (as per Chomsky, 2001), in the structure in (7a), the search will come across number before reaching gender – as graphically represented in (8). The value of number encountered can influence how agreement proceeds. An effect in the opposite direction is predicted to be impossible to obtain.

(8)


The long-dashed line with arrows represents the direction of search for a gender and number feature.

The structure in (7b) is even more straightforward: declension class is a feature residing on number, and therefore is expected to be sensitive to the narrow value of number. In this case, however, dependencies in the opposite direction are not excluded either.

[3]We refrain from committing to either of the possible analyses of the plural number in these examples: as the value specified on the entire conjunction, labeled ConjP, or as the default value assigned in the absence of a specified value; we refer to it descriptively throughout the paper as default number agreement.

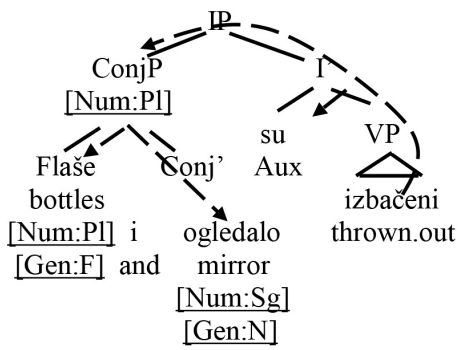
[4]Including such an overview was suggested to us by an reviewer, for which we express our gratitude.

In the meantime, arguments have been provided that even in Romance, number and declension class are represented separately. Fuchs et al. (2015) provide experimental evidence for a separate representation of number and gender in Spanish.

Serbo-Croatian is a language in which what is referred to as gender agreement is sensitive to both the semantic gender and the declension class of the noun [see Bošković (2009) for an argument that the two behave differently regarding conjunct agreement]. Findings like those in Arsenijević and Mitić (2016a,b), illustrated in (6) above, suggest that in SC the relevant features are specified separately from number – and it is exactly the reliability of these findings that are tested in the present paper.

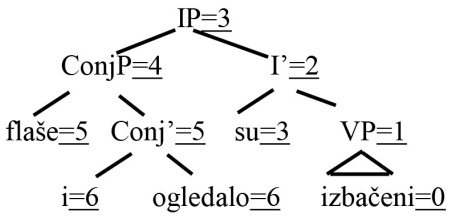
A hierarchical ordering similar to that in (7a) obtains with coordinated subjects. Conjunction of nominal expressions is known to derive semantically plural referents [but see Heycock and Zamparelli (2005) for a somewhat more complex view]. This can be modeled in terms of a plural feature in the conjunction phrase (ConjP, also referred to in the literature as the Boolean phrase, BoolP).[5] As illustrated in (9), conjunction itself has no effect on the interpretation of gender. Moreover, at least when the conjuncts are of different gender values, there is no single gender value that can be specified on the ConjP. Again, number – in this case plural – ends up hierarchically more local to the verb, and therefore with the capacity to trigger attraction effects regarding gender.

(9)



A range of different accounts of agreement have been proposed in the literature. Analyses of conjunct agreement in SC can be roughly classified in two families. One considers agreement a purely syntactic phenomenon. Bošković (2009), or Puškar and Murphy (2015), only use the syntactic operations Merge, Move (including pied-pipe), and Agree to derive the empirically attested patterns and eliminate the ungrammatical ones, exclusively relying on hierarchical structures, in particular on hierarchical locality, illustrated in (10a) for the relevant structural positions (by the underlined specification in the form = N, where the N component specifies the relative locality of the node to the verb from which the search originates). Marušič et al. (2007, 2015), on the other hand, argue that the linear locality of a conjunct to the verb is the strongest factor in Slovenian, a close relative of SC. They propose an account in which in agreement involves a crucial role of the interface with phonology, at which point linear locality plays an important role. Linear locality is illustrated in (10b).

(10) a.



b. Flaša = 4 i = 3 ogledalo = 2 su = 1
bottle.FSg and mirror.NSg AuxPl
izbačen-i = 0
thrown.out-MPl
"The bottle and the mirror have been thrown out."

The purely syntactic accounts have the locus of complexity in the syntactic operations involved in agreement (a complex interaction of different syntactic operations determines agreement), but avoid involving phonological considerations. The accounts involving a role of the interface with phonology place the complexity at the modular level, while dealing with simpler structural relations (agreement is determined by plain hierarchical and/or linear locality). Rather than resorting to complex computations within the module of syntax, they distribute them between two modules: syntax and phonology, with relatively simple computations within each, but with two modules involved rather than only one.

Marušič et al. (2015) Model of Conjunct Agreement

As noted in the section "Relevance of the research," effects that can be explained as attraction exhibited by the value of number specified on a conjunct over the gender agreement with that conjunct are observed in SC and in Slovenian. In order to account for them, while still deriving mixed agreement (referred to in their article as partial agreement), Marušič et al. (2015: 25–26) state the generalization that "[mixed] Agreement in Gender is allowed only when the Agreement value registered by the targeted conjunct Cx matches the Number value already on the verb (acquired from [ConjP])"[6] and argue for the following agreement procedure:

Step 1a. Agree: Participle Number([ConjP])
Step 1c. Copy-value: Participle Number([ConjP])
Step 2a. Agree: Participle Gender([ConjP]) → No Value on [ConjP]
Step 2b. Choose a Conjunct Cx where Number(Cx) = Number(Participle) Agree: Participle Gender(Conjunct Cx)
Step 2c. Copy-Value: Participle Gender(Cx)

We illustrate this in (11), on the example originally introduced in (2).

[5]In approaches like Citko (2004), the plural feature is not in ConjP, but on a pronominal element generated on top of it.

[6]They refer to (the projection standing for) the entire conjunction as the BoolP, but it has been replaced here with the notation ConjP, which is used in the present paper as fully synonymous.

(11) [$_{ConjP}$ Flaše i ogledala] *SC*
bottle.FPl and mirror.NPl
su izbačen-a.
AuxPl thrown.out-NPl
"The bottles and the mirrors have been thrown out."

Step 1a. *Agree: Participle Number(ConjP)*
In this step, the ConjP (the conjoined subject) is simply marked as the source of the number feature to occur on the participle.

Step 1c. *Copy-Value: Participle Number(ConjP)*
In this step, the number feature of the ConjP is copied onto the verb. ConjP is plural, since it involves two conjoined members (*bottles* and *mirrors* – that they are also plural only strengthens the plural status of the ConjP), hence the participle also receives the plural value.

Step 2a. *Agree: Participle Gender(ConjP) → No Value on ConjP*
In this step, ConjP is marked as the source of the gender feature to occur on the participle. However, no such feature is specified on ConjP due to the conflict among the gender values of the conjuncts (the first conjunct is feminine, the second is neuter).

Step 2b. *Choose a Conjunct Cx where Number(Cx) = Number(Participle) Agree: Participle Gender(Conjunct Cx).*
In this step, one conjunct is found, which matches the already copied value of number on the participle, and it is marked as the source of the gender to occur on the participle – in the example above, it is the last conjunct.

Step 2c. *Copy-Value: Participle Gender(Cx)*
In this step, the gender feature of the last conjunct is copied onto the verb. This conjunct is neuter, hence the participle also receives the neuter value.

The verb first agrees in number with the entire conjunction, thus receiving the value plural. Then it attempts to agree in gender with the entire conjunction – but fails since the ConjP is unspecified for a gender value due to the mixed gender values of its conjuncts. It then attempts to agree with the most local conjunct (in some grammars hierarchical locality matters, yielding FCA; in others linear locality, yielding LCA). However, conjunct agreement is not free – it is conditioned by the identity of the number value already acquired by the verb and the number value on the targeted conjunct. Since the value already acquired by the verb is plural, then as a result, plural number on the conjunct facilitates gender agreement with that conjunct.

A similar view is advocated by Arsenijević and Mitić (2016b), who investigate agreement with conjoined singulars. They observe that even singular agreement is attested on the verb at significant rates. As this pattern is unexpected on Marušič et al. (2015) model, where the verb must acquire the plural value of number, Arsenijević and Mitić (2016b) offer an alternative based on three soft constraints:

1. The verb should agree in number with the entire conjunction,
2. The verb should agree in gender with the local conjunct, and
3. The verb should agree with the same constituent in both gender and number.[7]

Plural conjuncts are more likely gender-agreement controllers than singular conjuncts because they allow for plural number on the verb to be interpreted both as agreement with the entire conjunction (hence avoiding a violation of the constraint 1 above) and as agreement with the plural conjunct (thus avoiding a violation of the constraint 3 above). With singular controllers of gender, if the verb is singular, it does not agree with the entire conjunction (violating constraint 1 above), and if it is plural, it does not have the same control as gender (violating constraint 3 above). In both cases, one of the constraints gets violated, and it is the ordering of constraints that decides the winner. On this view, plural number on the conjunct facilitates agreement in gender because when the verb agrees in gender with a plural conjunct – it satisfies both the constraint that it matches the number of the controller of gender agreement, and the one that requires it to match the number on the ConjP.

Both these investigations suffer from failing to control for one potential confound variable which is expected to have effects similar to those reported. Since masculine is the default gender in South Slavic conjunct agreement, in order to clearly attest FCA and LCA, the conjuncts must bear a combination of a feminine and a neuter gender value. In that case, each of the three gender values can in principle occur on the verb and signal a different agreement pattern: feminine and neuter the two different patterns of conjunct agreement, and masculine the default agreement. Both investigated languages, Slovenian and SC, display syncretism between FSg and NPl [compare (12a vs. 12d)], as well as between FPl and about a half of NSg nouns [compare (12b vs. 12c)]. This substantially undermines the findings of these two investigations: it is possible that the mixed agreement is simply an effect of the syncretism.

(12) a. žen-a knjiga slik-a *SC*
woman-FSg book-FSg picture-FSg
stolic-a
chair-FSg
b. žen-e knjig-e slik-e stolic-e
woman-FPl book-FPl picture-FPl chair-FPl
c. sel-o let-o polj-e mor-e
village-NSg summer-NSg field-NSg sea-NSg
d. sel-a let-a polj-a mor-a
village-NPl summer-NPl field-NPl sea-NPl

[7] Arsenijević and Mitić (2016a) provide evidence that the ordering of these constraints depends on other grammatical and semantic properties, such as the agentivity and animacy of the subject.

In Marušič et al. (2015), syncretism may be facilitating the ending that phonologically matches both conjuncts [the ending *-e* on the verb in (13a)]. In Arsenijević and Mitić (2016b) it is possible that the ending on the verb is supported by its phonological match with one conjunct in the form used, and with the plural form of the other [see (13b), where the feminine conjunct has *zakletv-e* as its plural form, and the neuter conjunct *obećanj-a*].[8] Since the verb tends to, or must be plural – it is reasonable to expect that this latent syncretism also plays a role. Especially considering that if the verb were plural and agreed in gender with one singular conjunct, its ending would be syncretic with that on the other singular conjunct. Therefore, in the results of both Marušič et al. (2015) and Arsenijević and Mitić (2016a,b), when the verb receives the ending *-e* or the ending *-a*, it is impossible to reliably determine whether it only does it due to bearing the respective features (and which features: NPl or FSg?), or it is, partially at least, because it phonologically matches the ending on one or both of the conjoined nouns.[9]

(13) a. Tel-e in krav-e so *Slovenian*
calf-NSg and cow-FPl AuxPl
se skril-e/skril-a za grmièevje.
Refl hid-FPl/NPl behind shrubs
"The calf and cows hid behind the shrubs."

b. Zakletv-a i obećanj-e su *SC*
oath-FSg and promise-NSg AuxPl
prekršen-e/prekršen-a.
broken-FPl/NPl
"The oath and the promise have been broken."

Hypotheses and Predictions

The null hypothesis predicts that the three types of agreement, FCA, LCA, and DEF, will be equally represented in the results, both with and without syncretism. However, since substantial research has already been done on some of the variables that have been controlled in our experiment, we can formulate a more informed, and more relevant, relative null hypothesis – as well as several competing alternative hypotheses and their predictions.

The reports in the literature before Marušič et al. (2015) and Arsenijević and Mitić (2016a,b) predict that due to the different number values on the conjuncts, only DEF will be produced. A significant level of production of FCA and/or LCA in gender would reject this view.

Hypotheses predicting conjunct agreement in gender need to be informed about the general ratio between the three agreement strategies, FCA, LCA, and DEF, in the configurations in which they are not suppressed or asymmetrically facilitated by additional factors. The best candidate for such a configuration is one with coordinated subjects involving only plural conjuncts. Willer-Gold et al. (2016, 2018) show that with this type of conjoined subjects, when the first conjunct is neuter and the last is feminine – DEF is the strongest strategy, followed by LCA – with FCA as the least produced pattern.[10] This is shown in **Figure 1**.

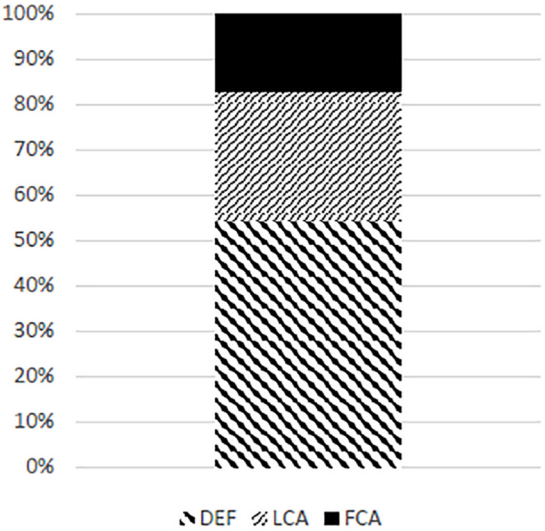


FIGURE 1 | The ratio between DEF, LCA, and FCA for all plural conjuncts.

These results can be taken as base-line expectations for the gender combination F&N, if we accept the conclusion based on the reports in Marušič et al. (2015) and Arsenijević and Mitić (2016a,b), namely that the agreement in mixed number and gender conjunctions is a special case of mixed gender conjunction agreement, with an additional facilitating effect of the plural number. Deviations from the distribution in **Figure 1** in that case indicate the effects of the two factors that we are investigating: facilitation of conjunct agreement in gender by the plural value of number and syncretism. This allows us to formulate the following alternative hypotheses and their predictions.

Hypothesis 1: As suggested in the literature (Bošković, 2009), a mixed value of number triggers DEF in number, there is no effect of syncretism whatsoever.

Prediction 1: In both conditions, only DEF will be produced, with zero instances of either FCA or LCA.

Hypothesis 2: Plural facilitates gender-agreement with the conjunct that bears it, because it matches the value of number of the entire conjunction. This hypothesis is an alternative to Hypothesis 1, as it makes the assumption

[8]In both investigations, the role of the confound variable is expected to be somewhat reduced for those stimuli in which one of the conjuncts is a NSg noun ending in *-o* (rather than *-e*). In these stimuli, the NSg noun is not syncretic with the FPl noun, which has the ending *-e*. However, since this variable was not controlled, this does not rescue the results of the experiments. It only predicts a somewhat smaller role of the confound variable than if syncretism was full.

[9]The fact that verbs with the ending *-o*, the unique NSg ending for the verb, were not produced in Marušič et al. (2015) supports the FPl analysis. Yet, since Arsenijević and Mitić (2016a,b), who conduct an experimental investigation of Sg&Sg conjunctions, do attest the NSg ending *-o* – the fact that it is not attested in Marušič et al. (2015) is probably due to a strong facilitating effect of the Pl conjunct.

[10]Participants in both Willer-Gold et al. (2016, 2018) and the present experiment were from the same population: first and second year students (19–23 years old) of non-linguistic majors from the University of Niš who have lived in the area the last 5 years or longer (we only discuss the results from Willer Gold et al. from the University of Niš – which was one among six sites where the experiment was administered).

that mixed gender and number conjuncts are a special case of mixed gender conjuncts. Therefore, it takes the production of agreement with all plural conjuncts, given in **Figure 1**, as a baseline.

Prediction 2: The ratio between LCA and FCA will change in favor of LCA in the condition without syncretism, in comparison to the base-line ratio in **Figure 1**.

Hypothesis 3: Syncretism facilitates conjunct agreement because the verb can then be interpreted both as showing FCA and LCA.

Prediction 3a: Combined with Hypothesis 1, it predicts that the condition without syncretism will elicit only DEF, with zero FCA and LCA, while the condition with syncretism will possibly elicit some LCA in addition to DEF.

Prediction 3b: Combined with Hypothesis 2, it predicts that on top of the effect of plural-facilitation (more LCA, less FCA in both conditions), syncretism will cause an additional increase of LCA and at the expense of DEF compared to the non-syncretic condition.

We conducted an experimental study to test whether indeed syncretism facilitates the production of the respective endings on the verb. Controlling for syncretism allowed us to examine our central question, i.e., to test whether the facilitation effect of the plural number on one of the conjuncts is real regarding the production of mixed agreement.

In the section "Elicited Production Study," we report and discuss the design and results of this experiment. The section "Design and Materials" describes the methodology, the experimental material used, and the fitting of the design, and the section "Participants" provides the information about the participants. The section "Procedure" summarizes the competing generalizations and hypotheses, and their predictions, and section "Procedure" reports the results. In the section "Results," we discuss how the results bear on the predictions outlined in the section "Procedure." The section "Conclusion" is the conclusion.

ELICITED PRODUCTION STUDY

Design and Materials

In order to investigate the effect of syncretism and facilitation of conjunct agreement by plural number in SC, we have designed and conducted an elicited production experiment, adopting the methodology implemented and reported in Marušič et al. (2015), Arsenijević and Mitić (2016a,b), Willer-Gold et al. (2016, 2018), and Mitić and Arsenijević (2019) and several other experimental works. The experiment was developed and administered using the Internet portal Ibex Farm[11].

[11]We express our gratitude to the administrators of Ibex Farm, in particular to its author Alex Drummond, for making our work considerably simpler.

Independent and Dependent Variables Adopted

We only had one dependent variable: the gender agreement pattern produced, with three levels: FCA (N), LCA (F), and DEF (M). Due to the mixed combination of genders, true resolved agreement (RES) from Willer-Gold et al. (2016), where the aggregate conjunction has the gender value shared by all the conjuncts, was not an option. There was only one manipulated independent variable: the presence vs. absence of syncretism between the conjuncts, i.e., whether the two conjuncts had homophone endings.

Properties of the Stimuli

All the sentences had preverbal subjects, were of approximately the same length in syllables and characters (mean length in syllables = 8.83, standard deviation = 0.70, mean length in characters = 26.00, standard deviation = 0.59), and involved nouns of similar frequency (average frequency 0.05 tokens per 1000 words, standard deviation 0.02, as per the Corpus of Contemporary Serbian Language, Krstev and Vitas, 2005)[12]. All the stimuli involved substitute subjects consisting of two conjoined disyllabic bare nouns (SC has no articles, hence bare nouns are fully unmarked), where the first member of conjunction was a NSg noun and the second a FPl noun. All substitute subjects had the identical length in syllables (five syllables each), and their length in characters ranged from 11 to 13, with a mean at 12.75, standard deviation: 0.61.

Out of the four possible combinations (NSg&FPl, NPl&FSg, FSg&NPl, FPl&NSg) – we included only one (NSg&FPl), for two reasons. One was that we wanted to keep as many variables controlled rather than tested, and avoid overcomplicating the experiment. Testing both variables – the order of gender values [shown to be a factor in Arsenijević and Mitić (2016a) and Willer-Gold et al. (2016, 2018)] and the gender value (in particular feminine or neuter) which is combined with the plural value of number [cf. the results in Marušič et al. (2015) for Slovenian] – is a task for further research. The other reason requires more details of the experiment to be introduced, and is elaborated below, and illustrated in (21). All the predicates in the stimuli were passive forms of transitive verbs.

The Stimuli

The experiment involved 60 stimuli: 12 critical (6 for each condition) and 48 fillers. The stimuli for each of the two conditions are illustrated in (14).

(14) Illustration examples for the two conditions

a. Condition with syncretism (both the NSg and the FPl noun end in *-e*):
Model sentence (i.e., first screen):
Ručak je pojeden na brzinu.
lunch.MSg is eaten.MSg on speed
"The lunch was eaten in rush."
Substitute subject (i.e., second screen):

[12]Frequencies were additionally tested in the SC Word Frequency Corpus (Arsenijević, 2018) with a more contemporary and less formal register-based sample.

jaje i šljive
egg.NSg and plum.FPl

b. Condition without syncretism (the NSg noun ends in *-o*, the FPl noun in *-e*)
Model sentence (i.e., first screen):
Dokaz je ukraden iz torbe.
evidence.MSg is stolen.MSg from bag
"The evidence was stolen from the bag."
Substitute subject (i.e., second screen):
pismo i mape
letter.NSg and map.FPl

There were two types of fillers. They were all identical in design like the critical examples (a model sentence with a MSg subject followed by a substitute subject), except that they had different substitute subjects. One group (N = 18), illustrated in (15a), involved conjoined substitute subjects with both plural conjuncts: FPl&NPl, such that one or both of the conjuncts were modified by an agreeing adjective. In the other (N = 30), the substitutes were nouns with a special behavior regarding number and gender, falling in five different sub-types, each represented with six items, illustrated in (15b–f).[13] The complete list of the stimuli is provided in the **Supplementary Table S1**.

(15) a. *Model sentence (i.e., first screen):*
konac je donet kod krojačice.
thread.MSg AuxSg brought at tailor
"The thread was brought to the tailor's."
Substitute subject (i.e., second screen):
ljubičaste igle i zrna.
violet.FPl needle.FPl and bead.NPl
"violet needles and violet beads"

b. *Model sentence:*
čuvar je obišao zgradu.
guard.MSg AuxSg visited building
"The guard visited the building."
Substitute subject:
Julijin komšija
Julija's.MSg neighbor.FSg/MSg

c. *Model sentence:*
vlasnik je došao u pekaru.
owner.MSg AuxSg come in bakery
"The owner came to the bakery."
Substitute subject:
moje cerekalo
my.NSg laugher.NSg/MSg

d. *Model sentence:*
konj je trčao po polju.
horse.MSg AuxSg run on field
"The horse ran around the field."
Substitute subject:
prvo prase
first.NSg pig.NSg

e. *Model sentence:*
rezultat je ohrabrio studente.
result.MSg AuxSg encouraged students
"The result has encouraged the students."
Substitute subject:
uspeh dekana
success.MSg dean.GenMSg

f. *Model sentence:*
kolega je zvao u podne.
colleague.MSg AuxSg called at noon
"My colleague called at noon."
Substitute subject:
prijatelj iz škole
friend.NSg from school GenFSg.

[13]The five sub-types of fillers were: hybrid agreement nouns with a possessive adjective (*Julijin komšija* "Julija's neighbor," where *komšija* can trigger M or F agreement), as in (15b); hybrid agreement nouns with a possessive pronoun (*moje cerekalo* "my laugher," where *cerekalo* can trigger M or N agreement), as in (15c); NSg animate nouns with an ordinal number [*prvo prase* "(the) first pig," where *prase* has a hybrid plural form, triggering FSg or NPl agreement], as in (15d); regular agreement nouns with a MSg nominal complement [*uspeh dekana* "(the) success (of the) dean," where the genitive complement is homonymous with the paucal, and the paucal allows paucal and MPl agreement], as in (15e); regular agreement MSg nouns with a PP complement (*prijatelj iz škole* "the friend from school," where attraction effects could be expected), as in (15f).

Participants

The experiment was conducted at the University of Niš. Thirty-six native speakers of B/C/S who had spent at least the past 5 years within the area in which this language is spoken participated in the experiment, with 18 per list (age range 19–23, average age 20.61, standard deviation 1.13). Participants included 28 (77.78%) females and 8 males (22.22%). The participants were all students in their first or second year of undergraduate programs which do not involve linguistic courses. A written informed consent was obtained from each participant. An ethics approval was not required for this research as per applicable institutional and national guidelines and regulations.

Procedure

The experimental procedure involved two steps for each stimulus. In the first step, the participant reads aloud a model sentence involving a masculine singular non-coordinated subject as in (16a), which is displayed on the first screen. In the second step, the second screen shows a substitute subject as in (16b), and the participant pronounces the sentence again, but with the substitute subject instead of the original one – adapting also the morphosyntax of the verb to it.

(16) a. *FIRST SCREEN*
ulaz je očišćen prošlog petka
entrance.MSg is cleaned.MSg last Friday
"The entrance was cleaned last Friday."

b. *SECOND SCREEN*
kupatilo i kuhinje
bathroom.NSg and kitchen.FPl

The agreement pattern used by the participant in the pronounced sentence is coded as Sg or Pl for number and as FCA, LCA, or DEF (Default) for gender.

The experiment begins with six training examples, used by the administrator to instruct the participants about the experimental procedure. The training examples involved, both in model sentences and as substitutes, only non-conjoined subjects of various, yet balanced number–gender combinations.

The details of the experiment most closely matched the methodology in Mitić and Arsenijević (2019). Critical items were organized in two lists, so that each stimulus occurred exactly once in each condition. The purpose was to control for a possible effect of the particular lexical items, or of various other idiosyncratic properties of the particular stimuli. The lexical items were selected from reference dictionaries, such that the resulting sentences could saliently be used in a natural conversation.[14] All participants completed the experiment, and were included in the results.

RESULTS

Data analysis was determined by the design of the experiment. Since both the predictor and the dependent variable are categorical, we had originally implemented a χ^2 test to assess the significance of the relevant differences. One reviewer suggested that we could obtain more reliable insights if we used a linear mixed effects model. Indeed, this test turned out to be partly applicable after we observed that in spite of the principled multi-level nature of the categorical variable of the gender-agreement pattern – the results instantiated only two of the three levels: LCA and DEF, without a single instance of FCA. Effectively, thus, both our categorical variables had two levels, and could be coded as pseudo-scalar variables, where one level is coded as 0 and the other as 1. For the comparisons involving datasets with three levels of the dependent variable (FCA, LCA, and DEF) – we were forced to stick to the χ^2-test. As the probabilities for all the effects that were significant are of a very low level ($p < 0.0001$ in all of them), we consider the χ^2-test sufficiently reliable as well.

The Results of the experiment, as mentioned above, included only LCA and DEF agreement (see the **Supplementary Table S1** for the aggregate raw results). All the produced sentences displayed unambiguous plural number, and no FCA was produced in either condition [i.e., no verbs were produced with the ending *-a*, as in (17d) and (18d), which is ambiguous between FSg and NPl, or with the NSg ending *-o*, as in (17c) and (18c)]. The actual results in percentages are given in **Table 1** and graphically represented in **Figure 2**, followed by illustration examples for each type of result data obtained.

TABLE 1 | Results of the experiment.

	DEF (%)	LCA (%)	Error (%)
Syncretism	63	36	1
No syncretism	80.5	15	4.5

(17) Illustration of examples for results in the condition with syncretism

a. Condition with syncretism, default agreement (63%):
Jaje i šljive su pojedeni na brzinu.
egg.NSg and plum.FPl are eaten.MPl on speed
"The egg and the plums were eaten in rush."

b. Condition with syncretism, LCA, and/or syncretism (36%):
Jaje i šljive su pojedene na brzinu.
egg.NSg and plum.FPl are eaten.FPl on speed
"The egg and the plums were eaten in rush."

c. Condition with syncretism, FCA in number, and gender (0%):
Jaje i šljive su/je pojedeno na brzinu.
egg.NSg and plum.FPl are/is eaten.NSg on speed
"The egg and the plums were eaten in rush."

d. Condition with syncretism, Pl, and FCA gender or Sg and LCA in gender (0%):
Jaje i šljive su/je pojedena na brzinu.
egg.NSg and plum.FPl are/is eaten.NPl/FSg on speed
"The egg and the plums were eaten in rush."

(18) Illustration of examples for results in the condition without syncretism

a. Condition without syncretism, default agreement (80.5%):
Pismo i mape su ukradeni iz torbe.
letter.NSg and map.FPl are stolen.MPl from bag
"The letter and the maps were stolen from the bag."

b. Condition without syncretism, LCA without syncretism (15%):
Pismo i mape su ukradene iz torbe.
letter.NSg and map.FPl are stolen.FPl from bag
"The letter and the maps were stolen from the bag."

c. Condition without syncretism, FCA in number, and gender (0%):
Pismo i mape su/je ukradeno iz torbe.
letter.NSg and map.FPl are/is stolen.NSg from bag
"The letter and the maps were stolen from the bag."

d. Condition without syncretism, Pl, and FCA gender or Sg and LCA in gender (0%):
Pismo i mape su/je ukradena iz torbe.
letter.NSg and map.FPl are/is stolen. NPl/FSg from bag
"The letter and the maps were stolen from the bag."

The χ^2-test has confirmed a significant difference between the distribution of agreement patterns in the two conditions:

[14]The salience of the examples was controlled in the following way. First, one author formulated the examples according to her intuition, and the other evaluated them; where disagreement emerged, examples were replaced and the procedure was repeated for the newly introduced examples.

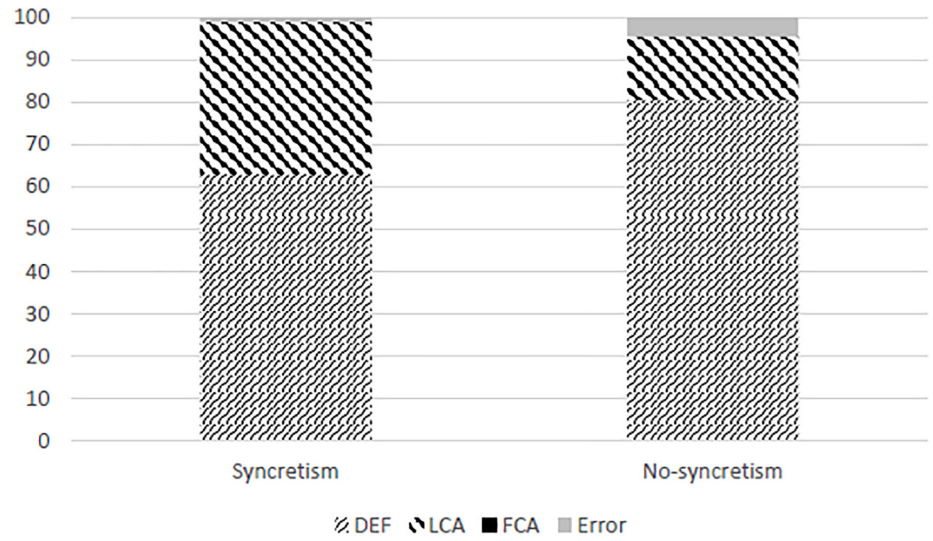


FIGURE 2 | A graphical representation of the produced patterns and errors.

there was significantly more LCA and less DEF in the syncretic than in the non-syncretic condition [$\chi^2(2, N = 421) = 22.79$, $p < 0.00001$]. Our results hence match the Prediction 2 from the section "Participants": the effect of syncretism in facilitating non-default agreement is clearly confirmed.

Even though both our variables were categorical (with levels syncretic and non-syncretic for the predictor, and FCA, LCA, and DEF for the dependent variable), due to the absence of FCA observations in the dependent variable both were effectively two-level variables in the data-set. As pointed out by an reviewer, this allows to code them as (pseudo-)scalar variables. We took advantage of this opportunity, and report this test as well. We used the lme4 package (Bates et al., 2012) in R (R Core Team, 2012) to subject the difference between the syncretic and non-syncretic conditions to a linear mixed effects model test. As the predictor we entered presence vs. absence of syncretism (as 0 and 1, respectively), and as the observations for the dependent variables we coded DEF as 0 and LCA and 1. As random effects, we entered items and participants, and we specified the binomial family, without random slopes: *glmer*(*AgreePattern* ∼ *Syncretism* + (1| *Participant*) + (1| *Item*), *family* = *binomial*, *data* = *SyncrAgree*). The test confirmed a significant difference between the distribution of agreement patterns in the two conditions ($\beta = -0.205$, $t = -4.897$, $p < 0.0001$, where the reference level of the intercept was LCA and the syncretic condition). The absence of syncretism thus resulted in a significantly lower rate of LCA, i.e., there was significantly more LCA and less DEF in the syncretic than in the non-syncretic condition. Our results hence match Prediction 3 from the section "Procedure": the effect of syncretism in facilitating non-default agreement is clearly confirmed.

In order to assess the significance of *the differences between the two patterns of agreement* produced within conditions, we compared each of the conditions to the null hypothesis regarding the rate of DEF and LCA (i.e., an equal number of elicited sentences for the two patterns). To achieve this, we used the same methodology as above. We pseudo-randomly distributed an equal number of LCA and DEF observations (coded as 1 and 0) across the aggregate number of observations for each level of the predictor variable. Hence as the predictor, we entered the null hypothesis and the relevant condition (i.e., syncretic and non-syncretic in independent applications of the test). We coded them as 0 for the null hypothesis and 1 for the respective condition – syncretic in one application of the test, and non-syncretic in the other). The dependent variable with two levels, DEF and LCA, was again coded as 0 for DEF and 1 for LCA. The linear mixed effects model attested a significant difference between the prediction of the null hypothesis and the result of the experiment for the non-syncretic condition ($\beta = -0.5$, SE = 0.03, $t = -17.64$, $p < 0.0001$, Intercept = 0.5). It did not, however, confirm the significance of the difference between the prediction of the null hypothesis and the syncretic condition ($\beta = -0.053$, SE = 0.04, $t = -1.307$, $p = 0.193$, Intercept = 0.5). Since syncretism is the marked level, the straightforward interpretation is that the difference between the two patterns of agreement (LCA vs. DEF) is confirmed for conjoined subjects involving conjunct with mixed both number and gender values, but syncretism strengthens LCA to the extent that this difference ceases to be visible.[15]

Even in the non-syncretic condition, there were 15% of produced sentences exhibiting unambiguous LCA. This is a relatively high rate of production, compared with the complete absence of FCA, and with Prediction 1 that no LCA will be produced. Note also that LCA is produced at rates much higher than typical error rates: the level of erroneous productions for

[15]The significance of the difference between agreement patterns within the condition is orthogonal to the hypotheses in the focus of the paper, as they only make predictions about the relative quantities between the conditions, or between the conditions and the base-line ratios. We report these tests in the interest of completeness of the report, as advised by the editors of the volume.

this type of task is typically below 5% (as is the case with the clear errors in the present experiment, as well as with the error rates attested in other experiments using similar methodology: Marušič et al., 2015; Arsenijević and Mitić, 2016a,b; Willer-Gold et al., 2016, 2018; Mitić and Arsenijević, 2019).

DISCUSSION

Our experiment clearly shows that not only is DEF available for conjoined subjects when conjuncts have different number values, but that LCA was present in both conditions (see **Table 1** and **Figure 2**). This clearly rejects the generalizations in the earlier literature, formulated in the section "Procedure" as Hypothesis 1, as well as the hybrid Hypothesis 3a based on the same generalization. A mismatch in number indeed decreases conjunct agreement in favor of default, but it does not eliminate it. Considering the reports of Marušič et al. (2015), Arsenijević and Mitić (2016a,b), and Willer-Gold et al. (2016, 2018) – this decreasing effect probably does not need to be restricted to a mismatch in number, but can also come from a difference in the gender values of the conjuncts within conjoined subjects – which is a topic for a separate investigation.

The results confirm Hypothesis 2, that plural number on a conjunct facilitates agreement with that conjunct, in congruence with the models by Marušič et al. (2015) and Arsenijević and Mitić (2016b). Recall Prediction 2, derived from this hypothesis in the section "Procedure," that conjoined subjects of the type NSg&FPl used in our experiment will elicit relatively more LCA and less FCA than the all-plural conjuncts in Willer-Gold et al. (2016, 2018); see **Figure 1**. Our results displayed a significant difference between the condition without syncretism (15% of LCA and 0% of FCA) and the NPl&FPl condition in Willer Gold et al. (30% of LCA and 17.78% of FCA), as well as between the condition with syncretism (36% of LCA and 0% of FCA) and the NPl&FPl condition in Willer Gold et al. Hypothesis 3 from the section "Procedure," more precisely its version 3b, was also confirmed. The prediction was that the syncretic condition will elicit more LCA and less DEF than the non-syncretic condition, and this difference was attested as significant.

In spite of the negative effect of the double mismatch between the conjuncts, both in gender and in number, the rate of LCA for NSg&FPl was the same or higher than for NPl&FPl subjects in the base-line data-set from Willer-Gold et al. (2016, 2018). The rate of FCA – the condition which was facilitated neither by plural number nor by syncretism, dropped to zero in our experiment, both with and without syncretism. We can conclude that both syncretism and plurals display clear facilitating effects on conjunct agreement in SC.

This means that while syncretism may have been a confounding variable in Marušič et al. (2015) and Arsenijević and Mitić (2016a,b), it was not solely responsible for the results. The generalization that conjunct agreement is not impossible with mixed number conjuncts and that plural on conjuncts facilitates agreement with them was still correct.

A curious question emerges from these results: Why was no FCA at all produced in the present experiment? In the experiment conducted by Marušič et al. (2015), syncretism was not controlled for, but otherwise there is a condition fully matching the type of stimuli in the present experiment: their condition NSg&FPl. This condition yields 5% of produced sentences with FCA. The obvious explanation is that Slovenian and SC are not that similar when it comes to conjunct agreement. Moreover, since the Sg&Sg conjunction tested in Arsenijević and Mitić (2016a,b) also rendered a considerable level of FCA (at the rate of 19%, which is not far from the level of 17% of FCA with Pl&Pl conjunctions reported in Willer-Gold et al. (2016), there seems to be a particularly strong negative effect of the double mismatch in feature values (both number and gender) in SC. Still, no definite conclusion regarding the question why FCA is so strongly suppressed can be offered based on our experiment, and therefore we leave it for further research.

Our results provide support for the models of agreement in which agreement is not a purely syntactic phenomenon, but partly takes place at the interface with phonology [Arregi and Nevins (2012); Marušič et al. (2015), and Willer-Gold et al. (2016, 2018) for South Slavic]. Syncretism is a phenomenon which involves phonological identity of the exponents of different feature–value combinations. If agreement were fully determined by syntactic structure, then syncretism would be less likely to have effect on agreement than if agreement extends to the interface with phonology. In views which distinguish between competence and performance, it is, however, possible that this effect is a matter of performance, and hence orthogonal to the question of modularity of agreement.

This possibility opens up a more general question which has not yet been convincingly answered in the literature: is conjunct agreement a grammatical agreement strategy, or an error similar to agreement attraction? The fact that in conjunct agreement the controller belongs to the subject and carries the relevant morphosyntactic features has made most researchers maintain the former option. This was further supported by the fact that conjunct agreement is produced, and its acceptability is judged, at the levels similar to, or often even significantly higher than those of the agreement plausibly interpreted as targeting the entire ConjP.

The sensitivity of conjunct agreement to syncretism as a property it shares with agreement attraction (Bader and Meng, 2002; Hartsuiker et al., 2003; Slioussar, 2018) calls for reconsidering this view. This is highly compatible with restricting the role of the syntactic structure to narrowing down the retrieval space for the agreement features – while the actual retrieval takes place at the interface with phonology (Arregi and Nevins, 2012; Marušič et al., 2015). Therefore, it is actually expected that those attractors which sit within the narrowest retrieval space will have a sigificantly stronger attraction power – which would explain the higher acceptability and rates of production compared to agreement with attractors which are outside the subject constituent, or with those in peripheral (i.e., modifier) positions within the subject constituent. In this view, an important question is where the line should be drawn between competence and performance within the phonological component of agreement.

Finally, our experiment does not provide decisive evidence for or against the view that number and gender are represented and processed as a bundle, rather than apart. Since in our experiment

no FCA was produced (recall that the first conjunct was singular, and the last conjunct was plural) – instances of LCA could be interpreted as LCA in both gender and number, and DEF as default in both features. Counterexamples would be those where number is plural, and gender has the value of the conjunct which has the singular value for number – which were not produced in our experiment (see the **Appendix** for the reasons we chose the distribution of features NSg&FPl in our experiment). This pattern is more likely to occur when the last conjunct is singular and the first conjunct is plural[16].

Our experiment provides evidence for a dependence of gender on number, and no such dependency in the other direction. However, as it was not designed to capture the latter, the strongest conclusion we can make in this respect is that number can be processed without gender (no indications regarding the processing of gender without number), and that the processing of gender is dependent on number.

CONCLUSION

Two recently proposed models of the interaction of number and gender agreement build on results attesting facilitation of conjunct agreement in gender by a plural value of number on the conjunct. As both experiments that these investigations are based on involve a possible confound variable of syncretism between the conjuncts – we tested both the effect of syncretism, and the facilitation effect of the plural number in the absence of syncretism. Our results are doubly confirming. Syncretism is indeed a factor that facilitates conjunct agreement, but the facilitating effect of the plural is also real. The research lends support to the models of agreement extending to the interface between syntax and phonology, and opens some new questions about conjunct agreement within and between the South Slavic varieties.

[16] We have designed and administered a new experiment along these lines, and we are currently analyzing the data.

ETHICS STATEMENT

As specified in the manuscript, at the time of administration of the experiment, no approval of an ethical committee was required by the positive regulations of the University of Niš or of the Republic of Serbia.

AUTHOR CONTRIBUTIONS

Both authors designed the experiment and analyzed the results. IM coded the experiment in Ibex, administered it, and coded the results. BA wrote the manuscript.

FUNDING

Leverhulme Trust grant Coordinated Research in the Experimental Morphosyntax of South Slavic Languages. Ministry of Education, Science and Technology, Republic of Serbia, grant Dinamika struktura srpskog jezika (OI 178014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00942/full#supplementary-material

TABLE S1 | Stimuli and row results of the experiment on the role of plural number and syncretism in conjunct agreement.

REFERENCES

Aoun, J., Benmamoun, E., and Sportiche, D. (1994). Agreement, word order, and conjunction in some varieties of Arabic. *Linguist. Inq.* 25, 195–220.

Arregi, K., and Nevins, A. (2012). *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Dordrecht: Springer.

Arsenijević, B. (2018). *Morphological Corpus of Serbo-Croatian*. Graz: University of Graz.

Arsenijević, B., and Mitić, I. (2016a). "Effect of animacy and agentivity on the processing of agreement in Serbo-Croatian," in *Studies in Languages and Mind, Selected papers from third Novi Sad Workshop on Psycholinguistics, Neurolinguistic and Clinical Linguistic Research*, eds S. Halupka-Rešetar and S. Martínez-Ferreiro (Novi Sad: Faculty of Philosophy), 41–77.

Arsenijević, B., and Mitić, I. (2016b). On the in dependence of gender with respect to number in agreement with coordinated subjects: an experimental study. *J. Slavic Linguist.* 24, 41–69.

Babić, S. (1998). *Sročnost (Kongruencija) u Suvremenom Hrvatskom Književnom Jeziku*. Zagreb: Matica hrvatska.

Badecker, W., and Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: evidence from gender and case in Slovak. *J. Mem. Lang.* 56, 65–85.

Bader, M., and Meng, M. (2002). *Case Attraction Phenomena in German*. Jena: University of Jena.

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *J. Cogn. Neurosci.* 17, 137–153.

Bates, D. M., Maechler, M., and Bolker, B. (2012). *Lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version 0.999375-42*.

Bhatt, R., and Walkow, M. (2013). Locating agreement in grammar: an argument from agreement in conjunctions. *Nat. Lang. Linguist. Theory* 31, 951–1013.

Bojović, D. (2003). Neki slučajevi kongruencije koordinirane subjekatske sintagme s preponiranim i postponiranim predikatom u srpskom jeziku. *Srpski Jezik* 8, 539–555.

Bošković, Ž (2009). Unifying first and last conjunct agreement. *Nat. Lang. Linguist Theory* 27, 455–496.

Carminati, M. (2005). Processing reflexes of the feature hierarchy (person > number > gender) and implications for linguistic theory. *Lingua* 115, 259–285.

Chomsky, N. (2001). "Derivation by Phase," in *Ken Hale. A Life in Language*, ed. M. Kenstowicz (Cambridge: MIT Press), 1–52.

Citko, B. (2004). "Agreement asymmetries in coordinate structures," in *Formal Approaches to Slavic linguistics: The Ottawa Meeting 2003*, eds A. Olga, W. Browne, cpsfnmM. Luisacpefnm Rivero, and D. Stojanović (Ann Arbor: Michigan Slavic Publications), 91–108.

De Vincenzi, M. (1999). Differences between the morphology of gender and number: evidence from establishing coreferences. *J. Psycholinguist. Res.* 28, 537–553.

De Vincenzi, M., and Di Domenico, E. (1999). A distinction among Φ-features: the role of gender and number in the retrieval of pronoun antecedents. *Ital. J. Linguist.* 11, 41–74.

Doron, E. (2000). "VSO and left-conjunct agreement : biblical Hebrew vs. modern Hebrew," in *The Syntax of the Verb Initial Languages*, eds A. Carnie and E. Guilfoyle (New York, NY: Oxford University Press), 75–96.

Faussart, C., Jakubowicz, C., and Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Ital. J. Linguist.* 11, 75–101.

Fuchs, Z., Polinsky, M., and Scontras, G. (2015). The differential representation of number and gender in Spanish. *Linguist. Rev.* 32, 703–737.

Hartsuiker, R., Schriefers, H., Bock, K., and Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Mem. Cogn.* 31, 1316–1326.

Heycock, C., and Zamparelli, R. (2005). Friends and colleagues: plurality, coordination, and the structure of DP. *Nat. Lang. Semant.* 13, 201–270.

Hinojosa, J., Martín-Loeches, M., Casado, P., Muñoz, F., and Rubia, F. (2003). Similarities and differences between phrase structure and morphosyntactic violations in Spanish: an event-related potentials study. *Lang. Cogn. Process.* 18, 113–142.

Igoa, J. M., García-Albea, J. E., and Sánches-Casas, R. (1999). Gender-number dissociation in sentence production in Spanish. *Ital. J. Linguist.* 11, 163–196.

Krstev, C., and Vitas, D. (2005). "Corpus. and Lexicon - Mutual Incompletness," in *Proceedings of the Corpus Linguistics Conference*, eds P. Danielsson and M. Wagenmakers, Liverpool.

Malko, A., and Slioussar, N. (2013). "Attraction errors in gender agreement: Evidence from Russian," in *Proceedings Formal Approaches to Slavic Linguistics: The Third Indiana Meeting*, eds S. Franks, et al. (Ann Arbor: Michigan University Press), 162–175.

Maretić, T. (1899). *Gramatika i Stilistika Hrvatskoga ili Srpskoga Književnog Jezika*. Zagreb: Štampa i naklada Knjižare L. Hartmana (Kugli i Deutsch).

Marušič, F., Nevins, A., and Badecker, W. (2015). The grammars of conjunction agreement in Slovenian. *Syntax* 18, 39–77.

Marušič, F., Nevins, A., and Saksida, A. (2007). "Last conjunct agreement in Slovenian," in *Proceedings of Formal Approaches to Slavic linguistics 2006,* eds R. Compton, M. Goledzinowska, and U. Savchenko (Ann Arbor: Michigan Slavic Publications), 210-227.

McCloskey, J. (1986). Inflection and conjunction in modern Irish. *Nat. Lang. Linguist. Theory* 4, 245–281.

Mitić, I., and Arsenijević, B. (2019). Structural ambiguity and optionality of agreement patterns in Bosnian/Croatian/Serbian conjunct agreement. *A J. Gen. Linguist.* 4:6.

Munn, A. (1999). First conjunct agreement: against a clausal analysis. *Linguist. Inq.* 30, 643–668.

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94.

Puškar, Z., and Murphy, A. (2015). "Closest conjunct agreement in serbo-croatian: a rule-ordering account," in *Topics at InfL. Linguistische Arbeitsberichte (LAB)*, Vol. 92, eds A. Assmann, S. Bank, D. Georgi, T. Klein, P. Weisser, and E. Zimmermann (Leipzig: Universität Leipzig), 441–482.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ritter, E. (1993). Where's gender? *Linguist. Inq.* 24, 795–803.

Slioussar, N. (2018). Forms and features: the role of syncretism in number agreement attraction. *J. Mem. Lang.* 101, 51–63.

Vigliocco, G., Butterworth, B., and Garrett, M. F. (1996). Subject-verb agreement in Spanish and English: differences in the role of conceptual constraints. *Cognition* 61, 261–298.

Willer-Gold, J., Arsenijević, B., Batinić, M., Becker, M., Čordalija, N., Kresić, M., et al. (2018). When linearity prevails over hierarchy in syntax. *PNAS* 115, 495–500. doi: 10.1073/pnas.1712729115

Willer-Gold, J., Arsenijević, B., Batinić, M., Čordalija, N., Kresić, M., Leko, N., et al. (2016). Morphosyntactic production of coordination agreement in South Slavic–a comparative study. *J. Slavic Linguist.* 24, 187–224.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX: ON THE CHOICE OF THE COMBINATION OF GENDER–NUMBER COMBINATIONS

This section clarifies some technical issues about the choices made in the design of our experiment, in light of the special properties of SC morphology. It is aimed primarily for those interested in the theoretical and descriptive linguistic, rather than psycholinguistic aspects of the research.

In our experimental design, we have capitalized on the fact that in SC NSg nouns end either in *-e* or in *-o*, and that their plural ends in *-a*, while at the same time FSg nouns end in *-a*, and their plural forms end in *-e*. This yields a crossed, yet incomplete syncretism.

(19) The crossed incomplete syncretism between F and N nouns in Sg and Pl

	Sg	**Pl**
N	-e, -o	-a
F	-a	-e

The combination of NSg and FPl allows for the formation of minimal pairs between a syncretic and a non-syncretic pair of nouns, while the combination of NPl and FSg allows for only one possibility, which is syncretic.

(20) a. NSg + FPl: NSg-e & FPl-e vs. NSg-o & FPl-e
polje i livade pismo i olovke
field.NSg and meadow.FPl letter.NSg and pen.FPl

b. NPl + FSg: only NPl-a & FSg-a
pisma i olovka
letter.NPl and pen.FSg

We used minimal pairs as in (20a) in our critical stimuli. The selected option, however, allows for two sub-options, depending on which gender–number combination comes as the first, and which as the last conjunct. This was decided by another similar consideration.

The verb bears the endings: *-o* for NSg, *-a* for NPl and FSg, or *-e* for FPl – i.e., it is possible to distinguish NSg from FPl on the verb. This means that, apart from the unambiguously default masculine ending *-i*, when a verb in *-e* was produced, we were sure that it was FPl, and when a verb in *-o* was produced, we knew that it was NSg. However, when a verb in *-a* was produced – it was uncertain whether it was plural, agreeing in gender with the first conjunct (NPl), or it was singular and agreed in gender with the last conjunct (FSg). The four logically possible combinations and their properties are illustrated in (21).

(21)

NSg	&	FPl	FCA (verb-NPl)	LCA (verb-FPl),	facilitated: LCA in FPl
-e/-o		*-e*	*-a*	*-e*	*-e*
FPl	&	NSg	FCA (verb-FPl)	LCA (verb-NPl),	facilitated: FCA in FPl
-a		*-e/-o*	*-e*	*-a*	*-a*
NPl	&	FSg	FCA (verb-NPl)	LCA (verb-FPl),	facilitated: FCA in NPl
-a		*-a*	*-a*	*-a*	*-a*
FSg	&	NPl	FCA (verb-FPl)	LCA (verb-NPl),	facilitated: LCA in NPl
-a		*-a*	*-a*	*-a*	*-e*

It was in the interest of the experiment to minimize the amount of the patterns of agreement realized as the ending *-a*, in order to also minimize the amount of ambiguously interpretable results. Tendencies reported in the literature (one of which is the topic of this paper) – that the verb agrees with the conjunct which bears the plural number and that it rather agrees with the last than with the first conjunct (Marušič et al., 2015; Arsenijević and Mitić, 2016a,b) – imply that if we have a plural conjunct in *-e*, and in particular if it is in the position of the last conjunct, there will be few instances of verbs with the ending -a produced (or even none, as it turned out to be the case).

All combinations other than the selected NSg&FPl would include a considerable participation of the ambiguous ending *-a* on the verb and/or would locate the plural value of number, whose facilitating effects are tested, on the first conjunct – which is a less likely controller of agreement (Willer-Gold et al., 2016, 2018).

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 May 2019
doi: 10.3389/fpsyg.2019.01002

Error-Driven Retrieval in Agreement Attraction Rarely Leads to Misinterpretation

Zoe Schlueter[1,2], Dan Parker[3] and Ellen Lau[1]*

[1] Department of Linguistics, University of Maryland, College Park, MD, United States, [2] Department of Linguistics and English Language, The University of Edinburgh, Edinburgh, United Kingdom, [3] Linguistics Program, Department of English, College of William & Mary, Williamsburg, VA, United States

Previous work on agreement computation in sentence comprehension motivates a model in which the parser predicts the verb's number and engages in retrieval of the agreement controller only when it detects a mismatch between the prediction and the bottom-up input. It is the error-driven second stage of this process that is prone to similarity-based interference and can result in the illusory licensing of a subject–verb number agreement violation in the presence of a structurally irrelevant noun matching the number marking on the verb ('*The bed by the lamps were*. . .'), giving rise to an effect known as 'agreement attraction'. Here we ask to what extent the error-driven retrieval process underlying the illusory licensing alters the structural and thematic representation of the sentence. We use a novel dual-task paradigm that combines self-paced reading with a speeded forced choice task to investigate whether agreement attraction leads comprehenders to erroneously interpret the attractor as the thematic subject, which would indicate structural reanalysis. Participants read sentence fragments ('*The bed by the lamp/lamps was/were undoubtedly quite*') and completed the sentences by choosing between two adjectives ('*comfortable*'/'*bright*') which were either compatible with the subject's head noun or with the attractor. We found the expected agreement attraction profile in the self-paced reading data but the interpretive error occurs on only a small subset of attraction trials, suggesting that in agreement attraction agreement checking rarely matches the thematic relation. We propose that illusory licensing of an agreement violation often reflects a low-level rechecking process that is only concerned with number and does not have an impact on the structural representation of the sentence. Interestingly, this suggests that error-driven repair processes can result in a globally inconsistent final sentence representation with a persistent mismatch between the subject and the verb.

Keywords: sentence processing, comprehension, grammatical agreement, memory retrieval, similarity-based interference, agreement attraction

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Laurel Brehm,
Northwestern University,
United States
Titus von der Malsburg,
Universität Potsdam, Germany

****Correspondence:***
Zoe Schlueter
zoe.schlueter@ed.ac.uk;
zschluet@umd.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *12 December 2018*
Accepted: *15 April 2019*
Published: *07 May 2019*

Citation:
Schlueter Z, Parker D and Lau E (2019) Error-Driven Retrieval in Agreement Attraction Rarely Leads to Misinterpretation. Front. Psychol. 10:1002. doi: 10.3389/fpsyg.2019.01002

INTRODUCTION

Much recent work has asked whether the interpretation comprehenders arrive at always tracks the syntax. We pursue this issue by investigating whether the illusory licensing of an agreement violation ('*The key to the cabinets are rusty*'), known as agreement attraction, reflects a change in the structural and thematic representation of the sentence or a low-level rechecking operation.

Previous work has shown that when comprehenders receive input that cannot be integrated into the current parse, they often engage in structural reanalysis of the previous input. This illustrates that an error signal can cause restructuring, but does a grammatical illusion like agreement attraction also reflect structural reanalysis? If the error signal from an agreement violation triggers similar reanalysis, the structural representation would be consistent with the grammar and the attractor would be misinterpreted as the subject. Although the interpretation would differ from the input, it would be consistent with the structure of the mental representation. However, if agreement attraction is the result of a simple rechecking operation the final representation contains an agreement violation. Here, we show that the illusory licensing of subject–verb number agreement generally does not lead to the misinterpretation of the attractor as the thematic subject, suggesting that most instances of agreement attraction do not reflect a structural reanalysis when the attractor is misretrieved in the search for the agreement controller in memory. Instead, we propose that error-driven retrieval of the agreement controller generally involves a low-level number rechecking operation.

Structure and Interpretation

In the past 15 years there has been mounting evidence that the interpretations comprehenders arrive at are not always uniformly consistent with the linguistic input (for recent reviews see Ferreira, 2003; Christianson, 2016; Karimi and Ferreira, 2016). Renewed interest in this question was first sparked by work by Ferreira and colleagues, who showed that after reading garden-path sentences like '*While Anna dressed the baby played in the crib*', participants would frequently accept interpretations not consistent with the input, answering 'yes' when asked if Anna had dressed the baby (Christianson et al., 2001; Ferreira et al., 2001). Ferreira and colleagues initially considered an 'erroneous structure' view, concluding that comprehenders do not always recover completely from the initial misparse in garden-path sentences. However, more recent research (Slattery et al., 2013) suggests that the lingering misinterpretation observed with garden-path sentences is not a result of the parser's failure to completely reanalyze the structural representation, but a failure to suppress the initial interpretation. In other words, if both parsing and interpretation are incremental, then the initial (erroneous parse) will have been interpreted even if the syntactic parse is successfully reanalyzed at the point of disambiguation. Therefore, the interpretation of the initial misparse is not licensed by the final input, but it is consistent with an interpretation derived from the structure during processing. Slattery et al. (2013) argued that this interpretation lingers in memory and can impact end-of-sentence judgments, even if the ultimate syntactic parse – and the ultimate sentence-level interpretation – is consistent with the input.

Misinterpretations have recently also been observed for implausible but syntactically unambiguous sentences. Gibson et al. (2013) found that participants frequently answered comprehension questions about implausible sentences (like '*The mother gave the candle the daughter*') not based on the grammatically licensed interpretation, but rather on a plausible alternative (here '*The mother gave the candle to the daughter*'). Gibson et al. (2013) argued that such effects can be explained by a noisy channel model of language comprehension (e.g., Levy, 2008; Levy et al., 2009). Interestingly, there is evidence that comprehenders not only generate a plausible interpretation that is not licensed by the linguistic input, but that they actually build a syntactic representation of the unlicensed interpretation. For instance, implausible sentences with a double object construction have been found to syntactically prime the prepositional dative construction of the plausible alternative (Slevc and Momma, 2015). This finding is consistent with a speech error reversal system proposed by Frazier and Clifton (2015); Frazier (2015). According to this account, comprehenders use their knowledge of the production system – specifically, what kind of speech errors frequently occur – to repair the input they receive. Similar proposals have also been made to account for the systematic misinterpretation of antecedent-ellipsis mismatches (Arregui et al., 2006; Frazier, 2013; but cf. Parker, 2018).

Misinterpretations are not random and arise systematically: garden-path sentences, implausible sentences, and other types of mismatches present instances in which the interpretation is not licensed by the actual linguistic input, but is licensed by the structure that is assigned to the input at some stage during processing. In these cases, the parser engages in structural reanalysis when it encounters an error signal from the bottom-up input. For instance, in the case of garden-path sentences, the misinterpretation arises before the parser engages in reanalysis of the input and then lingers, whereas for implausible sentences the error signal is semantic in nature (the comprehender arrives at an interpretation that they believe was not the intended speaker meaning) and leads to reanalysis that is not consistent with the actual input. Importantly for us, this suggests that the parser frequently engages in structural reanalysis in response to error signals and that misinterpretations are systematically linked to structures assigned to the input which are consistent with the misinterpretation.

In summary, there is clear evidence that under certain circumstances comprehenders systematically generate interpretations that are not faithful to the linguistic input. However, it seems possible that this involves building grammatically well-formed structural representations that are consistent with the misinterpretation, though not completely faithful to the input. Here, we ask whether misretrieval due to similarity-based interference in subject–verb agreement attraction is another source of systematic misinterpretation. In the following sections we outline the mechanisms underlying agreement attraction and how they might interact with interpretation.

Subject–Verb Agreement Attraction

Subject–verb agreement in English is a morphosyntactic dependency in which the number feature on the verb has to match the number feature of the subject. This dependency is susceptible to so-called "agreement attraction" errors, in which the number marking on the verb matches a structurally inaccessible plural noun rather than the singular subject ('*The key*

to the cabinets are rusty'). Agreement attraction occurs not only in production (Bock and Miller, 1991), but also in comprehension, where these sentences are often perceived as grammatical and do not show the processing cost normally associated with agreement violations (e.g., Pearlmutter et al., 1999; Wagers et al., 2009). This facilitation can be accounted for by a memory architecture based on cue-based retrieval (Wagers et al., 2009; Dillon et al., 2013; Tanner et al., 2014; Lago et al., 2015; Tucker et al., 2015). Sentence processing frequently requires comprehenders to establish dependencies between items that are not directly adjacent to each other, which means that retrieving items from memory is central to language comprehension. According to cue-based retrieval models (e.g., McElree, 2000; Van Dyke and Lewis, 2003; Lewis and Vasishth, 2005), items are encoded in memory as bundles of features and are content-addressable based on the features they contain (Lewis et al., 2006). When retrieval is triggered, the retrieval cues available at the retrieval site are used to access the target item in memory. Activation from each cue is transferred to each item with a matching feature and the item with the highest activation level is retrieved. When the target is a perfect match for all the retrieval cues, a partial match between the cues and a non-target item will not prevent it from being retrieved. However, when there is a partial mismatch between the target's features and the cues, the presence of a partially matching non-target item can lead to the misretrieval of this non-target item, in what is known as "similarity-based interference".

In the case of subject–verb agreement, the retrieval cues on the verb include both structural and number cues, e.g., [+subject] and [+plural] (see Arnett and Wagers, 2017, for discussion of the subject cue). When there is a number mismatch between the subject and the verb in the presence of a plural non-subject attractor (i.e., ungrammatical sentences like, '*The key to the cabinets are. . .*'), the activation from the number cue raises the level of activation of the attractor, but not the subject. In a subset of cases, this leads to the misretrieval of the number-matching attractor instead of the number-mismatching subject. This is reflected in higher acceptance rates and an amelioration of the processing difficulty associated with agreement violations in online measures.

In a cue-based retrieval model of agreement attraction there are two theoretical possibilities about when retrieval of the agreement controller is triggered. In principle, it is possible that subject–verb agreement processing in comprehension always involves retrieval of the agreement controller from memory, regardless of whether the verb and subject match in number. In grammatical sentences, the subject's features are a perfect match for the retrieval cues on the verb: it fulfills both the structural cue of being the subject and its number feature matches the number cue. Even if there is a structurally irrelevant noun that matches the number marking on the verb, this item only receives activation from one of the retrieval cues. Its activation level is therefore lower than that of the subject (modulo effects of noise). Consequently, the appropriate target is retrieved from memory. Retrieval in a sentence with an agreement violation would be triggered in the same way (by default), but the outcome would be different.

The second possibility under a cue-based retrieval account is that the retrieval-process underlying agreement attraction is an error-driven phenomenon (Wagers et al., 2009; Lago et al., 2015) that occurs only when the verb and subject mismatch in number (i.e., ungrammatical sentences). There is overwhelming evidence that language comprehension is not exclusively driven by bottom-up input and that comprehenders deploy top-down mechanisms to make use of existing information to predict upcoming input (see Kutas et al., 2011, for review). In the case of subject–verb agreement, this motivates a view in which comprehenders predict the number of the upcoming verb based on the number feature of the subject. If the bottom-up input matches their prediction, the verb's number marking is licensed and there is no need to retrieve the agreement controller. However, when the prediction is violated, this triggers error-driven retrieval of the agreement controller. Under this model, grammatical sentences without an agreement violation do not involve cue-based retrieval. Instead, agreement checking is a two-stage process and the second step (retrieval) is limited to instances where an agreement violation has been detected.

An important type of evidence in favor of this two-stage model are data suggesting that comprehenders initially show sensitivity to the agreement violation even in the presence of a number-matching attractor. Recent research has shown that attraction effects occur in the right tail of the reading time distribution, compared to the effect of grammaticality which also exerts an influence on faster reading times (Staub, 2009, 2010; Lago et al., 2015). Moreover, in eye-tracking studies, agreement violations have been observed in early reading time measures, while attraction effects were only found in late reading time measures (Dillon et al., 2013; Parker and Phillips, 2017). This suggests that during the initial processing of the verb comprehenders are sensitive to the agreement violation even in the presence of a plural attractor. The amelioration of the processing disruption associated with this violation does not occur until a later stage of processing.

Agreement and Interpretation

While this study focuses on the question whether agreement attraction leads to the misinterpretation of the local noun as the thematic subject, it should be noted that a separate question relating to agreement and interpretation is whether attraction cases reflect instances where the number of the subject is misrepresented as plural. Representational models relying on feature percolation or spreading activation like those often assumed for agreement attraction in production (e.g., Bock and Eberhard, 1993; Pearlmutter et al., 1999; Bock et al., 2004; Eberhard et al., 2005) have sometimes been proposed to extend to comprehension (Pearlmutter et al., 1999). The question whether comprehenders mistakenly interpret the subject as plural is central to representational accounts of agreement attraction in comprehension but has only rarely been directly addressed in previous studies.

One study that did investigate the subject's number representation in agreement attraction was conducted by Patson and Husband (2016). This study used self-paced reading followed by comprehension questions that explicitly probed

participants' interpretation of the subject's number feature: a sentence like '*The key to the cabinets are on the table*' was followed by the question '*Was there more than one key?*'. Comprehenders were more likely to agree that there were multiples of the entity denoted by the singular head noun when there was a plural attractor or a plural verb. This effect was strongest in agreement attraction configurations, in which both the attractor and the verb were plural. This study was recently replicated and extended by Brehm et al. (2019), who observed the same pattern of results to the comprehension questions, and additionally found that non-literal interpretations were more likely when the sentence was assumed to be produced by a native speaker of standard American English compared to an L2 speaker or a speaker of a regional dialect. Based on these studies, it does seem that comprehenders do indeed sometimes misrepresent the number of the complex subject noun phrase.

However, for both Patson and Husband (2016) and Brehm et al. (2019), non-literal answers about the number of the subject occurred not only in agreement attraction configurations, but whenever there was a plural feature present on the attractor or the verb. While a non-literal answer in the presence of a plural attractor would support a representational account of agreement attraction in comprehension, there are two reasons why the data overall suggest a somewhat different explanation. First, non-literal answers were also more common when the local noun was singular and only the verb was plural, which is not predicted under representational accounts of agreement attraction. As Brehm et al. (2019) point out, this is consistent with a noisy channel model of comprehension, in which comprehenders make rational inferences about the intended meaning of anomalous utterances. Second, Patson and Husband's self-paced reading data are not consistent with the automatic misrepresentation of complex noun phrases, as it shows no evidence of disrupted processing at the verb in grammatical sentences when the attractor was plural ('*The key to the cabinets was*. . .'). If comprehenders misrepresent the number feature of the subject in the presence of a plural attractor, this should be reflected in processing difficulties at the verb in grammatical sentences with plural attractors. One alternative explanation of the comprehension results in these studies is that answers to explicit comprehension questions are not always an accurate reflection of the representation built during the earlier processing of the sentence.

In fact, a recent series of experiments by Dempsey et al. (2016) and Tanner et al. (2018) is consistent with this alternative explanation. In a self-paced reading task, they used items in which a complex noun phrase with a singular head noun and either a singular or plural noun inside a prepositional modifier was introduced as the object in the first sentence and then referred back to by a singular or plural noun phrase as the subject of the second sentence ['*My husband placed the newspaper with the perfume ad(s) on the kitchen table. The newspaper(s) looked muddy* . . .']. They did not find any facilitation in the processing of a co-referential plural noun phrase when the noun inside the prepositional modifier was plural. This indicates that the complex NP's number information had not been misrepresented as plural by virtue of containing a plural element. In spite of this, a quasi-replication of Patson and Husband's study with the same materials as the self-paced reading task showed that follow-up comprehension questions about the number of the entity denoted by the complex NP were affected by the presence of a plural noun inside the prepositional modifier. Tanner et al. (2018) argued that, when taken together with the self-paced reading data, this shows that comprehension question accuracy might not directly reflect the misrepresentation of NP number during processing. Instead, they proposed a feature misbinding account according to which direct metalinguistic questions might lead to the retrieval of "floating" plural features that are not bound to their lexical hosts in memory. In the Discussion, we return to the question of number misinterpretation and whether agreement attraction in comprehension might result in, rather than stem from, misrepresenting the subject as plural.

Although representational models can account for the agreement attraction data in production, they fail to capture some of the comprehension data. If agreement attraction is a result of misrepresenting the number feature of the subject, this predicts that grammatical sentences should sometimes be perceived as ungrammatical in the presence of a plural attractor ('*The key to the cabinets is*. . .'). However, that does not seem to be the case (Wagers et al., 2009; Lago et al., 2015; Tucker et al., 2015; but cf. Pearlmutter et al., 1999). Cue-based memory retrieval models provide a good account of the formation of morphosyntactic dependencies such as subject–verb agreement in sentence processing. However, the ultimate goal of comprehension is not to establish dependencies between items to check formal features, but to derive the intended interpretation by building a structural representation of the input. We therefore ask if the output of memory retrieval operations for checking formal features changes the structural representation and interpretation of a sentence.

Under a two-step model of agreement attraction, encountering an agreement violation is an error signal from the bottom-up input. As previously discussed, the parser frequently engages in structural reanalysis when it encounters error signals, for example at the point of disambiguation in garden-path sentences. However, it should be noted that the proposed reanalysis in agreement attraction would be fundamentally different from reanalysis in garden-path sentences. In a garden-path sentence, it is simply impossible to integrate the new input into the existing structure without violating structural constraints. In contrast, when the parser encounters a subject–verb agreement violation, the structural configuration for integrating the verb is there. There is only a mismatch between one of the predicted features (number) and the bottom-up input. If reanalysis is costly, it might only be deployed when the error-signal is triggered by a severe violation. Moreover, in garden path sentences, the parser assigns a different analysis to the entire previous input. In agreement attraction, misrepresenting the attractor as the subject would require excluding some of the previous input from the newly built structure. In a sentence like '*The key to the cabinets are old*,' if the attractor ('*the cabinets*') is misanalyzed as the subject due to misretrieval in agreement

checking, there is no clear way for the subject's actual head noun to be incorporated into this revised structure. Reanalysis might only be possible if the input that has already been assigned a structure can be completely integrated into the new structure.

If agreement attraction involves reanalysis and the retrieval output is integrated in the subject position, this would lead to misinterpretation of the attractor as the thematic subject. The interpretation would not be consistent with the linguistic input, but not because comprehenders are engaging in shallow parsing. Instead, the misinterpretation would be a systematic result of the basic properties of the memory system subserving language comprehension. Here, we briefly review the studies that we are aware of that address the question of whether the attractor is misanalyzed as the subject in agreement attraction.

Thornton and MacDonald (2003) conducted a series of experiments examining the impact of whether the attractor was also a plausible subject for the verb. In two production studies, participants were presented with a preamble containing two nouns ('*The album by the classical composers*') and a verb that had to be used to form a complete sentence. They manipulated whether the verb could have both the head noun and the attractor or only the head noun as a plausible (passive) subject and found that agreement attraction error rates were increased when the plural attractor was a plausible subject. The comprehension experiment also showed plausibility effects as reflected in an increase in reading time at the verb in the presence of a plural attractor when both the head noun and the attractor were a plausible subject, which is reminiscent of the semantic interference found by Van Dyke and McElree (2006). However, the comprehension experiment did not include ungrammatical sentences to test for agreement attraction effects. Therefore, the data is not directly informative about how misretrieval for formal feature checking can alter interpretations in comprehension.

Pittman and Smyth (2005) replicated Thornton and MacDonald's production results and added a new component to the elicited production task in order to investigate whether participants had misrepresented the attractor as the subject in cases where they produced agreement errors. After repeating the preamble and completing the sentence using the given predicate, which was either plausible with both the head noun and the local noun or only with the head noun, participants were presented with a choice of two predicates. They had to continue the sentence using '*and*' followed by whichever of the two predicates they chose. One of the predicates was always a semantic match for the head noun and the other for the attractor. For example, for a preamble like '*The boy by the trees*' with the first predicate '*tall*' (matching both head and attractor) or '*playful*' (matching only the head), the choice would be between '*chubby*' and '*green*.' As in previous studies on agreement attraction in production, preambles with a singular head noun and a plural local noun led to the production of more agreement errors. The agreement error rate was higher when the local noun was a plausible subject of the first predicate and the selection error rate for the second predicate was higher in trials in which participants had produced an agreement error. According to Pittman and Smyth, this shows that participants sometimes got confused about which of the nouns was the thematic subject during the planning stage of production and a subset of the agreement errors were a reflection of this confusion. While this suggests that in an elicited production task the attractor might sometimes be misinterpreted as the thematic subject, these data do not allow us to draw conclusions about the impact of misretrieval on the structural representation of the sentence in comprehension. Not only are agreement attraction in production and comprehension often attributed to different mechanisms (Acuña-Fariña, 2009, 2012; Acuña-Fariña et al., 2014; Tanner et al., 2014), but the misinterpretation in this case arose during the message planning stage, which does not apply to comprehension. However, as outlined above, if the retrieval output for agreement checking is used to change the existing parse of the sentence, a possible consequence of misretrieval in agreement attraction is that comprehenders might misinterpret the attractor as the subject of the sentence.

Lau et al. (2008) used inverted pseudoclefts in a self-paced reading experiment to address the question whether the attractor is misinterpreted as the subject by testing for plausibility effects at the thematic verb. They used sentences like '*The phone by the toilets was/were what Patrick used/dialed/flushed/embarrassed*,' in which they manipulated grammaticality as well as the plausibility of the head noun and the attractor as thematic subjects by varying the verb. If agreement attraction triggers structural reanalysis and the misrepresentation of the attractor as the thematic subject, the plausibility match between the attractor and the verb should matter. However, the results only show a main effect of head noun plausibility with participants exhibiting a slow-down at the thematic verb when the head noun of the subject was not a plausible match. There was no interaction with attraction context or the plausibility of the attractor. Lau et al. (2008) conclude that the misretrieval of the attractor does not lead to thematic subject reassignment, meaning that the misretrieval is selective for formal feature satisfaction. However, this study used inverted pseudoclefts, which is not a structure used in other agreement attraction studies. It requires retrieval of the subject not just for agreement checking at the inflected auxiliary, but again at the wh-word before the main verb is encountered, which might have influenced their results. We address this question by using a dual-task design that provides a very clear measure of which noun phrase comprehenders took to be the subject.

THE PRESENT STUDY

We used a novel dual-task paradigm to investigate whether agreement attraction leads comprehenders to erroneously interpret the attractor as the subject of the sentence. Misinterpretation of the attractor as the thematic subject would indicate that the retrieval output for agreement checking is used to alter the structural representation of the sentence. We developed a dual-task paradigm combining self-paced reading with a forced-choice task. Participants read sentence fragments and had to complete them by selecting an adjective that was either compatible with the head noun of the subject or the attractor noun. The choice of adjective on each trial is indicative of whether the attractor

was misrepresented as the subject. If erroneously retrieving the attractor in the process of agreement checking leads to the reanalysis of the attractor as the subject, we expect to see a higher rate of participants choosing the adjective that matches only the attractor in an agreement attraction configuration, i.e., with an ungrammatical verb and a plural attractor. If, however, the error-driven retrieval process in agreement checking is has no impact on the structural representation, comprehenders should not be more likely to choose the attractor-matching adjective in the agreement attraction condition.

The nature of the dual-task paradigm also makes it possible to analyze not only adjective choice and overall reading times, but to take adjective choice on each trial into consideration when analyzing reading times. Overall, we expected to find a typical agreement attraction profile for the self-paced reading data, i.e., a slow-down in ungrammatical conditions, ameliorated by the presence of a plural attractor. If agreement attraction causes comprehenders to mistake the attractor for the subject, this should be reflected by choosing the attractor-matching adjective. Consequently, in the reading time data we would expect an attraction effect for trials on which the attractor-matching adjective was chosen. In contrast, we would expect to see less attraction for trials that culminated in a head-matching adjective choice. However, if misretrieval of the attractor does not result in reanalysis, the reading time data should show agreement attraction regardless of adjective choice.

Participants

Sixty-four native speakers of American English were recruited via Amazon Mechanical Turk for monetary compensation. All participants in this experiment and both norming studies described below provided informed consent and underwent a screening for native speaker abilities. This screening probed knowledge of the constraints on English morphology, tense, modality, ellipsis, and syntactic islands.

Materials

There were 48 items sets in 4 conditions. Each item consisted of a sentence fragment for self-paced reading and two adjectives for the sentence-final adjective-choice task. The sentence fragments all had a complex subject with a singular head noun and a prepositional modifier containing the attractor. The subject was followed by an inflected form of '*be*' and two adverbs. The sentence-final adjective was displayed as a forced-choice task: one adjective was a plausible match only for the head noun of the subject and the other only for the attractor, as illustrated in (1). We manipulated attractor number (singular/plural) and grammaticality (grammatical/ungrammatical). The full set of experimental items can be found in the **Supplementary Materials**.

(1)

(a) The boy by the tree is really very CHUBBY/GREEN
(b) The boy by the tree are really very CHUBBY/GREEN
(c) The boy by the trees is really very CHUBBY/GREEN
(d) The boy by the trees are really very CHUBBY/GREEN

The items were distributed across four lists in a Latin Square design. In addition to the 48 experimental items, each list also contained 72 filler items of similar syntactic complexity for which participants also had to choose between two possible sentence-final completions.

Plausibility Norming

Since the premise of the dual-task paradigm is that the adjective choice is informative about whether the participant has misinterpreted the attractor as the thematic subject, it is crucial that one of the adjectives is semantically plausible only for the head noun and the other only for the attractor. We conducted a plausibility rating study of simple sentences with potential head nouns and attractor nouns in subject position, varying the predicative adjective. The aim was to select 48 item sets in which one of the adjectives was rated highly plausible only for the head noun and the other only for the attractor.

Thirty native speakers of English participated in an adjective norming study on Ibex in which they rated 66 items in 6 conditions for plausibility on a scale from 1 (very implausible) to 7 (very plausible). These participants did not participate in the other norming study or the main experiment. All items were grammatical and the task also included 18 plausible fillers, 16 implausible fillers and 7 control items. We constructed 66 preliminary items containing a complex subject with a prepositional modifier, followed by an inflected form of *be*, two adverbs, and a sentence-final adjective. For each item, there were 8 conditions, crossing attractor number, grammaticality, and adjective plausibility. Based on these preliminary items, we constructed 66 item sets for norming, manipulating whether the subject was the head noun or the attractor noun in the 66 preliminary items. Apart from subject type (head noun vs. attractor), we also manipulated adjective type (head-match vs. attractor-match), and subject number. Since in the materials for the dual-task paradigm the head noun of the subject is always singular, the norming study included plural versions only of the attractors. This led to a total of six conditions, as illustrated in (2). The ratings were used to calculate the average plausibility ratings for the plausible conditions (a, d, f) and the implausible conditions for each item (b, c, e). We then selected the 48 items with the greatest difference between plausibility ratings for the plausible and the implausible conditions.

(2)

(a) The **boy** is really very *chubby*.
(b) The **boy** is really very *green*.
(c) The **tree** is really very *chubby*.
(d) The **tree** is really very *green*.
(e) The **trees** are really very *chubby*.
(f) The **trees** are really very *green*.

Agreement Attraction Norming

The 48 chosen items were then used in a speeded acceptability judgment task to confirm that they caused the expected agreement attraction effect. 24 native speakers of American

English read sentences presented word-by-word in the center of the screen with a stimulus onset asynchrony of 400 ms (inter-stimulus interval: 100 ms). None of these subjects participated in the other norming study or the main experiment. Following each sentence, participants had 2,000 ms to indicate whether the sentence had been acceptable. The instructions explicitly asked them to judge sentences based on whether they sounded like natural English. There were 72 fillers (half grammatical) in addition to the 48 experimental items. In order to avoid exposing participants to a large number of implausible sentences, the sentence-final adjective was always the one compatible with the head noun of the subject. In the dual-task paradigm, the attraction effect in self-paced reading is measured on the verb and its spillover regions, before participants are presented with the adjectives.

The acceptance rates across conditions were analyzed with a mixed-logit model (Jaeger, 2008), excluding trials on which no response was made within 2,000 ms (2.5% of all trials). The acceptance rates for each condition are plotted in **Figure 1**. **Table 1** contains the results of the mixed-logit model with grammaticality and attractor model as fixed effects (sum-coded). The random effects structure included by-subject and by-item random intercepts and by-subject random slopes for grammaticality.[1] As expected, grammatical sentences were more likely to be judged acceptable than ungrammatical sentences (89.4% vs. 16.7%). Sentences with a plural attractor were also more likely to be accepted than sentences with a singular attractor (49.5% vs. 57.1%), but this effect was driven by the higher rate of acceptance of ungrammatical sentences with plural attractors. Participants were more likely to accept an ungrammatical sentence when the number of the attractor was plural (25.1% for ungrammatical sentences with a plural attractor compared to 8.2% for those with a singular attractor). This indicates that comprehenders indeed experience attraction with this particular item set, making these materials suitable for the novel dual-task paradigm.

TABLE 1 | Results of the mixed logit model in the speeded acceptability judgment task.

Parameter	Estimate	Std. error	*z*-value	*p*-value
Intercept	0.13	0.23	0.57	0.57
Grammaticality	2.41	0.17	13.89	<0.001
Attractor number	−0.39	0.10	−3.79	<0.001
Grammaticality × attractor number	0.45	0.10	4.31	<0.001

[1]The model also converged with by-subject and by-item random intercepts and by-subject random slopes for attractor number, but the significance of the effects does not depend on which of these models is used.

Procedure

The sentences were presented in a self-paced reading paradigm with centered display using Ibex Farm (Drummond, 2019). Participants had to press the spacebar to see each new word and only one word at a time was visible. When they pressed the spacebar to reveal the final word of the sentence, the two adjectives for the forced-choice task appeared on the screen simultaneously, one to the left of the center and one to the right. The order in which the adjectives were displayed was randomized for each participant. Once the two adjectives appeared, participants had 3,000 ms to choose one of them by pressing the 'f'-key for the one on the left or the 'j'-key for the one on the right. If no response was made within 3,000 ms, the adjective-choice task timed out and the experiment moved on to the next trial.

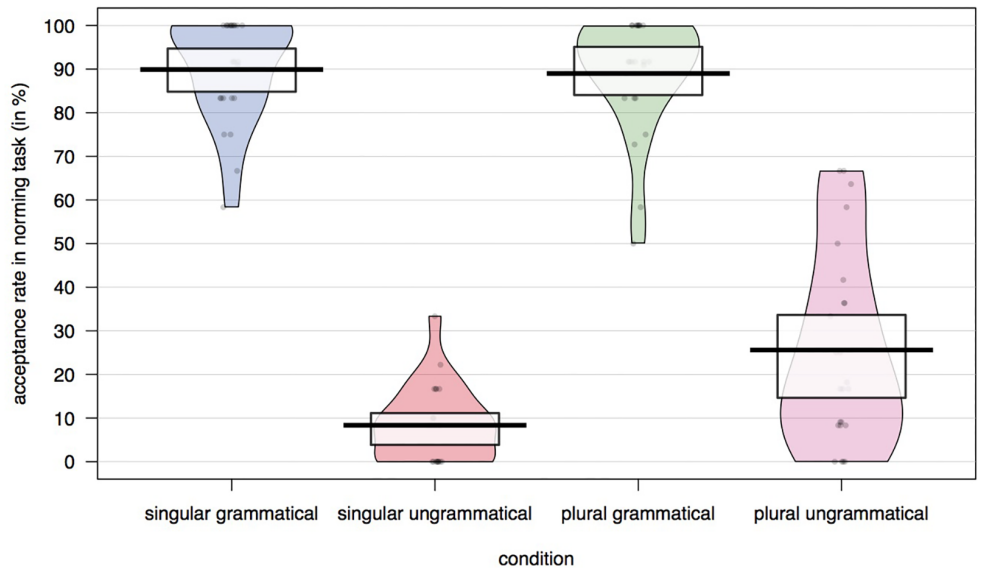


FIGURE 1 | Acceptance rates across conditions in the speeded acceptability judgment task.

Analysis

Trials on which there was no response within the 3,000 ms time limit were excluded from all analyses reported here (1.4% of experimental trials, 42 of 3,072 trials). We analyzed responses to the adjective-choice task with a mixed logit model (Jaeger, 2008) using the lme4 package (Bates et al., 2015) in the R computing environment (R Core Team, 2018). The model included attractor number and grammaticality as fixed effects (sum-coded) and by-subject and by-item random intercepts. The model was initially fitted with the maximal random effects structure, which was then simplified until the model converged (Barr et al., 2013).

Although the main focus of the experiment was the adjective-choice task, we also analyzed the self-paced reading data. The regions of analysis were the verb and its spillover region (first adverb). Reading times of 0 ms and reading times exceeding a threshold of 2,000 ms were not included in the analysis, leading to the exclusion of less than 0.2% of experimental trials in each region of analysis.[2] RTs were log transformed and analyzed using linear mixed effects models with attractor number, grammaticality and adjective choice as fixed effects. The final model included random by-subject and by-item intercepts. In addition, we also split the SPR data based on adjective choice on each trial and conducted a response-contingent RT analysis.

Results

Adjective-Choice Task

The percentage of trials on which a head-noun matching adjective was chosen for each of the experimental conditions is plotted in **Figure 2** and the results from the model are presented in **Table 2**. There was a significant main effect of grammaticality ($p < 0.01$): participants were more likely to choose the adjective that matched only the subject's head noun in grammatical than in ungrammatical sentences. There was also a significant interaction between grammaticality and attractor number ($p = 0.03$). In ungrammatical sentences participants were less likely to choose the head-matching adjective when the attractor was plural. As can be seen in **Figure 2**, the overall accuracy rate in the forced-choice task was very high. The rate of choosing the attractor-matching adjective was only 5.6% higher in the attraction condition (ungrammatical with a plural attractor: 16.6%) than in the grammatical condition with a plural attractor (10.8%), and only 3.2% higher than in the ungrammatical condition with a singular attractor (13.3%).

TABLE 2 | Results of the mixed logit model for adjective choice.

Parameter	Estimate	Std. error	*z*-value	*p*-value
Intercept	2.57	0.23	11.39	<0.001
Grammaticality	0.18	0.06	3.00	<0.01
Attractor number	0.04	0.06	0.63	0.53
Grammaticality × attractor number	−0.13	0.06	2.16	0.03

Figure 3 plots raw RTs for head-matching and attractor-matching adjective responses across conditions. Results of the linear mixed effects model with fixed effects of grammaticality, attractor number and adjective choice are presented in **Table 3**. There was a significant effect of adjective choice ($t = -3.17$), with a slowdown in trials on which the attractor-matching adjective was chosen compared to when the head-matching adjective was chosen. The RT difference between head-compatible and attractor-compatible adjective responses was larger in the grammatical than the ungrammatical conditions. However, this interaction between grammaticality and adjective choice was only marginally significant ($t = -1.95$).

[2]This led to the exclusion of 3 out of 6057 observations in the regions of analysis.

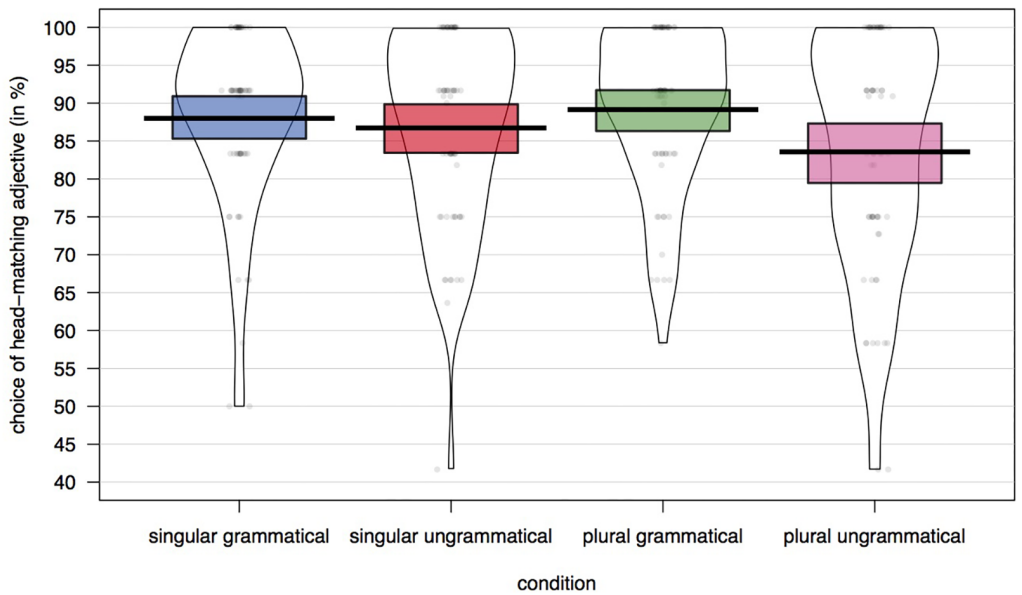


FIGURE 2 | Percentage of trials with a head-matching adjective choice across conditions.

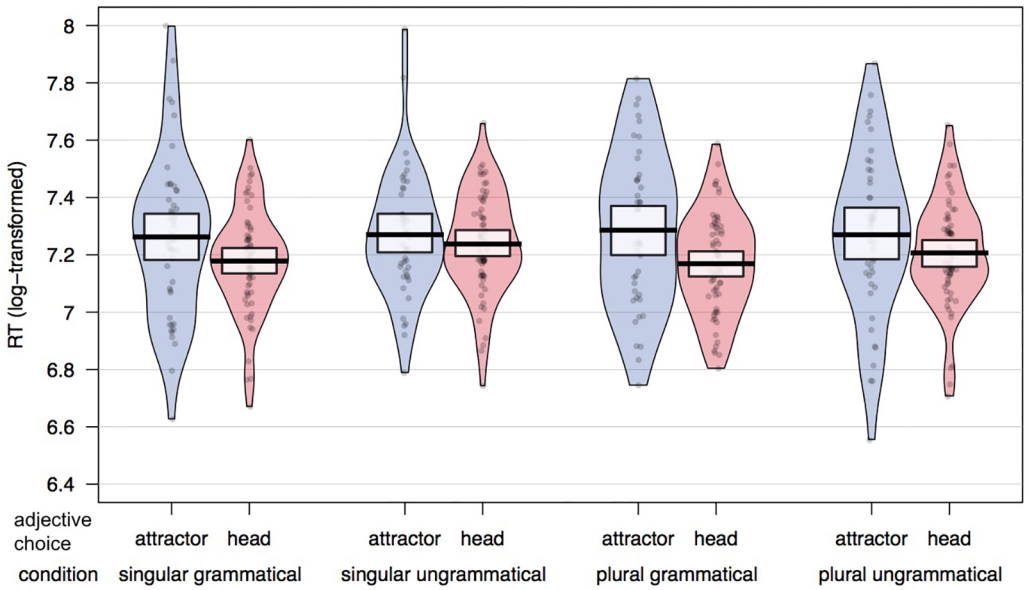


FIGURE 3 | Mean RTs split by adjective choice (attractor-matching response in blue; head-matching response in red) in each experimental condition. Proportion of head noun compatible responses beneath condition labels.

Self-Paced Reading

All analyses were performed on log transformed RTs. **Table 4** contains the results of the linear mixed effects models for the verb region and the spillover region. The region-by-region average (log-transformed) reading times are plotted in **Figure 4**. The only significant effect in the verb region was a three-way interaction between grammaticality, attractor number and adjective choice ($t = 2.48$): Grammaticality had a larger effect on adjective-choice when the attractor was plural compared to when it was singular. In the spillover region, there was a main effect of grammaticality ($t = -4.08$), with increased reading times for ungrammatical sentences. There was also a main effect of attractor number ($t = 2.02$), with increased reading times for sentences with singular attractors, but the interaction between grammaticality and attractor number was not significant.

TABLE 3 | Results of linear mixed effects model of response time on the adjective-choice task (using log transformed RTs).

Parameter	Estimate	Std. error	*t*-value
Intercept	7.25	0.028	262.20
Grammaticality	<−0.01	0.012	−0.01
Attractor number	<−0.01	0.012	−0.14
Adjective choice	0.04	0.014	−3.17
Grammaticality × attractor number	−0.01	0.012	−0.85
Grammaticality × adjective choice	−0.03	0.013	−1.95
Attractor number × adjective choice	0.01	0.013	1.04
Grammaticality × attractor number × adj. choice	0.01	0.013	0.38

Response-Contingent Self-Paced Reading

The nature of the dual-task paradigm allows us to examine reaction time profiles of trials based on adjective choice. **Figure 5** shows the average log-transformed reading time per region for each condition for trials on which the (correct) head-matching adjective was chosen. The plot looks almost identical to the overall SPR plot. Visually, there is a very clear slow-down for the ungrammatical conditions in the verb's spillover region, which is ameliorated for ungrammatical sentences with a plural attractor. Statistical analysis confirms this: While there is no significant effect in the verb region, in the verb's spillover region grammaticality, attractor number and their interaction had a significant effect on reading times (see **Table 5**). As expected, agreement violations led to a slowdown in the verb's spillover region compared to sentences with correct subject–verb agreement, as reflected in the main effect of grammaticality ($t = -6.67$). Reading times in the spillover region were longer for sentences with a singular than a plural attractor ($t = 2.78$). This result was not expected and seems to be attributable to the large slowdown in the ungrammatical condition with a singular attractor: the large slow-down in the ungrammatical singular condition means that the average RT of the two singular conditions is significantly slower than the average RT of the two plural conditions. Crucially, reading times show an agreement attraction pattern with the slowdown associated with a subject–verb number agreement violation being much reduced in the presence of a plural attractor (interaction between grammaticality and attractor number: $t = -3.18$).

Average log-transformed reading times for trials on which participants chose the attractor-matching adjective are plotted in **Figure 6**. It should be noted that the high accuracy on the adjective choice task meant that the sample size for this analysis was extremely small, so we do not present a statistical analysis.

TABLE 4 | Results of the linear mixed effects model (using log transformed RTs).

Parameter	Estimate	Std. error	*t*-value
Verb region			
Intercept	5.849	0.043	136.49
Grammaticality	−0.003	0.008	−0.33
Attractor number	−0.004	0.008	−0.51
Adjective choice	−0.004	0.008	−0.51
Grammaticality × attractor number	0.014	0.008	1.91
Grammaticality × adjective choice	−0.004	0.008	−0.53
Attractor number × adjective choice	−0.002	0.008	−0.31
Grammaticality × attractor number × adjective choice	0.019	0.008	2.48
Spillover region			
Intercept	5.907	0.043	136.17
Grammaticality	−0.030	0.007	−4.08
Attractor number	0.015	0.007	2.02
Adjective choice	0.013	0.008	1.72
Grammaticality × attractor number	−0.008	0.007	−1.04
Grammaticality × adjective choice	0.006	0.008	0.75
Attractor number × adjective choice	<−0.001	0.008	−0.02
Grammaticality × attractor number × adjective choice	0.009	0.007	1.22

Visual inspection of the plot reveals a very different pattern than for the head noun compatible adjective response trials with a slowdown for ungrammatical sentences with a plural attractor in the verb region. However, this data is suggestive at best and we refrain from interpreting it.

DISCUSSION

As expected, participants showed a clear agreement attraction effect in the overall self-paced reading data. If misretrieval of the attractor triggers structural reanalysis, this should be reflected in participants' choosing the attractor-matching adjective. In fact, we did find that participants chose the attractor-matching adjective more frequently in the agreement attraction configuration. However, the subset of trials on which this happened was small across all conditions. If we take the speeded-acceptability data from the attraction norming study as a very rough proxy of how frequently participants experienced attraction in these materials, we can compare this to the rate of misinterpretation in the adjective-choice task. In the norming study, the rate of accepting ungrammatical sentences with a plural attractor was 16.9% higher when the attractor was plural (25.1% acceptance rate) compared to when it was singular (8.2% acceptance rate). In contrast, the rate of choosing the attractor-matching adjective was only 3.3% higher in ungrammatical sentences when the attractor was plural (16.6%) compared to ungrammatical sentences in which the attractor was singular (13.3%). While we acknowledge that this is a very rough estimate, we do think it suggests that misretrieval of the attractor during agreement processing frequently occurs without resulting in the misinterpretation of the attractor as the subject.

Further evidence against the idea that agreement attraction generally results in reanalysis comes from the response contingent analysis of the self-paced reading data. There is a clear pattern of agreement attraction in the trials on which the correct head-matching adjective was chosen (the majority of trials). Under a view in which misretrieval of the attractor leads the parser to reanalyze it as the subject, we would expect less attraction on these trials than in the overall data since misretrieval should result in choosing the attractor-matching adjective. Unfortunately, the subset of trials on which the attractor-matching adjective was chosen is too small for statistical analysis and we cannot easily compare the rate of attraction in the self-paced reading data based on adjective choice.

Although the results demonstrate that error-driven retrieval for agreement checking is not inextricably linked to reanalysis, they also suggest that misretrieval and misinterpretation are not completely independent. The advantage of the dual-task paradigm is that it provides an explicit measure of what participants interpreted as the subject on each individual trial: while comprehenders very rarely chose the adjective compatible with the attractor, they did so significantly more frequently in ungrammatical sentences with plural attractors. This suggests that the attractor is at least occasionally misrepresented as the subject and that error-driven retrieval in response to the detection of an agreement violation might contribute to the likelihood of structural reanalysis.

The nature of the task meant that the number marking always had to appear on copular '*be*,' which is semantically impoverished, but it is possible that misretrieval of the attractor triggers restructuring if the verb simultaneously contains additional semantic cues in favor of the alternative structure [see Cunnings and Sturt (2018) for data suggesting implausible verb-object combinations are susceptible to semantic facilitative similarity-based interference]. Moreover, the type of materials could have made reanalysis less likely: the subject's head noun was always the first noun in the sentence, making it very salient. In fact, participants could have used a task-specific strategy in which they rely on sentence-initial position to establish subjecthood in the adjective choice task. In future research, this potential task-specific heuristic could be prevented by including items in which subjecthood and sentence-initial position are dissociated.

While the results of the present study point toward an interaction between error-driven retrieval for agreement checking and misinterpretation, it should be acknowledged that a potential explanation for this pattern can be provided without assuming that it is directly linked to agreement attraction as such. The average reading times for trials with an attractor-matching response were faster than for trials on which the head-matching adjective was chosen. Again, it needs to be noted that this was only a very small subset of trials. Nevertheless, this would be compatible with a situation in which attractor-matching responses might occur on trials on which participants were not paying attention. In that case, the mental representation

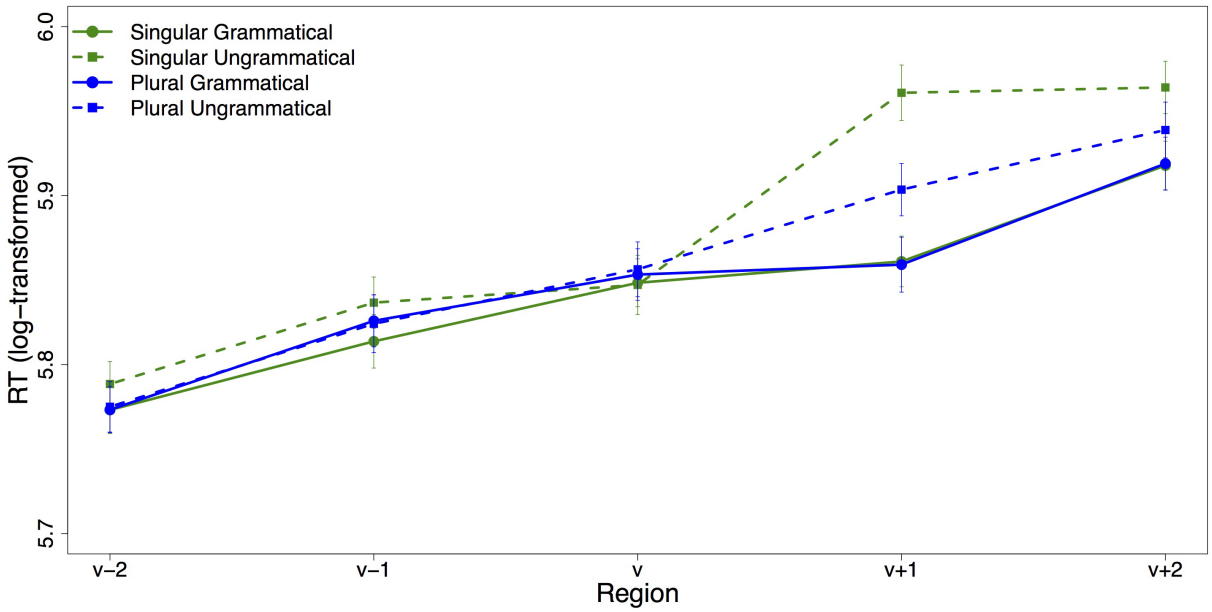


FIGURE 4 | Region-by-region mean reading times. Error bars indicate standard error of the mean.

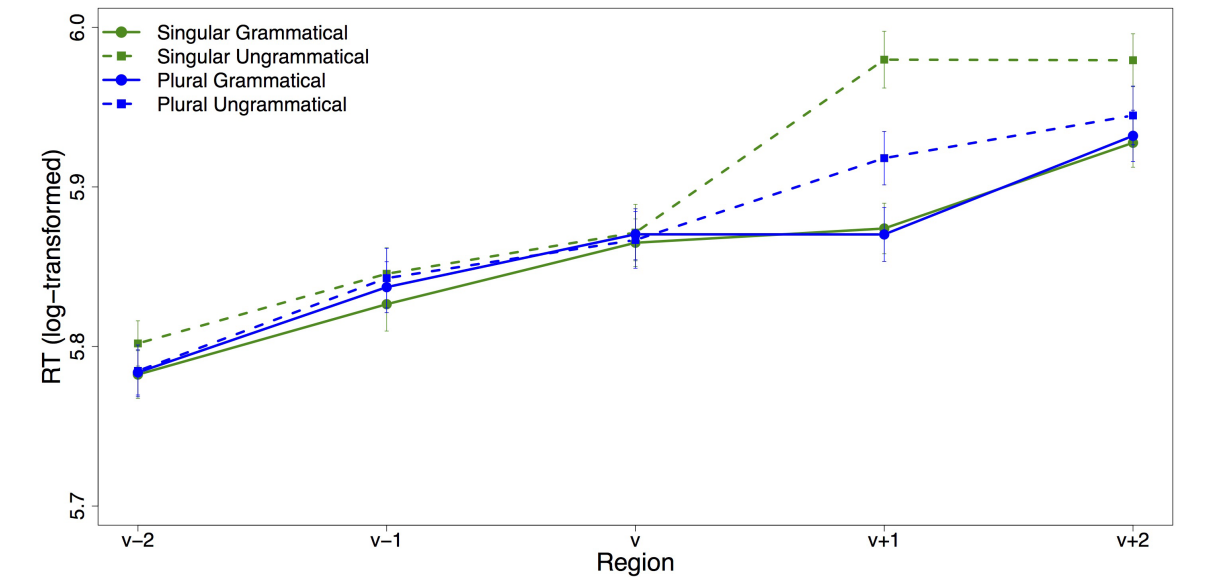


FIGURE 5 | Region-by-region mean reading times for trials on which the (correct) head-matching adjective was chosen. Error bars indicate standard error of the mean.

of the subject might be less well encoded and less stable than usual. On some of these trials, the attractor might even have been analyzed as the subject before the verb was encountered. Without a robust structural representation of the input prior to the verb, it is possible that in these cases neither of the NPs is in subject position when the verb is encountered. The plural marking on the verb could then have served as a cue to pick the NP with the matching number feature as the subject, explaining why attractor-matching adjectives were chosen more frequently in ungrammatical sentences with plural attractors. Although this relies on a match between the attractor's number feature and the retrieval cues of the verb, it is not identical to the mechanism we usually assume for agreement attraction. Unfortunately, we have no data on how confident participants were about their adjective choices. If attractor compatible adjective choices really were due to inattention, participants would be expected to be less confident about their choice on these trials.

TABLE 5 | Results of the linear mixed effects model for trials on which the head-matching adjective was chosen (using log transformed RTs).

Parameter	Estimate	Std. error	*t*-value
Verb region			
Intercept	5.853	0.043	137.87
Grammaticality	0.002	0.005	0.35
Attractor number	−0.002	0.005	−0.28
Grammaticality × attractor number	−0.004	0.005	−0.85
Spillover region			
Intercept	5.894	0.043	136.18
Grammaticality	−0.036	0.005	−6.77
Attractor number	0.015	0.005	2.78
Grammaticality × attractor number	−0.017	0.005	−3.18

Overall, the results of this study indicate that error-driven retrieval triggered by the detection of a subject–verb agreement violation only sometimes results in the misinterpretation of the attractor as the subject. This suggests that attraction effects in comprehension might reflect two different processes: In some cases, misretrieval of the attractor triggers structural reanalysis and results in the misinterpretation of the attractor as the subject. However, agreement attraction seems to often index a low-level feature checking operation in the following sense: Comprehenders predict the number marking of the verb based on the subject and retrieve the agreement controller if the verb does not match this prediction to check whether its number feature can license the number marking on the verb. If, it is no longer perceived as an agreement violation. This relies on a low-level morphosyntactic checking mechanisms in which only the retrieved item's number feature is checked, since the misretrieved attractor does not match all of the verb's retrieval cues.

A reviewer notes that one possible alternative explanation of these data is that misinterpretation does occur in tandem with agreement attraction, but that participants 'fix' the misinterpretation at a later stage process when the adjective is encountered. In other words, participants could have initially integrated the adjective with the misinterpretation driven by the agreement configuration, but then re-checked the interpretation by retrieving the initial noun in the sentence, such that this reanalysis would yield the correct interpretation. Although we don't have any evidence for this two-stage strategy in the current data, we agree that it will be important for future work to more carefully evaluate this possibility with a more time-sensitive interpretation measure.

The Final Representation of Agreement Attraction Sentences

The question whether the misretrieval of the attractor in agreement processing triggers reanalysis has important implications for whether grammatical illusions can arise with mental representations that are not actually grammatical. If misretrieval of the attractor necessarily triggers restructuring, agreement attraction would only occur when the verb's number marking is actually licensed by the final representation: with the plural attractor misrepresented in subject position, there would be no agreement violation. This would suggest that grammatical illusions arise on the basis of final representations that are not consistent with the input, but are consistent with the grammar.

In contrast, if the output of retrieval is only used to check that the number marking on the verb is consistent with the number feature of the agreement controller, misretrieval of a number matching attractor would simply signal that there is no agreement violation after all. However, the final structural representation in memory would still contain a number mismatch between the

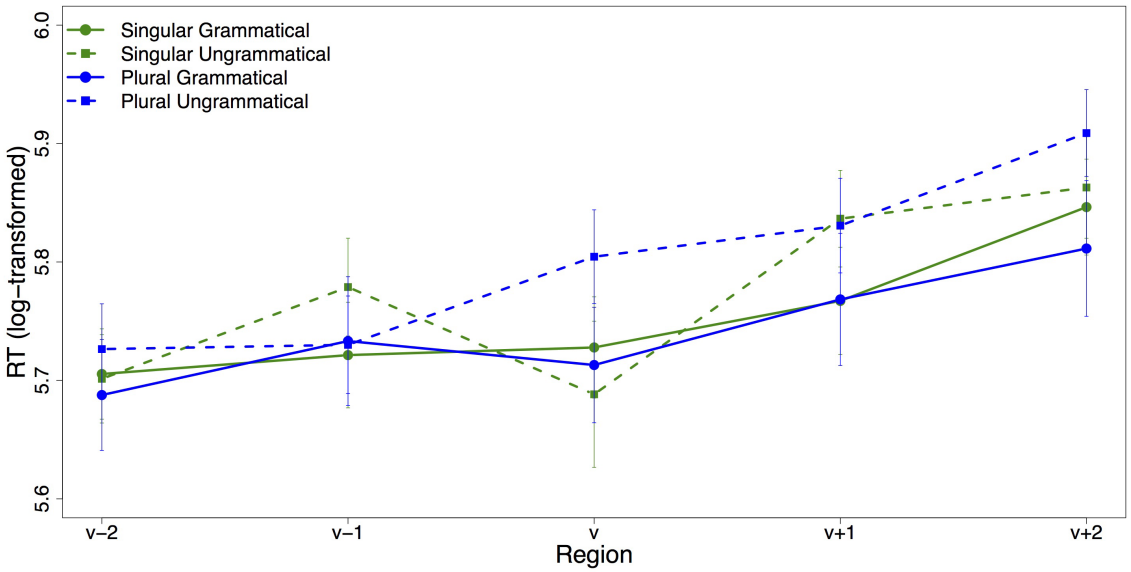


FIGURE 6 | Region-by-region mean reading times for trials on which the attractor-matching adjective was chosen. Error bars indicate standard error of the mean.

actual subject and the verb and would therefore be consistent with the input but not the grammar.

If a number matching attractor is retrieved instead of the number-mismatching subject, that signals that there is no agreement violation after all. Due to this illusory licensing of the verb's number marking by the attractor, the comprehender does not perceive the sentence to be ungrammatical. Consequently, there is no additional repair process to revise the subject's or the verb's number and the final representation remains inconsistent with the grammar. That might be considered a problem for a low-level feature checking account if we assume a framework in which interpretations have to be derived from structural representations consistent with the grammar. However, it very much depends on when exactly we think agreement has to be licensed in online processing. If the verb's number only matters at the point at which it is integrated into the structure, illusory checking due to misretrieval of the number-matching attractor would be entirely sufficient and the discrepancy between the structure and the features that were checked does not matter.

The results of our study suggests an account of agreement attraction that does not necessarily involve reanalysis. This means that the illusory licensing of an agreement violation must be possible without a final mental representation of the sentence in which it is actually licensed. However, the slightly higher proportion of attractor-matching adjective choices in agreement attraction configurations suggests that a subset of trials on which the attractor is misretrieved does lead to the misrepresentation of the attractor as the subject. In this subset, the final mental representation does actually license the verb's number marking. This suggests that what we observe as the phenomenon of agreement attraction in measures such as speeded acceptability judgments and self-paced reading may not reflect exactly the same underlying process on all trials.

A Third Possibility: Revising the Subject's Number Feature

The results of the present study suggest that the error-driven retrieval process that results in agreement attraction is a low-level rechecking process that does not usually have any structural impact. However, one could imagine a third possibility that falls in between a structural reanalysis account and a simple feature-checking model. It is possible that the representation of the sentence is altered based on the retrieval output, but without structural reanalysis. In particular, the parser could use the number feature of the erroneously retrieved attractor to substitute the number feature of the subject as it was originally encoded in memory. For example, in a sentence with an agreement violation and a number-matching attractor, such as '*The key to the cabinets are rusty*,' the process would be the following: The subject is correctly encoded as singular and the parser predicts a singular verb. Upon encountering '*are*,' there is a mismatch between the number feature of the prediction and the bottom-up input, which triggers a search for the agreement controller in memory. If the number-matching attractor is erroneously retrieved, it's number feature is used to "correct" the subject's current number feature. Unlike the pure rechecking process, this account predicts interpretive consequences of misretrieval, but would result in a final representation that is consistent with the grammar as a whole and does not contain an agreement violation.

If misretrieval of the number matching attractor results in the change of the subject's number feature, this could in a sense be considered a representational account since it involves misrepresenting the number of the subject. However, it would be fundamentally different from other misrepresentation accounts: In representational accounts like feature percolation (Bock and Eberhard, 1993; Vigliocco et al., 1995; Eberhard, 1997; Franck et al., 2002) and the Marking and Morphing model (Bock et al., 2001; Eberhard et al., 2005), agreement attraction is a consequence of misencoding the subject's number feature prior to encountering the verb. In contrast, if the parser changes the subject's number feature based on the output of retrieval in agreement processing, misrepresenting the subject's number information would be a consequence of agreement attraction, rather than the cause of it. The main argument against representational accounts of agreement attraction in comprehension is the grammatical asymmetry (Wagers et al., 2009): If the subject's number is misrepresented in the presence of a plural attractor, we would expect grammatical sentences to sometimes be perceived as containing an agreement violation. This illusion of ungrammaticality has generally not been found in the literature (Wagers et al., 2009; but cf. Hammerly et al., 2018). However, if misrepresentation of the subject's number feature occurs not before the verb is encountered but as a consequence of encountering a plural verb, this would account for the lack of an illusion of ungrammaticality.

As discussed in the section on agreement and interpretation, there is some data that suggest that comprehenders do misinterpret the subject as plural in agreement attraction configurations (Patson and Husband, 2016; Brehm et al., 2019). However, this was measured by non-literal plural responses to comprehension questions, which were also higher when the local noun was singular and only the verb was plural. This is not predicted by representational accounts of agreement attraction and is more consistent with a noisy-channel model of comprehension, or an account in which the answers to explicit comprehension questions do not necessarily show an accurate reflection of the representation built during the earlier processing of the sentence. Although the data suggest that agreement attraction does not arise as the consequence of number misrepresentation, they do not speak to the question whether number misrepresentation might arise as a result of misretrieving the attractor.

Consequently, while the results from Patson and Husband (2016) and Brehm et al. (2019) are intriguing, they do not provide conclusive evidence that agreement attraction arises from comprehenders misrepresenting the subject's number feature due to the presence of a plural attractor. Nevertheless,

in light of the recent evidence that comprehenders sometimes carry out structural repairs on anomalous input, the possibility that comprehenders end up misrepresenting the subject's number information in agreement attraction cannot be dismissed without further research.

CONCLUSION

We explored the relationship between the output of error-driven retrieval in agreement processing and the final structural representation of the sentence. We used a novel dual-task design to assess whether comprehenders misinterpret the attractor as the subject when they experience agreement attraction. The results suggest that comprehenders do not misinterpret the attractor as the subject on all trials on which agreement attraction occurs, indicating that misretrieval of the attractor does not necessarily trigger restructuring. While this implies that subject–verb agreement attraction is not a straightforward reflection of reanalysis, misretrieval of the attractor does appear to increase the likelihood of misinterpreting the attractor as the subject. This suggests that the error-driven retrieval process in agreement checking often involves low-level feature checking without integrating the output of retrieval into the agreement controller's position in the mental representation. Nevertheless, in a subset of cases, this low-level feature checking does serve as an impetus for structural reanalysis.

Since the data suggest that structural reanalysis is not necessarily triggered when the attractor is misretrieved, this indicates that illusory licensing can occur even if there is no actual licensing in the final mental representation. Whether this discrepancy will hold for other grammatical illusions is unclear; agreement as such does not contribute to the interpretation of a sentence and, unlike grammatical illusions involving dependencies that cannot be predicted such as reflexives or VP-ellipsis, it is an error-driven phenomenon. This potential difference between agreement attraction and non-error driven grammatical illusions certainly warrants further investigation.

ETHICS STATEMENT

The protocol was approved by the University of Maryland Institutional Review Board. All subjects gave written informed consent.

AUTHOR CONTRIBUTIONS

ZS initiated the project and designed the experiments in collaboration with EL and DP. ZS ran the study and performed the analysis under the supervision of EL. ZS wrote the first draft of the manuscript. All authors critically revised the drafts.

FUNDING

This work was supported in part by an NSF Doctoral Dissertation Research Improvement grant (NSF DDRI 1651058) awarded to ZS.

ACKNOWLEDGMENTS

We would like to thank Jamie Lebovics and Rebekah Senderling for their assistance with constructing experimental materials and the two reviewers for helpful comments on the manuscript. We are also grateful to the members of the Cognitive Neuroscience of Language Laboratory at UMD and the audience at CUNY 2017 for feedback on various stages of this project. This manuscript is based on dissertation work conducted by ZS at the University of Maryland.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01002/full#supplementary-material

REFERENCES

Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: a tutorial overview. *Lingua* 119, 389–424. doi: 10.1016/j.lingua.2008.09.005

Acuña-Fariña, J. C. (2012). Agreement, attraction and architectural opportunism. *J. Linguist.* 48, 257–295. doi: 10.1017/s0022226712000084

Acuña-Fariña, J. C., Meseguer, E., and Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua* 143, 108–128. doi: 10.1016/j.lingua.2014.01.013

Arnett, N., and Wagers, M. (2017). Subject encodings and retrieval interference. *J. Mem. Lang.* 93, 22–54. doi: 10.1016/j.jml.2016.07.005

Arregui, A., Clifton, C., Frazier, L., and Moulton, K. (2006). Processing elided VPs with flawed antecedents. *J. Mem. Lang.* 55, 232–246. doi: 10.1016/j.jml.2006.02.005

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48.

Bock, K., and Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Lang. Cogn. Process.* 8, 57–99. doi: 10.1080/01690969308406949

Bock, K., Eberhard, K. M., and Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *J. Mem. Lang.* 51, 251–278. doi: 10.1016/j.jml.2004.04.005

Bock, K., Eberhard, K. M., Cutting, J. C., Meyer, A. S., and Schriefers, H. (2001). Some attractions of verb agreement. *Cogn. Psychol.* 43, 83–128. doi: 10.1006/cogp.2001.0753

Bock, K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Brehm, L., Jackson, C. N., and Miller, K. L. (2019). Speaker-specific processing of anomalous utterances. *Q. J. Exp. Psychol.* 72, 764–778. doi: 10.1177/1747021818765547

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: good-enough, underspecified, or shallow language processing. *Q. J. Exp. Psychol.* 69, 817–828. doi: 10.1080/17470218.2015.1134603

Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cogn. Psychol.* 42, 368–407. doi: 10.1006/cogp.2001.0752

Cunnings, I., and Sturt, P. (2018). Retrieval interference and semantic interpretation. *J. Mem. Lang.* 102, 16–27. doi: 10.1016/j.jml.2018.05.001

Dempsey, J., Christianson, K., and Tanner, D. (2016). "Misrepresentations of plurality in late processing: evidence from self-paced reading," in *Proceedings of the Poster Presented at the 29th CUNY Conference on Human Sentence Processing*, (Gainesville, FL: University of Florida), 3–5.

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Drummond, A. (2019). *Ibex Farm*. Available at: http://spellout.net/ibexfarm (accessed April 1, 2019).

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *J. Mem. Lang.* 36, 147–164. doi: 10.1006/jmla.1996.2484

Eberhard, K. M., Cutting, J. C., and Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychol. Rev.* 112, 531–539.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cogn. Psychol.* 47, 164–203. doi: 10.1016/s0010-0285(03)00005-7

Ferreira, F., Christianson, K., and Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of sentence processing and reanalysis. *J. Psycholinguist. Res.* 30, 3–20.

Franck, J., Vigliocco, G., and Nicol, J. (2002). Subject-verb agreement errors in French and English: the role of syntactic hierarchy. *Lang. Cogn. Process.* 17, 371–404. doi: 10.1080/01690960143000254

Frazier, L. (2013). "A recycling approach to processing ellipsis," in *Diagnosing Syntax*, eds L. Lai-Shen Cheng and N. Corver (Oxford: Oxford University Press), 485–501. doi: 10.1093/acprof:oso/9780199602490.003.0024

Frazier, L. (2015). Two interpretive systems for natural language? *J. Psycholinguist. Res.* 44, 7–25. doi: 10.1007/s10936-014-9328-0

Frazier, L., and Clifton, C. Jr. (2015). Without his shirt off he saved the child from almost drowning: interpreting an uncertain input. *Lang. Cogn. Neurosci.* 30, 635–647. doi: 10.1080/23273798.2014.995109

Gibson, E., Bergen, L., and Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8051–8056. doi: 10.1073/pnas.1216438110

Hammerly, C. M., Staub, A., and Dillon, B. (2018). *The Grammaticality Asymmetry in Agreement Attraction Reflects Response Bias: Experimental and Modeling Evidence*. Amherst, MA: University of Massachusetts Amherst.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Karimi, H., and Ferreira, F. (2016). Informativity renders a referent more accessible: evidence from eyetracking. *Psychon. Bull. Rev.* 23, 507–525. doi: 10.3758/s13423-015-0917-1

Kutas, M., DeLong, K. A., and Smith, N. J. (2011). "A look around at what lies ahead: prediction and predictability in language processing," in *Predictions in the Brain: Using our Past to Generate a Future*, ed. M. Bar (New York, NY: Oxford University Press).

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., and Phillips, C. (2015). Agreement attraction in Spanish comprehension. *J. Mem. Lang.* 82, 133–149. doi: 10.1016/j.jml.2015.02.002

Lau, E., Wagers, M., Stroud, C., and Phillips, C. (2008). "Agreement and the subject of confusion," in *Proceedings of the Talk given at 21st Annual CUNY Sentence Processing Conference*, (Chapel Hill, NC: University of North Carolina).

Levy, R. (2008). "A noisy-channel model of rational human sentence comprehension under uncertain input," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (Stroudsburg, PA: Association for Computational Linguistics), 234–243.

Levy, R., Bicknell, K., Slattery, T., and Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21086–21090. doi: 10.1073/pnas.0907664106

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., and Van Dyke, J. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *J. Psycholinguist. Res.* 29, 111–123.

Parker, D. (2018). A memory-based explanation of antecedent-ellipsis mismatches: new insights from computational modeling. *Glossa* 3:129. doi: 10.5334/gjgl.621

Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002

Patson, N. D., and Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Q. J. Exp. Psychol.* 69, 950–971. doi: 10.1080/17470218.2014.992445

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

Pittman, C., and Smyth, R. (2005). "The effect of the predicate on agreement error rates," in *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*, Ontario.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., and Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *J. Mem. Lang.* 69, 104–120. doi: 10.1016/j.jml.2013.04.001

Slevc, L. R., and Momma, S. (2015). "Noisy evidence and plausibility influence structural priming," in *Proceedings of the Poster Presented at the Annual Architectures & Mechanisms for Language Processing (AMLaP) Meeting*, Valetta.

Staub, A. (2009). On the interpretation of the number attraction effect: response time evidence. *J. Mem. Lang.* 60, 308–327. doi: 10.1016/j.jml.2008.11.002

Staub, A. (2010). Response time distributional evidence for distinct varieties of number attraction. *Cognition* 114, 447–454. doi: 10.1016/j.cognition.2009.11.003

Tanner, D., Dempsey, J., and Christianson, K. (2018). "Does attraction lead to systematic misinterpretation of NP number? Probably not," in *Proceedings of the Poster Presented at the 31st CUNY Conference on Human Sentence Processing*, Davis.

Tanner, D., Nicol, J., and Brehm, L. (2014). The time-course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *J. Mem. Lang.* 76, 195–215. doi: 10.1016/j.jml.2014.07.003

Thornton, R., and MacDonald, M. C. (2003). Plausibility and grammatical agreement. *J. Mem. Lang.* 48, 740–759. doi: 10.1016/s0749-596x(03)00003-2

Tucker, M. A., Idrissi, A., and Almeida, D. (2015). Representing number in the real-time processing of agreement: self-paced reading evidence from Arabic. *Front. Psychol.* 6:347. doi: 10.3389/fpsyg.2015.00347

Van Dyke, J. A., and Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: a cue-based parsing account of recovery from misanalyzed ambiguities. *J. Mem. Lang.* 49, 285–316. doi: 10.1016/s0749-596x(03)00081-0

Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2006.03.007

Vigliocco, G., Butterworth, B., and Semenza, C. (1995). Constructing subject-verb agreement in speech: the role of semantic and morphological factors. *J. Mem. Lang.* 34, 186–215. doi: 10.1006/jmla.1995.1009

Wagers, M. W., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.3389/fpsyg.2015.00349

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 May 2019
doi: 10.3389/fpsyg.2019.00978

The Attractions of Agreement: Why Person Is Different

Marcel den Dikken[1,2]*

[1] Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary, [2] Department of English Linguistics, Eötvös Loránd University, Budapest, Hungary

This paper establishes the generalization that whenever agreement with the finite verb is controlled by a constituent that is not in a Spec–Head relation with the inflectional head of the clause, this agreement cannot affect *person*. A syntactic representation for person inside the noun phrase and on the clausal spine is proposed which, in conjunction with the workings of agreement and concord, accommodates this empirical generalization and derives Baker's Structural Condition on Person Agreement. The proposal also provides an explanation for the φ-feature agreement facts of specificational copular sentences. The paper places its findings on person vs. number agreement in the context of recent psycho- and neuro-linguistic investigation of number/person dissociation.

Keywords: agreement, person, number, agreement attraction, long-distance agreement, relativization, specificational copular sentences, concord

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Matthew Wagers,
University of California, Santa Cruz,
United States
Ad Neeleman,
University College London,
United Kingdom

***Correspondence:**
Marcel den Dikken
marcel.den.dikken@nytud.mta.hu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *11 October 2018*
Accepted: *12 April 2019*
Published: *22 May 2019*

Citation:
den Dikken M (2019) The Attractions of Agreement: Why Person Is Different. Front. Psychol. 10:978. doi: 10.3389/fpsyg.2019.00978

INTRODUCTION

Agreement remains a highly complex matter, empirically as well as theoretically. With particular reference to agreement in specificational copular sentences, various 'agreement attraction', and long-distance agreement constructions, this paper addresses the question of why agreement phenomena systematically make a distinction between person and the other φ-features. Baker's, (2008, 2011) Structural Condition on Person Agreement (SCOPA) was formulated to account for this, but by itself it offers no explanation for it. After a survey of the empirical territory I devote the core of the paper to deriving SCOPA and its effects from the syntactic representation of person in the noun phrase (as a specifier of the number phrase) and on the clausal spine (as a functional head in the complement of the number head), and from the workings of agreement and concord.

I close the paper by placing the findings regarding the difference in behavior between number and person agreement in the context of the recent psycho- and neuro-linguistic literature on number/person dissociation. Significant differences in behavior have been found between number agreement and person agreement in a suite of psycho- and neuro-linguistic studies — especially those conducted by Mancini and her co-workers on various Romance languages (see Mancini et al., 2011 for ERP experiments, Mancini et al., 2014b for self-paced reading experiments, and Mancini et al., 2017 for an event-related fMRI experiment). In their 2011 study on the ERP patterns evinced by subject-verb agreement violations in Spanish, number and person were found to differ in two ways: *(a)* person agreement violations give rise to N400 effects, which are 'seldom reported in the literature' (p. 69) for 'mere' agreement mismatches; and *(b)* person but not number agreement violations produce an early increased P600 effect at frontal (rather than posterior) sites. Mancini and colleagues interpret the frontal P600 effect as a

reflex of 'discourse-related integration difficulties' (p. 73), and reinforce the semantic-pragmatic role played by person agreement by their understanding of the N400 effect (usually associated with problems of interpretation) as an indication that person mismatch causes an interruption of 'the establishment of interpretive relations among constituents' (p. 72) — particularly, of the association of the morphosyntactic person marking with the representation of the discourse participant (speaker/hearer) in the left periphery of the clause.

Mancini et al. (2017) recast their findings in terms of the postulation of two different mechanisms involved in agreement phenomena, which they call 'feature-checking' and 'feature-mapping.' Number and person agreement are argued to involve a common φ-feature-checking mechanism but to differ in their feature-mapping options, with number mapping to cardinality and person to the discourse. The present paper bears marginally on feature-mapping (the interpretive side of number and person marking), in its discussion of number agreement between the relativized head and the finite verb of a relative clause. But the main impact of this paper lies in what it has to say regarding Mancini and colleagues conclusion that number and person agreement share the same feature-checking mechanism. The material reviewed in this paper argues for a key difference between the feature-checking processes involved in number and person agreement: number agreement is possible under both Agree and the Spec–Head relation; person agreement, on the other hand, cannot transpire under (downward) Agree, being establishable only in a Spec–Head configuration.

PERSON IS DIFFERENT

Agreement in Specificational Copular Sentences

Specificational Pseudoclefts

It is often said that copula agreement distinguishes neatly and reliably between the predicational and specificational readings of pseudoclefts of the type in (1): Declerck (1988, p. 79), the source of these particular examples, asserts that (1a) is unambiguously specificational, and (1b) is predicational.[1]

(1) (a) what you have bought *is* fake jewels
(b) what you have bought *are* fake jewels

But while it is true that (1a) only supports a specificational reading (equivalent to *you have bought fake jewels*), (1b) is not quite as unambiguous as Declerck makes it out to be. Similarly, in *what John brought was/*[?] *were the crackers*, plural inflection on the copula is (marginally) possible on a specificational reading of the pseudocleft. Declerck (1988, pp. 79–80) himself points out that '[i]n specificational sentences the number of the copula can apparently be determined by that of either the superficial subject NP or the variable NP.' The examples in (2a,b) are from Declerck, with his judgments (or those of his informants) provided; the ones in (2c–e) I have taken from Heycock (2012), with her original judgments included (see also Den Dikken, 2017).

(2) (a) what I need {*is*/[??]*are*} more books
(b) what we can't have here {*is*/[?]*are*} theft and robbery
(c) what he saw behind him {*was*/*were*} two men
(d) what makes something a pencil *are* superficial characteristics such as a certain form and function
(e) all I could see {*was*/*were*} two staring eyes

No such oscillation is found for person, however: the sentences in (3) (also due to Heycock, 2012) are ungrammatical with person agreement between the copula and the post-copular focus.

(3) (a) what he saw behind him {*was*/**were*} you
(b) what makes this party go {*is*/**are*} you
(c) all I could see {*was*/**were*} you

This is our first indication that number and person should be treated distinctly in the morphosyntax of English.

[1] Declerck (1988, p. 79) also claims that the copula of a specificational pseudocleft cannot agree with the focus when this focus is an NPI, as in (i) (reproduced with Declerck's judgments). But pseudoclefts with NPI-connectivity can be found for which at least some speakers allow for number agreement between the copula and the focus. The examples in (ii) (with (iic) taken from the internet) illustrate. (Here and throughout this paper, '[!]' marks 'unusual agreement.')

(i) what the book does not offer {is/*are} any solutions to the problems that are noted

(ii) (a) what nobody has bought {is/[!]are} any cups and glasses
(b) what isn't available {is/[!]are} papers that say anything about clefts
(c) what nobody has seen are any of the bonuses (including playmats, tokens, tins, counters) which were supposed to be shipped with the cards (http://boardgamegeek.com/thread/900269/stop-wulven-from-cheating-their-customers)

In the analysis of NPI-connectivity in specificational pseudoclefts presented in Den Dikken et al. (2000), the examples in (i)–(ii) have as their post-copular constituent a full clause stripped down to the focus. If this is to involve constituent ellipsis (as effects of Merchant's, 2001 P-stranding generalization in Dutch cases of this type suggest it must), we are dealing with a case of non-*wh* sluicing (IP ellipsis), with the focused constituent in the left periphery of the answer clause:

(iii) [$_{\text{TopP}}$ [$_{\text{Question}}$ *what nobody bought*] [$_{\text{Top}'}$ Top = *be* [$_{\text{Answer}}$ *any cups and glasses*$_i$ [$_{\text{IP}}$ ~~*nobody bought* t_i~~]]]]

On this analysis, (i)–(ii) remind us of long-distance agreement in Tsez (Polinsky and Potsdam, 2001) and several other languages [incl. Innu-aimûn (Branigan and MacKenzie, 2002), Passamaquoddy (Bruening, 2001), and Itelmen (Bobaljik and Wurmbrand, 2005)]. The example in (ivb) illustrates long-distance agreement in Tsez:

(iv) (a) eni-r [už-ā magalu
mother-DAT boy-ERG bread.III.ABS
b-āc'-ru-łi] r-iyxo
III-eat-PTC-NOMINAL.IV IV-knows
'the mother knows that they boy ate the bread'
(b) eni-r [už-ā magalu
mother-DAT boy-ERG bread.III.ABS
b-āc'-ru-łi] b-iyxo
III-eat-PTC-NOMINAL.IV III-knows

Polinsky and Potsdam (2001) argue that long-distance agreement in Tsez necessarily involves movement to an A′–position in the high left periphery of the subordinate clause — specifically, in the case of Tsez, SpecTopP, as in (v). The syntactic relationship between V and the lower topic in (v) is fully parallel to the relation between *be* and the lower focus in (iii).

(v) V [$_{\text{LowerClause}}$ *magalu*$_i$ [$_{\text{IP}}$ *užā* t_i *bāc'ruli*]]

Double-NP Specificational Copular Sentences

In English double-NP specificational copular sentences such as (4)–(5), the copula agrees with the precopular noun phrase for both number and person (Heycock, 1992; Moro, 1997):[2]

(4) the biggest problem {*is*/**are*} the agreement facts
(5) the biggest problem {*is*/**are*} you

Dutch, German and Italian seem to return judgments that are the exact opposite of the ones reported for English: in (6)–(7), the copula must agree with the post-copular focus in both number and person.

(6)(a) de oorzaak van het ongeluk {*waren*/**was*} kapotte remmen (Dutch)
the cause of the accident were/was broken brakes
(b) die Unfallsursache {*waren*/**war*} defekte Bremsen (German)
the accident-cause were/was defective brakes
(c) la causa della rivolta {*sono*/**è*} le foto del muro (Italian)
the cause of.the riot are/is the pictures of.the wall

(7)(a) de schuldige {*ben*/**is*} ik de schuldige {*ben*/**is*} jij (Dutch)
the culprit am/is I the culprit are/is you
(b) der Schuldige {*bin*/**ist*} ich der Schuldige {*bist*/**ist*} du (German)
the culprit am/is I the culprit are/is you
(c) il colpevole {*sono*/**è*} io il colpevole {*sei*/**è*} tu (Italian)
the culprit am/is I the culprit are/is you

For Italian, these facts are systematic. Moro (1997), who first discussed them in detail, has a syntax for them that makes them fall out without causing trouble for any extant account of agreement: *la causa della rivolta* in (6c) and *il colpevole* in (7c) are base-generated as left-adjuncts to IP, with a pro-predicate (*pro*) raising to the structural subject position; this *pro* copies the φ-features of the referential noun phrase of which it is predicated (i.e., the focus), so with I agreeing with *pro* we automatically derive full φ-agreement with the post-copular focus (8) makes this clear.

(8) [IP *il colpevole* [IP *pro*{φ}i [T′ COPULA{φ}i [SUBJECT{φ}i (...)]]]]

The Dutch and German facts are more problematic — first because (as Den Dikken, 1998 shows) they are not amenable to an account along Moro's (1997) lines; and secondly because they are not nearly as straightforward as the Italian facts are. As a matter of fact, the examples in (6a,b) and (7a,b) are a red herring. For these root sentences, there are derivations available that treat the sentence-initial noun phrase as a topic in the left periphery and place the post-copular subject in the structural subject position, SpecIP. On such a derivation, the φ-agreement facts in (6a,b) and (7a,b) are parallel to the φ-agreement found in (9), Verb Second constructions with a non-subject in the left periphery and the subject occupying the structural subject position and agreeing with the finite verb.

(9)(a) op de vensterbank {*staan*/**staat*} twee vazen (Dutch)
on the window-sill stand.PL/stand.3SG two vases
(b) bananen {*zul*/**zullen*} je daar niet vinden
bananas will.2SG/will.PL you there not find

To avoid the confounding effect of Verb Second, we should look at non-root clauses (which do not show Verb Second), as in Dutch (10) and (11):

(10) ze denken/betwijfelen dat de oorzaak van het ongeluk kapotte remmen {*waren*/**was*}
they think/doubt that the cause of the accident broken brakes were/was

(11)(a) ze denken/betwijfelen dat de schuldige ik {**ben*/**is*}
they think/doubt that the culprit I am/is
(b) ze denken/betwijfelen dat de schuldige jij {**bent*/**is*}
they think/doubt that the culprit you are/is

The result is grammatical with number agreement but bad with person agreement. In the case of (11) this yields ineffability: with this linear order, I find that there is no φ-feature inflection on the copula that comes out grammatical.[3] To get a grammatical output, we must refrain from predicate inversion, as in (11′), which has person agreement between the subject pronoun and the finite verb.

(11′)(a) ze denken/betwijfelen dat ik de schuldige {*ben*/**is*}
they think/doubt that I the culprit am/is
(b) ze denken/betwijfelen dat jij de schuldige {*bent*/**is*}
they think/doubt that you the culprit are/is

These facts present us with two questions: *(i)* why is *person* agreement with the focus impossible when predicate inversion takes place, and *(ii)* why is agreement with the inverted predicate barred? Question *(i)* bears directly on the main theme of this paper, and will be answered the section entitled "Why Person Is Different." The second question is strictly speaking tangential to my concerns here — but for completeness' sake, I will address it briefly in the remainder of this section.

Heycock (2012):fn. (3) suggests that the oscillation between singular and plural number inflection on the copula seen in (2), repeated below, is 'likely... due to the possibility of *what* and *all* (or the empty noun it modifies) being underspecified for

[2]Occurrences of number agreement in double-NP specificational copular sentences between the copula and the post-copular subject of predication are nonetheless attested. The examples in (i) (from Frances, 1986, p. 315) present two recorded cases of this type.

(i) (a) the weather to watch *are* those rains
(b) the cause of layoffs such as these *are* not the taxes

[3]For sentences like *dat het echte probleem jij/jullie*___ 'that the real problem you (SG/PL) BE', my own intuitions reveal that ineffability arises precisely where the form of the copula is *explicitly* person-marked (present-tense 2SG *bent* vs 3SG *is*), while the result is acceptable with forms that are syncretic for person (*jij was* and *jullie zijn/waren*). Hartmann & Heycock's (under review) experiments, revealing no person or syncretism effects here, may not have been sufficiently fine-grained to pick them up. These effects are perhaps even stronger in clefts (see also Ackema and Neeleman, 2018, who report intuitions matching mine): *dat het jij/jullie*___ *die S* 'that it you (SG/PL) BE who S' works with *was* (for *jij*) and *zijn/waren* (for *jullie*), but not with explicitly 2SG *bent*. Hereinafter, I will base myself on my own judgements for Dutch.

number,' which she thinks allows them to pick up their number specification from their associate (presumably under concord).

(2)(a) what I need {*is*/??*are*} more books
(b) what we can't have here {*is*/?*are*} theft and robbery
(c) what he saw behind him {*was*/*were*} two men
(d) what makes something a pencil *are* superficial characteristics such as a certain form and function
(e) all I could see {*was*/*were*} two staring eyes

Heycock's parenthesis 'or the empty noun it modifies' points us toward an answer to the question of why agreement with the fronted predicate in (10)–(11) is impossible. In Den Dikken (2006), it is proposed that in copular inversion constructions, what raises to the structural subject position is consistently a projection of a silent noun. Thus, double-NP copular inversion sentences such as those in (10) and (11) have a syntax of the following sort:

(12) $[_{IP}\ [_{PRED}\ \emptyset\ [\textit{the cause/culprit}]]_i\ [_{I'}\ I{+}\textsc{relator} = \textit{be}\ [_{RP}\ \textsc{focus}\ [_{R'}\ t_{REL}\ t_i]]]]$

With copular inversion constructions analyzed as in (12), the fact that the copula cannot φ-agree with the fronted predicate in (10)–(11) can be attributed to the absence of inherent φ-features on the silent noun. The fact that in double-NP specificational copular sentences such as *my favorite authors are/*is Austen and Heller* (from Heycock, 2009) we find plural inflection on the copula follows from the silent noun's ability to show number concord (with either the conjoined subject of predication or plural *authors*): *the* PERSONS *who are my favorite authors are Austen and Heller*.[4]

Special Agreement and Person

Before moving on to the analysis the section entitled "Why Person Is Different," let me present a further set of contexts in which person agreement behaves markedly differently from number agreement: contexts that I will group together under the rubric of 'special agreement.'

Agreement Attraction

The variants of the sentences in (13) and (14) with a plural-inflected finite verb (*think*, *are*) are well-known from the syntax and psycholinguistics literature (see Kimball and Aissen, 1971 on the former, and Bock and Miller, 1991 on the latter) as examples of number agreement between the finite verb of the clause and the 'wrong' target: in each case, finite verb agreement fails to target the entire subject of the clause; instead, agreement is 'attracted' to the relativized noun in (13) or to a subpart of the complex subject noun phrase in (14) (modeled on examples given in Kayne, 1998) — whence the name 'agreement attraction' [coined for cases of the type in (14), but apt for (13) as well].[5]

(13)(a) the people who Clark {*thinks*/![5] *think*} are in the garden
(b) how many people {*does*/! *do*} Clark think are in the garden?
(14)(a) the identity of these people {*is*/! *are*} to remain a secret
(b) these people's identity {*is*/! *are*} to remain a secret

Like agreement in specificational copular sentences, agreement of this type involves *number*, not *person*: person agreement between the finite verb and a non-subject is impossible in English:[6]

(15)(a) I, who Clark {*is*/**am*} hoping will marry his daughter
(b) you, who Clark {*is*/**are*} hoping will marry his daughter
(16)(a) the identity of me {*is*/**am*} to remain a secret
my identity {*is*/**am*} to remain a secret
(b) the identity of you {*is*/**are*} to remain a secret
your identity {*is*/**are*} to remain a secret

Long-Distance Agreement

Baker (2011, sect. 2.3.3) points out that the kind of long-distance (cross-clausal) agreement found in Tsez (Polinsky and Potsdam, 2001; see fn. 1, above) likewise sets person apart — this time not just from number but from gender as well. Baker brings up the case of Lokąą. In (17a), agreement between the matrix predicate and the object of the gerund that serves as its subject involves noun class (gender) and number; (17b) shows that such long-distance agreement is impossible for person.

(17)(a) [ę-sau kę-dęi] e-tum ę-tawa (Lokąą)
7-fish GER/5-buy 7SG-be.very 7SG-be.difficult
'buying fish is very difficult'
(b) *[min ke-funna] n-tum n-tawa
1SG GER/5-surprise 1SG-be.very 1SG-be.difficult
'surprising me is very difficult'

[4]For the text account, it is important that '∅' not be taken to be *pro* (which is plainly in possession of φ-features) but a silent noun (PERSON or THING, *à la* Kayne, 2005). See also Den Dikken and Griffiths (to appear) for relevant discussion.

The English equivalents of (10)–(11) do not give rise to ineffability, thanks to the fact that English allows for *default* (3SG) inflection on the copula. Béjar and Kahnemuyipour (2017) argue that in Eastern Armenian and Persian sentences of the type *the problem is the children*, the copula has default inflection as well. In Dutch this is impossible.

[5]Dillon et al. (2017, p. 90) point out that the Kimball and Aissen effect [at least in *wh*-questions, such as (13b)] 'stands apart from other forms of attraction [such as (14)], either in strength or in kind.' In "Why Person Is Different," it will turn out that the syntax of the attraction configuration in (13) is also different in detail from that in (14).

[6]Baker (2011) notes this for the Kimball and Aissen facts (15) (see also Dillon et al., 2017, whose test items consistently feature a form of *be* as the agreeing auxillary, *pace* Kayne, 2005, p. 264), and Nevins (2011) for 'agreement attraction' cases of the type in (16), both focusing on the first-person singular pronoun. I gave examples with second-person *you* as well to avoid interference coming from the case form of the pronoun — only *me*, not *I*, is possible in (4); and since *me* is not nominative, it will experience more difficulty in controlling agreement for independent reasons (see esp. Hartsuiker et al.'s, 2003 observation that the rate of agreement attraction is highest in cases in which the attractor is explicitly nominative or syncretic with the nominative). The second-person pronoun *you* shows case syncretism for nominative and accusative, hence should in principle be eligible for attracting agreement. The fact that it nonetheless fails to so do in (16b) is therefore interesting and significant. In this connection, note also the Hungarian examples in (i), where the possessive pronoun has the same form as the pronominal subject of a finite clause ('nominative,' or absence of morphological case-marking), yet person-agreement attraction remains sharply ungrammatical.

(i) (a) az én identitásom titok {*volt*/**voltam*} (Hungarian)
the I identity.1SG secret was(3SG)/was.1SG
(b) a te identitásod titok {*volt*/**voltál*}
the you identity.2SG secret was(3SG)/was.2SG
'my/your identity was a secret'

The facts in (13)–(17) solidify the conclusion reached in "Agreement in Specificational Copular Sentences" on the basis of the data of specificational copular sentences, and confirm the existence of an important dichotomy within the set of φ-features, setting person aside from the rest. The next section seeks to explain this dichotomy.

WHY PERSON IS DIFFERENT

Structural Condition on Person Agreement

Baker (2008, 2011) codifies the specialness of person agreement as his SCOPA, reproduced in (18). Baker (2011, p. 877, fn. 3) suggests (building on but modifying Franck et al.'s, 2006 work on agreement) that 'agreement for first- and second-person can never take place under mere Agree,' but requires the Spec–Head relation. I believe this is on the right track. In "The Place of Person in the Structure of the Noun Phrase and on the Clausal Spine," I will present an analysis of the place of person in the structure of the complex noun phrase and on the clausal spine which is mobilized in "The Syntax of Agreement: Agree Versus the Spec–Head Relation," "Person Agreement as Attraction," "Agreement Attraction "Long-Distance Agreement," "Long-Distance Agreement," "Copular Inversion and Agreement," and "Relativization and Agreement" to explain how person agreement is different from number agreement, and to derive the main effects of SCOPA.

(18) *Structural Condition on Person Agreement (SCOPA)*
a category F can bear the features +1 or +2 if and only if a projection of F merges with a phrase that has that feature and F is taken as the label of the resulting phrase.

The Place of Person in the Structure of the Noun Phrase and on the Clausal Spine

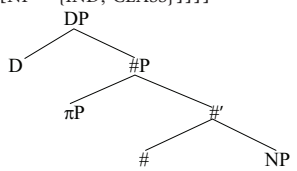
For the functional heads for person (Harley and Ritter's, 2002 class node PARTICIPANT) and number (Harley and Ritter's INDIVIDUATION), I will henceforth use the Greek letter π and the symbol #, resp. Like Harley and Ritter, I will take π to exclusively make the distinction between speaker ([+AUTHOR]) and addressee ([–AUTHOR]). For 'third person.' the feature [–PARTICIPANT] can be assigned to the D-head of the nominal phrase. But importantly, 'third person' is not a possible specification for π.[7] The following subsections address the place of π and # in the complex noun phrase (see "The Place of Person in the Internal Structure of the Noun Phrase") and on the clausal spine (see "The Place of Person on the Clausal Spine").

[7]Absence of any specification for person is often a viable option for 'third person' (Benveniste, 1966; Harley and Ritter, 2002; Nevins, 2007; Harbour, 2016; Ackema and Neeleman, 2018); but for English, third person is arguably $D_{[-PART]}$. Relevant here is the discussion (in the section entitled "The Featural Specification of *Who* as Relative Operator") of the English relative operator *who* as radically unspecified for person.

The Place of Person in the Internal Structure of the Noun Phrase

As a starting point, I will build up the structure of the complex noun phrase, along the lines of (19), which is effectively a 'syntactic translation' of Harley and Ritter's (2002) feature geometry for the set of φ-features — person, number, and gender.[8]

(19) *the structure of the noun phrase*
(a) $[_{NP}\ N_{\{IND,CLASS\}}]$
(b) $[_{\#P}\ \#_{\{IND,CLASS\}}\ [_{NP}\ N_{\{IND,CLASS\}}]]$
(c) $[_{\#P}\ \pi P\ [\#_{\{IND,CLASS\}}\ [_{NP}\ N_{\{IND,CLASS\}}]]]$
(d) $[_{DP}\ D_{\{IND,CLASS\}}\ [_{\#P}\ \pi P\ [\#_{\{IND,CLASS\}}\ [_{NP}\ N_{\{IND,\ CLASS\}}]]]]$



At the bottom of the noun phrase, we find a projection of the head noun, N. The gender specification of the noun (CLASS, again following Harley and Ritter's, 2002 terminology) is inherent to N. The noun is also specified for number, but its number properties are environmental, not genetic: the value for the feature [IND] is determined by a functional head labeled #, projecting outside NP. On top of #P, a projection for the definite determiner (D) can be built. This D-head establishes an Agree relation with # for [IND] and [CLASS], which is how articles get specified for number and gender. Importantly, person is represented inside the structure of the noun phrase not as a head on the nominal spine but as a *specifier* in the nominal extended projection — the specifier of #P, to be precise. It occupies the same structural position (*mutatis mutandis*) as the subject of a clause: with D corresponding to C, and # corresponding to I, the πP in (19d) occupies the equivalent of SpecIP in the clause.

One thing that the proposal in (19) helps explain is the well-known fact (see Postal, 1966) that (20a,b) are grammatical while (20c) is not (regardless of the case form of the pronoun):

(20) (a) we/us linguists
$[_{\#P}\ \pi P = \textit{we/us}\ [\#_{\{IND:PL,CLASS\}}\ [_{NP}\ N_{\{IND:PL,CLASS\}} = \textit{linguists}]]]$
(b) you linguists
$[_{\#P}\ \pi P = \textit{you}\ [\#_{\{IND:PL,CLASS\}}\ [_{NP}\ N_{\{IND:PL,CLASS\}} = \textit{linguists}]]]$
(c) *they/them linguists

The pronouns in (20a,b) are interpreted as the subjects of the predicate *linguists*, with # as the RELATOR of the predication relation (in the sense of Den Dikken, 2006). To be able to form

[8]For my purposes here, it is immaterial whether N is inherently endowed with a categorial feature or has its category label determined by a *n*-head merged outside NP. In the latter case, it will be *n*, not N, that is endowed with the gender feature. I have very little to say in this paper about gender agreement (attraction). If gender agreement involves feature valuation in syntax, it should behave very much like number agreement: both IND and CLASS are present on D. But I am not convinced that there is a gender probe on the clausal spine. I will proceed on the assumption that gender agreement involves concord (on which see "Relativization and Agreement"). For experiments and discussion (incl. a literature review) of gender agreement attraction effects (with some surprising results from Russian), I refer the reader to Slioussar and Malko (2016).

a grammatical pronoun–noun construction of this sort, # must be present in the structure and explicitly specified for number to serve as a RELATOR. In simple binary number systems such as English, 'singular' is absence of an explicit specification for number (i.e., a 'bare' class node [IND]). This explains the fact that (20a,b) do not have singular counterparts (**I linguist*, **you linguist*). And to be eligible for occupying Spec#P in (19), the pronoun must be specified for person, and no larger than πP.[9] The English third person plural pronouns *they* and *them* fail to meet these requirements. I have taken the position that 'third person' never instantiates a feature specification for the person head π (which is only specifiable for [± AUTHOR]), but instead is marked on D as [–PART] (or not marked at all; see fn. 7). The fact that the English third person plural pronouns *they* and *them* are introduced by the same voiced dental fricative that represents the definite article (*the*) confirms that these pronouns project full-fledged DPs, too large for Spec#P. This explains why in English, *they* and *them* cannot be combined with the projection of a common noun, as in (20c).[10]

Cross-linguistically as well, first- and second-person pronouns show a tendency to be relatively small in size, whereas third-person pronouns pattern with DPs.[11] Thus, in the Romance languages, while the third-person object clitic pronouns typically feature a token of the definite article (D = *l*-) in their morphology (cf. French *le* 'him,' *la* 'her,' *les* 'them'), the first- and second-person clitics do not. And in Hungarian (21), where full DPs and third-person pronouns serving as objects invariably trigger definiteness inflection on the transitive verb, first- and second-person object pronouns combine with indefinite inflection, due to their limited size (no larger than #P).

(21)(a) szereted a fiút / őt (Hungarian)
love.2SG.DEF the boy.ACC (s)he.ACC
'you$_{SG}$ love the boy/him/her'

(b) szeretsz minket
love.2SG.INDEF us.ACC
'you$_{SG}$ love us'

(c) szeretünk titeket
love.1PL.INDEF you$_{PL}$.ACC
'we love you$_{PL}$'

In the structure of nominal expressions, person/[PART] finds itself in the specifier position of number/[IND]. Inside the noun phrase, there is agreement for number and gender, but never for [PART]. Similarly, # (spelled out by the indefinite article, simple numerals, perhaps certain existential quantifiers) inflects for number and gender, but never for [PART]. That D, # and N share their specifications for number and gender is a straightforward reflex of the fact that all three are in an Agree-chain ('head-head agreement'; cf. also 'feature inheritance' or 'extended projection'), all having matching number and gender properties. The πP, as a left branch, is not a member of this chain.

The Place of Person on the Clausal Spine

On the clausal spine, # and π are also separate entities. But this time around, they find themselves in a complementation configuration, with the #-head embedding πP as its complement:[12]

(22) [$_{CP}$ C [$_{\#P}$ #$_{\{IND\}}$ [$_{\pi P}$ π$_{\{PART\}}$ (...) [$_{VP}$ V$_{\{IND,PART\}}$]]]]

In the clause, the finite verb shows agreement with the subject for number and person. The fact that person is a player in the clausal agreement system (unlike inside the noun phrase) indicates that it must be able to serve as a probe, adorned with unvalued feature [*u*PART]. This motivates the decision to represent π as a head on the clausal spine. Number has that status as well, bearing [*u*IND]. In addition, the head # is responsible for the assignment of nominative case to the subject. Nominative case is associated with φ rather than tense (as we know from inflected infinitives with nominative subjects in Portuguese; Raposo, 1987). For reasons discussed in the section entitled "The Syntax of Agreement: Agree Versus the Spec–Head Relation," the clausal π-head cannot serve as a probe in (downward) Agree relations, so in constructions in which the nominative subject appears below the inflectional domain (sentences with 'VP-internal subjects') it is inevitable to pin the nominative case feature on #. In constructions in which the nominative subject appears in the structural subject position ('SpecIP'), it surfaces in the higher of the two φ-related functional projections in (22): the contrast between *probably he isn't the culprit* and **probably isn't he the culprit* shows that, with *is* in the higher inflectional head, the nominative subject *he* must be placed to its left, in Spec#P. This in turn tells us that # is structurally higher than π (something that, for Indo-European, is impossible to verify on morphological grounds: person and number form portmanteaux in the verbal inflectional system of IE). The #-over-π structure in (22) is further supported on the basis of the syntax of number and person agreement in these languages, as I will now show.

The Syntax of Agreement: Agree Versus the Spec–Head Relation

The hypothesis that person is represented as a specifier in the noun phrase and as a functional head on the clausal spine

[9]On why a DP cannot serve as the specifier of the constituent occupying the complement position of a higher D, see Den Dikken and Dékány (2018).

[10]From Postal (1966), I retain the idea that the pronoun in *we/us/you linguists* occupies D — not through base-insertion (as Postal had it) but via raising. For a related proposal regarding the syntax of 'pluringulars' (as in *the committee are deliberating*), see Den Dikken (2001). See also Spanish 'unagreement,' briefly discussed in fn. 28, below.

[11]A reviewer mentions that in Cheke Holo (an Oceanic language spoken on the Solomon Islands; see Bosma, 1981; Palmer, 2009), first- and second-person pronouns can co-occur with determiners (see, e.g., *ta-hati-a* 'we$_{INCL}$-PL-ART'). A cursory inspection of the data suggests to me that this happens only when they are emphasized (with the emphasis particle *egu*, even in vocatives) or focused (with the particle *si*, as in *si iago ia* 'FOC you$_{SG}$ ART'), although appearing with *si* does not seem to require the presence of the article (*si go-tilo* 'FOC you-PL'). Whatever the determinants of the presence of the article with first-/second-person pronouns in Cheke Holo may turn out to be, it is noteworthy that articled personal pronouns do not seem to trigger person agreement on the finite verb in this language. This may follow from the proposal presented in this paper: when #P is encapsulated in a DP, the πP in Spec#P becomes very difficult (perhaps impossible) to access as an Agree-goal for the clausal π-head.

[12]See esp. Preminger (2011) for a defence of this structure, aimed, like the present paper, at an understanding of SCOPA and the restrictions on person agreement. Hartmann & Heycock (under review) likewise have number and person project autonomously, but they follow Sigurðsson and Holmberg's (2008) lead in placing person above number.

below number has important consequences for the distribution of person agreement in the clause.

In the structure of the noun phrase, person is not represented on D or #.[13] So how does person agreement in the clause come about? Let us first examine (downward) Agree. The person head on the clausal spine has nothing to probe for: the πP of the pronominal subject in the verbal core is not directly accessible to the clausal π-head because it is contained within the pronominal subject, occupying the specifier position of the subject, which is itself a specifier. Subparts of specifiers are not directly accessible to higher probes: specifiers are merged into the structure as fully built structural chunks (see Uriagereka, 1999 for the origins of this idea); no outside probe can by itself reach into the innards of a specifier. So the clausal π-head cannot directly target the πP inside the subject. The clausal π-head cannot target the entire subject pronoun (i.e., #P) integrally either because #P, specified for [IND] but not for [PART], is not a match for the π-head's [*u*PART] feature. So person agreement cannot happen under (downward) Agree.[14] And since the clausal π-head cannot probe the pronominal subject of the clause, it cannot attract it to its specifier position either, so it also cannot establish a Spec–Head relation with the pronominal subject in the clausal πP.

But the next higher head, #, does manage to Agree with and attract the pronominal #P, provided that the clausal π-head with its [*u*PART] feature gets out of the way. Locality of probing makes it impossible for the #-head's [*u*IND] feature to probe past an intervening unvalued feature [*u*PART] on the clausal spine.[15] But if the clausal π-head raises and adjoins to #, then # will find a match without obstruction: it can engage in an Agree relation with the pronominal #P; and if the EPP so dictates, the clausal #-head can also attract the pronominal #P to its specifier, which results in a Spec–Head relation between the clausal #-head and the pronominal subject. This Spec–Head relation involves not just number but person as well. The clausal π-head must raise to # in order for # to be able to attract the pronominal subject to Spec#P, adjoining to # and forming a complex probe [$_{\#}$ π [#]] with it. Under the Spec–Head relation between this complex probe and the subject, a total match between the two must be forged, in concert with (23) (from Den Dikken and Dékány, 2019; see also Guasti and Rizzi, 2002; Shlonsky, 2004, p. 1496; Franck et al., 2006 for relevant facts and discussion):

(23) *the* TOTAL MATCH *constraint on Spec–Head agreement*
feature checking under the Spec–Head relationship requires total matching of the features of the head and the features of its specifier.

Under (downward) Agree, the functional head # probes the subject just for its own unvalued [*u*IND] feature. But once the #-head has probed the subject and attracted it to the specifier of the complex probe [$_{\#}$ π [#]], a total match must be established between this probe and the subject, by (23). The probe–goal relation between the clausal #-head and #P lifts the opacity of the latter, rendering the πP in the specifier position of the pronominal subject an accessible goal to the π-portion of the complex probe.[16] The structure in (24) illustrates, for first- and second-person pronominal subjects.

(24) [$_{\#P}$ [$_{\#P}$ πP$_{\{\text{PART: } \pm \text{ AUTHOR}\}}$ [$_{\#'}$ #$_{\{\text{IND}\}}$ [$_{NP}$ N]]]$_i$ [$_{\#'}$ [$_{\#}$ π$_{\{u\text{PART}\}}$ [#$_{\{u\text{IND}\}}$]]... t_{π}... t_i...]]

The result of (24) is agreement for both number and person, with the latter contingent on the former, as desired: it is impossible for the finite verb of a clause to agree with a pronoun in person but not in number, but the converse IS possible. Directly relevant to the unidirectional contingency relation between person and number agreement are the facts in (25) (Akmajian, 1970, p. 154), involving highest-subject relativization, and (26) (Baker, 2011, p. 887), illustrating Kimball and Aissen (1971)-type relatives in which the head (a non-subject within the relative clause) attempts to control agreement with the finite verb of the relative clause.

(25)(a) I, who am tall, was forced to squeeze into that VW
(b) we, who are/*am tall, were forced to squeeze into that VW

(26)(a) *I, who Clark am hoping will come,..
(b) !we, who Clark are hoping will come,..

The ungrammaticality of (25b) with *am* tells us that person agreement in the absence of number agreement is illegal. And the fact that *are* is possible (for speakers who have 'Kimball and Aissen effects') in (26b) indicates that the verb can agree with the head in number without agreeing in person: after all, from the ungrammaticality of (26a) [recall (15a)] we learn that person agreement between the finite verb of the relative clause and the head of the relative is impossible when the head is not the finite verb's subject.

To summarize, there can be no person agreement under downward Agree between the clausal π-head and the πP of the subject pronoun because, the latter being encapsulated inside an opaque #P, π cannot itself peek inside the subject and target its specifier (πP). But, provided that π raises to #, the clausal #-head can probe the entire subject pronoun, #P, and attract it to its specifier. Once #P has been probed, its specifier becomes accessible, and hence, in compliance with the constraint in (23), which demands that all the features of the complex probe [$_{\#}$ π [#]] find a match, the π-portion of this complex probe values the

[13]Here and in what follows, whenever I talk about 'person,' I am referring to first- or second-person. Recall that 'third person,' whenever it involves an explicit feature specification [–PART], is marked on D, not on π; 'third person pronouns' are DPs, behaving in relevant respects like common noun phrases.

[14]Preminger (2009; 2011, p. 920) discusses examples of long-distance agreement from 'substandard Basque' which he takes to instantiate person-feature valuation under downward Agree. I do not have the space here to engage in a discussion of these examples. I would seek to reanalyse them in terms of object shift into the matrix clause, with person valuation under the Spec–Head relation.

[15]If the intervening feature had been valued antecedently, # would have had no trouble probing past it. But since (for reasons discussed in the previous paragraph) π cannot probe and value its unvalued feature by itself, this causes this feature to be a harmful intervener for the establishment of probe–goal relations by functional heads higher on the clausal spine. Taking π out of the way (by raising it up to #) is the only way around this intervention effect.

[16]For a defense of the idea that probe–goal relations open up otherwise opaque domains, see Den Dikken (2018).

[*u*PART] feature of the subject's πP-specifier. The raised π-head MUST probe the subject when the latter is in Spec#P. By contrast, when the subject does not raise, π CANNOT probe it when π is *in situ* (because #P is not a match for it, and the subject's πP is not accessible); and when π moves and adjoins to #, it lies dormant as an inactive subpart of [$_{\#}$ π [#]] unless it is activated by the constraint in (23), which applies only when [$_{\#}$ π [#]] is in a Spec–Head relation with the raised subject. From this it emerges that person agreement with pronominal subjects is possible if and only if the subject is in a Spec–Head relationship established in the #P on the clausal spine. Person agreement under (downward) Agree is impossible.

Person Agreement as Attraction

In the approach taken in the section entitled "The Syntax of Agreement: Agree Versus the Spec–Head Relation," person agreement between the finite verb and a first- or second-person pronominal subject involves a relationship of feature valuation targeting the specifier of the structural subject, itself occupying a specifier position [see (24), repeated below]. This reminds us of agreement attraction cases of the type in (27b) [recall (14b), above]. Like (24), (27b) instantiates an agreement relation between the finite verb and the specifier (here, the possessor) of the structural subject: see (28).

(27) (a) these people's identity is to remain a secret
(b) !these people's identity are to remain a secret

(28) [$_{\#P}$ [$_{DP1}$ DP2$_{\{-PART,IND:PL\}}$ [$_{D'}$ D$_{\{-PART,IND\}}$... [$_{NP}$ N]]]$_i$ [$_{\#'}$ [$_{\#}$ π$_{\{u\text{PART}\}}$ [#$_{\{u\text{IND}\}}$]]... t_π... t_i...]]

(24) [$_{\#P}$ [$_{\#P}$ πP$_{\{PART:\pm AUTHOR\}}$ [$_{\#'}$ #$_{\{IND\}}$ [$_{NP}$ N]]]$_i$ [$_{\#'}$ [$_{\#}$ π$_{\{u\text{PART}\}}$ [#$_{\{u\text{IND}\}}$]]... t_π... t_i...]]

In (24), the clausal π-head gets a chance to agree with the πP embedded in the pronominal subject thanks to the fact that the #-head of the clause values its [*u*IND] feature against that of the #P in its specifier. Similarly, in (28) the #-head gets a chance to value its [*u*IND] feature against the [IND:PL] specification of the possessor DP2 embedded in the possessive DP2 thanks to the fact that the π-portion of the complex probe [$_{\#}$ π [#]] can establish a feature-valuing relationship with the [–PART] feature on the head of the possessive DP1, opening it up for # probing the plural feature of DP2. With (27b) commonly referred to as a case of agreement attraction, we come to the conclusion that person agreement with first- or second-person subject pronouns is a form of agreement attraction.

This is a *prima facie* rather surprising conclusion in light of the fact that whereas (27b) is usually considered an error, person agreement with the subject is perfectly flawless. Why does person agreement not have the acceptability status of familiar agreement attraction cases? The answer lies in competition. In the case of person agreement (24), there is just a single [PART]-specified node in the Spec–Head domain of the complex probe [$_{\#}$ π [#]], meeting no competition and serving as the only possible match for the probe's [*u*PART] feature. In (28), on the other hand, there are two instances of [IND] present in the complex subject: one on DP1 and another on DP2. Each is a potential match for an agreement relation with the finite verb. When such competition presents itself, the structurally closest agreement relation is the unmarked one. In (27a), the clausal #-head agrees directly with DP1 in the structure in (28); in (27b), valuation of [*u*IND] is postponed until after π-probing has opened up DP1 and made DP2 available as a goal for #. The unmarked option of these two is (27a); (27b) is the marked case. But in the case of (24), there is no competition — indeed, probing the πP in Spec#P is the clausal π-head's only chance (its last resort, if you will) at getting its [*u*PART] feature valued. Hence markedness does not come into play in (24).

Agreement Attraction

The bulk of the literature on number agreement attraction effects has concentrated, not on cases in which the attractor occupies the 'Saxon genitive' position [as in (14b)], but instead on cases in which the attractor is contained in a post-nominal PP or relative clause, as in (14a) [adapted from Kayne, 2000, and repeated here as (29a)], (29b,c) (Bock and Miller, 1991) and (29d) (Dillon et al., 2013).

(29) (a) the identity of these people {is/!are} to remain a secret
(b) the key to the cabinets {is/!are} rusty
(c) the path to the monuments {is/!are} littered with bottles
(d) the new executive who oversaw the middle managers {was/!were} dishonest about the company's profits.

In Den Dikken (2001), I suggested (following Kayne, 1998) that the DP-contained plural makes its way up to SpecDP (the 'Saxon genitive' position) at LF, via an operation akin to or identical with Quantifier Raising. This would help account for the distributive interpretation of (29a) (for each person, there is a different identity) and possibly of (29b) as well. But a QR-style approach does not carry over to (29c,d), for which there is neither a Saxon-genitival paraphrase nor a distributive reading — and at any rate, QR out of a relative clause would be syntactically very difficult to uphold. I will not pursue this line of thinking further, therefore.

For (29a–c), the idea that probe–goal relations make otherwise opaque domains transparent (Den Dikken, 2018; recall fn. 16) may be put to good syntactic use, with the clause-level π-probe agreeing with the subject-DP for [–PART] and allowing the #-probe to target the DP-contained plural noun phrase. But for (29d), it is inconceivable that the matrix #-head could be given syntactic access to the plural object of the relative clause construed with *executive*. For examples of this type, it seems to me vanishingly likely that syntax could assist in providing an account. So although syntax can make major strides in the understanding of agreement attraction, there remains to my mind an irreducible residue of linear string effects in the realm of agreement attraction phenomena. (Relevant here as well is the discussion of Dillon et al., 2017 at the end of section "Feature Sharing in Non-subject Relativization: The Kimball and Aissen Facts Revisited," below.)

But neither structurally nor linearly is the person specification of a subpart of the complex subject ever local to the finite verb. As a consequence, agreement attraction never involves person,

as we saw in (16) (repeated below): the πP embedded inside the specifier of the clausal #-head cannot be engaged in an agreement relation with the #-adjoined π-head of the clause.

(16)(a) the identity of me {is/*am} to remain a secret
my identity {is/*am} to remain a secret
(b) the identity of you {is/*are} to remain a secret
your identity {is/*are} to remain a secret

In (30), I illustrate the structure of the second example in (16a). The clausal #-head's [*u*IND] can find a match in the number specification for DP1, the possessive noun phrase. And π can value its [*u*PART] feature against DP1's [–PART], contributed by the D-head. The result of these feature valuations is *my identity is to remain a secret*, which is grammatical. As π finds a match in [–PART] ('third person') on DP1, it cannot probe beyond this point. Hence, the possessor's $\pi_{[PART:+AUTH]}$ never comes into the picture. Even though both (30) and (24) (the latter repeated once more below, for ease of direct comparison) feature a πP in the specifier domain of the clausal [# π [# #]] probe, only in (24) is this πP accessible to the π-portion of the complex probe: in (24), the specifier of the clausal #-head is not itself specified for [PART], enabling π to pick the person specification of the subject pronoun as its goal; but in (30), DP1 bears [–PART], rendering a valuation relationship between the clausal π-head and the person features of DP1's pronominal possessor impossible.

(30) [#P [DP1 [#P πP{PART:+AUTH} [#′ #{IND}...] [D′ D1{–PART,IND}...]] [#′ [# π{*u*PART} [#{*u*IND}]]...]]
(24) [#P [#P πP{PART:±AUTHOR} [#′ #{IND} [NP N]]] [#′ [# π{*u*PART} [#{*u*IND}]]...]]

For the versions of (16) in which the personal pronoun occurs in a post-nominal *of*-phrase, agreement between the finite verb and the person feature of the pronoun is also impossible. Syntactically, the fact that the container-DP is specified as [–PART] once again renders a probe–goal relation between the clausal π-head and the pronoun's πP impossible. And because the pronoun's πP is the specifier of the pronominal #P, it is not linearly adjacent to the finite verb either. All avenues toward person agreement attraction in constructions of the type in (16) are thus blocked, as desired.

Long-Distance Agreement

Now that we have an answer to the question of why person-agreement attraction fails in (16), let us verify that long-distance person agreement of the type in (17b) [repeated below, along with grammatical (17a)] is also correctly ruled out.

(17)(a) [ẹ-sau kẹ-dẹi] e-tum ẹ-tawa (Lokąą)
7-fish GER/5-buy 7SG-be.very 7SG-be.difficult
'buying fish is very difficult'
(b) *[min ke-funna] n-tum n-tawa
1SG GER/5-surprise 1SG-be.very 1SG-be.difficult
'surprising me is very difficult'

The number feature of *ê-sau* 'fish' in (17a) is directly represented on DP, and accessible to the complex [# π [# #]] probe in the matrix clause after the π-portion of this probe has established a feature valuation relation with CP, which I assume is, like D, specified for [–PART].[17] But the person feature of *min* in (17b) is not a possible goal for the matrix π-probe: after π has valued its [*u*PART] feature against CP's [–PART], it is no longer active as a probe. The structures in (31a) and (31b) (in which I treat the gerund as the structural of the matrix clause[18]) illustrate, for (17a) and (17b), respectively.

(31)(a) [#P [CP [DP D{–PART,IND}...] [C′ C{–PART}...]] [#′ [# π{*u*PART} [#{*u*IND}]]...]]
(b) [#P [CP [#P πP{PART: ± AUTHOR} [#′ #{IND} [NP N]]] [C′ C{–PART}...]] [#′ [# π{*u*PART} [#{*u*IND}]]...]]

Copular Inversion and Agreement

Next, let us take a closer look at the specificational copular sentences of the section entitled "Agreement in Specificational Copular Sentences." In these sentences, person agreement with the post-copular subject of predication is impossible. The examples in (32) and (33) (repeated from above) show this clearly.

(32)(a) all I could see {*was*/*were*} two staring eyes
(b) all I could see {*was*/**were*} you
(33)(a) *ze betwijfelen dat de schuldige ik ben (Dutch)
they doubt that the culprit I am
(b) *ze betwijfelen dat de schuldige jij bent
they doubt that the culprit you are
(c) ze betwijfelen dat de schuldige Jan is
they doubt that the culprit Jan is

This again falls out from the proposal in "The Syntax of Agreement: Agree Versus the Spec–Head Relation," given the analysis of inverse specificational copular sentences first presented in Moro (1997) and developed in further detail in Den Dikken (2006), according to which their syntax involves fronting of the underlying predicate into the structural subject position, as illustrated in (34):

(34)(a) [SC=RP [SUBJECT] [R′ RELATOR [PREDICATE]]]
⇒ PREDICATE INVERSION ⇒
(b) [TP [PREDICATE]$_i$ [T′ T+RELATOR = *be* [SC=RP [SUBJECT] [R′ t_{REL} t_i]]]]

Predicate inversion results in a syntactic structure in which the only way in which the copula can establish an agreement

[17] For Indo-European, it is not difficult to argue that finite C is specified for person in the same way that D is: the finite complementisers of Indo-European derive from nominal elements (demonstratives, *wh*-words). In the Lokąą examples in (17), there is no C-element to which we can attribute properties on independent grounds. But the fact that we are dealing with a gerund (well-known to be a hybrid of nominal and clausal properties) makes it plausible to assume that its C is specified for 'third person.'

I will ignore the gender (noun-class) agreement found in (17). See fn. 8 for some remarks on gender agreement.

[18] The familiar long-distance number/gender agreement cases of Tsez [from Polinsky and Potsdam, 2001; see (i)] involve a clause in complement (rather than subject) position. See Den Dikken (2018) for discussion of how this long-distance agreement comes about.

(i) eni-r [už-ā magalu b-āc'-ru-łi] b-iyxo
mother-DAT boy-ERG bread.III.ABS III-eat-PTC-NOMINAL.IV III-knows
'the mother knows that they boy ate the bread'

relationship with the post-copular subject is via (downward) Agree. Agree with the entire post-copular subject, as in (32a) and (33c), is perfectly fine; but person agreement with a subpart of the post-copular subject (in particular, with its πP) is impossible, for reasons discussed in "The Syntax of Agreement: Agree Versus the Spec–Head Relation."

It also follows from the approach to syntactic agreement taken in this paper that in contexts of the type in (35b) and (36b), agreement attraction is impossible even for number. In the a-sentences, the noun phrase of *people* is the specifier of the specifier of the clausal #-head, just as in (27b), whose structure was given in (28). But in (35b) and (36b), the noun phrase of *people* is not in an agreement relation with #. This plural noun phrase is invisible to # both under the Spec–Head relation and for Agree purposes: though it is in the c-command domain of the probe #, the #-head can Agree directly only with the complex singular possessive noun phrase as a whole, which leads unequivocally to singular verb inflection.

(35) (a) two people's silhouette {*was*/[!]*were*} all I could decipher
(b) all I could decipher {*was*/**were*} two people's silhouette
(36) (a) these people's information {*is*/[!]*are*} the cause of the computer glitch
(b) the cause of the computer glitch {*is*/**are*} these people's information

The prediction made by the proposal accords well with the facts: while *are* is possible in the a-examples under attraction, it does not work at all in the copular inversion constructions in (35b) and (36b).[19]

Relativization and Agreement

Finally, I will now return to the Kimball and Aissen (1971) facts, further enhanced by Baker (2011). The key contrast here is between (37a,b) and (37c) [adapted from (13) and (15), above]:[20]

(37) (a) I, who Clark {is/*am} hoping will be the finalist,..
(b) you, who Clark {is/*are} hoping will be the finalist,..
(c) these people, who Clark {is/[!]are} hoping will be the finalists,..

While the finite verb in the relative clause can be attracted to the number specification of the head of the relative clause, its person feature cannot be matched by the finite verb when the head is not its subject.

This observation is significant because, as we saw already in (25) [repeated below as (38)], when the head IS the subject of the relative clause, person agreement between the head and the finite verb of the relative clause is grammatical:

(38) (a) I, who am tall, was forced to squeeze into that VW
(b) we, who are/*am tall, were forced to squeeze into that VW

But even when the head is itself a subject, its person agreement behavior has an interesting twist: as Morgan (1972, p. 284) points out, long-distance relativization makes person agreement with the verb in the downstairs clause impossible:

(39) *I, who John says (the FBI thinks) am an anarchist/responsible,…

The empirical picture for person agreement under relativization is further complicated when we take the case form of the head of the relative clause (determined in the external syntactic context) into account. Kimball and Aissen (1971, p. 241) note that number agreement between the non-subject head and the finite verb is possible even when the head is not in a nominative case environment in the matrix clause:

(40) Mark knows/wants to talk to the people who Clark {thinks/[!]think} are in the garden

For number agreement between the head and the finite clause in cases of highest-subject relativization, the case-form of the head is also inconsequential, as (41) shows. However, Akmajian (1970) notes that person agreement in this context is possible only if the head is itself nominative: see (42).

(41) he had the nerve to say that to them, who have made him what he is today
(42) *he had the nerve to say that to me, who have made him what he is today

What I would like to present in this section is a comprehensive account of this entire picture. To my knowledge, this has never been undertaken previously. Analyses of the facts in (38) and (39) are themselves quite few and far between (since Akmajian, 1970; Ross, 1970; Morgan, 1972 first unearthed them, the generative literature has largely set them aside, with a moderate resurgence of attention in recent works by Heck and Cuartero, 2012; Douglas, 2015). But as far as I am aware, these subject relativization data have never been coupled with an analysis of the (extended) Kimball and Aissen facts.[21]

The following are the key players in the discussion to follow:

(a) the representation of person and number in the complex noun phrase presented in "The Syntax of Agreement: Agree Versus the Spec–Head Relation"
(b) the properties of the relative operator *who*
(c) an analysis of relativization involving predication inside the noun phrase (Den Dikken, 2006)
(d) feature sharing between the relative CP and the head noun phrase under concord

[19] The *there*-existential in (ib) makes the same point, given a predicate inversion approach to *there*-sentences (Hoekstra and Mulder, 1990). (ia) with *are* is identical with a naturally occurring sentence taken from the internet; but in (ib) *are* is entirely impossible.

(i) (a) some people's information {*is*/[!]*are*} stored in a database
(b) there {*is*/**are*} some people's information stored in a database

[20] For the sake of uniformity, (37) gives a triple of non-restrictive relatives. Kimball and Aissen's (1971) original examples involve restrictive relatives; but with first- and second-person heads, only non-restrictives are possible. As Baker (2011, p. 887, fn. 10) points out, number attraction to plural is possible in non-restrictive relatives as well.

[21] Douglas (2015, p. 46, fn. 6) does in fact mention one the Kimball and Aissen facts in passing, but never draws these data into the analysis.

(e) feature sharing between CP, its head C, and the inflectional system of the clausal spine

With these players, we can gain a complete understanding of the facts in (37)–(42). The account is entirely deterministic, based in its entirety on assumptions defended in the foregoing and standard or independently plausible ingredients of the theory.

The Featural Specification of *Who* as Relative Operator

Let me begin by stating and supporting my assumptions regarding the featural specification of the relative operator *who* (which converge with those in Douglas, 2015, *contra* Heck and Cuartero, 2012).

The operator *who* projects a DP. As we know from the section entitled "The Place of Person in the Structure of the Noun Phrase and on the Clausal Spine," D is not specifiable for the features [PART: ± AUTHOR] (i.e., for first- or second-person): πP finds itself on a left-branch position inside the structure of the complex noun phrase, and its specification for [PART] does not 'percolate' up to D. We expect it to be universally impossible for *wh*-operators to be inherently marked for [PART: ± AUTHOR]. I assume that inherently, the D-head of *who* is radically unspecified for [PART].

But D IS specifiable for the feature [IND]. In English *wh*-questions, *who* is systematically singular (*who is/*are coming?*, *who is/*are eligible?*) unless it is in a predication relation with a plural-marked nominal, as in *who are the finalists?* (a question enquiring about the identity of the individuals to which *the finalists* applies, NOT a question asking the interlocutor to name the property that the finalists share — the latter would require the use of *what* as the *wh*-operator). In light of the fact that *who*, even in English *wh*-questions, can be plural-marked under the appropriate circumstances, I assume that the *wh*-word *who* is capable of bearing the feature specification [IND:PL]. (I return in the section entitled "Feature Sharing in Non-subject Relativization: The Kimball and Aissen Facts Revisited" to the way in which this comes about.) In relative clauses with a plural-marked human head, it is this plural-specified *who* that serves as the relative operator.

The Syntax of Relativized Noun Phrases

My outlook on the syntactic structure of relativized noun phrases is anchored in Den Dikken's (2006) general theory of the syntax of predication. In this theory, relations that are traditionally treated in terms of modification and its structural correlate of adjunction are brought into the predicational fold, with adjectival attributive modification constructions of the type in (43a) involving reverse predication (i.e., a structure in which the predicate finds itself in the specifier position of the RELATOR phrase), and their counterparts in (43b) being instances of canonical predication (with the predicate in the complement of the RELATOR head).

(43)(a) the visible stars [RP [AP A] [R′ RELATOR [#P # [NP N]]]]
the responsible person
(b) the stars visible [RP [#P # [NP N]] [R′ RELATOR [AP A]]]
the person responsible

For relative clause constructions, this procures a straightforward analysis, with the relative clause in the position of AP in (43b), as shown in (44).

(44) the stars that are visible
[RP [#P # [NP N]] [R′ RELATOR [CP RELCLAUSE]]]
the person who is responsible

The difference between restrictive and non-restrictive relativization can be made in familiar terms, as a function of the size of the relativized constituent (i.e., the nominal in SpecRP). I will not take a specific stand on this issue. I will say only that the familiar ban on restrictive relativization of first- and second-person pronouns can be made to follow if restrictive relatives are necessarily in the scope of the D-head whereas non-restrictives are not: recall from "The Place of Person in the Internal Structure of the Noun Phrase" that first- and second-person pronouns are mere #Ps, hence ineligible for restrictive relativization except when a D is merged with them, as in *the me you're seeing now is different from the me people see in public* (see also fn. 22).

The Raising Approach to Relativization Cannot Make the Right Cut

From (44) it is apparent that I am adopting a head-external approach to relative clauses: the head does not originate inside the relative clause. The person facts reviewed in the introduction to this section supply us with a cogent argument against existing head-internal or 'raising' analyses of relativization, at least for non-restrictive relatives with a first-person pronominal head.

Both Kayne's (1994) version of the raising approach and Bianchi's (1999) development thereof treat the head of the relative clause and the relative operator (*who* or *which*, depending on the humanness of the head) as a single constituent. At some point before the end of the syntactic derivation, the head moves around the relative operator into SpecDP, which is its terminus for Kayne; Bianchi subsequently splits the head and the relative operator apart via onward movement of the head into a position in the high left periphery of the relative clause. But such onward movement happens well and truly after the DP in (45) has already vacated its A-position in the clausal core — the structural subject position in the cases under consideration here. So the difference between Kayne's and Bianchi's versions of the raising analysis is of no consequence to us here: the two analyses share (45a) and (45b).

(45)(a) [DP [D′ D = *who*/*which* HEAD]]
(b) [DP HEADi [D′ D = *who*/*which* t_i]]

Let us investigate what the predictions made by (45) are for person agreement with the head.

To make the examination easier, (46) presents an update of (45) for the specific case of a first-person relativized head:

(46)(a) [DP [D′ D = *who* [#P πP[PART:+AUTHOR] [#′ #...]]]]
(b) [DP [#P πP[PART:+AUTHOR] [#′ #...]]i [D′ D = *who* t_i]]

An immediate question we face is whether the movement of the head to SpecDP happens before or after the DP has made its way into the Â-domain of the relative clause. There is no immediately obvious answer to this question; so I will do the

exercise of verifying the possibility of person agreement with the head for both logically possible scenarios. If the head remains *in situ* in the complement position of D while the DP is still in the A-domain of the clause, we get (47a); if movement to SpecDP happens early, we get (47b).

(47)(a) $[_{\#P}\ [_{DP}\ [_{D'}\ D = \textit{who}\ [_{\#P}\ \pi P\ [_{\#'}\ \#...]]]]]\ [_{\#'}\ [_{\#}\ \pi\ [_{\#}\ \#]]]$..
(b) $[_{\#P}\ [_{DP}\ [_{\#P}\ \pi P\ [_{\#'}\ \#...]]_i\ [_{D'}\ D = \textit{who}\ t_i]]]\ [_{\#'}\ [_{\#}\ \pi\ [_{\#}\ \#]]]$..

Person agreement between the complex probe $[_{\#}\ \pi\ [_{\#}\ \#]]$ and the pronoun contained inside the subject-DP will be possible in (47a) and (47b) provided that the πP of the pronoun, which is quite deeply embedded in the subject, can be made accessible to the π-portion of the complex probe. We can give the π-part of the probe access to πP inside the subject only if the complex probe establishes a feature-valuation relationship with DP and #P. The #-portion of the $[_{\#}\ \pi\ [_{\#}\ \#]]$ probe can value its [*u*IND] feature against that of DP, and D and # share their [IND] specification within the extended nominal projection ('feature inheritance' or 'head-head agreement'). By the logic of Den Dikken's (2018) theory of locality, this should probably be sufficient to render both DP and #P transparent for the purposes of a probe–goal relationship between the clausal π-probe and the πP inside DP. I have assumed (see "The Featural Specification of Who as Relative Operator") that the D-head of *who* is itself radically unspecified for [PART]. So provided that the #-portion of the clausal $[_{\#}\ \pi\ [_{\#}\ \#]]$ probe matches its [*u*IND] feature against that of D, it should be technically possible for the π-portion of this probe to match the [PART:+AUTHOR] specification of the πP inside DP in the structures in (47), regardless of whether movement of #P to SpecDP happens early or late in the derivation.

This sounds like good news for the analysis of the examples in (38), repeated below, where first-person agreement in conjunction with number agreement is obligatory inside the relative clause.

(38)(a) I, who am tall, was forced to squeeze into that VW
(b) we, who are/*am tall, were forced to squeeze into that VW

But the problem for the raising analysis is that it makes person+number agreement with the head of the relative clause behave the same way in highest-subject relatives such as those in (38) and in long-distance relativization cases. As we know from (39), agreement with the head actually fails in examples of long relativization.

(39) *I, who John says (the FBI thinks) am an anarchist/responsible,…

The fact that long-distance relativization cannot give rise to person agreement on the finite verb of the clause of which the head is the subject is unexpected on the raising approach, assuming that (47) can deliver person agreement in principle. The head of the relative clause originates, on the raising approach, in the subject position of the most deeply embedded clause, where we know that, in the absence of relativization, it would certainly control person agreement with the finite verb; and we also know from (38) that in highest-subject relatives the complex structure in (47) that the raising analysis postulates within the relative clause manages to control full agreement with the finite verb.

I conclude, based on (39), that at least for non-restrictive relatives with a first-person head, a raising analysis is not tenable.[22] I will work hereinafter with (44), taken from Den Dikken (2006).

Feature Sharing Between the Head and the Relative Clause: Concord

In the structure in (44), the relative clause and the projection of the head are in a predication relationship. Predication relations are well-known to give rise to feature sharing between the predicate and its subject. In the Russian example in (48a), for instance, the predicative adjective is inflected for the same gender, number, and case as its subject.[23] This feature-sharing relationship between predicates and their subjects is commonly referred to as concord. Den Dikken and Dékány (2019) argue explicitly that concord does not involve the syntactic relationship of Spec–Head agreement. The short version of the argument is that concord (unlike Spec–Head agreement) frequently does not involve complete matching of features: there can be case concord without φ-concord; and there can be φ-concord without case concord. The examples in (48b,c) demonstrate this for Russian.

(48)(a) devočka krasivaja (Russian)
girl.F.SG.NOM beautiful.F.SG.NOM
'the girl is beautiful'
(b) devočka byla krasivoj
girl.F.SG.NOM was beautiful.F.SG.INST
'the girl was beautiful'
(c) eti fakty problema
these fact.M.PL.NOM problem.F.SG.NOM
'these facts are a problem'

From this, I conclude that concord in predication structures (RELATOR phrases) does not involve the syntactic relationship of agreement — it is not a feature-valuation relation but instead a copying operation, arguably [see esp. (50), below] taking place in the post-syntactic component (i.e., at PF).

[22]If one considers connectivity effects (for binding and idiomaticity) to furnish the only compelling argument for a raising approach to relativization, this conclusion, in the narrow version of the text formulation, is innocuous: relative clauses with a first- or second-person head cannot exhibit any such connectivity effects. But if one considers the distribution of articles with particular noun-phrase types to be a critical argument for the raising analysis (as Kayne, 1994 does), the text conclusion is more consequential: (ib) patterns with (ia) in this regard.

(i) (a) the Paris *(that I knew as a boy)
(b) the me *(that people don't often see)

Readers who, based on their general theoretical assumptions regarding the locality restrictions on agreement relations, have a different perspective regarding the possibility of person and number agreement between the structures in (47) and the finite verb of the relative clause will still come to the same *general* conclusion drawn here, viz., that the raising analysis is not capable of accounting for the person agreement facts in non-restrictive relative clauses with a first-person pronominal head. For me, it is (39) rather than (38) that stands out as the problem; but for those readers who find that (47) cannot control person agreement inside the relative clause, the problem is (38) rather than (39). Either way, the raising analysis is bound to run into trouble in the account of (38) vs. (39).

[23]The Russian examples in this section were kindly provided by Irina Burukina (p.c.).

More specifically, for cases of full concord such as (48a), I will assume that the sum total of the features present on the subject is copied over to the predicate. In the specific case of a relativized noun phrase, full concord takes place between the head and the relative CP in a structure of the type in (44). This causes CP to have exactly the same φ- and case-feature set as the head:

(49) $[_{RP}\ [_{\#P}\ \#\ [_{NP}\ N]]]_{\{\varphi,\text{CASE}\}}\ [_{R/}\ \text{RELATOR}\ [_{CP}\ \text{RELCLAUSE}]_{\{\varphi,\text{CASE}\}}]]$

COPYING UNDER CONCORD

Concord targets the full set of features of the head of the relativized noun phrase as a batch, regardless of where these features are represented in the internal syntax of the nominal constituent — blind, that is, to the question of whether the person feature is represented on the syntactic head of the relativized nominal or only on its specifier (as in the case of first- and second-person pronouns). In the post-syntactic component, with internal syntactic bracketing erased, the head of the relative clause is one single cluster of features. All of the relativized head's features will thus be involved in concord — including [PART: ± AUTHOR].

Since C is the head of CP, by standard feature percolation along X-bar projection lines this entire feature set is present on C as well. C and I (i.e., #+π) are in a feature-sharing relationship ('feature inheritance,' 'head-head agreement'), so the local I has the head's features communicated under concord as well. It is via this concordial chain that the local I of a highest-subject relative clause ends up agreeing with the head for all φ- and case-features.

This predicts that person agreement between the subject-head and the finite verb of the relative clause is possible only when the head has the appropriate case (i.e., nominative): otherwise there is a clash with I. As we saw in (42) [see again (50a)], this prediction is borne out, in a structural accusative case context. It is worth emphasizing that person agreement in the relative clause remains ungrammatical in environments in which the accusative case form of the head is not the fruit of a structural case-assignment relationship but instead the default case (see Schütze, 2001), as in (50b) (Akmajian, 1970) and (50c) (not previously discussed in the literature, to my knowledge).

(50) (a) he had the nerve to say that to me, who {has/*have} made him what he is today
(b) it is me who {is/*am} responsible
(c) (A) who's going to climb up the ladder?
(B) definitely not me, who {has/*have} vertigo
(B′) me, who {has/*have} vertigo, climb up that ladder?! no way!

Neither in *it*-clefts (50b) nor in fragment answers (50c.B) or '*Mad Magazine* sentences' (50c.B′) does the syntax assign structural accusative case to the pronoun. The accusative case form of the pronoun is not the exponent of a structural accusative case feature valued in the course of the syntactic derivation: the default accusative is a purely phonological (PF) property of the pronouns in question. Concord is a PF operation, so it copies not just structural accusatives but also default accusatives over onto the relative clause and, ultimately, onto the I-head of the relative clause, which has a nominative case feature. Resulting in a feature clash at I, the result of this copying is correctly rejected, not just in (50a) but also in (50b,c).

Unlike concord for person, number agreement between the head of a highest-subject relative clause and the finite verb is not ruled out in non-nominative environments [see (41)] because, as I mentioned in the section entitled "The Featural Specification of Who as Relative Operator," the relative operator *who* is specifiable for [IND:PL] independently of concord. Person-feature sharing between the head and the finite verb, by contrast, is entirely dependent on concord, which entails case-feature identity.[24]

Though, as we have seen, the concord relationship between the head of the relativized noun phrase and the relative CP can stretch all the way down to the I-domain of the relative clause (via the feature-sharing relation between C and the local I), it cannot reach beyond this point. It is entirely impossible for concord to penetrate a clause embedded inside the relative clause: there is no path from the matrix I down into the subordinate clause along which the cluster of features of the head could be copied into the lower clause and reach its I-domain. In non-highest-subject relative clauses, in fact, the I of the relative clause itself is in a feature-valuation relationship with the subject of its clause, which is not the *wh*-operator linked to the head. So concord between the head of the relativized noun phrase and the I-domain is restricted to highest-subject relatives; the I of a clause embedded within the relative CP cannot be the beneficiary of a concord relationship between the head and the relative CP.[25] This explains the locality effect seen in (38) vs. (39): I in the lower clause in (39) can only get default person inflection ('third person'). Note that, because *who* is itself specifiable for [IND:PL], independently of CONCORD, it is expected that number agreement should be possible in the downstairs clause — as is indeed the case: (51) is grammatical (see Douglas, 2015).

(51) we, who John says (the FBI thinks) are anarchists/responsible,..

[24]The facts of person agreement in German relative clause constructions are more complex than the English ones. Heck and Cuartero (2012) report a singular/plural split for finite verb agreement in relative clauses with a pronominal head: see (i). It is likely that the form of the relative pronoun plays a role in this, as Heck and Cuartero (2012) suggest. But the exact way in which the text analysis can mobilize the form of the German relative pronouns in an analysis of the agreement facts in (i) is something that I have not figured out at this time.

(i) (a) *weil du es bist, der die ganze Arbeit machst (German)
since you$_{SG}$ it are.2SG RELPRON the whole work do.2SG
(b) weil ihr es seid, die die ganze Arbeit macht
since you$_{PL}$ it are.2PL RELPRON the whole work do.2PL
'since it's you who do all the work'

[25]One might ask whether concord between the head and the relative CP is possible at all when the I-domain is in a feature-valuing Agree relationship with a non-relativized subject. If C and I engage in an automatic feature-sharing relationship ('feature inheritance,' 'head-head agreement') and if feature-sharing between a head and its maximal projection is likewise automatic (as is standardly assumed: it is in fact a linchpin of the X-bar principle), it follows that when I is in an Agree relationship for its φ- and case-features with a non-relativized subject, it is impossible for the head and the relative CP to engage in concord for any features that do not match those of the subject of the relative clause. I will accept this conclusion, leaving a fuller investigation of the workings of concord for another occasion.

Feature Sharing in Non-subject Relativization: The Kimball and Aissen Facts Revisited

With the results of the discussion in the previous section in mind, let us return to the facts noted by Kimball and Aissen (1971) and Baker (2011):[26]

(37)(a) I, who Clark {is/*am} hoping will be the finalist,..
(b) you, who Clark {is/*are} hoping will be the finalist,..
(c) these people, who Clark {is/[!]are} hoping will be the finalists,..

Concord does not help create person agreement in these kinds of sentences: the I-domain of the relative clause is in an Agree relationship with the subject of the relative clause (*Clark*) in syntax, which values I's person feature — the DP of *Clark* is specified as [–PART]. Concord between the head and the relative clause could not interfere with this. By the time that the effects of concord could kick in (at PF), the φ-features of the relative clause's finite verb have already been fixed. The null hypothesis is that PF cannot undo or override specifications for φ-features established by valuation under Agree in syntax. So for (37a,b), concord would come too late: it cannot impose the head's [PART: ± AUTHOR] specification onto the finite verb of the relative clause (already valued as [–PART] by *Clark*) anymore.

Things are different in the case of number, for which the relative operator *who* is inherently specifiable. The feature [IND:PL] can be present on *who* in the syntax of the relative clause, and under the right circumstances, it can impose itself on the finite verb of the relative clause. There are, logically speaking, three points in the structure at which number agreement between the finite verb and *who* could come about in the case of (37c): in CP, under Spec–Head agreement; in IP, with *who* as an adjunct to or outer specifier of IP (in the case of densely successive-cyclic movement), again under a form of Spec–Head agreement; or under (downward) Agree, when the *wh*-operator is adjoined to the phase in the complement of I. In each of these configurations, the finite verb should be able to establish a feature-valuation relationship with *who*'s [IND:PL]. For my purposes in this paper, it does not matter which of these options is the right one. I will leave the matter open.[27]

One thing that I think is worth noting is that the ungrammaticality of *are* in (37b) not only confirms that person agreement is impossible in Kimball and Aissen-style constructions but also compels us to be precise about the number-matching relationship between the head of the relative clause and *who*, the relative operator. The null hypothesis for English *you*, whose form does not covary with the number of addressees, is that its morphological feature specification is constant regardless of its reference. Since *you*, when it is itself the subject of a finite verb, always triggers a plural form of the verb (also in the case of the copula: *you are*), this leads Kayne (2000) to assume that *you* is morphologically plural even in contexts of singular reference. Adopting this assumption leads to the conclusion that the number specification for the operator *who* in the relative clause is based on semantic numerosity, not morphological number. We know from (37c) that *who*, when specified as [IND:PL], is capable of controlling plural agreement in the relative clause. The fact, then, that plural agreement on the finite verb (*are*, the blanket plural form of the present-tense copula) is impossible in (37b) tells us that English *you*, when it has a singular referent [as is clear in (37b) from the form of the predicate nominal, *the finalist*], cannot be construed with [IND:PL]-specified *who* in the relative clause. The data in (52) (from Douglas, 2015, p. 36) make the same point for highest-subject relatives:

(52)(a) he had the nerve to say that to you$_{SG}$, who {has/*have} made him what he is today
(b) he had the nerve to say that to you$_{PL}$, who {have/*has} made him what he is today.

In highest-subject relatives with a non-nominative pronominal head (where person agreement is impossible, for reasons discussed in "Feature Sharing Between the Head and the Relative Clause: Concord"), there is a clear difference between (52a) (with a single addressee) and (52b) (with a plurality of addressees) in the inflection on the finite verb. We know from (42) that person agreement with the pronominal head is excluded when the head is in a non-nominative environment (because concord between the head and the C–I cluster of the relative clause would result in a case clash in this context); so *have* in (52b) is a reflex of number agreement alone, between the finite verb and the relative operator *who*. The question is what determines the number specification of this *who*.

[26]The unusual agreement pattern in (37c) is particularly well-known for cases of relativization, but Kimball and Aissen (1971) themselves also report such effects for root *wh*-questions: recall (13b). I will turn to this at the end of this section. Note in this connection as well that Hartsuiker et al. (2001) point out agreement attraction effects triggered by objects in Dutch SOV constructions.

[27]General assumptions regarding the directionality of syntactic structure building (see Den Dikken, 2018 and references cited there) will make particular approaches to the way in which *who* in (37c) establishes its number agreement relation with the finite verb more plausible than others. But this is not the place to delve into these matters.

All three approaches can, it seems to me, account for Kimball and Aissen's (1971) observation that number agreement between the non-subject relative operator and the finite verb cannot happen when the subject of the relative clause is non-pronominal: (i). There are good grounds for believing that weak pronominal arguments must establish a Spec–Head relationship with the verb (as witness, for instance, the fact that object shift, which establishes a Spec–Head relation, is obligatory, in languages that have it, whenever the object is a weak pronoun). Forging an agreement relationship between *who* and the finite verb (whether it be via the Spec–Head relation or under Agree) would make it impossible for *he* to engage in a Spec–Head agreement relationship with the inflectional system of the relative clause.

(i) these people, who he {is/*are} hoping will be the finalists, ..

I would like to point out in this context that though Dillon et al. (2017, p. 81) did find 'a non-significant numerical trend' in this direction in object *wh*-questions (such as *which basketball players {is/[!]are} he planning to use this season?*), they 'failed to observe any interaction of subject type [pronoun vs. lexical noun phrase; MdD] and the mismatch effect.' In their conclusion (p. 90), however, they call attention to the fact that the approach to the attraction facts that they favor 'predicts that highly marked controllers should outcompete less marked controllers, minimizing the amount of interference they contribute. Intuition suggests that this prediction is correct' — based on examples of the type in (ii) [their (9), not explicitly tested in any of their experiments], involving a first-person singular subject.

(ii) which flowers (am/*are) I planting in the garden today?

If, as Kayne (2000) argues, *you* is always morphologically plural, the fact that 'singular *you*' resists construal with plural *who* is surprising if the number specification for *who* is determined on the basis of morphological feature matching. After all, 'singular *you*' then has what it takes, morphologically, to license plural *who*. This does not necessarily mean, however, that Kayne's morphological analysis of *you* is ill-founded (which is what Douglas, 2015, p. 36 takes the facts in (52) to show). What the ill-formedness of *are* in (37b) and the distribution of *has* and *have* in (52) show, on Kayne's approach to *you*, is that the determination of the number specification for the relative operator *who* is based, not on the morphological number specification (i.e., the [IND]-feature) of the head of the relative clause, but on the numerosity of the referent of the head of the relative clause. Succinctly put, on Kayne's analysis of English *you*, these inflection facts would have to be a reflex of 'semantic agreement.' The question of whether 'semantic agreement' exists and how it works is by no means an easy one to answer (see, e.g., Wechsler, 2011 and references cited there). I will not take a stand on the matter because it is orthogonal to my concerns in this paper. But the lie of the land is clear: for those who believe independently that 'semantic agreement' exists and is applicable in the context of relativization, the facts reviewed above do not pose a threat to Kayne's (2000) argument that English *you* is always grammatically plural; but to those who reject 'semantic agreement' (in general, or in the specific context at hand), these facts suggest that English has two homophonous forms of the second-person pronoun *you*, only one of them morphologically specified as [IND:PL].

One final remark is in order. I have argued in this section that concord is not at play in Kimball and Aissen-style number agreement cases of the type in (37c), nor in examples such as (52b): the plural number inflection of the finite verb is determined in these cases by the [IND:PL] of *who* itself. This leads us to expect that plural agreement cases of these types should be replicable in *wh*-questions as well, with *wh*-operators that are specified as [IND:PL]. Consider the pair in (53)–(54):

(53) which people {are/*is} hoping Clark will be a finalist?
(54) which people {is/!are} Clark hoping will be finalists?

The *wh*-phrase *which people*[IND:PL] of course controls plural agreement with the finite verb of the *wh*-question when serving as its subject, as in (53). But Kimball and Aissen (1971, p. 245) already showed that *wh*-questions can also give rise to agreement attraction, as shown in (54) (recall also (13)). Interestingly, however, Richard Kayne tells me that he rejects such agreement in questions with *who*, even when *who* is construed through predication with an explicitly plural nominal in the lower clause, as in *who* {*does/*do*} *Clark think will be the finalists*?. This should be investigated further: it is likely to be revealing regarding the precise circumstances (incl. possible locality restrictions) under which a [IND:PL] specification can be assigned to the bare *wh*-operator *who* under concord.

Dillon et al. (2017) report on a recent series of experiments they ran on Kimball and Aissen-effects in English *wh*-questions.[28] Their results show that although 'mismatch effects are largest when the *wh*-object is adjacent to the verb' (p. 86), string adjacency between the finite verb and the *wh*-constituent is nonetheless 'neither necessary nor sufficient to generate a mismatch effect' (p. 85). The answer to the question of whether the structural subject or the *wh*-constituent controls finite verb agreement turns out to be determined for the most part by syntactic (configurational) factors. Dillon et al. (2017) review the spectrum of extant theoretical approaches to agreement attraction errors (in terms of feature transmission, subject confusion, and syntactic interference), and the results of their extensive empirical studies land them on the side of syntactic interference approaches such as Franck et al. (2006). This is a conclusion I welcome. However, Dillon and colleagues also caution that the role played by the Spec–Head relation may be less robust than Franck and colleagues (and the present paper) have made it out to be: in particular, Dillon et al. (2017, p. 80) 'failed to find any reliable effect of preposition fronting on [their] ratings' — agreement attraction was rated roughly equally in (55a) and (55b):

(55) (a) which trees {is/!are} the hiker resting under?
(b) under which trees {is/!are} the hiker resting?

Future research should look into this at greater length, against the background of syntactic analyses of PP pied-piping (see esp. Heck, 2008 for important discussion of the syntax of pied-piping).

Beyond the P-stranding/PP pied-piping dichtomy, Dillon et al.'s (2017) study leaves room for additional experimentation regarding the nature and scope of Kimball and Aissen-style effects as well. I would particularly encourage future work that juxtaposes *wh*-relatives and *wh*-questions in a way that makes them more directly comparable. All of and colleagues five acceptability-judgment experiments involve ROOT *wh*-questions, with subject–auxiliary inversion, which maneuvres the finite verb into a position in between the *wh*-constituent (in SpecCP) and the structural subject (in SpecIP). This makes these *wh*-questions different from relative clause constructions, which do not feature movement of the finite verb to C. A direct comparison of *wh*-relatives with NON-ROOT *wh*-questions, leveling the playing field with regard to the placement of the finite verb, would be able to tell us with more precision whether the 'Kimball and Aissen effect' is the same or different, qualitatively and/or quantitatively, in relatives and *wh*-questions.

Summary

After this (unavoidably) rather elaborate discussion of agreement in relative clause constructions, let me summarize our findings in this section regarding person versus number agreement.

I started out by giving an explicit syntax for the representation of person inside the noun phrase and on the clausal spine. For first- and second-person pronouns, a structure was presented in which person (π) projects a phrase occupying the specifier

[28] In their title and in various places throughout their paper, they refer to this as 'object agreement.' But as is clear from cases of the type in (54) (where the controller of finite verb agreement in the matrix clause is not an object of that clause but instead the subject of the subordinate finite clause), this term should not be taken too literally.

position of number (#). This not only gives us a natural syntax for pronoun–noun constructions such as *us linguists*, but also paves the way for an explanation for the range of ways in which person behaves differently from number in the realm of agreement phenomena. Effectively, whenever person agreement is possible in syntax, it comes about in a configuration that is similar to the number agreement attraction effect seen in !*these people's identity are to remain a secret*. Both can materialize only when the subject is in a Spec–Head relation with the inflectional cluster on the clausal spine. This gives us an account of the majority of person agreement contexts, and explains the restrictiveness of person agreement in comparison to number agreement (without, however, giving person agreement the flavor of an attraction error).

For cases of person agreement between the pronominal head of a relativized noun phrase and the finite verb inside the relative clause, an appeal to a feature-sharing relationship different from Agree or Spec–Head agreement needed to be exploited: concord, a post-syntactic copying operation. The feature bundle of the head of the relativized noun phrase can be copied wholesale onto its predicate (the relative clause), and can make it from there to the inflectional cluster of the relative clause, but not beyond. Via this feature-copying process, full feature-sharing between the head and the finite verb becomes possible in highest-subject relatives — but not in long-distance subject relativization constructions, nor in cases of non-subject relativization. Number agreement between the finite verb of the relative clause and the head is always possible, giving rise to attraction effects in the case of long-distance and non-subject relativization, *à la* Kimball and Aissen (1971). That number never comes up empty-handed is thanks to the fact that the English relative pronoun *who* is specifiable for plural number (probably under 'semantic agreement,' as discussed in "Feature Sharing in Non-subject Relativization: The Kimball and Aissen Facts Revisited"). But as a DP, *who* is not specifiable for first- or second-person: [PART: ± AUTHOR] is exclusively the province of a πP in the specifier position of #P.

The main effects of SCOPA have now been successfully derived.[29]

CONCLUSION

The empirical spotlight in this paper has been on 'out of the ordinary' agreement phenomena and the circumstances under which they are found in Universal Grammar. A key property common to all 'agreement attraction' and 'long-distance agreement' cases is that they cannot involve first- or second-person. It is this person restriction (hitherto poorly understood) that has been at center-stage here.

Baker's, (2008, 2011) SCOPA captures the specialness of person, but because it is a condition that is itself left underived, it cannot explain it. I have derived SCOPA from *(a)* the syntactic representation of person in the noun phrase (with person structurally represented as a *specifier*) and on the clausal spine (with person and number each projecting X-bar structures, the former's phrase embedded in the latter's) and *(b)* the workings of agreement and concord. Central to the syntax of *(b)* is a distinction between (downward) Agree and Spec–Head agreement, with only the latter capable of effecting number and person agreement conjointly.

This, in combination with the representation of person in the internal structure of pronouns, successfully rules out all person agreement attraction and long-distance person agreement, as desired. The Spec–Head relation creates just the sort of niche needed for person agreement where it is legal.

With Franck et al. (2006), this paper thus affirms the existence of two syntactic mechanisms for the establishment of φ-feature agreement: Agree and the Spec–Head relation. Agree is more liberal in the sense that it can potentially establish long-distance agreement dependencies, whereas Spec–Head relations are by definition more local. But in another way, Agree is also more restrictive: it is impossible for an Agree-probe to target a subpart of a specifier in its c-command domain; so whenever we find agreement dependencies between a head and a subpart of a specifier (including cases of person agreement with subjects), a Spec–Head configuration must be involved.

Mancini et al. (2017) also argue for the postulation of two different mechanisms involved in agreement phenomena, which they call 'feature-checking' and 'feature-mapping.'[30] Number and person agreement are argued to involve a common φ-feature-checking mechanism ('of which the similar left-anterior negative effect could be evidence'; p. 142) but to differ in their feature-mapping options, with number mapping to cardinality and person to the discourse ('which would be behind the different posterior negative effect elicited by the two violations'; p. 142).

I have little to add to Mancini and her co-workers' findings regarding the interpretive side of person marking (what they call 'feature-mapping'), and find their interpretation thereof (which appeals to the representation of the speaker and the hearer in the left periphery) eminently plausible. For the discussion in the present paper, this is of no immediate concern.

[29]What Baker (2011) calls 'two-and-a-half agreement' (found in ditransitive constructions in which V agrees for all φ-features with the subject and the indirect object but cannot agree for person with the direct object) falls out from the text proposal: the direct object of a ditransitive, unlike the subject and the indirect object, is never in a Spec–Head relation with the verb. Baker's (2011, sect. 3.1) analysis of partial agreement in Sakha is also directly compatible with my derivation of SCOPA.

I will not talk here about the fact that a Spanish predicate adjective can reflect the number and gender features of its subject but never its person feature [Spanish *somos gorda(*mo)s* 'we are fat']. Baker derives this from SCOPA; for me this does not involve agreement but concord, which for reasons unknown to me can only be partial in this context. For the apparent person mismatch in Spanish 'unagreement' constructions (*los profesores somos inteligentes* 'the professors are.1PL intelligent'), DP-internal predication between the common noun phrase and a silent pronoun (contributing '1PL') is plausible. Mancini et al. (2014a) present three eye-tracking experiments and a grammaticality judgment task contrasting Spanish 'unagreement' cases and erroneous person agreement, showing that although '[s]imilarly to Person Mismatch, Unagreement elicited an early negative effect, suggesting rapid recognition of a subject–verb mismatch ..., [a]t later stages, the effect of true person anomalies persisted, while the patterning of Unagreement with Standard Agreement clearly evidenced that the grammaticality of the apparent mismatch had been acknowledged' (p. 143).

[30]Mancini et al. (2017) is currently the most recent installment in a series of psycho- and neuro-linguistic studies conducted by Mancini et al. (2011; 2014a; 2014b, for earlier reports), all of which have found significant differences in behavior between number agreement and person agreement.

More to the point of the current discussion is Mancini et al.'s (2017) conclusion that number and person agreement share the same feature-checking mechanism. If what I have argued in the foregoing is on target, the feature-checking processes involved in number and person agreement are not, in fact, systematically identical: while number agreement is possible under both Agree and the Spec–Head relation, the facts reviewed above suggest that person agreement is established exclusively under the Spec–Head relation. The neurological measurements reported in Mancini and colleagues work do not suggest that there is a grammatical-processual difference between number and person agreement — but this may very well be an effect of the choice of constructions and languages studied: preverbal and silent subjects in Romance pro-drop languages (Italian, Spanish). It is likely that preverbal subjects and pro-dropped subjects in these languages are systematically in a Spec–Head relation with the T-head. Recall that both number and person agreement are possible under the Spec–Head relation. To probe into the question of whether the feature-checking mechanism(s) involved in person and number agreement are neurologically different, one would need to look at data like the ones studied in the present paper. These data define a research agenda that can take the interesting results of the work done by Mancini's team further.

The picture resulting from the present paper (in particular, the division of labor between Agree, the Spec–Head relation, and the post-syntactic copying operation called concord) is principled and descriptively adequate in the complex realm of agreement attraction and long-distance agreement constructions discussed in this paper. The attractions of agreement are a boundless resource for morphosyntacticians. My hope is that the perspectives on the workings of agreement and the structural representation of the φ-features within the noun phrase will prove their mettle well beyond the range of facts reviewed here.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

An earlier version of this paper was presented as an invited talk at the Getegra workshop 'Agreement Φ — In honor of Maria Denilda Moura,' organized in Recife (Brazil) by the Universidade Federal Rural de Pernambuco in February 2014. I thank the organizers of this splendid event (in particular, Marcelo Sibaldo and Dorothy Brito) for the invitation, and the participants for their excellent comments. Apart from reruns at the Linguistics Departments of NYU, Rutgers University, the University of Texas at Arlington, and MIT in the spring of 2014, nothing happened with the paper until the invitation extended to me in early 2018 by the editors of the present collection of articles to contribute my paper to it. The helpful feedback of two reviewers (gratefully acknowledged here) and the developments since 2014 in my own thinking about agreement have resulted in an output that, while sharing the key ingredients with the original paper, looks quite different from its predecessor. All errors (agreement- or otherwise) in the text are my responsibility.

REFERENCES

Ackema, P., and Neeleman, A. (2018). *Features of Person: From the Inventory of Persons to Their Morphological Realization*. Cambridge, MA: MIT Press.

Akmajian, A. (1970). On deriving cleft sentences from pseudo-cleft sentences. *Linguist. Inq.* 1, 149–168.

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge: Cambridge University Press.

Baker, M. (2011). When agreement is for number and gender but not person. *Nat. Lang. Linguist. Theory* 29, 875–915. doi: 10.1007/s11049-011-9147-z

Béjar, S., and Kahnemuyipour, A. (2017). Non-canonical agreement in copular clauses. *J. Linguist.* 53, 463–499. doi: 10.1017/s002222671700010x

Benveniste, E. (1966). *Problèmes de Linguistique Générale*. Paris: Gallimard.

Bianchi, V. (1999). *Consequences of Antisymmetry: Headed Relative Clauses*. Berlin: Mouton de Gruyter.

Bobaljik, J., and Wurmbrand, S. (2005). The domain of agreement. *Nat. Lang. Linguist. Theory* 23, 809–865. doi: 10.1007/s11049-004-3792-4

Bock, K., and Miller, C. (1991). Broken agreement. *Cognit. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Bosma, D. (ed.) (1981). *Life in our Village: Short Stories From Solomon Islands*. Honiara: Translation Committee, Solomon Association.

Branigan, P. H., and MacKenzie, M. (2002). Altruism, A'–movement, and object agreement in Innu-aimûn. *Linguist. Inq.* 33, 385–407. doi: 10.1162/002438902760168545

Bruening, B. (2001). *Syntax at the Edge: Cross-Clausal Phenomena and the Syntax of Passamaquoddy*. Ph.D. thesis, MIT, Cambridge.

Declerck, R. (1988). *Studies on Copular Sentences, Clefts and Pseudo-Clefts*. Leuven: Leuven University Press.

Den Dikken, M. (1998). Review of Moro (1997). *Linguistische Berichte* 174, 246–263.

Den Dikken, M. (2001). 'Pluringulars', pronouns and quirky agreement. *Linguist. Rev.* 18, 19–41.

Den Dikken, M. (2006). *Relators and Linkers*. Cambridge, MA: MIT Press.

Den Dikken, M. (2017). "Pseudoclefts and other specificational copular sentences," in *The Companion to Syntax*, 2nd Edn, eds M. Everaert and H. van Riemsdijk (Oxford: John Wiley & Sons, Inc.).

Den Dikken, M. (2018). *Dependency and Directionality*. Cambridge: Cambridge University Press.

Den Dikken, M., and Dékány, E. (2018). A restriction on recursion. *Syntax* 21, 37–71. doi: 10.1111/synt.12149

Den Dikken, M., and Dékány, E. (2019). Adpositions and case: alternative realisation and concord. *Finno-Ugric Lang. Linguist*. 8.

Den Dikken, M., Meinunger, A., and Wilder, C. H. (2000). Pseudoclefts and ellipsis. *Stud. Linguist.* 54, 41–89. doi: 10.1111/1467-9582.00050

Den Dikken, M., and Griffiths, J. (to appear). Ellipsis and Spec–Head agreement.

Dillon, B., Levy, J., Staub, A., and Clifton, C. (2017). Which noun phrases is the verb supposed to agree with? Object agreement in American English. *Language* 93, 65–96. doi: 10.1353/lan.2017.0003

Dillon, B., Mishler, A., Slogget, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Douglas, J. (2015). Agreement (and disagreement) among relatives. *Cam. Occas. Papers Linguist.* 7, 33–60.

Frances, W. N. (1986). Proximity concord in english. *J. English Linguist.* 19, 309–318.

Franck, J., Lassi, G., Frauenfelder, U., and Rizzi, L. (2006). Agreement and movement: a syntactic analysis of attraction. *Cognition* 101, 173–216. doi: 10.1016/j.cognition.2005.10.003

Guasti, M. T., and Rizzi, L. (2002). "Agreement and tense as distinctive syntactic positions: Evidence from acquisition," in *Functional Structure in DP and IP.*

The Cartography of Syntactic Structures, Vol. 1, ed. G. Cinque (New York, NY: Oxford University Press), 167–194.
Harbour, D. (2016). *Impossible Persons*. Cambridge, MA: MIT Press.
Harley, H., and Ritter, E. (2002). Person and number in pronouns: a feature-geometric analysis. *Language* 78, 482–526. doi: 10.1353/lan.2002.0158
Hartsuiker, R. J., Antón-Méndez, I., and van Zee, M. (2001). Object attraction in subject-verb agreement construction. *J. Mem. Lang.* 45, 546–572. doi: 10.1006/jmla.2000.2787
Hartsuiker, R. J., Schriefers, H., Bock, K., and Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Mem. Cogn.* 31, 1316–1326. doi: 10.3758/BF03195814
Heck, F. (2008). *On Pied-Piping: Wh-Movement and Beyond*. Berlin: De Gruyter Mouton.
Heck, F., and Cuartero, J. (2012). "Long distance agreement in relative clauses," in *Local Modelling of Non-Local Dependencies in Syntax*, eds A. Alexiadou, T. Kiss, and G. Müller (Berlin: Mouton de Gruyter), 49–83.
Heycock, C. (1992). *Layers of Predication. The Non-Lexical Syntax of Clauses*. Ph.D. thesis, University of Pennsylvania, New York, NY.
Heycock, C. (2009). Agreement in specificational sentences in Faroese. *Nordlyd* 36, 57–77.
Heycock, C. (2012). Specification, equation, and agreement in copular sentences. *Can. J. Linguist.* 57, 209–240. doi: 10.1353/cjl.2012.0033
Hoekstra, T., and Mulder, R. (1990). Unergatives as copular verbs: locational and existential predication. *Linguist. Rev.* 7, 1–79.
Kayne, R. (2005). *Movement and Silence*. New York, NY: Oxford University Press.
Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
Kayne, R. S. (1998). *Class lectures, NYU, fall 1998*.
Kayne, R. S. (2000). "Notes on English agreement," in *Parameters and Universals*, ed. R. S. Kayne (Oxford: Oxford University Press), 187–205.
Kimball, J., and Aissen, J. (1971). I think, you think, he think. *Linguist. Inq.* 2, 241–246.
Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject–verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055
Mancini, S., Molinaro, N., Davidson, D., Avilés, A., and Carreiras, M. (2014a). Person and the syntax–discourse interface: an eye-tracking study of agreement. *J. Mem. Lang.* 76, 141–157. doi: 10.1016/j.jml.2014.06.010
Mancini, S., Postiglione, F., Laudanna, A., and Rizzi, L. (2014b). On the person-number distinction: subject-verb agreement processing in Italian. *Lingua* 146, 28–38. doi: 10.1016/j.lingua.2014.04.014
Mancini, S., Quiñones, I., Molinari, N., Hernandez-Cabrera, J., and Carreiras, M. (2017). Disentangling meaning in the brain: left temporal involvement in agreement processing. *Cortex* 86, 140–155. doi: 10.1016/j.cortex.2016.11.008
Merchant, J. (2001). *The Syntax of Silence: Sluicing, Islands, and the Theory of Ellipsis*. Oxford: Oxford University Press.
Morgan, J. L. (1972). "Verb agreement as a rule of English," in *Papers From the Eighth Regional Meeting of the Chicago Linguistic Society*, eds P. M. Peranteau, J. N. Levi, and G. C. Phares (Chicago: CLS), 278–286.
Moro, A. (1997). *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge: Cambridge University Press.
Nevins, A. (2007). The representation of person and its consequences for person–case effects. *Nat. Lang. Linguist. Theory* 25, 273–313. doi: 10.1007/s11049-006-9017-2
Nevins, A. (2011). Multiple agree with clitics: person complementarity vs. omnivorous number. *Nat. Lang. Linguist. Theory* 29, 939–971. doi: 10.1007/s11049-011-9150-4
Palmer, B. (2009). Clause order and information structure in Cheke Holo. *Ocean. Linguist.* 48, 213–249. doi: 10.1353/ol.0.0038
Polinsky, M., and Potsdam, E. (2001). Long-distance agreement and topic in Tsez. *Nat. Lang. Linguist. Theory* 19, 583–646.
Postal, P. (1966). "On so-called 'pronouns' in English," in *Report on the Seventeenth Annual Round Table Meeting on Linguistics and Languages Studies*, ed. F. Dinneen (Washington, DC: Georgetown University Press), 177–206.
Preminger, O. (2009). Breaking agreements: distinguishing agreement and clitic-doubling by their failures. *Linguist. Inq.* 40, 619–666. doi: 10.1162/ling.2009.40.4.619
Preminger, O. (2011). Asymmetries between person and number in syntax: a commentary on baker's SCOPA. *Nat. Lang. Linguist. Theory* 29, 917–937. doi: 10.1007/s11049-011-9155-z
Raposo, E. (1987). Case theory and Infl-to-comp: the inflected infinitive in European Portuguese. *Linguist. Inq.* 18, 85–109.
Ross, J. R. (1970). "On declarative sentences," in *Readings in English Transformational Grammar*, eds R. A. Jacobs and P. S. Rosenbaum (Waltham, MA: Ginn and Company), 222–277.
Schütze, C. (2001). On the nature of default case. *Syntax* 4, 205–238. doi: 10.1111/1467-9612.00044
Shlonsky, U. (2004). The form of semitic noun phrases. *Lingua* 114, 1465–1526. doi: 10.1016/j.lingua.2003.09.019
Sigurðsson, H. Á., and Holmberg, A. (2008). "Icelandic dative intervention: Person and Number are separate probes," in *Agreement Restrictions*, eds R. D'Alessandro, S. Fischer, and G. H. Hrafnbjargarson (Berlin: Mouton de Gruyter), 251–279.
Slioussar, N., and Malko, A. (2016). Gender agreement attraction in Russian: production and comprehension evidence. *Front. Psychol.* 7:1651. doi: 10.3389/fpsyg.2016.01651
Uriagereka, J. (1999). "Multiple spell-out," in *Working Minimalism*, eds S. D. Epstein and N. Hornstein (Cambridge, MA: MIT Press), 251–282.
Wechsler, S. (2011). Mixed agreement, the person feature, and the index/concord distinction. *Nat. Lang. Linguist. Theory* 29, 999–1031. doi: 10.1007/s11049-011-9149-x

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 May 2019
doi: 10.3389/fpsyg.2019.00983

Attachment and Concord of Temporal Adverbs: Evidence From Eye Movements

Nicoletta Biondo[1,2,3], Francesco Vespignani[3] and Brian Dillon[4]*

[1] Basque Center on Cognition, Brain and Language (BCBL), San Sebastián, Spain, [2] Fondazione Marica De Vincenzi ONLUS, Rovereto, Italy, [3] Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy, [4] Department of Linguistics, University of Massachusetts (UMass), Amherst, MA, United States

The present study examined the processing of temporal adverbial phrases such as "last week," which must agree in temporal features with the verb they modify. We investigated readers' sensitivity to this feature match or mismatch in two eye-tracking studies. The main aim of this study was to expand the range of concord phenomena which have been investigated in real-time processing in order to understand how linguistic dependencies are formed during sentence comprehension (Felser et al., 2017). Under a cue-based perspective, linguistic dependency formation relies on an associative cue-based retrieval mechanism (Lewis et al., 2006; McElree, 2006), but how such a mechanism is deployed over diverse linguistic dependencies remains a matter of debate. Are all linguistic features candidate cues that guide retrieval? Are all cues given similar weight? Are different cues differently weighted based on the dependency being processed? To address these questions, we implemented a *mismatch paradigm* (Sturt, 2003) adapted for temporal concord dependencies. This paradigm tested whether readers were sensitive to a temporal agreement between a temporal adverb like *last week* and a linearly distant, but structurally accessible verb, as well as a linearly proximate but structurally inaccessible verb. We found clear evidence that readers were sensitive to feature match between the adverb and the linearly distant, structurally accessible verb. We found no clear evidence on whether feature match with the inaccessible verb impacted the processing of a temporal adverb. Our results suggest syntactic positional information plays an important role during the processing of the temporal concord relation.

Keywords: tense, temporal adverbs, temporal concord, attachment, eye movements, memory retrieval, sentence comprehension

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Stephani Foraker,
University at Buffalo, United States
Dan Parker,
College of William & Mary,
United States

****Correspondence:***
Nicoletta Biondo
n.biondo@bcbl.eu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *12 December 2018*
Accepted: *15 April 2019*
Published: *31 May 2019*

Citation:
Biondo N, Vespignani F and Dillon B (2019) Attachment and Concord of Temporal Adverbs: Evidence From Eye Movements. Front. Psychol. 10:983. doi: 10.3389/fpsyg.2019.00983

INTRODUCTION

Sentences are routinely made of words whose formal properties need to covary in order to reach grammaticality. This relation among words, which has been generally called agreement or concord (Corbett, 2003), can involve several elements such as the subject noun phrase and the verb of a sentence (e.g., *The man is washing the car*), and/or the subject noun phrase and an anaphoric pronoun (e.g., *The man is washing himself*). In addition to being pervasive features of human language, concord phenomena have attractive properties for researchers investigating the interplay

between memory and sentence comprehension. For example, consider a sentence such as *the pasta recipe from the northern provinces tastes amazing*. Comprehending this sentence requires the reader to integrate the subject phrase, headed by *the pasta recipe,* with the verb *tastes*. Because these terms are not linearly adjacent in the input, this process plausibly requires memory retrieval: the comprehenders must encode the subject noun phrase, and have some mechanism for reactivating or retrieving the information in that encoding when it is needed, at a later point in processing (Lewis et al., 2006; McElree, 2006). This intuition lies at the heart of *cue-based parsing models*, which hypothesize that incremental sentence processing relies on a fast, associative, cue-based retrieval mechanism to reactivate linguistic encodings in memory when those encodings are necessary to parse or interpret the current input (for overviews, see Lewis et al., 2006; McElree, 2006; Foraker and McElree, 2011; Van Dyke and Johns, 2012; Wagers and McElree, 2013).

From this perspective, concord phenomena are useful to study in sentence processing, because the linguistic features marked on one element may provide important *retrieval cues* that can help comprehenders retrieve previously processed linguistic encodings. For example, the agreement morphology on the verb *tastes* in the example above might provide a (SING) feature that could be used to reactivate or retrieve the subject phrase *the pasta recipe* at the verb *tastes*. There are many empirical and theoretical questions raised by this hypothesis. Are all linguistic features that participate in agreement relations used as retrieval cues? Do all potential linguistic constraints belong to the set of cues that guide retrieval? If so, are all linguistic cues given similar weight, or do some specific cues, such as structural cues, have a leading role in the set of available cues used during memory retrieval (Van Dyke and McElree, 2006, 2011; Dillon et al., 2013; Patil et al., 2016; Parker and Phillips, 2017; Kush et al., 2018)? Researchers addressing these questions have largely focused on how comprehenders implement agreement and anaphoric dependencies in online comprehension and how different cues, such as structural cues and morphological cues, are differently weighted during the processing of these dependencies (Felser et al., 2017; see Jäger et al., 2017 for a comprehensive summary and meta-analysis).

In this paper, we try to extend the empirical basis of this literature by investigating a different and less typical concord phenomenon, namely the relationship between a deictic temporal adverb such as *last month* and its match with the temporal information expressed by the verb of the sentence. An example is given in (1):

(1) The postman who **used to work** in Yonville **delivered** a nice gift to me **last month**.

The processing of the adverb-verb temporal concord dependency is a good place to investigate the role of memory retrieval in syntactic processing. The successful attachment of a deictic temporal adverb such as *last month* in (1) would require finding a grammatically accessible verb phrase to modify (a *structural* constraint). Moreover, the adverb must express temporal information that is coherent with the temporal information expressed by the finite verb (a *morphosyntactic* constraint). Both these types of constraints - structural constraints that determine *where* the adverb can attach, and morphosyntactic constraints that determine *what temporal features* that attachment site must have - could plausibly be used as retrieval cues during memory retrieval in a cue-based parsing model. The morphological (i.e., temporal) cue provided by the adverb is triggered/available at the same time in which a structural cue is initiated (in order to find an appropriate structural placement). In addition, it seems very plausible that retrieval processes would be necessary to fully integrate a temporal adverb into an unfolding parse. This is because adverbs constitute optional constituents that cannot be reliably anticipated; the processing of an adverb might therefore not receive much facilitation due to the predictive computation of syntactic structure (e.g., the left-corner parsing framework in Lewis and Vasishth, 2005). It would thus seem that processing of temporal adverbs is ideal to study the interplay between structural and morphological cues during memory retrieval.

In this paper, we will first consider the processes necessary to integrate a temporal adverbial into a sentence. We will then turn to a consideration of how cue-based parsers realize these different processes. We then present two eye-tracking experiments that investigate the role of morphosyntactic and structural constraints on the processing of deictic temporal adverbs.

Adverb-Verb Attachment and Concord: Previous Studies

Each time a temporal adverb is encountered there are two potentially distinct processes that need to occur in order to reach a complete and coherent temporal interpretation of the event expressed in the sentence. One, a structural attachment site must be found to integrate the adverb into the syntactic structure. This attachment site is provided by the maximal projection of the phrase modified by the adverb (Chomsky, 1986, 1995; Sportiche, 1988 among others), such as the Temporal Phrase (TP) or the Verb Phrase (VP).[1] Two, a temporal feature match must be established between the deictic temporal adverb[2] and the tensed verb it modifies, in order to successfully define the temporal location of the event expressed by the verb. Existing experimental evidence suggests that both processes—attachment and concord—occur during the incremental processing of the adverb-verb relation.

[1]Other theoretical approaches posit a different syntactic position for adverbs, such as the specifier in a dedicated functional projection for tense (e.g., Alexiadou, 1997, 2000; Cinque, 1999, 2004). On this view, the temporal adverb is located in the specifier of the TP, and the tense in the head T. This local syntactic configuration then is what permits the two constituents to concord in temporal features and build up a coherent temporal interpretation of the event. However, the differences between these syntactic accounts are not crucial for the present study.

[2]Deictic temporal adverbs belong to a specific category of temporal adverbs that need to be anchored to the time of utterance (i.e., yesterday defines the 24-hour time interval preceding the time of utterance "now"). They differ from other adverbs such as clock-calendar adverbs (e.g., at noon, at 5 PM) or dependent adverbs (e.g., previously, afterward) that may be, or never are, anchored to the time of utterance (Smith, 1978, 1981). Deictic temporal adverbs and clock-calendar adverbs do not lead to similar temporal inconsistencies (e.g., I left/*will leave yesterday; I left/will leave at 5 PM).

Evidence concerning the attachment process of temporal adverbs in incremental comprehension comes from studies on syntactic ambiguity resolution. For example, Altmann et al. (1998) measured the reading times on a temporal adverb such as *next week* in syntactically ambiguous sentences (e.g., Fiona *implemented* the plan she *proposed next week*). Altmann et al. (1998) manipulated the temporal features of these two attachment sites to force the low attachment of the adverb (e.g., Fiona *implemented* the plan she *will propose next week*) or the high attachment of the adverb (e.g., Fiona *will implement* the plan she *proposed next week*). This eye-tracking study showed longer reading times (from early measures on) for the high attachment condition compared to the low attachment condition (see similar results in Van Gompel et al., 2005). In sum, there is evidence that low attachment of the temporal adverb was generally preferred (i.e., more easily processed) than high attachment. The low attachment preference has been related to general recency effects: the parser attaches the adjunct to the most recent and/or active verb phrase (e.g., MacDonald et al., 1994; Gibson et al., 1996). Alternatively, the low attachment preference has also been related to parsing principles such as the *Late Closure* principle (Frazier, 1979), or *Construal* (Frazier and Clifton, 1996) which holds that new phrases (e.g., adverbs) are preferably attached to the current phrase (or thematic domain) being processed.

Evidence that comprehenders evaluate temporal concord between a verb and a temporal adverb in real-time comes primarily from event-related potential (ERP) studies investigating the electrophysiological activity triggered by a grammatical violation during sentence processing. These studies have shown that a violation of the concord relationship between a deictic temporal adverb and the verb tense (e.g., *Yesterday I sailed/*sail*) yields ERP components characteristically associated with both syntactic and semantic anomaly detection. Relative to an acceptable baseline, sentences containing a verb that mismatches in temporal features with a deictic temporal adverb yields negative ERP deflections in early time windows (e.g., 300–500 ms) and positive ERP deflections in later time windows (e.g., 600–900 ms) after the verb onset. This is sometimes characterized as a LAN-P600 complex (Steinhauer and Ullman, 2002; Baggio, 2008), sometimes as an N400-P600 complex (Dillon et al., 2012; Qiu and Zhou, 2012). More recently, it has also been shown that when a temporal mismatch occurs between a deictic temporal adverb and a distal verb (e.g., *Yesterday afternoon the tired traveler *will come/came back home*) longer reading times are found compared to the correct control condition, both in early and late eye-tracking measures (Biondo et al., 2018).

Adverb-Verb Attachment and Concord: A Cue-Based Perspective

The evidence briefly reviewed above lends support to the idea that the parser needs to find a structurally appropriate attachment site for an adverb, and that it evaluates temporal concord between an adverb and a verb. Some parsing models treat sentence structure building (e.g., attachment) and the check of feature consistency as two independent and temporally ordered operations, potentially subserved by distinct processing mechanisms (e.g., Frazier, 1987; Friederici, 2002). The distinction between these different processes is less clear-cut in cue-based parsing models (e.g., Lewis and Vasishth, 2005). This is because both syntactic and morphological features can be used as retrieval cues that guide the memory retrieval processes necessary to integrate the adverb into the sentence. In this sense both constraints are "enforced" at the same time (i.e., at retrieval). More specifically, in a cue-based parser, concord and structural constraints are both used in tandem to retrieve a potential attachment site.

Is this a good model of how comprehenders process temporal adverbs? This is the central question of this paper. The cue-based model parsing makes several predictions, which we test in our experiments. Consider the sentences in (3). On our hypothesis about the processing of temporal adverbs, both temporal features and structural features will be used to retrieve an attachment site for the temporal adverbial *last week*. In (3b) there is only one potential attachment site that agrees in all features: the first verb, *taught*. In (3a), however, the syntactically inaccessible verb matches the temporal features of the adverb. This creates the possibility of *similarity-based interference* in the retrieval process (Lewis and Vasishth, 2005). Specifically, the Lewis and Vasishth model predicts *inhibitory interference* in these configurations: the presence of a feature-matched distractor verb *shocked* will slow down retrieval of the target verb *taught,* because in this configuration the feature-matched distractor reduces the amount of activation spread to the target encoding (Jäger et al., 2017). Note that this is only predicted if both tense features and structural features are used as retrieval cues during the processing of the temporal adverb.

(3) a. The musician **taught** the song [that **shocked** everyone] to his new bandmates **last week.**
b. The musician **taught** the song [that will shock everyone] to his new bandmates **last week.**

However, interference can sometimes be *facilitatory* (Jäger et al., 2017). Consider (4). In (4a), no verb agrees with the temporal features of the adverb. Because there is no item in memory that matches the features of the adverb, retrieval will be slow (or may fail), and processing is expected to be difficult. In (4b) however, the distractor verb now matches the tense features of the verb. Thus the target verb *will teach* matches the structural cues, and the distractor verb *shocked* matches the tense cues. This means that the processing of the adverb in (4b) is expected to be faster on average than (4a). This occurs because when there are two verbs that are equally well-matched to the retrieval cues, the overall time to identify a single attachment site is reduced (Jäger et al., 2017; see also Logačev and Vasishth, 2016).

(4) a. The musician will teach the song [that will shock everyone] to his new bandmates **last week.**
b. The musician will teach the song [that **shocked** everyone] to his new bandmates **last week.**

The Current Study

In the present study, we measured the processing of adverb-verb temporal coherence in sentences as the ones in **Table 1** where a structurally accessible attachment site (V1) and an

TABLE 1 | Experimental conditions of Experiment 1.

V1:match, V2:match	(a) The musician *taught* the song that *shocked* everyone to his new bandmates last week during the dress rehearsal.
V1:match, V2:mismatch	(b) The musician *taught* the song that *will shock* everyone to his new bandmates last week during the dress rehearsal.
V1:mismatch, V2:match	(c) The musician *will teach* the song that *shocked* everyone to his new bandmates last week during the dress rehearsal.
V1:mismatch, V2:mismatch	(d) The musician *will teach* the song that *will shock* everyone to his new bandmates last week during the dress rehearsal.

inaccessible attachment site (V2) matched or mismatched the temporal features of the temporal adverb. The resulting four experimental conditions are known as the *mismatch paradigm* (Sturt, 2003).

If comprehenders use only structural information to restrict the retrieval of an attachment site when processing the temporal adverb, then they should be only sensitive to the mis/match in temporal features between the temporal adverb and the structurally accessible verb V1. This should result in longer reading times for the *V1:mismatch* condition compared to the *V1:match* condition. Given that past eye-tracking studies investigating the attachment of temporal adverbs show attachment preferences from early measures on (e.g., Altmann et al., 1998; Van Gompel et al., 2005), we can expect the effect of *V1:match* to be visible from the first-pass to later measures.

Alternatively, if both structural and featural constraints are deployed during the processing of the temporal adverb, the presence of a distractor mis/matching the temporal features of the adverb should affect the retrieval of the licit attachment site. In particular, cue-based parsing models would predict two types of interference: an inhibitory interference effect, with longer reading times for the *V1:match,V2:match* condition when comparing the two *V1:match* conditions, and a facilitatory interference effect, with smaller reading time for the *V1:mismatch,V2:match* condition when comparing the two *V1:mismatch* conditions (Lewis and Vasishth, 2005). A graphic representation[3] of the four tested conditions, relative retrieval cues and predicted effects from a cue-based perspective is provided in **Table 2**.

EXPERIMENT 1

Methods

Participants

Thirty-five undergraduate students from the University of Massachusetts Amherst (31 female, mean age = 20 years, ranging from 18 and 21) participated in this experiment. Participants gave informed consent under an experimental protocol approved by the University of UMass Amherst Institutional Review Board and received course credit for their participation. They were all native speakers of English and had normal or corrected-to-normal vision. Given the absence of past studies addressing our research question, the selection of the sample size was based on past eye-tracking studies[4] investigating memory retrieval during sentence processing.

Materials

A sample of the experimental sentences is provided in **Table 1**. The experimental material consisted of 24 experimental sentences that were randomly assigned to different lists according to a Latin Square design, so that each subject could see only one version of each item set. Thus, each subject read 6 sentences in each of the four experimental conditions, in addition to 76 grammatical filler sentences (24 of this filler sentences contained a different manipulation that is not reported here). All sentences had the similar length (18–22 words) and the same syntactic structure. The main clause always contained a lexical subject and a ditransitive main verb in either the past tense form, or in the future with *will*. The matrix verb was always followed by two complements of the verb, respectively the direct object (e.g., *the song*) and the indirect object (e.g., *to his new bandmates*), and a temporal adverb followed by some continuations as prepositional phrases or locative adverbs. The embedded relative clause was always attached to the direct object of the main clause and consisted of the complementizer "that" and a past or future verb (e.g., *shocked*/*will shock*) occasionally followed by a direct object (e.g., everyone). The indirect object of the main verb (e.g., *to his new bandmates*) was always a prepositional phrase that followed the relative clause. In order to prevent the prepositional phrase from incorrectly attaching into the relative clause, the verb inside the relative clause was chosen to be syntactically incompatible with this specific prepositional phrase.

In each experimental condition the temporal specification of the deictic temporal adverb (the target word) was held constant; only the temporal features of the two preceding verbs were manipulated. Moreover, to be sure that the two temporal forms were not recognized as always leading to correct (e.g., past) or wrong (e.g., future) verb forms, the experimental material contained the 50% of items with past temporal adverbs (e.g., last month, yesterday) and the other 50% with future adverbs (e.g., next week, tomorrow).

Procedure

Eye-movements were recorded using an EYELINK 1000 eye-tracker, with a sampling rate of 1000 Hz. Participants had binocular vision while movements were measured, but only the right eye was tracked. A chin rest bar and a forehead restraint were provided for each participant to minimize head movements. Before the experiment, and whenever necessary during the experiment, the experimenter calibrated the eye-tracker asking participants to fixate nine positions indicated by a black dot, linearly distributed along the central line of the screen. The

[3]For the realization of this table we took inspiration from Figure 1 in Jäger et al. (2017)

[4]The search was conducted in the database Web of Science by using the following keywords "memory AND retrieval AND eye AND movements AND sentence." The list of papers published in the last 10 years was then integrated with the list of eye-tracking studies reported in the recent review by Jäger et al. (2017).

TABLE 2 | Graphic representation of the tested conditions and expected effects from a cue-based perspective (Lewis and Vasishth, 2005).

Condition	V1	V2	ADVERB	*Prediction*
(a) V1:match, V2:match	*Full match*	*Partial match*		
	+PAST	+PAST	+PAST	
	main clause domain	relative clause domain	main clause domain	Inhibitory
(b) V1:match, V2:mismatch	*Full match*	*No match*		interference
	+PAST	−PAST	+PAST	(a > b)
	main clause domain	relative clause domain	main clause domain	
(c) V1:mismatch, V2:match	*Partial match*	*Partial match*		
	−PAST	+PAST	+PAST	
	main clause domain	relative clause domain	main clause domain	Facilitatory
(d) V1:mismatch, V2:mismatch	*Partial match*	*No match*		interference
	−PAST	−PAST	+PAST	(c < d)
	main clause domain	relative clause domain	main clause domain	

The structural constraint is indicated by the main/relative clause domain. The morphological constraint is represented by +/−PAST. The highlighted cells indicate a match between the retrieval cues provided by the adverb and the target item V1/distractor item V2.

monitor was positioned 66.3 cm away from the participant, and three characters were subtended by each degree of visual angle. Sentences were presented in 11 point Monaco font via EyeTrack Software[5]. Participants initiated each trial by fixating on a black box on the left side of the screen, specifically where the first word of the sentence would have appeared. Once a fixation in the target region reached a stable value, the entire sentence was displayed, on one single line. After reading, participants ended the presentation of each sentence using one of the buttons of the response pad. Each sentence was followed by a comprehension question concerning the content of the sentence just read (e.g., *Who is going to learn the new song?*). Participants answered by pressing either one of two buttons placed on the response pad corresponding, respectively to the answer on the left (e.g., *The musician*) or on the right (e.g., *The bandmates*) of the screen. The experimental session was preceded by three practice trials to familiarize the participant with the procedure. Testing sessions lasted approximately 1 h, including practice, calibration, break and debriefing.

Data Analysis

Sentences were divided into nine regions as shown in (5) separated by the vertical pipe (|). The post-target area was divided in two regions (i.e., post-target, end of the sentence) to divide possible spill-over effects (Just et al., 1982; Mitchell, 1984) in the post-target area due to the experimental manipulation, from general wrap-up effects (Mitchell and Green, 1978; Just and Carpenter, 1980) generally visible at the end of the sentence. Eye-movements were analyzed in three regions of interest: the critical region (e.g., *last week*), the pre-critical region (e.g., *to his new bandmates*) and the post-target region (e.g., *during*).

(5) The musician | taught | the song that | shocked | everyone | to his new bandmates | last week | during | the dress rehearsal.

We report four measures for each region of interest. First, we analyzed *first-pass reading times*, defined as the sum of all fixations on a region of interest before leaving it either to the left or the right. We also analyzed *go-past times* (sometimes called *regression path duration*), defined as the sum of all fixations made once a region of interest has been fixated before moving to the right. Thus, go-past times include time spent re-reading previous regions in addition to the critical region itself. The last reading time measure we report is *total time*, which is the sum of all fixations made on a region of interest, including refixations made after the region has been exited to the right. In addition to these reading time measures, we also report the *probability of regression out*, that is the proportion of times a backward regression was made out of a given region.

Prior to statistical analysis, trials with track loss or blinks in first-pass reading at the critical region were excluded. In this experiment, only one participant was excluded from the analysis because of more than 25% of data loss. The remaining 34 participants (with less than 6% of missing data) reached at least 75% accuracy on the comprehension questions so no participants were excluded due to poor accuracy.

The analysis was carried out fitting linear mixed-effect models to our data, using the R package *lme4* (Bates et al., 2014) and the package *lmerTest* (Kuznetsova et al., 2017) which provides *p*-values in the summary table of each model. The models were built adding *V1:match* as fixed-effects factor, as well as two nested contrasts to test the effect of interference from the illicit distractor V2 both in the V1:match conditions (*c1*) and in the V1:mismatch conditions (*c2*), and crossed random intercepts and random slopes for all fixed-effect parameters both for subject and item grouping factors (Barr et al., 2013). In order to select a parsimonious model which was properly supported by the data, the complexity of the random effect structure of the maximal model was reduced by performing a principal component analysis (PCA; Bates et al., 2015). Only the principal components that were sufficient to cumulatively account for 100% of variance were included in the simplified model. Moreover, the correlation parameters were forced to zero, but only when this further simplification of the model did not significantly decrease the goodness of fit, according

[5]http://blogs.umass.edu/eyelab/software/

to a likelihood ratio test ($\alpha_{LRT} = 0.2$). The final structure of the best-fitting models is provided in **Appendix C**.

Our categorical fixed effects predictors were coded using sum-contrast coding. In *V1:match* [V1:match] = 1 and [V1:mismatch] = −1; in *c1*, [V1:match,V2:match] = 1, [V1:match,V2:mismatch] = −1 and [V1:mismatch] = 0; in *c2*, [V1:mismatch,V2:match] = 1, [V1:mismatch,V2:mismatch] = −1 and [V1:match] = 0. For the analysis of the probability of regression measure, logistic mixed-effect models were employed (Jaeger, 2008) using the same coding scheme. The Bonferroni correction was applied to correct the *p*-values for multiple comparisons (von der Malsburg and Angele, 2017). After this correction, a fixed effect was considered significant if its *p*-value was equal or smaller than 0.006.

Results

Bar plots of mean reading times and probability of regressions in each (pre-target, target, post-target) region are illustrated in **Figure 1** while numeric values are given in **Appendix A**. In **Table 3**, we report the estimated regression coefficient (Estimate), the standard error (SE) and *t*/Wald's *z* and *p*-values resulting from the linear mixed-effect model analysis on log-transformed reading times (Baayen and Milin, 2010), for each region.

Analyses on the target region revealed a significant effect of the *V1:match* fixed effect factor, in total reading times, while no significant effects were found in other regions (i.e., pre-target and post-target areas) or measures (i.e., first-pass, go-past, probability of regressions out of a region).

Discussion

In Experiment 1, reading times significantly increased when the adverb temporal features mismatched the tense features of the main verb of the clause (V1), in late measures (i.e., total time). We found no clear evidence of a significant modulation of the reading times on the adverb as a result of match to the tense features of the embedded verb (V2). However, we note that there is a non-reliable numerical trend that we observed in the go-past measure. At the critical region and spillover region, numerically longer mean go-past times were observed when both verbs matched the temporal features of the adverb. In the spillover region, numerically shorter go-past times were observed when the embedded verb V2 matched the tense features of the adverb.

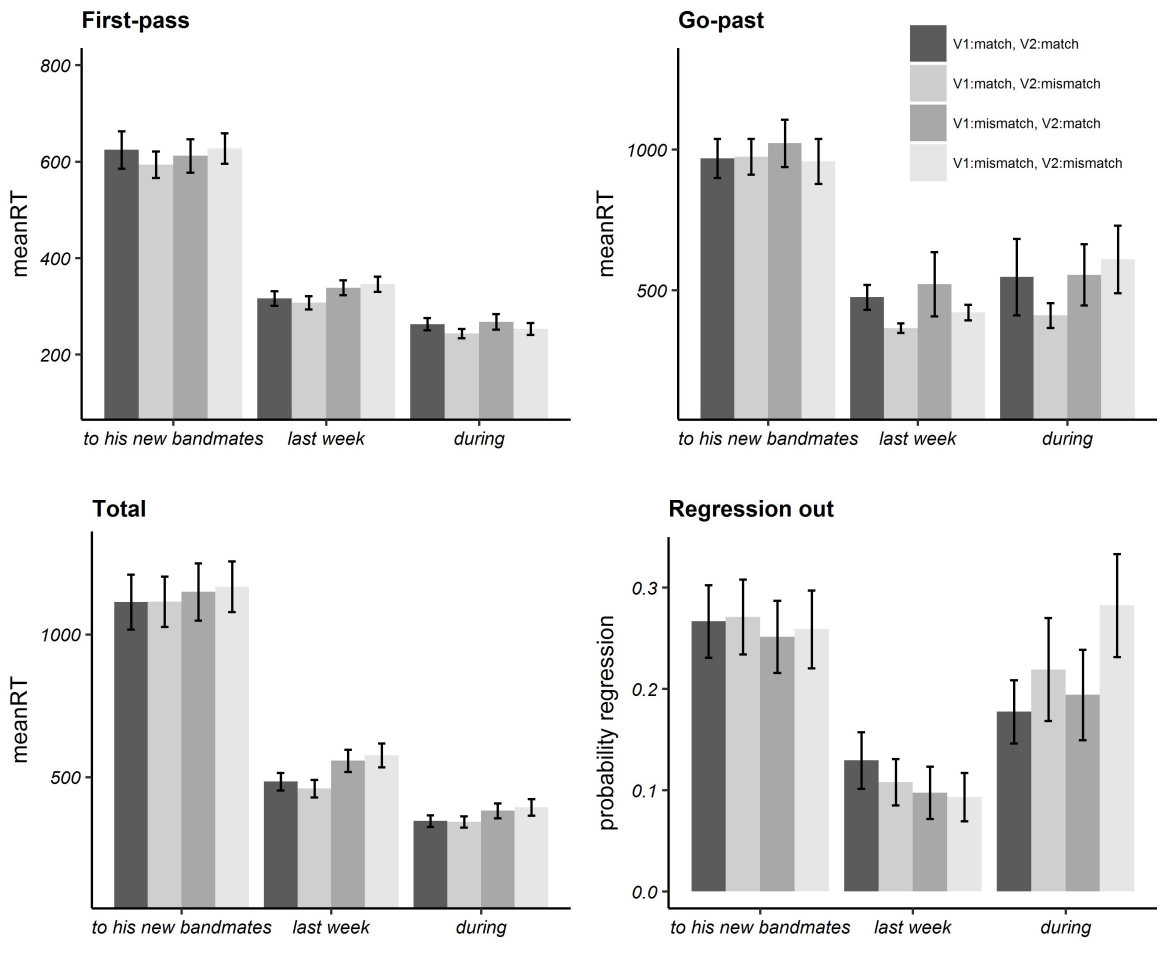


FIGURE 1 | Bar plots of mean reading times in milliseconds in eye-tracking latency measures and mean probabilities of regression out for Experiment 1. Error bars represent standard errors by participant.

TABLE 3 | Summary of LME analyses of log first-pass, go-past and total time, and probability of regression out for Experiment 1.

	to his new bandmates			*last week*			*during*		
First-pass	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	0.002 (0.02)	0.11	0.90	−0.04 (0.01)	−2.21	0.04	−0.004 (0.02)	−0.20	0.85
c1	0.01 (0.03)	−0.28	0.78	0.01 (0.02)	0.51	0.61	−0.027 (0.02)	1.23	0.22
c2	−0.02 (0.03)	−0.80	0.42	−0.01 (0.02)	−0.52	0.60	0.005 (0.02)	0.24	0.81
Go-past	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	0.006 (0.02)	0.28	0.78	−0.02 (0.01)	−1.04	0.30	−0.045 (0.03)	−1.45	0.16
c1	−0.009 (0.03)	−0.33	0.74	0.05 (0.03)	1.56	0.12	0.006 (0.04)	0.15	0.88
c2	0.004 (0.03)	0.15	0.89	−0.01 (0.03)	−0.37	0.72	−0.038 (0.04)	−0.87	0.38
Total	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	−0.01 (0.02)	−0.55	0.59	−0.08 (0.02)	−3.59	0.001	−0.04 (0.03)	−1.78	0.08
c1	−0.01 (0.03)	−0.41	0.69	0.02 (0.03)	0.59	0.56	0.001(0.04)	−0.04	0.97
c2	0.02 (0.03)	−0.94	0.35	−0.01 (0.03)	−0.51	0.61	−0.006 (0.04)	−0.24	0.81
Reg. out	**prop.**	***z***	***p***	**prop.**	***z***	***p***	**prop.**	***z***	***p***
V1:match	0.02 (0.09)	0.27	0.79	0.10 (0.13)	0.78	0.44	−0.10 (0.13)	−0.82	0.41
c1	−0.02 (0.13)	−0.15	0.88	0.11 (0.18)	0.64	0.52	0.11(0.18)	−0.59	0.56
c2	−0.03 (0.13)	−0.26	0.80	−0.07 (0.19)	−0.38	0.71	−0.22 (0.17)	−1.24	0.22

Standard errors are reported in parentheses.

These patterns may be consistent with an inhibitory interference effect and a facilitatory interference effect, respectively (Lewis and Vasishth, 2005, see also Jäger et al., 2017). However, neither of these trends was reliable; we return to these findings in Experiment 2 below.

The main finding from Experiment 1 is that comprehenders are primarily sensitive to the agreement between the temporal features of the adverb and of the matrix verb V1 in incremental sentence processing: reading times were slower when the temporal concord relationship was violated. We interpret these results as evidence that comprehenders retrieve the structurally licit attachment site for the adverb in incremental sentence processing in order to check temporal concord consistency, despite the fact that this verb phrase is linearly more distant than the more recent but more syntactically embedded verb phrase.

Still, the data from Experiment 1 leave open several questions. First, no significant effects were found in early measures (i.e., first-pass) while we observed apparent trends of a V2 match effect in go-past measures and a clear effect of V1 match in total reading times; this leaves open the question of how much interference V2 creates for the attachment of the temporal adverb, at least during sentence re-readings. Second, it is not clear if readers were confident of the appropriate attachment site of the indirect object PP that immediately preceded our temporal adverb, since inflated reading times were found at the PP region. In Experiment 2, we seek to address both of these open questions by replicating and extending our primary finding. We tested the same experimental material of Experiment 1 but added an extra-sentential context preceding each experimental sentence to actively disambiguate the attachment site of the pre-critical region.

TABLE 4 | Sample of the experimental material of Experiment 2.

V1:match, V2:match	Tell me more about the musician. To whom did he teach the song that shocked everyone? (a) The musician *taught* the song that *shocked* everyone to his new bandmates last week during the dress rehearsal.
V1:match, V2:mismatch	Tell me more about the musician. To whom did he teach the song that will shock everyone? (b) The musician *taught* the song that *will shock* everyone to his new bandmates last week during the dress rehearsal.
V1:mismatch, V2:match	Tell me more about the musician. To whom will he teach the song that shocked everyone? (c) The musician *will teach* the song that *shocked* everyone to his new bandmates last week during the dress rehearsal.
V1:mismatch, V2:mismatch	Tell me more about the musician. To whom will he teach the song that will shock everyone? (d) The musician *will teach* the song that *will shock* everyone to his new bandmates last week during the dress rehearsal.

EXPERIMENT 2

The goal of Experiment 2 was twofold. First, we wanted to pursue a replication of the primary finding of Experiment 1, namely that readers are primarily sensitive to the V1-adverb match during incremental processing. Second, we decided to extend the paradigm of Experiment 1 adding an extra-sentential context before each sentence, as shown in **Table 4**.

The goal of this manipulation was to use context to disambiguate the attachment of the prepositional phrase in the pre-critical region, in order to ensure that the effects observed in Experiment 1 were not contaminated by garden-pathing that may have occurred prior to the critical adverb.

We followed Altmann et al. (1998, experiment 2B), who used an interrogative context to guide the attachment of a temporal adverb in sentences such as "She*'ll implement* the plan he *proposed next week*, of course." In their experiment, the extra-sentential context was manipulated to either focus the temporal adverb *next week* and promote high attachment, e.g., *When will Fiona implement the plan she proposed? – She'll implement the plan* [*she proposed*] *next week, of course*, or to focus a complex noun phrase and therefore favor the low attachment of the temporal adverb, e.g., *Which of the plans she proposed will Fiona implement? – She'll implement the plan* [*she proposed next week*], *of course*.

In our study, we adopted this approach to clarify the attachment of the prepositional phrase *to his new bandmates* to the matrix clause. In our experimental sentences, the prepositional phrase was intended to attach to V1, but it is linearly positioned after the embedded verb V2. We cannot thus exclude that the parser could have been garden-pathed, and temporarily associated this prepositional phrase to V2, although the prepositional phrases were specifically chosen to be incompatible with V2, as outlined above; this could occur either as the result of a structural parsing principle such as *Late Closure* (Frazier, 1979), or as the result of a more general recency preference (MacDonald et al., 1994; Gibson et al., 1996). If the readers were garden-pathed in this fashion—temporarily associating the pre-critical PP to V2—then this could have partially masked the effect of our manipulation or otherwise interfered with the adverb attachment process that immediately follows the PP. This is especially true in rereading measures such as go-past duration: recall that in Experiment 1 we observed a numerical trend toward an interference effect from a structurally inaccessible attachment site. While not reliable, this trend raises the possibility that the V2 distractor matching the temporal cues of the adverb could in fact modulate reading times at the target region, at least in later measures.

To test whether our context manipulation effectively facilitated the interpretation of our experimental sentences, we added 18 filler sentences in which we manipulated the pre-sentential context (see **Appendix B** for a complete description of this study). The results of this manipulation indicated that the contexts we adopted in Experiment 2 did facilitate the reading of PP and adverb regions, in particular in rereading measures such as go-past and total reading time, thus minimizing any parsing difficulty that may have occurred prior to the critical region.

Methods

Participants

Forty-eight undergraduate students from the UMass Amherst participated in this experiment. They were all native speakers of English and had normal or corrected-to-normal vision, and none had taken part in Experiment 1. Participants gave informed consent under an experimental protocol approved by the UMass Amherst IRB and received course credit for their participation.

Materials

The materials in Experiment 2 followed the same design as Experiment 1. However, in Experiment 2 the critical sentences were preceded by a context whose role was to lead the readers expect a PP indirect object of the main verb V1, as shown in **Table 4**.

Participants were asked to read small dialogues in which there was a character A asking a question (i.e., the pre-sentential context) to a character B. The answer of character B represented the experimental sentence. The context had always the same structure, namely *Tell me more about X. To whom did/will he/she . . ..?* The first character introduced by the context (e.g., *the musician*) was also the subject/agent of the experimental sentence expressed by a pronoun (i.e., *he* or *she*), while the wh- phrase (i.e., to whom) of the pre-sentential context always referred to the PP of the experimental sentence (e.g., *to his new bandmates*). The experimental sentences provided the answer to a question posed in the extra-sentential context. Comprehension questions targeted information that could have been deduced from various parts of the sentence, aside from the prepositional phrase. Thus, only participants reading the entire target sentence were expected to achieve high comprehension accuracy.

As in Experiment 1, the experimental material consisted of 24 experimental sentences that were randomly assigned to different lists according to a Latin Square design, so that each subject read six sentences in each of the four experimental conditions, in addition to 76 filler sentences (58 simple filler sentences, and 18 filler sentences containing a context manipulation whose description and analysis is reported in **Appendix B**).

Procedure

The same facilities and calibration procedure of Experiment 1 were adopted for the follow-up experiment. However, the procedure for the presentation of the stimuli was different since each trial was composed by a context sentence, an experimental sentence and a comprehension question. Participants initiated each trial by reading the context sentence. After reading the context, participants proceeded to the reading of the experimental sentence using one of the buttons of the response pad. They were asked to fixate on a black box on the left side of the screen, specifically where the first word of the sentence would have appeared. Once a fixation in the target region reached a stable value, the sentence was displayed. After reading, participants ended the presentation of each sentence using one of the buttons of the response pad. Each sentence was followed by a comprehension question concerning the content of the sentence just read. Participants answered by pressing either one of two buttons placed on the response pad corresponding, respectively to the answer on the left or on the right of the screen. The experimental session was preceded by three practice trials to familiarize the participant with the procedure. Testing sessions lasted approximately 1 h, including practice, calibration, break and debriefing.

Data Analysis

All features of the analysis were identical to Experiment 1. In this experiment, five participants were excluded from the analysis because of more than 25% of data loss. The remaining 43 participants reached at least 75% accuracy on the comprehension questions; no participants were excluded due to poor accuracy.

Results

Bar plots of mean reading times and probability of regressions in each (pre-target, target, post-target) region are illustrated in **Figure 2** while numeric values are given in **Appendix A**. In **Table 5** we report the estimated regression coefficient (Estimate), the standard error (SE), *t*/Wald's *z* and *p* values resulting from the linear mixed effects model analysis on log-transformed reading times, for each region.

Analyses on the target region revealed a significant effect of the *V1:match* fixed effect factor in total reading time, while analyses on the post-target region revealed a significant effect of *V1:match* both in go-past and total reading time.

Discussion

The results of Experiment 2 present many similarities to, but some differences from Experiment 1. In Experiment 1 we observed inflated reading times on the target adverb when the adverb mismatched the tense features of the main verb of the clause (V1) in late measures (i.e., total time). In Experiment 2 we replicate the same pattern of results on the target region, together with an additional V1 match effect in late measures (i.e., go-past, total time) on the post-target region. Thus like Experiment 1, readers seemed to mainly consider the structurally accessible attachment site for the adverb and the match in features between the verb V1 and the adverb. Overall, the results of Experiment 2 largely confirm the general picture suggested by Experiment 1. Readers were primarily sensitive to the concord between the temporal adverb and the linearly distant, but structurally accessible V1, resulting in a significant effect of *V1:match*. This finding suggests that the parser reliably makes use of structural information to find the right attachment site for the temporal adverb.

Conversely, no statistically significant interference effects were found in the *V1:match* and *V1:mismatch* conditions. According to this analysis, there is not enough evidence to state that the processing of the adverb-verb relation can be modulated by the presence of a structurally illicit but feature matching verb phrase.

BAYESIAN ANALYSIS

The clearest result from both Experiments 1 and 2 is that comprehenders are sensitive to a match between the temporal adverb and the (structurally available) verb V1. In Experiment 1, this resulted in a significant reading time slowdown on the critical temporal adverb in total time measures; in Experiment 2, the slowdown was observed in go-past and total times at the spillover, in addition to total times at the critical region. From this, we can confidently conclude that comprehenders incrementally construct a dependency between the temporal adverb and the linearly distant, but structurally accessible, V1.

However, it is less clear whether there is any interference from the features of the grammatically inaccessible V2, as predicted by cue-based parsing models. To evaluate the strength of these findings, we performed a supplementary Bayesian analysis of our data from Experiments 1 and 2. Instead of asking the binary, categorical question "is there interference from V2, or not?" familiar from null hypothesis significance testing (NHST), the Bayesian approach we employ here allows us to ask instead the inherently gradient question "what is the strength of the evidence for interference from V2?" (Nicenboim and Vasishth, 2016).

For this analysis, we used the *rstanarm* package (Stan Development Team, 2016) to fit Bayesian linear mixed effects models to our data. Rather than adopting a parsimonious random effects structure as we did above, we fit 'maximal' random effects structures (i.e., varying intercepts and slopes for all fixed effects by subjects and items, along with their correlations; see Barr et al., 2013). This decision was made because maximal random effects structures can be fit without yielding unreasonable results in Bayesian analysis (Nicenboim and Vasishth, 2016). For models of reading times, log-transformed reading times were used as the dependent variable; for models of percent regressions out, we fit a logistic mixed effects model. For each model reported, we fit four Markov chain Monte Carlo (MCMC) chains of 2000 iterations each; unless otherwise noted, the convergence statistic *R-hat* was 1.0 for all parameters estimated. We used the default, weakly informative specification of the prior distributions on model parameters in *rstanarm*, with one exception: we followed Nicenboim and Vasishth (2016) in setting the regularization parameter on the covariance matrix to 2 to promote more conservative estimates of the intercept-slope correlations. In modeling the results in the spillover region of Experiment 2 we performed a prior sensitivity analysis to evaluate whether the choice of prior distribution substantially modified posterior estimates over parameter values (Nicenboim and Vasishth, 2016). We did not find that the choice of prior distribution had a substantial impact on our posterior estimates.

Table 6 summarizes the results as 95% credible intervals over parameter estimates for the models described above. Overall, there is a close alignment between these parameter estimates and those from the planned mixed effects model analysis. For example, in all the regions and measures where we found a statistically significant effect of *V1:match* in the planned linear mixed effects model analysis, we find that the credible intervals in our Bayesian model are quite far from overlapping with 0; we interpret this as evidence that there is clearly an effect of *V1:match* in our data. However, the strength of this Bayesian analysis lies not in making categorical decisions about the presence of absence of an effect, but instead, in quantifying the range of plausible values associated with that effect.

To aid in this interpretation of our Bayesian analysis, **Figure 3** presents a histogram of the posterior samples for models of total reading times at the critical region in both Experiments 1 and 2. As in the NHST analysis, the posterior distribution reveals clear evidence for a V1 match effect, such that reading times were slower when V1 mismatched the adverbial's temporal features.

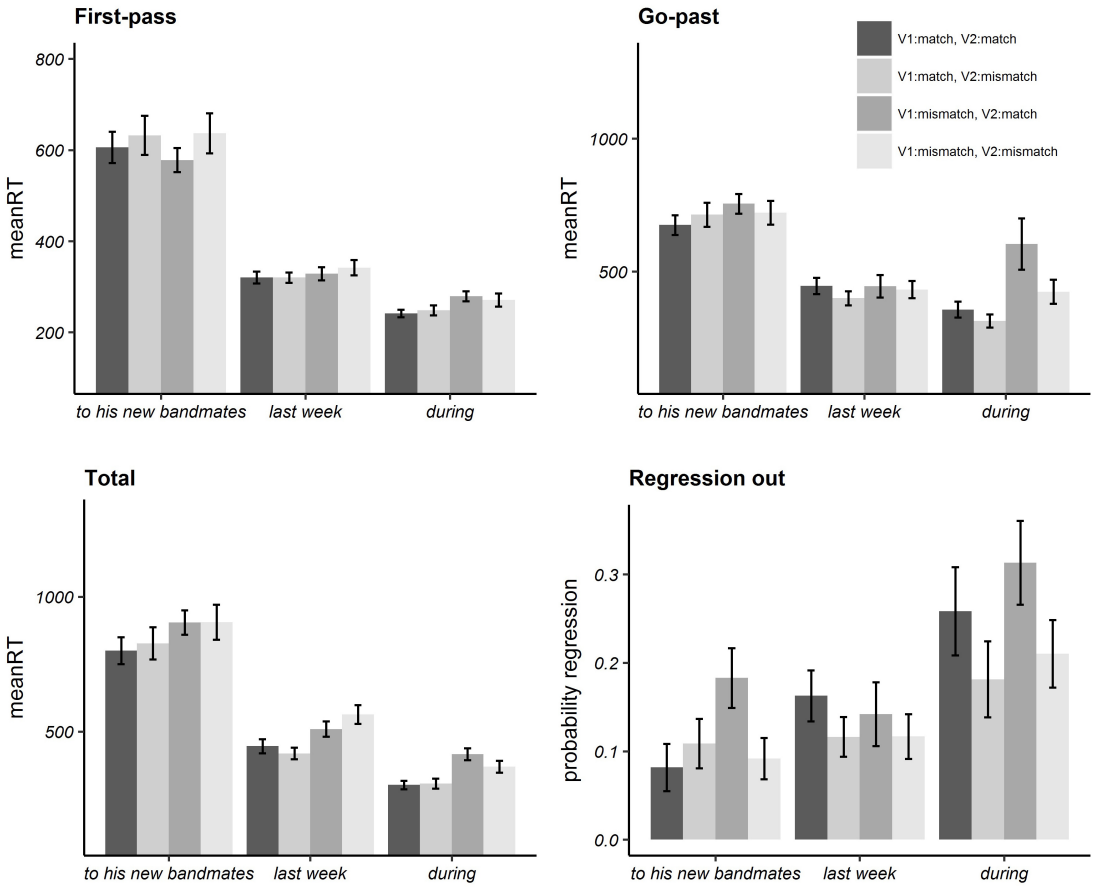


FIGURE 2 | Bar plots of mean reading times in milliseconds in eye-tracking latency measures and mean probabilities of regression out for Experiment 2. Error bars represent standard errors by participant.

TABLE 5 | Summary of LME analyses of log first-pass, go-past and total time, and probability of regression out for Experiment 2.

	to his new bandmates			*last week*			*during*		
First-pass	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	0.02 (0.02)	0.94	0.35	−0.03 (0.01)	−0.93	0.35	−0.05 (0.02)	−2.66	0.01
c1	−0.02 (0.02)	−0.77	0.45	−0.003 (0.02)	−0.17	0.87	−0.02 (0.02)	−0.72	0.47
c2	−0.05 (0.02)	−1.86	0.07	−0.02 (0.02)	−1.25	0.21	0.02 (0.02)	1.17	0.25
Go-past	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	−0.02 (0.01)	−1.25	0.21	−0.004 (0.02)	−0.23	0.82	−0.12 (0.03)	−4.26	0.0001
c1	−0.03 (0.02)	−1.37	0.17	−0.04 (0.03)	1.30	0.19	0.04 (0.03)	1.13	0.26
c2	0.02 (0.02)	0.76	0.45	−0.01 (0.03)	0.15	0.88	0.06 (0.03)	1.66	0.10
Total	**logRT**	***t***	***p***	**logRT**	***t***	***p***	**logRT**	***t***	***p***
V1:match	−0.04 (0.02)	−2.81	0.01	−0.09 (0.02)	−5.20	2.52e-07	−0.12 (0.02)	−5.73	6.18e-07
c1	−0.02 (0.03)	−0.63	0.53	0.02 (0.02)	0.74	0.46	−0.01 (0.03)	−0.25	0.81
c2	0.01 (0.02)	0.66	0.51	−0.04 (0.02)	−1.60	0.11	0.06 (0.03)	2.43	0.02
Reg. out	**prop.**	***z***	***p***	**prop.**	***z***	***p***	**prop.**	***z***	***p***
V1:match	−0.25	−1.74	0.08	0.07 (0.13)	0.55	0.58	−0.20 (0.11)	−1.86	0.06
c1	−0.17	−0.88	0.38	0.24 (0.15)	1.58	0.11	0.37 (0.16)	2.25	0.03
c2	0.42	2.65	0.01	0.11 (0.16)	0.71	0.48	0.09 (0.15)	0.60	0.55

TABLE 6 | Summary of Bayesian mixed effects analysis of critical fixed effects.

	Last week		*During*	
	E1	**E2**	**E1**	**E2**
First-pass				
V1:match	−0.04 [−0.08,0.00]	−0.01 [−0.04,0.02]	0.00 [−0.04,0.04]	−0.05 [−0.08,−0.01]
c1	0.01 [−0.03,0.06]	0.00 [−0.05,0.04]	0.03 [−0.02,0.08]	−0.02 [−0.06,0.03]
c2	−0.01 [−0.06,0.04]	−0.02 [−0.07,0.02]	0.01 [−0.04,0.06]	0.02 [−0.02,0.07]
Go-past				
V1:match	−0.02 [−0.07,0.02]	0.00 [−0.05,0.04]	−0.05 [−0.12,0.03]	−0.12 [−0.18,−0.06]
c1	0.05 [−0.02,0.11]	0.03 [−0.03,0.09]	0.01 [−0.09,0.11]	0.04 [−0.04,0.11]
c2	−0.01 [−0.07,0.06]	0.00 [−0.06,0.07]	−0.04 [−0.13,0.06]	0.06 [−0.03,0.14]
Total				
V1:match	−0.08 [−0.13,−0.03]	−0.09 [−0.13,−0.05]	−0.03 [−0.08,0.01]	−0.11 [−0.16,−0.07]
c1	0.02 [−0.05,0.08]	0.02 [−0.04,0.07]	0.00 [−0.06,0.06]	−0.01 [−0.07,0.05]
c2	−0.01 [−0.08,0.03]	−0.04 [−0.10,0.02]	−0.01 [−0.07,0.06]	0.06 [0.00,0.12]
Reg. out				
V1:match	0.12 [−0.16,0.42]	0.08 [−0.17,0.34]	−0.09 [−0.40,0.21]	−0.22 [−0.51,0.04]
c1	0.12 [−0.26,0.51]	0.24 [−0.07, −0.58]	−0.09 [−0.53,0.33]	0.41 [0.01,.85]
c2	−0.08 [−0.51,0.34]	0.11 [−0.26,0.48]	−0.23 [−0.66,0.19]	−0.24 [−0.66,0.48]

We report 95% credible interval of model estimates, rounded to the nearest two decimal places. We report analyses of first-pass, go-past and total time, and probability of regression out for critical region and spillover for both Experiments 1 and 2. Boxes indicate effects of particular interest discussed in text.

The picture becomes more interesting when we consider the posterior distribution for the V2 match effect for grammatical sentences (i.e., *V1:match* conditions) in total times at the critical region. On our NHST analysis, this coefficient did not reach statistical significance; correspondingly, the 95% credible intervals in our Bayesian analysis clearly include 0.

However, there are some aspects of our data that may be construed as weak evidence in favor of an inhibitory interference effect of the sort predicted by cue-based parsing models. First, in both experiments, in almost all (13/16) measures the mean effect of the c1 model coefficient is positive, the predicted direction (Jäger et al., 2017). Second, when we compare the posterior distribution for this parameter with the results of Jäger et al. (2017) meta-analysis on interference for subject-verb dependencies, we find that their 95% CI (see dotted lines in **Figure 3**) aligns with the region of highest density in our parameter estimates. Third, E1 and E2 are in almost complete agreement about the range of plausible values for this parameter in total time measures. Fourth, and finally, we note that there is some evidence that there is a positive value for this coefficient for regressions out in the spillover region of E2. In other words, based on the Bayesian analysis, our data seem to show some evidence for a V2 match effect in grammatical sentences of a magnitude comparable to that observed for inhibitory interference effects for subject-verb dependencies (Jäger et al., 2017). What about a V2 match effect for ungrammatical sentences (i.e., in *V1:mismatch* conditions)? Do we find clear evidence for the predicted facilitatory match effect of a V2 match in this condition? The Bayesian analysis presents less compelling evidence that this is the case. First, there is somewhat less consistency in the direction of this parameter: 10/16 parameter estimates go in the predicted, facilitatory direction. Second, when we do find evidence that the 95% credible interval for this parameter does not include zero (total times in the spillover for E2), the estimated effect goes opposite the predicted direction: we see inhibitory interference (but see Jäger et al., 2017, on a similar pattern observed for reflexive-antecedent dependencies). Third, there is somewhat less cross-experiment consistency in the estimates of this parameter.

It should be noted, however, that this discussion about the presence/absence of a V2 match effect both in grammatical and ungrammatical conditions can be only speculative in this context. The posterior distributions here discussed are too wide and compatible with a wide range of possible results. In order to safely conclude that there is evidence for interference or not, estimates with higher precision than the one presented here (i.e., smaller confidence intervals obtained through higher statistical power) would be needed. In all, our results do not allow us to clearly conclude that there is no effect of V2 match, nor do they clearly allow us to conclude that there is evidence for the (predicted) interference effect of V2 match.

GENERAL DISCUSSION

The main aim of this study was to expand the investigation of concord phenomena in order to understand how linguistic dependencies are processed during sentence comprehension. As we have argued, concord phenomena are useful to study in sentence processing, because the linguistic features marked on one element may provide important cues that can help comprehenders retrieve previously processed linguistic encodings.

In this study, we investigated a different and less typical concord phenomenon, namely the adverb-verb temporal concord relation. We posed some of the questions that are still

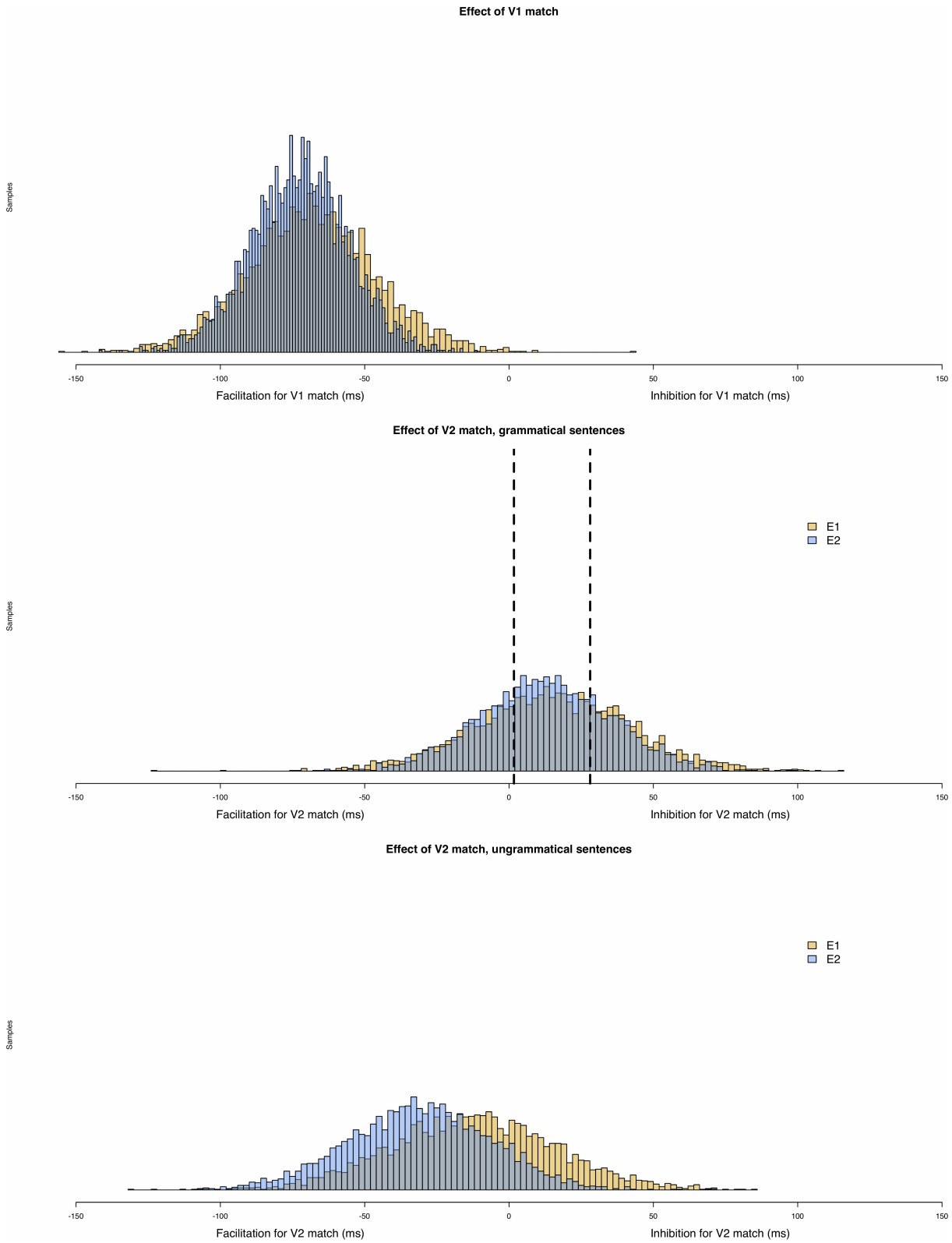


FIGURE 3 | Histogram of the posterior samples for models of total reading times at the critical region in Experiments 1 and 2. The first graph represents the posterior distribution for the V1:match effect, while the second and the third graph represents, respectively the effect of V2:match in *V1:match* conditions (c1) and in *V1:mismatch* conditions (c2).

debated within a cue-based perspective: are all linguistic features candidate cues that guide retrieval? Are all cues given similar weight? Are different cues differently weighted based on the dependency being processed?

We ran two eye-tracking studies in which we tested the processing of deictic temporal adverbs such as *last month* and their dis/agreement in features with two antecedent tensed verbs. Participants read sentences in which a deictic adverb such as *last week* could either agree or disagree in temporal features with a structurally accessible verb (V1) and/or a structurally inaccessible verb (V2) in temporal features. We wished to investigate to what degree the retrieval mechanisms implied during the processing of a temporal adverb are sensitive to structural and/or featural constraints in incremental sentence processing. We expected comprehenders to mainly show sensitivity to the V1-adverb match in the case in which structural information is used to process the adverb-verb relation at long distance. This should have resulted in a main effect *V1:match*. Conversely, we expected reading times to be modulated by the V2-adverb match if both structural and featural information are jointly deployed during the processing of the adverb-verb relation. In this case, two possible interference effect patterns would have been predicted by cue-based parsing models: inhibitory interference in *V1:match* conditions and facilitatory interference in *V1:mismatch* conditions.

Our results reveal two main findings. First, Experiment 1 and Experiment 2 showed that readers were sensitive to the temporal concordance between the adverb and the structurally accessible verb of the main clause V1. When the tense features of V1 mismatched the features of the adverb, longer RTs were observed on the adverb itself in late measures[6] (i.e., total time) in Experiment 1 and in Experiment 2, and on the word following the adverb (in go-past, total time) in Experiment 2. This pattern of results is further supported by an additional Bayesian analysis which shows that the credible intervals for the effect of *V1:match* do not overlap with 0 exactly in the same regions and measures described above, as well as in the first-pass measure of the post-target region, in Experiment 2.

Second, we did not find unambiguous evidence of interference from the structurally inaccessible verb phrase (V2) in any region or measure of the two experiments, neither in the linear effect model analysis nor in the parameter estimates of the Bayesian analysis. This fails to provide evidence in favor of the claim that both structural and featural cues are used to retrieve a verb to associate the adverb with, as predicted by cue-based parsing models (e.g., Lewis and Vasishth, 2005). However, this failure to find evidence cannot be taken as strong evidence against this view: the low precision of the parameter estimates provided by our Bayesian analysis does not allow us to definitively conclude that an interference effect from V2 is either present or absent. Further studies aiming at increasing the precision of the estimates (e.g., via higher statistical power) are necessary to confidently answer this question. It may be that, the effect size of V2 interference effect (unlike the V1 match effect) is too weak to be detectable in our data without a very large sample size (in the order of hundreds of participants; for a recent discussion on similar topics see Vasishth et al., 2018).

What can be safely concluded from these data is that readers do consider the structurally accessible attachment site (V1) during the processing of the temporal adverb, despite the fact that it is neither the most recent verb, nor the most linearly proximate. In this sense, we may conclude that structural cues guide the processing of the temporal adverb phrase. In what follows we take up some remaining questions from our study, as well as situate our findings on processing temporal adverbs in a broader theoretical context.

Relating Classical and Cue-Based Approaches to Processing Adverbials

In this paper, we have approached the problem of the attachment and concord between temporal adverbs and verbs from the perspective of cue-based parsing models. It is interesting to consider our theoretical conclusions in light of the broader literature on attachment and concord provided by other psycholinguistic models of sentence parsing.

As reported in the introduction, much of the experimental work on the processing of temporal adverbs has focussed on the processing of syntactically ambiguous sentences such as *John sold the guitar that he found on the beach last week.* There is general consensus in considering the attachment of the adverb to the second verb of the sentence because of general recency effects (e.g., MacDonald et al., 1994) or because of specific parsing principles such as *Late Closure* principle (Frazier, 1979) or *Construal* (Frazier and Clifton, 1996). In light of these earlier claims, our failure to find a strong interference effect from the most recent and linearly closer attachment site V2 may appear surprising. However, this discrepancy is only superficial.

In our study the most recent and linearly closer attachment site does not head the most current thematic domain or argument structure; V1 does. This is because the last phrase which is encountered before attaching the adverb is one of the arguments of the main verb V1 (i.e., the indirect object). Because comprehenders were overwhelmingly sensitive to a V1 match in our data, our findings show that the availability of an attachment site is not gated by simple recency or linear proximate of an attachment site, but it is gated by syntactic structure.

If this conclusion is correct, then one important question that remains is exactly how the parser determines what encodings in memory constitute structurally available attachment sites for the adverb (see Kush, 2013, for an extended discussion

[6]Previous eye-tracking studies have shown significant effects of adverb attachment/concord from early measures on (e.g., Altmann et al., 1998; Van Gompel et al., 2005). One possible explanation of this discrepancy could be methodological. We adjusted alpha for multiple comparisons while the aforementioned studies did not. Indeed, without the alpha adjustment our data would have also given a significant effect of *V1:match* in first-pass (see first-pass of the target region in Experiment 1, first-pass of the post-target region in Experiment 2).

of the theoretical issues). We are not in a position to offer a definitive answer to this question, but we see this as an exciting area of future research. One possible implementation of this idea is to borrow Construal's claim that the current thematic domain is what defines which attachment sites are syntactically available for the temporal adverb. This could perhaps be implemented as a retrieval cue that matches material in a current thematic domain. Such a model would integrate Construal's claim about what constitutes a licit attachment site (the current thematic/syntactic domain), with the cue-based retrieval mechanism for forming attachment relations in a model such as ACT-R (Lewis and Vasishth, 2005). This possibility is highly speculative at present, but this perspective could provide a useful avenue for further addressing the interplay between cue-based and structured processing models.

There is one important limitation of our study that bears further discussion. The type of sentence structure we adopted in this study allowed us to disentangle simply recency from structural factors in the investigation of adverb attachment. However, an open question is whether the same sentence structure was ideal to test interference effects from the illicit distractor. In the framework of a model such as the one proposed by Lewis and Vasishth (2005), the processing of the indirect object of the main verb of the sentence (i.e., to his new bandmates) may reactivate and strengthen the encoding of V1 or its associated verb phrase (Lewis et al., 2006; Vasishth and Lewis, 2006). This is because this constituent is still a dependent of V1, even if it is not directly involved in the long-distance temporal concord dependency. If this line of reasoning is correct, the increased activation of V1 (or its associated verb phrase) could have diminished the strength of any interference effect from V2, because V2 would be relatively less active. This line of reasoning is consistent with the presence of some residual/weak interference in the Bayesian analysis. This is a general design issue for studies looking at interference effects that bears closer scrutiny, since models such as Lewis and Vasishth's predict that covert reactivation of constituents boosts their activation in working memory. For example, the retrieval of the licit antecedent of a reflexive pronoun has been investigated in sentences such as "[*The surgeon* [*who treated Jennifer*] *had pricked himself*]. . ." (example provided by Jäger et al., 2017 in their meta-analysis, taken from Sturt (2003)). Based on the reactivation account, the VP (*had pricked*) of the sentence can have, in principle, re-activated its argument (i.e., the licit NP *the surgeon*), thus lowering the activation of the illicit distractor (*Jennifer*) before encountering the reflexive pronoun. Dillon et al. (2013) explored such a 'reactivation-based' account of the diminished reflexive intrusion effect but argued using computational simulations that the reactivation boost was not sufficient to predict the observed lack of intrusion effects. Dillon et al.'s (2013) results suggest that the reactivation of the target can diminish interference effects from the distractor, although in those simulations it did not totally eliminate those interference effects. It is difficult to compare the results of those simulations too directly to our present materials, although they do provide a proof of concept that the concern about target reactivation is well-placed, although in those simulations it did seem that interference from the distractor was still predicted. In general, it is difficult to reason about the impact of this reactivation process without the aid of an implemented computational model, and it is beyond the scope of the present project to simulate this process.

Interference Across Dependencies

We have highlighted the fact that in our data, comprehenders were clearly sensitive to a V1 match. We have interpreted this as evidence for the immediate application of structural constraints in the selection of an attachment site.

One important motivation for our study was to broaden the empirical base on which cue-based models are founded and include a novel linguistic relation: temporal adverb – verb dependencies. How does the processing of these adverbs compare to other dependencies that have been studied in this literature? Recent evidence has shown that the presence and the direction of the interference effect, both in grammatical and ungrammatical conditions, is not consistent across dependencies (Jäger et al., 2017). With respect to temporal adverbs, our Bayesian analysis has shown that the most compelling evidence for interference from an illicit attachment site seems to arise only in the grammatical conditions, in the form of inhibitory interference. There was not clear evidence for facilitatory interference in the ungrammatical conditions. This pattern should be supported by stronger experimental evidence in order to be extensively discussed. However, if we speculatively take this as the interference profile for temporal adverbs, then this sets them apart from both reflexive and subject-verb agreement dependencies. These dependencies show, respectively, small or strong evidence for interference (facilitatory in the case of subject-verb agreement, inhibitory in case of reflexives), but only in the ungrammatical conditions. Temporal adverbs seem more similar to subject-verb dependencies, which show an inhibitory effect in the grammatical conditions (Jäger et al., 2017). A more complete comparison between subject-verb attachment and verb-adverb attachment is not possible at present, however, because interference on subject-verb attachment has not been tested in ungrammatical conditions. Thus, further research is needed to show whether this dependency fully matches the results we found for adverb attachment, or whether the subject-verb dependency would show a reliable facilitatory effect in the ungrammatical conditions, as predicted by Lewis and Vasishth's (2005) model.

Our finding that structural constraints are immediately applied in the processing of temporal adverbs are also partially in line with some recent proposals that draw a link between the priority that structural constraints can have during retrieval and the predictability of the dependency under computation. In particular, Parker and Phillips (2017) proposed that the relative unpredictability of reflexives may be the source of their structure sensitivity found during the processing of the reflexive dependency, especially as compared to the processing of subject-verb agreement dependencies. Similarly, we could extend this

proposal to the processing of temporal adverb attachment. It may be that the strong role played by structural constraints during the processing of the adverb can be related to the unpredictability and optionality of the adverb constituent.

However, reflexives and temporal adverbs are unpredictable in somewhat different ways that complicate this comparison. While a reflexive and its linguistic features cannot be predicted, *per se,* the structural position of a reflexive *can* be predicted: this is typically the direct object position of a verb phrase, and this position can be easily predicted, at least for verbs with a specific subcategorization frame (e.g., transitive verbs). Conversely, the attachment site of the temporal adverb is not predictable and needs to be built "from scratch" once the adverb is encountered. In other words, for reflexives attachment processes may be selectively facilitated by predictive processing, leaving only concord and binding processes to be resolved through memory retrieval. But in the case of temporal adverbs, neither attachment nor concord processes are likely to benefit from predictive processes, as we have emphasized. It is currently hard to imagine whether this more subtle difference plays a role in the pattern of findings provided by the current experimental literature. Further research is primarily needed in order to better assess to which extent the attachment of temporal adverbs is prone to interference and to confirm or disconfirm the pattern of results provided by the current study.

Another potential—but at present speculative—explanation is that the attachment of a temporal adverb fails to show the complete predicted pattern of interference because of the specific sentential context we adopted in our study. In this study we tested adverb attachment in isolated sentences or in presence of a context whose role was to clarify the interpretation of the pre-target constituent. By varying syntactic configurations, or by varying the amount of distractor to target match in the items, it may still be possible to observe clearer inhibitory and facilitatory interference effects during adverb attachment by adopting different experimental designs that raise the salience of the distractor verb. This possibility seems particularly plausible if we consider experimental findings on the processing of reflexives in detail. Some studies on the processing of reflexives failed to find interference effects from illicit distractors (e.g., Nicol and Swinney, 1989; Sturt, 2003; Xiang et al., 2009; Dillon et al., 2013; Cunnings and Sturt, 2014), though more recent studies have shown that the prominence of the illicit distractor or the degree of feature match of the licit antecedent can increase the strength of interference effects during the processing of reflexives (Patil et al., 2016; Parker and Phillips, 2017; see also Sloggett, 2017). In other words, we cannot exclude that other factors, including non-syntactic ones, may increase the sensitivity to syntactically illicit attachment sites during the processing of a temporal adverb. Further research is needed to identify which factors may boost interference effects during adverb attachment.

CONCLUSION

The central question of the current study was whether comprehenders use syntactic positional information and/or (temporal) featural information to process the adverb-verb temporal concord relation at long distance during sentence processing. In two eye-tracking studies, we found consistent evidence that comprehenders use structural information to determine the attachment site for the temporal adverb and process the concord relation. Further research is needed to establish whether other, non-structural factors may also play a role during the processing of the adverb-verb temporal relation.

ETHICS STATEMENT

The protocol (nr. 2015–2661) was approved by the UMass Amherst Institutional Review Board. All participants gave written informed consent.

AUTHOR CONTRIBUTIONS

NB wrote the original draft and was responsible for the conception and the design of the work, as well as for the acquisition, analysis and interpretation of the data. FV contributed to the original draft and to the interpretation of the data. BD contributed to the original draft, as well as to the design, to the analysis and interpretation of the data.

ACKNOWLEDGMENTS

We would like to thank Chuck Clifton, Lyn Frazier, and Adrian Staub for their feedback during one of the UMass Psycholinguistics Workshops. We would also like to thank Amanda Doucette and Anthony Yacovone for their assistance in collecting the eye-tracking data. We would like to acknowledge that this study has been first described in NB's doctoral thesis (Biondo, 2017). Part of this research was also presented at the (refereed) conference "29th Annual CUNY Conference on Human Sentence Processing" (Gainesville, Florida, March 4, 2016).

SUPPLEMENTARY MATERIALS

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00983/full#supplementary-material

REFERENCES

Alexiadou, A. (1997). *Adverb Placement: A Case Study in Antisymmetric Syntax*, Vol. 18. Amsterdam: John Benjamins Publishing.

Alexiadou, A. (2000). On the syntax of temporal adverbs and the nature of Spec, TP: 1888. *Rivista di Lingüística* 12, 55–76.

Altmann, G. T., van Nice, K. Y., Garnham, A., and Henstra, J. A. (1998). Late closure in context. *J. Mem. Lang.* 38, 459–484. doi: 10.1006/jmla.1997.2562

Baayen, H. R., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28.

Baggio, G. (2008). Processing temporal constraints: an ERP study. *Lang. Learn.* 58, 35–55. doi: 10.1002/9781444309645.ch4

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. arXiv:1506.04967 [Preprint].

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv:1406.5823 [Preprint].

Biondo, N. (2017). *The Syntactic Side of Time: Processing Adverb-Verb Temporal Agreement*. Doctoral dissertation, University of Trento, Trento, TN.

Biondo, N., Vespignani, F., Rizzi, L., and Mancini, S. (2018). Widening agreement processing: a matter of time, features and distance. *Lang. Cogn. Neurosci.* 33, 890–911. doi: 10.1080/23273798.2018.1446542

Chomsky, N. (1986). *Barriers*, Vol. 13. Cambridge, MA: MIT press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT press.

Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.

Cinque, G. (2004). Issues in adverbial syntax. *Lingua* 114, 683–710. doi: 10.1016/s0024-3841(03)00048-2

Corbett, G. G. (2003). "Agreement: terms and boundaries," in *Proceedings of the 2001 Texas Linguistic Society Conference of the Role of Agreement in Natural Language*, Austin, 109–122.

Cunnings, I., and Sturt, P. (2014). Coargumenthood and the processing of reflexives. *J. Mem. Lang.* 75, 117–139. doi: 10.1016/j.jml.2014.05.006

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Dillon, B., Nevins, A., Austin, A. C., and Phillips, C. (2012). Syntactic and semantic predictors of tense in Hindi: an ERP investigation. *Lang. Cogn. Process.* 27, 313–344. doi: 10.1080/01690965.2010.544582

Felser, C., Phillips, C., and Wagers, M. (2017). Encoding and navigating linguistic representations in memory. *Front. Psychol.* 8:164. doi: 10.3389/fpsyg.2017.00164

Foraker, S., and McElree, B. (2011). Comprehension of linguistic dependencies: speed-accuracy tradeoff evidence for direct-access retrieval from memory. *Lang. Linguist. Comp.* 5, 764–783. doi: 10.1111/j.1749-818x.2011.00313.x

Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. Doctoral dissertation, University of Connecticut, Connecticut.

Frazier, L. (1987). "Theories of sentence processing," in *Modularity in Knowledge Representation and Natural-Language Processing*, ed. J. Garfield (Cambridge, MA: MIT Press), 291–307.

Frazier, L., and Clifton, C. (1996). *Construal*. Cambridge, MA: Mit Press.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/s1364-6613(00)01839-8

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., and Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition* 59, 23–59. doi: 10.1016/0010-0277(95)00687-7

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339. doi: 10.1016/j.jml.2017.01.004

Just, M., Carpenter, P., and Woolley, J. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol. Gen.* 111, 228–238. doi: 10.1037//0096-3445.111.2.228

Just, M. A., and Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychol. Rev.* 87, 329–354. doi: 10.1037//0033-295x.87.4.329

Kush, D., Johns, C. L., and Van, J. D. (2018). Prominence-sensitive pronoun resolution: new evidence from the speed-accuracy tradeoff procedure. *J. Exp. Psychol. Learn. Mem. Cogn.* [Epub ahead of print].

Kush, D. W. (2013). *Respecting Relations: Memory Access and Antecedent Retrieval in Incremental Sentence Processing*. Doctoral dissertation, University of Maryland, College Park.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). LmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26.

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

Logačev, P., and Vasishth, S. (2016). A multiple-channel model of task-dependent ambiguity resolution in sentence comprehension. *Cogn. Sci.* 40, 266–298. doi: 10.1111/cogs.12228

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101:676. doi: 10.1037//0033-295x.101.4.676

McElree, B. (2006). Accessing recent events. *Psychol. Learn. Motiv.* 46, 155–200. doi: 10.1016/s0079-7421(06)46005-9

Mitchell, D. C. (1984). "An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading," in *New Methods in Reading Comprehension Research*, eds D. Kieras and M. A. Just (Hillsdale, NJ: Erlbaum), 69–89.

Mitchell, D. C., and Green, D. W. (1978). The effects of context and content on immediate processing in reading. *Quart. J. Exp. Psychol.* 30, 609–636. doi: 10.1037/xlm0000526

Nicenboim, B., and Vasishth, S. (2016). Statistical methods for linguistic research: foundational Ideas—Part II. *Lang. Linguist. Comp.* 10, 591–613. doi: 10.1111/lnc3.12207

Nicol, J., and Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *J. Psychol. Res.* 18, 5–19. doi: 10.1007/bf01069043

Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002

Patil, U., Vasishth, S., and Lewis, R. L. (2016). Retrieval interference in syntactic processing: the case of reflexive binding in English. *Front. Psychol.* 7:329. doi: 10.3389/fpsyg.2016.00329

Qiu, Y., and Zhou, X. (2012). Processing temporal agreement in a tenseless language: an ERP study of Mandarin Chinese. *Brain Res.* 1446, 91–108. doi: 10.1016/j.brainres.2012.01.051

Sloggett, S. (2017). *When Errors Aren't: How Comprehenders Selectively Violate Binding Theory*. Doctoral dissertation, University of Massachusetts Amherst, Amherst, MA.

Smith, C. S. (1978). The syntax and interpretation of temporal expressions in English. *Linguist. Philos.* 2, 43–99. doi: 10.1016/j.neuroimage.2012.04.020

Smith, C. S. (1981). Semantic and syntactic constraints on temporal interpretation in tense and aspect. *Syn. Semant. Ann Arbor Mich.* 14, 213–237.

Sportiche, D. (1988). A theory of floating quantifiers and its corollaries for constituent structure. *Linguist. Inquir.* 19, 425–449.

Stan Development Team. (2016). *rstanarm: Bayesian Applied Regression Modeling Via Stan. R Package Version 2.13. 1*.

Steinhauer, K., and Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain Lang.* 83, 62–65.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/s0749-596x(02)00536-3

Van Dyke, J. A., and Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Lang. Linguist. Comp.* 6, 193–211. doi: 10.1002/lnc3.330

Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2006.03.007

Van Dyke, J. A., and McElree, B. (2011). Cue-dependent interference in comprehension. *J. Mem. Lang.* 65, 247–263. doi: 10.1016/j.jml.2011.05.002

Van Gompel, R. P., Pickering, M. J., Pearson, J., and Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *J. Mem. Lang.* 52, 284–307. doi: 10.1016/j.jml.2004.11.003

Vasishth, S., and Lewis, R. L. (2006). Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 82, 767–794. doi: 10.1353/lan.2006.0236

Vasishth, S., Mertzen, D., Jäger, L. A., and Gelman, A. (2018). The statistical significance filter leads to overoptimistic expectations of replicability. *J. Mem. Lang.* 103, 151–175. doi: 10.1016/j.jml.2018.07.004

von der Malsburg, T., and Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *J. Mem. Lang.* 94, 119–133. doi: 10.1016/j.jml.2016.10.003

Wagers, M., and McElree, B. (2013). "Working memory and language processing: theory, data and directions for future research," in *The Cambridge Handbook for Biolinguistics*, eds C. Boeckx and K. K. Grohmann (Cambridge, MA: Cambridge University Press), 203–232. doi: 10.1017/cbo9780511980435.014

Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain. Lang.* 108, 40–55. doi: 10.1016/j.bandl.2008.10.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 06 June 2019
doi: 10.3389/fpsyg.2019.01320

Proximity and Same Case Marking Do Not Increase Attraction Effect in Comprehension: Evidence From Eye-Tracking Experiments in Korean

Nayoung Kwon[1]* *and Patrick Sturt*[2]

[1] Department of English, Konkuk University, Seoul, South Korea, [2] Department of Psychology, University of Edinburgh, Edinburgh, United Kingdom

Previous studies have suggested that during the on-line sentence processing, relevant memory representations are directly accessed based on cues at retrieval (McElree et al., 2003). Under this hypothesis, retrieval cues activate any memory representation with matching features, leading to the so-called attraction effect. This predicts that attraction effects would be modulated by memory representation of a distractor. Here, we investigated this possibility, focusing on two factors (i.e., proximity to the retrieval point and the number of matching features) that would affect representation of a distractor in three Korean eye-tracking experiments. We predicted that if memory representation of a distractor decays over time, a distractor close to a retrieval point would lead to stronger attraction effects. We also predicted that a distractor would be more likely to lead to interference when it shares a higher number of matching features with the retrieval cues of a dependency, relative to the target of the dependency, due to multiple direct accesses based on multiple matching cues. However, the results did not show evidence that proximity of a distractor to the retrieval point enhanced attraction effects. Likewise, there was no evidence that a greater number of matching cues of a distractor alone would trigger more mis-retrieval, in contrast to a previous finding that a greater number of mismatching cues of a licit antecedent in addition to a greater number of matching cues of a distractor did so (Parker and Phillips, 2017). On the other hand, the results suggested that a distractor marked with nominative case was more likely to be mis-retrieved as the subject of a verb, compared to a distractor marked with a dative case, suggesting that the subject grammatical role is a critical cue for a subject-verb agreement. These results are best compatible with the hypothesis that retrieval cues are weighted, possibly depending on the nature of the dependency that is currently processed.

Keywords: Korean, attraction, honorific agreement, subject-verb agreement, eye tracking, proximity, case marking, memory representation

OPEN ACCESS

Edited by:
Simona Mancini,
Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Dan Parker,
College of William & Mary, United States
Sol Lago,
Universität Potsdam, Germany

***Correspondence:**
Nayoung Kwon
nayoung.kw@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *16 February 2019*
Accepted: *20 May 2019*
Published: *06 June 2019*

Citation:
Kwon N and Sturt P (2019) Proximity and Same Case Marking Do Not Increase Attraction Effect in Comprehension: Evidence From Eye-Tracking Experiments in Korean. Front. Psychol. 10:1320. doi: 10.3389/fpsyg.2019.01320

INTRODUCTION

Successful processing of a long-distance dependency requires retrieval of linguistic items in working memory. For example, in (1) the head NP (i.e., *the book*) of a relative clause should be retrieved at the embedded verb (i.e., *admired*), where it can be associated with a thematic role within the relative clause, and in (2) the head NP (i.e., *the key*) of a complex NP should be retrieved at the main verb (i.e., *was*) to form a subject-verb agreement dependency.

(1) This was the book that the editor admired.
(2) The key to the cabinet was rusty from many years of disuse.

Results of previous studies have been argued to support a content-addressable direct access model of retrieval (McElree, 2000; McElree et al., 2003; c.f., Kintsch, 1970), according to which items stored in working memory are activated in parallel, based on the matching of retrieval cues (see Lewis and Vasishth, 2005) for an implementation of such a content-addressable cue-based retrieval model of sentence processing).

Importantly to the goal of the study, this cue-based parallel retrieval mechanism predicts that the processing of a linguistic dependency can be affected by elements in memory that are not licit parts of that dependency. This arises because the model predicts that any item in memory that matches in features with a retrieval cue will be activated to a certain extent, even when the feature match is only partial. Such effects have been discussed in terms of both inhibitory and facilitatory mechanisms. For inhibitory effects, it has been shown that when NPs (noun phrases) in memory are of the same type, this can result in increased processing costs, and this effect has been argued to be due to partial activation of the illicit NP (similarity-based interference: Gordon et al., 2001). Thus, as Lee et al. (2006) discussed, rereading times of NP1 and NP2 in sentences of the form of (3) were longer after reading the verb region (i.e., retrieval point) when they were both descriptive nouns than when they were of different types (e.g., descriptive NP1 and a pronominal NP2), probably due to enhanced difficulty in establishing legitimate syntactic or semantic relations between NPs and the verbs. These results suggest that retrieval cues activated all memory representations of linguistic items with matching features in parallel and that the mis-retrieved NP interfered with the processing of a subject-verb dependency.

(3) Two NPs with the nominative caser marker
[MAINCLAUSE subject NP_1 [EMBEDDEDCLAUSE subject NP_2 EMBverb] MAINverb].

Turning to facilitatory effects, the so-called *attraction* effect is a case where the activation of an illicit element has been argued to lead to mis-retrieval of that element instead of the target on a proportion of trials, leading to an overall facilitation of processing (Lewis and Vasishth, 2005; Wagers et al., 2009). The attraction effect is a "grammatical illusion" where processing difficulty due to ungrammaticality is reduced when a feature-matching distractor is mis-retrieved. For example, sentences in (c) and (d) in (4) are both ungrammatical as the licit subject (*the key*) mismatches the verb (*were*) in its number feature, but they differ from each other in that sentence (d) has a distractor NP with a plural number feature (e.g., *the cabinets… were*) while sentence (c) does not (e.g., *the cabinet… were*).

(4) Experimental sentences in Pearlmutter et al. (1999)
a. The key to the cabinet was rusty from many years of disuse.
b. The key to the cabinets was rusty from many years of disuse.
c. *The key to the cabinet were rusty from many years of disuse.
d. *The key to the cabinets were rusty from many years of disuse.

If processing difficulty of the subject-verb agreement in sentences like (4) is only affected by the licit subject (i.e., *the key*), then an equal level of processing difficulty is predicted for sentences (c) and (d) in comparison to their grammatical counterpart sentences (a) and (b), regardless of different number features of a distractor in (c) and (d). However, experimental results have shown reduced processing difficulty sentences like (d) in comparison to sentences like (c), suggesting that the processing difficulty of the subject-verb agreement is also affected by a distractor item which does not participate in the dependency (see also Nicol et al., 1997; Pearlmutter et al., 1999; Thornton and MacDonald, 2003; Drenhaus et al., 2005; Vasishth et al., 2008; Xiang et al., 2009; Dillon et al., 2013; Tanner et al., 2014; Lago et al., 2015; Parker and Phillips, 2017; for related effects in production, see Bock and Miller, 1991; Bock and Cutting, 1992; Bock and Eberhard, 1993; Vigliocco and Nicol, 1998; Hartsuiker et al., 2001; Haskell and MacDonald, 2003; Thornton and MacDonald, 2003; for related inhibitory effects, see Lewis, 1996; Gordon et al., 2001, 2004, 2006).

Similar effects have been found in Korean as well. Although Korean does not have rich verbal agreement, and so verbs in Korean do not agree with their subjects in person or number in most cases (Sohn, 1999), the subject honorific marker *–si* is an exception to this. *–Si* attaches to the stem of a verb, marking the speaker's respect for its agreeing local subject as in (5). Accordingly, *–si* can only occur with a honorifiable NP such as *grandpa, uncle* or *teacher* but not with nouns such as *kid* or *burglar*, which would be regarded as denoting people who are low in their perceived social status as shown in (6). In certain aspects, the *–si* subject-verb agreement in Korean is different from the number or person subject-verb agreement in English, given that *–si* agreement is pragmatically motivated, and given that the use of *–si* is optional such that its omission does not render the sentence unacceptable as shown in (5). Yet, in a previous study, Kwon and Sturt (2016b) showed that retrieval processes underlying the processing of *–si* agreement in Korean are similar to those of the number or person subject-verb agreement in English. Thus, in a sentence where a second NP forms a dependency with an embedded verb as in (7), a distractor with a matching honorific feature (e.g., chair) has been shown to reduce the processing difficulty due to honorific feature mismatches between a verb and its licit subject (e.g., embedded verb with *–si* and *Inho*, personal names in Korean, without honorific feature; Kwon and Sturt, 2016b).

(5) Taythonglyeng hayngsacang-ey nathana-si/ø-ess-ta
President-NOM event-at appear-HON/ø-past-DECL
'The president appeared at the event.'

(6) *kkoma-ka hayngsacang-ey nathana-si-ess-ta
kid-NOM event-at appear-HON-past-DECL
'The kid appeared at the event.'

(7)

chairman-HON.NOM/Mary-TOP

[*Inho*-NOM/*president*-HON.NOM … *start*-HON-COMP] …
'The chairman/Mary … that Inho/the president starts…'

In many cases, attraction effects have been found in ungrammatical sentences at a relatively late processing stage after the ungrammaticality of sentences is detected (Pearlmutter et al., 1999; Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015; cf. Van Dyke, 2007). Given this, and given the fact that these studies did not find reliable evidence of attraction for grammatical sentences, it has been proposed that the attraction effects arise as an error-driven process, whereby distractors are retrieved as a repair strategy (cf. Wagers et al., 2009). However, the effect of a distractor found in Kwon and Sturt (2016b) is only partially compatible with this proposal. While the grammaticality effect (i.e., cost of ungrammatical relative to grammatical conditions) indeed preceded the effect of a distractor in the eye-movement record, the effect of the distractor was not limited to ungrammatical sentences. Instead, the processing of grammatical sentences was also affected by a distractor, leading to longer reading times when the distractor (e.g., Mary without honorific feature) mismatched the embedded verb in honorific features. These results were taken to suggest that the attraction effect is not an error-driven mechanism. Instead, Kwon and Sturt proposed that the effect of a distractor is likely to result from general working memory principles of activating of potential items in memory (see also Vasishth et al., 2008). Note, however, that the effect observed by Kwon and Sturt showed a facilitation for grammatical sentences where the distractor matched the honorific features of the verb, while the model proposed by Lewis and Vasishth (2005) would have predicted inhibition in this condition due to similarity-based interference. We return to this issue in the discussion.

Despite some differences in the interpretation, these results suggest the possibility that the processing of the retrieval process is modulated by memory representations of a distractor as well as that of a licit target item. For example, while a larger pool of items in memory does not affect the retrieval speed of a target item (McElree, 2000), assuming that retrieval of a linguistic item in memory is preceded by the reactivation of its memory representation which decays over time, it is possible that a distractor is more easily activated and thus interferes more with the processing of a dependency when it is closer to a retrieval point than when it is further away (Lewis and Vasishth, 2005)[1]. In fact, manipulating the linear distance between the subject and the verb as well as the number feature of the subject and its intervening object, Kaan (2002) showed that participants better remembered the number features of the licit subject when the linear distance of the subject and the verb was shorter, suggesting the relevance of linear distance in the context of retrieval processes. Likewise, if retrieval is based on a parallel feature matching process, it is also possible that a distractor is more prone to affect retrieval of a target item when it has a larger number of matching features than when it has fewer. Supporting evidence for this latter observation comes from a study on the processing of reflexives, reported by Parker and Phillips (2017). Previous studies had shown that unlike number/person subject-verb agreement or negative polarity items, the processing of reflexives is not easily affected by a distractor (Sturt, 2003; Xiang et al., 2009; Dillon et al., 2013). For example, in Sturt's (2003) study of the processing of reflexives in sentences like (8), readers slowed down in the early parsing stages only when the licit subject (*He/She*) mismatched the reflexive in gender features. On the other hand, the gender feature of the intervening distractor (*the surgeon*) did not affect the processing until later processing stages, during the final interpretation (Sturt, 2003). However, Parker and Phillips (2017) showed that, when a distractor is associated with a larger number of matching (and probably semantically more salient) features [i.e., animacy as well as gender features in (9)], the processing of an emphatic reflexive is affected by a distractor in a similar manner to that of number/person agreement. Thus, the processing difficulty due to animacy/gender mismatches in a dependency involving a reflexive (e.g., the discovery-himself) was significantly reduced when there was a feature matching distractor (e.g., the researcher). These results suggest a possibility that retrieval is sensitive to memory representations of a distractor.

(8) He/She remembered that the surgeon had pricked himself/herself with a used syringe.

(9) The doctor/discovery that the researcher/report described meticulously was certified after debunking the urban myth himself in the new scientific journal.

Thus, in the experiments described below, we aimed to further investigate the effect of memory representations of a distractor on retrieval. We examined the effect of proximity by manipulating the linear distance of a distractor to a retrieval point, and the effect of the degree of similarity by manipulating the number of matching features between retrieval cues and a distractor, across the three experiments. To this aim, we ran eye-tracking studies of the processing of a subject-verb *–si* agreement using subject control, object control, and center-embedded clause constructions in Korean. The target stimuli of Experiments 1 through 3 are schematically presented in sentences (10), (11), and (12) respectively. As illustrated in these sentences, the structurally legitimate subject NP for the embedded verb differs depending on the sentence type. For example, in the subject control construction in Experiment 1, NP1 is the licit subject for the embedded verb, while NP2 is the licit subject in the object control construction in Experiment 2 and in the center-embedded construction in Experiment 3. Thus, in Experiment 1, NP2 is a distractor (i.e., NP in the dotted square), while in Experiments 2 and 3, NP1 is. Accordingly, the distractor linearly intervenes in the subject-verb dependency in Experiment 1 but it does not in Experiment 2 or Experiment 3. Thus, the comparison of the results of Experiment 1 and 2 could reveal the effect of proximity of a distractor to its retrieval point. If a distractor whose memory representations are more highly activated is more likely to lead

[1]Our predictions are based on Lewis and Vasishth (2005), which assumes time-based decay in working memory. While it should be noted that the notion of temporal decay has been controversial, with a possibility that time-based forgetting could be due to representation-based interference (Nairne, 2002; Oberauer and Kliegl, 2006; Lewandowsky et al., 2009) rather than time-related decay (Berman et al., 2009; Barrouillet et al., 2011), for our experiment, the interference-based account also makes the same prediction with the time-based decay account, given that a distractor intervenes with the subject-verb agreement in the subject control construction but not in the object control construction.

to interference or attraction, then it can be predicted that an effect of a distractor will be stronger in Experiment 1 than in Experiment 2.

Experiment 1: Subject control construction

(10) NP1-NOM | NP2-DAT | AdvP1 AdvP2 Emb.Verb1
Adv3 Main.Verb2

Experiment 2: Object control construction

(11) NP1-NOM | NP2-DAT AdvP1 AdvP2 Emb.Verb1
Adv3 Main.Verb2

Experiment 3: Center-embedded construction

(12) NP1-NOM | NP2-NOM AdvP1 AdvP2 Emb.Verb1
Adv3 Main.Verb2

On the other hand, the comparison of Experiments 2 and 3 could reveal the effect of the number of matching features of a distractor. Previous production studies have shown that similarity in case marking of the licit subject and a distractor led to increased attraction effects (Hartsuiker et al., 2001). Likewise, assuming that case information is a retrieval cue in comprehension, the morphology on the verb in Experiment 2 encodes a retrieval cue of dative case for its (overt) subject argument, while in Experiment 3, the retrieval cue is for nominative case. Thus, the distractor in Experiment 3 has more retrieval cue features than the distractor in Experiments 1 or 2. If a greater degree of feature overlap between distractor and retrieval cues leads to greater interference or attraction, then it can be predicted that the effect of the distractor will be stronger in Experiment 3, where the distractor matches the case retrieval cue, than in Experiment 2, where it does not. Thus, comparison of the general patterns of the results of these experiments would further our understanding of how memory representations of a distractor affect the retrieval processes, leading to a fuller understanding of the mechanisms underlying the retrieval processes in general. Below, we first report these three eye-tracking experiments, and then go on to present cross-experiment comparisons to address these questions.

EXPERIMENT 1

Experiment 1 investigated the processing of subject-verb honorific agreement in the subject control construction with the suffix *–keyss* "will" or "plan to" in Korean. In this particular construction, as used in Experiment 1, the embedded verb, marked with *–keyss,* indicates that the matrix subject (NP1) is the controller for PRO, as shown in (13). The construction is roughly translated as the subject "planning to" execute the action predicated in the embedded verb. Thus, the dependency is formed between NP1 and the embedded verb (via PRO), as illustrated in (10) above. On the other hand, the dative marked indirect object cannot serve as a controller although it is linearly closer to the embedded verb, intervening with the subject-verb honorific agreement.

(13) Subject control construction used in Experiment 1
[$NP1_i$-NOM $NP2_k$-DAT
[PRO_i leave-HON-SUBJECT.CTRL-DECL-comp] said]
'NP_i told $NP2_k$ that $PRO_{i/*k}$ will leave.'

Personal names in Korean do not have honorific features. Thus, following Kwon and Sturt (2016b), to investigate the memory retrieval processes underlying the processing of a dependency, we manipulated the honorific features of NP1 and NP2 of the experimental sentences, by using either personal names (i.e., NH: non-honorifiable) or descriptive NPs (i.e., H: honorifiable)[2]. On the other hand, the embedded verb is always marked with an honorific marker *–si* (see R5 in **Table 1** below). Accordingly, there are two congruous conditions (i.e., H_{NP1}-NH_{NP2} and H_{NP1}-H_{NP2}) and two incongruous conditions (i.e., NH_{NP1}-H_{NP2} and NH_{NP1}-NH_{NP2}), as shown in **Table 1**.

Since the honorific suffix *–si* should agree with the verb's subject in Korean, we predict that mismatched honorific features of the incongruous conditions with the NH subject (NH-H and NH-NH conditions) will incur processing difficulty (Kwon and Sturt, 2016b). Thus, the incongruous conditions will show longer reading times at the embedded verb marked with *–si–* when compared with their congruous counterpart sentences with the H subject (H-H and H-NH conditions). Crucially, although control information is accessed early during on-line sentence processing (Kwon and Sturt, 2014, 2016a), it has been also shown that the processing of subject-verb honorific agreement can be affected by a feature-matching yet structurally illicit distractor in a control construction (Sturt and Kwon, 2015). If so, we predict that the reading time penalty for the incongruous condition will be reduced in the NH-H condition, where the distractor matches the honorific features of the verb, compared to the NH-NH condition, where it does not, resulting in an interaction between the honorific features of the subject and the dative object at the embedded verb position. This interactive pattern would be consistent with previous studies investigating subject-verb number agreement in English (e.g., Wagers et al., 2009; Dillon et al., 2013).

Participants

Twenty eight native speakers of Korean (mean age: 23.46; range 19–27) participated in the study. At the time of the experiment, they were all either undergraduate or graduate students at Konkuk University, Korea and received KRW 10,000 per hour for their participation. They all had normal or corrected-to-normal vision. Written informed consent was obtained from all the participants.

[2]We intentionally did not use non-honorific description nouns (e.g., baby) for the NH conditions. This is because two consecutively occurring description NPs can be misanalysed as an instance of a double nominative construction in Korean, where the NP2 can inherit the honorific features of the NP1 and thus can be predicated with a honorific verbal form even when NP2 is normally not honorifiable (Sohn, 1999; e.g., *Teacher-nom house-nom is.far*-SI-DECL, "The teacher, his place is far away").

TABLE 1 | Example experiment item in Experiment 1: Subject control construction.

NP1	NP2	R1	R2	R3	R4	R5	R6	R7
H	H	선생님이	위원장님에게	행사장	주변에서	기다리시겠다고	다시	한번	강조하셨다
		teacher-NOM	chair-DAT	event.place	near	wait-HON-SBJ.CTRL-COMP	again	once	emphasized
NH	H	*수진이가	위원장님에게	행사장	주변에서	기다리시겠다고	다시	한번	강조하셨다
		Swujin-NOM	chair-DAT	event.place	near	wait-HON-SBJ.CTRL-COMP	again	once	emphasized
H	NH	선생님이	인호에게	행사장	주변에서	기다리시겠다고	다시	한번	강조하셨다
		teacher-NOM	Inho-DAT	event.place	near	wait-HON-SBJ.CTRL-COMP	again	once	emphasized
NH	NH	*수진이가	인호에게	행사장	주변에서	기다리시겠다고	다시	한번	강조하셨다
		Swujin-NOM	Inho-DAT	event.place	near	wait-HON-SBJ.CTRL-COMP	again	once	emphasized

The teacher$_i$ /Swujin$_i$ emphasized to the chair/Indo that __$_i$ would wait near the event place.

Materials

Forty sets of experimental sentences like those in **Table 1** were created. All the experimental sentences contained two NPs with a + or – value for the honorific feature. On the other hand, the embedded verb was always marked with *–si*, and this verb formed a dependency with one of these NPs.

Before the main experiment, we first conducted a norming study to control for the plausibility of the event described with an embedded verb with NP1 or NP2 as a potential subject. For example, for the sentences in **Table 1** four norming sentences were created using NP1 or NP2 (H_{NP1}, NH_{NP1}, H_{NP2}, NH_{NP2}) and the embedded verb, where the verbs of the norming sentences were matched with these NPs in their honorific features, as shown in (14) and (15). Thirty-two native Korean speakers participated in the norming study, each receiving KRW 3,000. At the time of the study, they were undergraduate students at Konkuk University, Korea. The norming sentences were split into four lists based on a Latin-square design along with 40 filler sentences with similar complexity. They were pseudo-randomized such that no two sentences from the same condition appeared in a row. Participants were asked to rate the plausibility of the sentences on a scale of 1 (very unlikely) to 5 (very plausible). The rating results were then analyzed using Linear Mixed Effect Regression (LMER) analysis (Baayen, 2008; Baayen et al., 2008; Jaeger, 2008). Models were constructed with the maximal random effect structure and were only simplified when the model did not converge (Barr et al., 2013). The results showed that the plausibility of the four conditions did not significantly differ from each other ($|t| < 0.14$ for all comparisons: H-NP1 vs. NH-NP2, NH-NP1 vs. NH-NP2, H-NP2 vs. NH-NP2), with the mean ratings of 4.69 (*se*: 0.042), 4.71 (*se*: 0.04), 4.72 (*se:* 0.039), and 4.62 (*se:* 0.046) for the H-NP1 (e.g., *teacher*), NH-NP1 (e.g., *chair*), H-NP2 (e.g., Swujin), and NH-NP2 (e.g., Inho) conditions, respectively.

Honorifiable subjects

(14) 선생님/위원장님이 행사장 주변에서
Teacher/chair-NOM event.place near
기다리셨다
wait-*si*-past-decl
'The teacher/chair waited near the event place.'

Not-honorifiable subjects

(15) 수진이/인호가 행사장 주변에서
Swujin/Inho-NOM event.place near
기다렸다
wait-past-decl
'Swujin/Inho waited near the event place.'

Given the results of the norming study, the experimental stimuli were distributed over four lists based on a Latin square design, along with 80 filler sentences of similar length and complexity. No two experimental items from a same condition were presented in a row.

Procedure

There were three practice trials before the main experiment started. Participants' eye movements were recorded using an SR Research Eyelink 1,000 Plus eye-tracker at a rate of 1,000 Hz. Each recording session began with a calibration procedure, using a standard 9 point calibration routine before the experiment started, and recalibration was performed whenever necessary throughout the experiment. In each trial, a participant was asked to fixate on a black square on the left side of the screen, where the first character of the upcoming sentence would be presented. When a participant's fixations were successfully detected on the black square, the square was automatically replaced by the experimental stimuli. Participants were asked to read the sentences at their natural speed and answered a yes/no comprehension question for all the sentences. The comprehension questions probed general understanding of the sentences. For example, for the H-H condition sentence in **Table 1**, "Will the teacher wait around the event place?" was asked. The experiment took about 40 min.

Data Analysis

For data analysis, following standard eye-tracking data analyses (Rayner, 1998; Sturt, 2007) we first merged short fixations under 80 ms into longer fixations within the distance of the visual angle of 0.05. If there was no such fixation, we removed the short fixations. We also removed fixations longer than 1,200 ms. This procedure affected 1.5% of the trials. Three eye-movement measures are presented. First pass reading times are the sum of

all fixations in a target region from the first entry into the region until leaving the region either to the left or right. Go-past times (also called regression path duration) are the sum of all fixations on a given region from the first entry into the region from the left until leaving it to the right. Total time is the sum of all fixations in a given region. For First-pass reading time and Go-past times, we excluded the trials in which a target region was not fixated on in initial reading. For Total Time, we excluded the trials in which the target region received no fixation at all. The proportion of missing data points, due to zeros or track losses were less than 1%.

Statistical analyses were conducted for Region 5 as defined in **Table 1**, as the region is critical for retrieval processes. The region included the embedded verb and its following adverbial word (mean length = 2.5 syllables). To lower the rate of false positives due to multiple comparisons (von der Malsburg and Angele, 2017), we report statistical analyses for only one region, while reporting means for all regions. We first log-transformed reading times, and the resulting reading time data were analyzed based on Linear Mixed Effect Regression (LMER) analysis (Baayen, 2008; Baayen et al., 2008; Jaeger, 2008), using the lme4 R package (Bates et al., 2015; version 1.1–8). The comprehension accuracy rates were analyzed using a generalized LME model with a binomial distribution. The regression models incorporated two fixed-effect factors (the honorific features of the mrain and the embedded subject: H vs. NH), their interaction and crossed random effects for participants and items. The fixed-effect factors were coded numerically using sum coding, with the two levels of each factor coded as 1 and −1. Models had the maximal random effect structure whenever possible, including both intercepts and slopes, and were only progressively simplified when the model did not converge (Barr et al., 2013). In case of non-convergence, we simplified the model by backwards elimination, following the hierarchy principle, such that the interaction slope parameters were removed, and convergence checked, before attempting to remove either of the random main effect parameters. Also, at each stage of model simplification, convergence was checked both including and excluding random correlation parameters. Random slope parameters corresponding to fixed-effects are reported in in the "slope" column of **Table 4** if they were included in the model for participants or items. **Table 4** also shows coefficients, standard errors and *t*-values (*z*-values for the logit model) for each fixed effect and interaction from the analyses. *P*-values were obtained using LmerTest (Kuznetsova et al., 2017), and were corrected for multiple comparisons (9 comparisons: three eye-tracking measures × two main effects and one interaction) using Holm's correction (Holm, 1979; Abdi, 2010). For the analysis of comprehension accuracy based on a binomial logit model, *p*-values were calculated from the Z score, and were also corrected for multiple comparisons (three comparisons: two main effects and one interaction) using Holm's correction. Finally, planned (paired) contrasts were made based on the Tukey test (using the glht function of multcomp package: Hothorn et al., 2008; version 1.4–1) in R (R Core Team, 2018).

Results and Discussion

Comprehension accuracy and mean reading times for each condition are given in **Tables 2**, **3**, respectively. Statistical analysis results for reading time measures are given in **Table 4**.

Region 5 (the critical embedded verb and the spill over region; 'wait-HON again')

At R5, there was a main effect of NP1 with the (ungrammatical) NH_{NP1} conditions taking longer to read than the (grammatical) H_{NP1} conditions. The effect was marginal in First pass times and significant in Go-past and Total times. In addition, there was a marginal main effect of NP2 in Total times with longer reading times for the H_{NP2} condition than for the NH_{NP2} condition.

The grammaticality effect (i.e., the main effect of NP1) suggests that the subject control information was accessed from an early processing stage, allowing, and constraining dependency formation between the verb and its licit subject throughout the various processing stages. On the other hand, the marginal effect of NP2 in Total times suggests a tendency for a distractor with matching features to be activated regardless of grammaticality of the target sentences. For both grammatical and ungrammatical sentences, a distractor matching the verb in the honorific feature led to a slow-down in Total Time, relative to conditions where the distractor mismatched, presumably reflecting a later stage of processing.

The grammaticality effect shows that readers were sensitive to the relevant control information in forming the dependency. On the other hand, the (marginal) effect of the distractor was different from the so-called "attraction effect," where facilitation is limited to the ungrammatical conditions (Wagers et al., 2009). Instead, the direction of the effect was such that the distractor interfered with and slowed down reading, when it matched the honorific features of the verb, regardless of the grammaticality of the sentence. We note that this effect is different from the general pattern observed by Kwon and Sturt (2016b), where a matching distractor tended to facilitate processing. We return to this point in the general discussion.

In summary, in Experiment 1 the grammaticality effect preceded any effect of the distractor. That is, while the grammaticality effect was found both in Go-past and Total times on the critical verb, the effect of the distractor was found only in Total Time. In addition, the marginal main effect of the distractor suggests that the distractor may affect the processing of the subject control construction regardless of grammaticality of the target sentences.

EXPERIMENT 2

Experiment 2 investigated the processing of subject-verb honorific agreement in the object control construction with *–la* in Korean. The embedded verb marked with *–la* signals that the indirect object (NP2) marked with a dative case marker is the licit controller for PRO, as shown in (16). Thus, the dependency is formed between the dative marked NP2 and the embedded verb (via PRO), as illustrated in (11) and (16). On the other hand, the nominative marked main clause subject (NP1) cannot serve as a controller.

(16) Object control construction used in Experiment 2
[$NP1_i$-NOM $NP2_k$-DAT
[PRO_k leave-HON-OBJECT.CTRL-comp] said]
'NP_i told $NP2_k$ to $PRO_{*i/k}$ leave.'

TABLE 2 | Mean comprehension accuracy rates in Experiment 1.

	Mean (se)		Estimate	SE	*z*	Slope	*p*	Adjusted *p*
H & H	87.5% (0.019)	Intercept	3.16	0.38	8.22		<2e-16***	<2e-16***
NH & H	91.8% (0.016)	NP1	−0.14	0.11	−1.29		n.s.	n.s.
H & NH	88.5% (0.019)	NP2	0.07	0.11	0.63		n.s.	n.s.
NH & NH	88.2% (0.019)	NP1*NP2	−0.2	0.12	−1.79		0.08	n.s.

TABLE 3 | Means (and standard errors), aggregated by participants, for first pass, go-past, and total times in Experiment 1.

	Region 1 *teacher/ swujin*	Region 2 *chair/ Inho*	Region 3 *event*	Region 4 *near*	Region 5 *wait- HON again*	Region 6 *once*	Region 7 *emphasized*
First pass (msec)							
H & H	417 (24)	373 (14)	254 (7)	244 (7)	711 (23)	275 (10)	299 (15)
NH & H	403 (23)	365 (14)	244 (7)	249 (9)	810 (28)	288 (13)	284 (15)
H & NH	394 (24)	332 (14)	272 (9)	252 (9)	741 (23)	278 (10)	287 (13)
NH & NH	453 (28)	334 (13)	270 (10)	245 (8)	831 (32)	283 (14)	281 (15)
Go past (msec)							
H & H		525 (26)	340 (18)	290 (14)	1514 (76)	941 (79)	2484 (169)
NH & H		477 (22)	311 (16)	304 (18)	2007 (111)	1402 (117)	2797 (209)
H & NH		452 (23)	357 (23)	314 (16)	1350 (66)	853 (70)	2004 (113)
NH & NH		458 (21)	369 (25)	330 (20)	1850 (96)	1139 (81)	2541 (182)
Total time (msec)							
H & H	855 (38)	1082 (46)	573 (25)	474 (20)	1599 (64)	486 (22)	452 (28)
NH & H	863 (39)	1136 (47)	586 (27)	533 (25)	1925 (83)	596 (27)	432 (30)
H & NH	791 (35)	844 (36)	530 (21)	470 (20)	1462 (56)	483 (22)	398 (20)
NH & NH	903 (38)	982 (51)	623 (33)	501 (23)	1824 (82)	554 (29)	416 (25)

Grammatical conditions: H & H, H & NH; Ungrammatical conditions: NH & H, NH & NH.

As in Experiment 1, we predicted that mismatched honorific features of the incongruous conditions with the NH subject (NH dative object conditions: H_{NP1}-NH_{NP2} & NH_{NP1}-NH_{NP2}) would incur processing difficulty (Kwon and Sturt, 2016b), leading to longer reading times at the critical region in these conditions than in the congruous conditions (H-H & NH-H conditions). In addition to the grammaticality effect, we also predicted an effect of the distractor. In particular, if interference from a distractor is affected by memory representations of a distractor, it is likely that a memory representation that is more highly activated at the point of retrieval will be more likely to lead to interference. If so, we predict that the effect of a distractor will be weaker in Experiment 2 than in Experiment 1, given that the distractor position is further away from the embedded verb and so its memory representation could be more decayed at the retrieval point in Experiment 2.

Participants

Twenty eight native speakers of Korean (mean age: 23.96; range 19–31) participated in the study, receiving KRW 10,000 per hour. They were all either undergraduate or graduate students at Konkuk University, Korea, and had normal or corrected-to-normal vision. Written informed consent was obtained from all the participants.

Materials

Forty sets of object control sentences like those in **Table 5** were created based on the stimuli of Experiment 1, replacing the subject control suffix—*keyss* with the object control suffix –*la*. Lexical items remained the same as in Experiment 1, but main verbs were changed when necessary. Thus, at the point of the embedded verb, the plausibility of the target sentences with potential subject NPs was identical to that in Experiment 1. We used the same filler sentences used in Experiment 1, and other remaining procedures were also analogous to Experiment 1.

Procedures

The same eye-tracking procedure was used as in Experiment 1.

Data Analysis

As in Experiment 1, we first merged short fixations under 80 ms into longer fixations within the distance of the visual angle of 0.05. On the other hand, fixations longer than 1,200 ms were removed. This procedure affected 2.1% of the trials. On the other hand, the proportion of missing data points, due to zeros or track losses were <2%. Remaining procedures were analogous to those used in Experiment 1.

TABLE 4 | Generalized Linear Mixed Effects results for reading times in Experiment 1.

		Coeff.	SE	*t*	Slope	*p*	Adjusted *p*
First pass							
R5 *"wait-HON again"*	Intercept	6.478	0.055	118.94		<2e-16***	<2e-16***
	NP1	−0.044	0.017	−2.55	(p,i)	0.0116*	0.07
	NP2	−0.016	0.02	−0.76	(p,i)	n.s.	n.s.
	NP1*NP2	−0.014	0.017	−0.84		n.s.	n.s.
Go-past							
	Intercept	7.114	0.089	80.62		<2e-16***	<2e-16***
	NP1	−0.129	0.026	−5.04	(p,i)	0.00001***	0.0001***
	NP2	0.032	0.021	1.49	(p,i)	n.s.	n.s.
	NP1*NP2	0.009	0.019	0.46		n.s.	n.s.
Total time							
	Intercept	7.229	0.0881	82.11		<2e-16***	<2e-16***
	NP1	−0.093	0.0167	−5.53	(p,i)	0.00001***	0.0001***
	NP2	0.038	0.0142	2.65	(p,i)	0.0087*	0.062+
	NP1*NP2	0.005	0.0144	0.28	(p,i)	n.s.	n.s.

*Coefficients, standard errors, t or z-values and p-values are reported for the main effects of NP1and NP2 manipulation, as well as for the interaction of these two factors. Note that the effect of NP1 (the main subject) corresponds to the grammaticality effect, while the effect of NP2 (the dative object) corresponds to the distractor effect. The "Slope" column indicates whether the random slope parameter corresponding to the effect was included in the model for participants (p) or items (i). P-values were obtained using LmerTest (Kuznetsova et al., 2017), and adjusted p values were calculated using Holm's correction for multiple comparisons (Holm, 1979; Abdi, 2010). *p < 0.05, ***p < 0.0005.*

TABLE 5 | Example experiment item in Experiment 2: Object control construction.

Main embedded subj	R1	R2	R3	R4	R5	R6	R7	
H	H	선생님이	위원장님에게	행사장	주변에서	기다리시라고 다시	한번	강조하셨다
		teacher-NOM	chair-DAT	event.place	near	wait-HON-OBJ.CTRL-COMP again	once	emphasized
NH	H	수진이가	위원장님에게	행사장	주변에서	기다리시라고 다시	한번	강조하셨다
		Swujin-NOM	chair-DAT	event.place	near	wait-HON-OBJ.CTRL-COMP again	once	emphasized
H	NH	*선생님이	인호에게	행사장	주변에서	기다리시라고 다시	한번	강조하셨다
		teacher-NOM	Inho-DAT	event.place	near	wait-HON-OBJ.CTRL-COMP again	once	emphasized
NH	NH	*수진이가	인호에게	행사장	주변에서	기다리시라고 다시	한번	강조하셨다
		Swujin-NOM	Inho-DAT	event.place	near	wait-HON-OBJ.CTRL-COMP again	once	emphasized

The teacher/Swujin emphasized to the chair$_i$/Indo$_i$ that __$_i$ should wait near the event place.

Results and Discussion

Mean comprehension accuracy and reading times for each condition are given in **Tables 6**, **7**, respectively, and statistical analysis results are given in **Table 8**.

Region 5 (the critical embedded verb and the spill over region; 'wait-HON again')

At R5, there was a main effect of the NP1 (main subject: distractor) in Go-past times with longer reading times for the NH_{NP1} conditions, where the distractor mismatches the honorific features of the verb (i.e., NH_{NP1}-H_{NP2} & NH_{NP1}-NH_{NP2}) than for the H_{NP1} conditions, where it matches (i.e., H_{NP1}-H_{NP2} & H_{NP1}-NH_{NP2}). In addition, there was also a main effect of NP2 (dative marked object NP: licit subject for the embedded verb) in Total times with longer reading times for the ungrammatical NH_{NP2} conditions (i.e., H_{NP1}-NH_{NP2} & NH_{NP1}-NH_{NP2}) than for the grammatical H_{NP2} conditions (i.e., H_{NP1}-H_{NP2} & NH_{NP1}-H_{NP2}). In Total time, there was also an interaction of NP1 and NP2. *Post-hoc* pairwise comparisons showed that this was because reading times to the H_{NP1}-NH_{NP2} condition were significantly longer than those to the H_{NP1}-H_{NP2} condition ($p < 0.001$), reflecting a grammaticality cost when the distractor matched the honorific feature of the verb. On the other hand, reading times to the NH_{NP1}-NH_{NP2} condition were not significantly different from those to NH_{NP1}-H_{NP2} condition (n.s.), reflecting the lack of a grammaticality cost in this measure when the distractor mismatched the honorific feature of the verb. Note that the form of this interaction is different from what would be expected based on previous literature on subject-verb agreement attraction in English (e.g., Wagers et al., 2009; Dillon et al., 2013), where

TABLE 6 | Mean comprehension accuracy rates in Experiment 2.

	Mean (se)		Estimate	SE	*z*	Slope	*p*	Adjusted *p*
H & H	89.3% (0.018)	Intercept	3.27	0.39	8.36		<2e-16***	<2e-16***
NH & H	90% (0.017)	NP1	−0.019	0.11	−0.17		n.s.	n.s.
H & NH	90% (0.017)	NP2	0.004	0.11	0.04		n.s.	n.s.
NH & NH	89.3% (0.018)	NP1*NP2	0.046	0.11	−0.41		n.s.	n.s.

TABLE 7 | Means (and standard errors), aggregated by participants, for first pass, go-past, and total times in Experiment 2.

	Region 1 *teacher/ swujin*	Region 2 *chair/ Inho*	Region 3 *event*	Region 4 *near*	Region 5 *wait- HON again*	Region 6 *once*	Region 7 *emphasized*
First pass (msec)							
H & H	364 (23)	328 (12)	256 (7)	221 (7)	700 (24)	256 (9)	313 (19)
NH & H	340 (16)	357 (13)	250 (7)	240 (7)	776 (25)	277 (10)	290 (20)
H & NH	345 (17)	313 (11)	267 (11)	232 (8)	750 (25)	280 (11)	330 (22)
NH & NH	373 (26)	318 (11)	260 (8)	239 (8)	771 (24)	284 (13)	288 (24)
Go past (msec)							
H & H		468 (22)	319 (15)	315 (22)	1274 (58)	729 (58)	1921 (119)
NH & H		457 (18)	328 (17)	323 (21)	1447 (64)	832 (57)	1924 (136)
H & NH		445 (20)	323 (16)	287 (15)	1447 (66)	762 (60)	2324 (142)
NH & NH		447 (21)	319 (16)	343 (24)	1640 (72)	1001 (70)	1855 (126)
Total time (msec)							
H & H	753 (36)	880 (35)	508 (21)	441 (21)	1281 (50)	415 (19)	415 (23)
NH & H	649 (30)	894 (37)	480 (19)	419 (19)	1351 (47)	471 (23)	440 (31)
H & NH	782 (38)	874 (31)	561 (24)	453 (19)	1538 (57)	493 (23)	488 (27)
NH & NH	737 (37)	843 (34)	523 (21)	431 (19)	1370 (49)	476 (25)	373 (31)

Grammatical conditions: H & H, NH & H; Ungrammatical conditions: H & NH, NH & NH.

the grammaticality cost is typically found to be *reduced* by a matching distractor.

In summary, these results suggest that the processing of object control construction was not only affected by a licit dative object controller (NP2) but also by a structurally illicit subject controller (i.e., NP1). In addition, the effect of a distractor preceded the grammaticality effect and was detected from a relatively earlier eye-gazing measurement (Go-past times), and its effect was not limited to ungrammatical sentences.

EXPERIMENT 3

Experiment 3 investigated the processing of subject-verb honorific agreement in the center-embedded construction, where an embedded clause serves as a sentential complement of the main verb. While both NP1 and NP2 are marked with a nominative case marker, the embedded verb in this construction signals that the embedded subject NP (NP2 in this case) is its licit subject, as shown in (12) and in (17).

(17) Center embedded construction used in Experiment 3
[MAIN CL. NP1-NOM [EMBEDDED CL. NP2-NOM
leave-HON-PST-DECL-comp] said]
'NP1 said that NP2 left.'

As in Experiments 1 and 2, we predicted that the incongruous conditions with the NH subject (H_{NP1}-NH_{NP2} & NH_{NP1}-NH_{NP2} conditions) would show longer reading times at the critical region than in the congruous conditions (H-H & NH-H conditions). In addition, assuming that case information is a retrieval cue (cf. Hartsuiker et al., 2001), the distractor in Experiment 3 matches more retrieval cue features than the distractor in Experiments 1 or 2. Thus, if a greater degree of feature overlap between distractor and the retrieval cues leads to greater interference or attraction, then Experiment 3 is predicted to show a stronger effect of a distractor than Experiment 2.

Participants

Twenty eight native speakers of Korean (mean age: 23.89; range 21–31) received KRW 10,000 and participated in the study. They all had normal or corrected-to-normal vision, and attended Konkuk University, Korea at the time of the experiment. Written informed consent was obtained from all the participants.

Materials

Forty sets of center-embedded complement sentences like those in **Table 9** were created. As in Experiment 2, there was no change in lexical items before the main verb position. Thus, at the critical region (i.e., the embedded verb region) the plausibility of the

TABLE 8 | Generalized Linear Mixed Effects results for reading times in Experiment 2.

		Coeff.	SE	*t*	Slope	*p*	Adjusted *p*
First pass							
R5 *"wait-HON again"*	Intercept	6.485	0.056	116.76		<2e-16***	<2e-16***
	NP1	−0.043	0.018	−2.39*	(p,i)	0.0238	n.s.
	NP2	−0.014	0.016	−0.87	(p,i)	n.s.	n.s.
	NP1*NP2	−0.011	0.016	−0.69	(p,i)	n.s.	n.s.
Go-past							
	Intercept	7.054	0.057	122.81		<2e-16***	<2e-16***
	NP1	−0.067	0.02	−3.35*	(p,i)	0.0019	0.0172*
	NP2	−0.054	0.024	−2.3*	(p,i)	0.027	n.s.
	NP1*NP2	0.006	0.022	0.29	(p,i)	n.s.	n.s.
Total time							
	Intercept	7.069	0.067	105.51		<2e-16***	<2e-16***
	NP1	0.001	0.014	0.08	(p,i)	n.s.	n.s.
	NP2	−0.05	0.016	−3.2*	(p)	0.003	0.0235*
	NP1*NP2	−0.042	0.014	−3.07*		0.002	0.0177*

*Coefficients, standard errors, t or z-values and p-values are reported for the main effects of NP1 and NP2, as well as for the interaction of these two factors. Note that the effect of NP2 (the dative object) corresponds to the grammaticality effect, while the effect of NP1 (the main subject) corresponds to the distractor effect. The "Slope" column indicates whether the random slope parameter corresponding to the effect was included in the model for participants (p) or items (i). P-values were obtained using LmerTest (Kuznetsova et al., 2017) and adjusted p-values were calculated using Holm's correction for multiple comparisons (Holm, 1979; Abdi, 2010). *p < 0.05.*

TABLE 9 | Example experiment item in Experiment 3: Center-embedded construction.

Main embedded subj	R1	R2	R3	R4	R5	R6	R7	
H	H	선생님이	위원장님이	행사장	주변에서	기다리셨다고 다시	한번	강조하셨다
		teacher-NOM	chair-NOM	event.place	near	wait-HON-PST-DECL-COMP again	once	emphasized
NH	H	수진이가	위원장님이	행사장	주변에서	기다리셨다고 다시	한번	강조하셨다
		Swujin-NOM	chair-NOM	event.place	near	wait-HON-PST-DECL-COMP again	once	emphasized
H	NH	*선생님이	인호가	행사장	주변에서	기다리셨다고 다시	한번	강조하셨다
		teacher-NOM	Inho-NOM	event.place	near	wait-HON-PST-DECL-COMP again	once	emphasized
NH	NH	*수진이가	인호가	행사장	주변에서	기다리셨다고 다시	한번	강조하셨다
		Swujin-NOM	Inho-NOM	event.place	near	wait-HON-PST-DECL-COMP again	once	emphasized

The teacher/Swujin emphasized to the chair$_i$/Indo$_i$ that __$_i$ should wait near the event place.

target sentences with potential subject NPs remained the same as in Experiments 1 and 2. In addition, as in Experiment 2, the same filler sentences used in Experiment 1 were employed. Remaining procedures were also analogous to Experiments 1 and 2.

Procedures

The same eye-tracking procedure was used as in Experiments 1 and 2.

Data Analysis

As in Experiments 1 and 2, short fixations under 80 ms were first merged into longer fixations within the distance of the visual angle of 0.05. Then, we removed fixations longer than 1,200 ms. This procedure affected 2.2% of the total trials. On the other hand, the proportion of missing data points, due to zeros or track losses were <1%. Analogous statistical analysis procedures were applied as in Experiments 1 and 2.

Results and Discussion

Mean comprehension accuracy and reading times for each condition are given in **Tables 10**, **11**, respectively, and statistical analysis results are given in **Table 12**.

Region 5 (the critical embedded verb and the spill over region; 'wait-HON again')

At R5, there was a main effect of the NP1 (distractor effect) and NP2 (grammaticality effect) in Go-past times, and a main effect of NP2 in Total times. The pattern was similar to that seen in Experiment 2. Both NH_{NP1} and NH_{NP2} conditions elicited longer reading times than their counterpart H_{NP1} and H_{NP2} conditions, respectively, reflecting a cost for the mismatching of honorific

TABLE 10 | Mean comprehension accuracy rates in Experiment 3.

	Mean (se)		Estimate	SE	*z*	Slope	*p*	Adjusted *p*
H & H	93.2% (0.015)	Intercept	3.926	0.46	8.44		<2e-16***	<2e-16***
NH & H	92.5% (0.015)	NP1	0.229	0.13	1.71		0.09	n.s.
H & NH	95% (0.013)	NP2	−0.037	0.13	−0.28		n.s.	n.s.
NH & NH	91.1% (0.017)	NP1*NP2	−0.161	0.13	−1.21		n.s.	n.s.

TABLE 11 | Means (and standard errors), aggregated by participants, for first pass, go-past, and total times in Experiment 3.

	Region 1 *teacher/ swujin*	Region 2 *chair/ Inho*	Region 3 *event*	Region 4 *near*	Region 5 *wait- HON again*	Region 6 *once*	Region 7 *emphasized*
First pass (msec)							
H & H	393 (20)	327 (12)	258 (7)	238 (7)	700 (27)	265 (9)	333 (19)
NH & H	376 (17)	340 (13)	269 (8)	245 (8)	737 (26)	273 (11)	295 (16)
H & NH	391 (20)	312 (14)	260 (8)	229 (7)	728 (24)	261 (8)	331 (18)
NH & NH	356 (18)	316 (11)	257 (8)	241 (8)	809 (28)	271 (10)	298 (15)
Go past (msec)							
H & H		521 (26)	488 (27)	406 (26)	1279 (63)	731 (60)	1543 (98)
NH & H		489 (20)	407 (20)	311 (15)	1385 (65)	824 (55)	1791 (133)
H & NH		431 (21)	388 (19)	390 (27)	1381 (69)	785 (68)	1882 (117)
NH & NH		566 (31)	496 (26)	376 (24)	1757 (93)	1104 (100)	1978 (137)
Total time (msec)							
H & H	729 (31)	839 (34)	558 (20)	431 (19)	1226 (47)	401 (19)	448 (26)
NH & H	685 (32)	821 (34)	567 (23)	453 (23)	1280 (49)	444 (21)	408 (26)
H & NH	719 (33)	728 (31)	605 (24)	465 (21)	1375 (52)	457 (21)	501 (29)
NH & NH	825 (41)	899 (35)	677 (30)	465 (22)	1521 (53)	471 (23)	425 (24)

Grammatical conditions: H & H, NH & H; Ungrammatical conditions: H & NH, NH & NH.

features between the NP and the verb. These effects suggest that the processing of honorific agreement in the embedded verb is affected both by a licit (NP2) and illicit (NP1) subject. There was, however, no significant interaction between the two.

In summary, the processing of the center-embedded construction was affected by both NP1 (distractor) and NP2 (the licit subject), and the effect of a distractor was not limited to ungrammatical sentences. In addition, while both the grammaticality effect and the distractor effect were detected in the same early eye-tracking measure (Go-past times), only the grammaticality effect was found in Total times, which are a more general measure of processing.

GENERAL DISCUSSION AND CONCLUSION

The goal of the study was to investigate how memory representations of a distractor affect the retrieval processes. In particular, we were interested in the effect of proximity of a distractor to a retrieval point. Our reasoning was that if a distractor is temporarily closer to a retrieval point, then its memory representation is more likely to be highly activated, and thus is more likely to lead to stronger interference or attraction than when it is further away from the retrieval point. Thus, we predicted a stronger interference effect in the subject control construction (10) than in the object control construction (11). In addition, we aimed to examine whether a distractor would lead to stronger interference when there is a higher degree of feature match between the distractor and the retrieval cues. If a distractor is activated based on feature matches with retrieval cues, then it is possible that multiple feature matches could lead to stronger activation of the distractor, leading to a stronger interference effect. Thus, we predicted a stronger interference effect in the center-embedded construction (12) than in the object control construction (11). To address these questions, we directly compared the results of Experiments 1 and 2, and the results of Experiments 2 and 3.

We re-analyzed Go-past times of the critical region (Region 5 in Experiments 1–3), including Experiment as a fixed-effect factor in the models. In addition, since the licit subject differs for Experiment 1 (i.e., NP1) and Experiments 2 and 3 (i.e., NP2), NP1 and NP2 were re-coded as the licit or the illicit (distractor) subject, and were also incorporated in the regression models as such, so that the effect of a distractor can be better compared across the experiments. The remaining procedures were analogous to those reported in Experiments 1–3. The results of the statistical analyses comparing the results of Experiments 1

TABLE 12 | Generalized Linear Mixed Effects results for reading times in Experiment 3.

		Coeff.	SE	*t*	Slope	*p*	Adjusted *p*
First pass							
R5 *"wait-HON again"*	Intercept	6.455	0.045	142.67		<2e-16***	<2e-16***
	NP1	−0.04	0.019	−2.13*	(p,i)	0.043	n.s.
	NP2	−0.039	0.017	−2.27*	(p,i)	0.029	n.s.
	NP1*NP2	0.004	0.016	0.26		n.s.	n.s.
Go-past							
	Intercept	7.021	0.066	106.48		<2e-16***	<2e-16***
	NP1	−0.081	0.024	−3.42*	(p,i)	0.0018	0.014*
	NP2	−0.072	0.02	−3.53*	(p,i)	0.0015	0.013*
	NP1*NP2	0.03	0.019	1.59	(p,i)	n.s.	n.s.
Total time							
	Intercept	7.041	0.071	99.11		<2e-16***	<2e-16***
	NP1	−0.04	0.017	−2.33*	(p,i)	0.026	n.s.
	NP2	−0.077	0.016	−4.92*	(p,i)	0.0001	0.0004***
	NP1*NP2	0.015	0.015	1.05	(p,i)	n.s.	n.s.

*Coefficients, standard errors, t or z-values and p-values are reported for the main effects of NP1 and NP2 manipulation, as well as for the interaction of these two factors. Note that the effect of NP2 (the embedded subject) corresponds to the grammaticality effect, while the effect of NP1 (the main subject) corresponds to the distractor effect. The "Slope" column indicates whether the random slope parameter corresponding to the effect was included in the model for participants (p) or items (i). P-values were obtained using LmerTest (Kuznetsova et al., 2017) and adjusted p-values were calculated using Holm's correction for multiple comparisons (Holm, 1979; Abdi, 2010). *p < 0.05, ***p < 0.0005.*

and 2 on the one hand, and the results of Experiments 2 and 3 on the other are presented in **Tables 13**, **14**, respectively. *P*-values were corrected for multiple comparisons using Holm's correction (Holm, 1979; Abdi, 2010).

The comparison of Experiments 1 and 2 showed a main effect of Licit NP ($p < 0.0001$), but the effect was accompanied by a significant interaction of Licit NP and Experiment ($p < 0.032$). The main effect suggests that the processing of honorific agreement is constrained by honorific features of the licit subject both in the subject control (Experiment 1) and the object control (Experiment 2) constructions, but the interaction suggests that the magnitude of grammaticality effect varies by the function of the Experiment. Indeed, the grammaticality effect survived Holm's adjustment for multi-comparisons only in Experiment 1 (NP1 effect in **Table 4**; adjusted $p < 0.0001$) but not in Experiment 2 (NP2 effect in **Table 8**; adjusted $p =$ n.s.), suggesting that the grammaticality effect in Experiment 2 was relatively weaker than that in Experiment 1. On the other hand, there was no main effect of Illicit NP, but there was a significant interaction of Illicit NP and Experiment ($p < 0.001$). This seems to be due to a significant Illicit NP effect in Experiment 2 (NP1 effect in **Table 8**; adjusted *p*-value < 0.017) but not in Experiment 1 (NP2 effect in **Table 4**; adjusted $p =$ n.s.), suggesting that the Go-past reading times for honorific agreement were affected by an illicit NP in the object control construction but not in the subject control construction.

On the other hand, the analyses comparing Experiments 2 and 3 showed main effects of both Licit NP ($p < 0.0003$) and Illicit NP ($p < 0.002$), but there was no significant interaction with Experiment. Thus, we do not have evidence for a difference in the effect of the licit NP in the object control and the embedded clause construction, despite the fact that the significant grammaticality effect survived Holm's adjustment for multi-comparisons in Experiment 3 (NP2 effect in **Table 12**; adjusted $p < 0.014$) but not in Experiment 2. Likewise, there was no interaction of Illicit NP with Experiment, suggesting that the processing honorific *–si* dependency is sensitive to the properties of the illicit NP in the object control and embedded clause construction to a similar degree.

Overall, these results showed that (i) the effect of a licit NP was stronger in the subject control than in the object control construction, (ii) the effect of an illicit NP was stronger in the object control than in the subject control, and (iii) the effect of a licit NP and an illicit NP was found to a similar degree in the object control and the embedded clause construction. We address the implications of these findings in turn below.

First, the observations (i) and (ii) above suggests that the proximity of a distractor to a retrieval point does not modulate the attraction effect. While a marginal distractor effect was also found in the subject control construction (Experiment 1), the effect was only observed in the Total reading times, probably reflecting relatively late processing. In addition, we suspect that the distractor effect in Experiment 1 may have been spurious, given that it was inhibitory, while the overall pattern for the distractor effect was facilitatory in Experiments 2 and 3, as well as in the experiments reported by Kwon and Sturt (2016b). We therefore reserve judgment on the status of the distractor effect in Experiment 1, pending replications in further research. On the other hand, there are several possibilities why the distance manipulation did not affect the interference effect in the study. Previous studies have shown the effect of a temporal (or linear) distance during on-line sentence processing (Warren and Gibson, 2002). While it has been controversial, the claim is that when a linear distance is shorter between two linguistic

TABLE 13 | Generalized Linear Mixed Effects results comparing Go-past times of Experiments 1 and 2.

	Coeff.	SE	*t*	Slope	*p*	Adjusted *p*
(Intercept)	7.083	0.055	129.62		<2e-16***	<2e-16***
Licit NP	−0.092	0.013	−6.95*		4.78e-12***	3.346e-11***
Illicit NP	−0.019	0.013	−1.42		n.s.	n.s.
Experiments	−0.03	0.051	−0.59		n.s.	n.s.
Licit*Illicit	0.007	0.013	0.57		n.s.	n.s.
Licit*Experiments	0.036	0.013	2.73		0.0064*	0.032*
Illicit*Experiments	−0.05	0.013	−3.77		0.00016***	0.001**
Licit*Illicit*Experiments	−0.001	0.013	−0.09		n.s.	n.s.

*Coefficients, standard errors, t or z-values and p-values are reported for the main effects of the licit NP, the illicit NP and Experiment, as well as for the interactions of these three factors. Note that the effect of the illicit NP corresponds to the grammaticality effect, while the effect of illicit NP corresponds to the distractor effect. The "Slope" column indicates whether the random slope parameter corresponding to the effect was included in the model for participants (p) or items (i). P-values were obtained using LmerTest (Kuznetsova et al., 2017) and adjusted p-values were calculated using Holm's correction for multiple comparisons (Holm, 1979; Abdi, 2010). *p < 0.05, **p < 0.005, ***p < 0.0005.*

TABLE 14 | Generalized Linear Mixed Effects results comparing Go-past times of Experiments 2 and 3.

	Coeff.	SE	*t*	Slope	*p*	Adjusted *p*
(Intercept)	7.037	0.045	155.19		<2e-16***	<2e-16***
Licit NP	−0.075	0.015	−4.99	(p, i)	0.00003***	0.00021***
Illicit NP	−0.064	0.015	−4.22	(i)	0.00015***	0.0011**
Experiments	−0.015	0.042	−0.36		n.s.	n.s.
Licit*Illicit	0.017	0.013	1.36		n.s.	n.s.
Licit*Experiments	−0.007	0.013	−0.53		n.s.	n.s.
Illicit*Experiments	−0.009	0.013	−0.69		n.s.	n.s.
Licit*Illicit*Experiments	0.013	0.013	0.99		n.s.	n.s.

*Coefficients, standard errors, t or z-values and p-values are reported for the main effects of the licit NP, the illicit NP and Experiment, as well as for the interactions of these three factors. Note that the effect of the illicit NP corresponds to the grammaticality effect, while the effect of illicit NP corresponds to the distractor effect. The "Slope" column indicates whether the random slope parameter corresponding to the effect was included in the model for participants (p) or items (i). P-values were obtained using LmerTest (Kuznetsova et al., 2017) and adjusted p-values were calculated using Holm's correction for multiple comparisons (Holm, 1979; Abdi, 2010). **p < 0.005, ***p < 0.0005.*

items, they are easier to integrate together than when the distance between the two is longer (for details of such proposals, see Gibson, 1998, 2000; Lewis and Vasishth, 2005; Lewis et al., 2006; for a review, see Kwon et al., 2010). On the other hand, the results from the current study suggest that linear distance does not affect sentence processing during retrieval processes. Instead, it could be the case that as long as a distractor is re-activated above a certain threshold level during a retrieval process, it interferes with the processing of a subject-verb dependency, and the interference effect is not further modulated by the level of activation of the distractor at the point of retrieval. In fact, it should be also noted that there was no evidence that the linear distance affected the integration of the licit subject and its verb either. If the linear distance affected the integration difficulty, the subject control construction should have elicited longer reading times than the object control construction due to longer linear/temporal distance between the licit subject and the verb. There was, however, no evidence that integration of the licit subject and the verb was more difficult in the subject control construction (Experiment 1) than in the object control construction (Experiment 2). Thus, the overall results suggest that at least for the construction examined here, temporal (or linear) distance is not a factor affecting sentence processing during the retrieval (or integration) processes. Alternatively, it is possible that temporal (or linear) distance affects the retrieval process, but there are other factors, which are more important than the mere linear distance difference to affect interference effect, and thus the effect of the linear distance is overridden by other factors of importance. It is, of course, also possible that our design was not powerful enough to detect such differences. Our data do not distinguish between these possibilities, but we will return to this issue below.

As the observation (iii) above indicates, the results also did not provide evidence that the degree of feature overlap between a distractor and retrieval cues affected the interference effect of a distractor. The effects of the distractor were of a similar size in the two experiments, and showed the same facilitatory direction of effect. This contrasts with Parker and Phillips (2017), where a distractor associated with a larger number of matching features significantly affected the processing of a dependency involving a reflexive. However, unlike in our study, Parker and Phillips also manipulated retrieval cues of a licit target in addition to those of a distractor, and found the effect of a distractor when the number of matching features was reduced for the licit target but increased for the distractor. This suggests that interference from a distractor is not just sensitive to the memory representation of

a distractor, but also to activation of a distractor relative to that of a target.

On the other hand, lack of clear evidence in support of the case marker as a retrieval cue seems surprising given previous studies. For example, in a production study in Dutch Hartsuiker et al. (2001) showed that the attraction effect diminished when case marking of a distractor was clearly distinct from that of the licit subject, but increased when the case of the distractor was ambiguous. Likewise, importance of case information as a retrieval cue has been also discussed in several comprehension studies. For example, Van Dyke and McElree (2011) reported stronger interference effects when a distractor (e.g., *the neighbor*) appeared in a structurally similar position to that of the licit subject (e.g., *the worker* and *the resident* in "The worker was surprised that the resident who said the neighbor was dangerous was complaining about the investigation") than when a distractor (e.g., *the witness*) appeared in a structurally different position (e.g., the attorney *and the judge* in "The attorney who the judge realized had rejected the witness in the case compromised"). Similar results were also reported in Arnett and Wagers (2017). After examining the processing of subject-verb agreement in sentential complement, ECM, and object control construction sentences, Arnett and Wagers argued that the interference effect is modulated by the structural position as well as case properties of a distractor. All these results suggest that structural information (i.e., whether an NP is a subject or an object) is an important retrieval cue, but this argument was not confirmed in the current study. However, it should be noted that the clear interference effects reported in Van Dyke and McElree (2011) and Arnett and Wagers came from additional manipulations of comprehension difficulty, either involving semantic relatedness (Van Dyke and McElree, 2011) or semantic complexity (Arnett and Wagers). In contrast, reading time results of those studies were not straightforward. For example, Arnett and Wagers found different reading time results only for the comparison of sentential complement constructions with ECM construction, but no difference was found for the comparisons of object control constructions with other constructions. Given this, our results are not incompatible with these previous studies.

We consider how our results fit with the predictions of the cue-based retrieval model as proposed by Lewis and Vasishth (2005). The most straightforward version of this model predicts a small inhibitory effect of a matching (relative to a mismatching) distractor in grammatical conditions, and a larger facilitatory effect of a matching distractor in ungrammatical conditions. The reason for the inhibitory prediction in the grammatical conditions is that the partial match of the distractor with the retrieval cues decreases the activation of the target of the dependency, leading to a prediction of more processing difficulty, relative to a case where the distractor does not match in any features. In contrast, in ungrammatical conditions, the fact that both target and distractor partially match the retrieval cues leads to the distractor being mis-retrieved in a proportion of trials, leading to shorter average processing times (i.e., facilitation), relative to the situation where the distractor completely mismatches the retrieval cues.

Given these considerations, the Lewis and Vasishth (2005) model would predict an interaction between the matching of target and distractor NPs in our experiments. Specifically, for the grammatical conditions, where the target NP matches the honorific features of the verb, the model would have predicted longer reading times where the distractor also matches the honorific features, relative to when it does not. In contrast, for the ungrammatical conditions, where the target NP does not match the honorific features of the verb, the model would have predicted facilitation where the distractor is honorific, relative to when it is not. However, this specific form of interaction was not found in our experiments. Instead, in Experiments 2 and 3, we found a general effect of facilitation, which in most measures did not significantly differ between grammatical and ungrammatical conditions, while in Experiment 1, if anything, the evidence suggests a general inhibitory effect of the distractor, again not interacting with the matching of the licit NP subject.

On the basis of a large-scale meta-analysis, Jäger et al. (2017) point out that several studies of attraction in verb-subject agreement show facilitation for matching distractors, even in grammatical conditions, as we found in our Experiments 2 and 3, but contra the predictions of Lewis and Vasishth (2005). One suggestion that has recently been made by Engelmann et al. (2019) is that the direction of the distractor effect in grammatical conditions may depend on the relative prominence of the target of the dependency and the distractor. In Engelmann et al.'s extended version of the Lewis and Vasishth (2005) model, the prominence of the distractor affects its baseline activation, leading to a prediction of facilitatory interference in grammatical conditions, where the distractor is particularly prominent, for example when the distractor is a main clause subject. In fact, we believe that our experimental results are consistent with this prediction. The distractor in both of Experiments 2 and 3 was a main clause subject, and we found robust evidence of a distractor effect in both of the experiments, with a general facilitatory effect of the distractor. In contrast, when a distractor was a dative argument with low prominence, and when the licit target of the dependency is the main clause subject as in our Experiment 1, there was no clear effect of a distractor. These results suggest that relative prominence of distractor in comparison to that of a target could affect the retrieval processes during the processing of a dependency. If so, it is also possible that the relative prominence of distractor could have over-ridden any effect that a linear distance could have.

On the other hand, we believe that prominence of a distractor could vary depending the nature of a dependency. It is likely that a main clause subject is particularly more prominent than a dative marked object when processing a subject-verb dependency. It is an empirical question whether similar levels of prominence will be observed for a main clause subject or an object when a dependency is not relevant for subjecthood, for example, as in the case of the processing of object-verb agreement. That is, it is suggested that individual languages might differ in the relative weight assigned to various sources of information used for language processing, and that this could be due to typological

variations in the importance of those cues (Kwon and Sturt, 2013). Given this, it is also a possibility that cues employed for language processing are weighted in a given context. In other words, different levels of prominence could be associated with different cues in a different context. However, future research should examine the effect of distractor prominence more systematically.

Finally, the results of the current study are not compatible with the hypothesis that the attraction effect is an error-driven processing mechanism (Pearlmutter et al., 1999; Wagers et al., 2009; Dillon et al., 2013; Lago et al., 2015; cf. Van Dyke, 2007). First, attraction effect was not found just for ungrammatical sentences but also for grammatical sentences. Second, in Experiment 2, attraction effect even preceded grammaticality effect. These results are consistent with previous findings of honorific agreement in Korean, supporting the view that attraction effect arises from general working memory principles (Kwon and Sturt, 2016b). That is, during the processing of a dependency, items with matching retrieval features are activated, even the feature match is only partial, affecting the processing of the dependency.

In summary, the current study investigated whether and how attraction effects would be modulated by the memory representation of a distractor, by examining the subject-verb honorific agreement in Korean. Our study did not find evidence that proximity of a distractor to the retrieval point (i.e., higher activation level of a distractor) increased interference effects of the distractor. Similarly, we did not find evidence that a higher number of matching cues of a distractor triggered more misretrieval. Instead, the results suggested that interference is not just sensitive to memory representation of a distractor but rather to activation of a distractor relative to that of a target. Our results are also consistent with Engelmann et al. (2019)'s proposal that the prominence of the distractor affects the direction of the distractor effect in grammatical conditions.

DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

We obtained full IRB approval for the study from the Human Research Protections Programs at Konkuk University.

AUTHOR CONTRIBUTIONS

NK and PS were both involved in the design of the study, and prepared the manuscript. NK carried out the experiment and analyzed the data.

FUNDING

This research was supported by the faculty research fund of Konkuk University in 2017.

REFERENCES

Abdi, H. (2010). "Holm's sequential Bonferroni procedure," in *Encyclopedia of Research Design*, ed N. Salkind (Thousand Oaks, CA: Sage), 574–578.

Arnett, N., and Wagers, M. (2017). Subject encodings and retrieval interference. *J. Mem. Lang.* 93, 22–54. doi: 10.1016/j.jml.2016.07.005

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801686

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Barrouillet, P., Portrat, S., Vergauwe, E., Diependaele, K., and Camos, V., (2011). Further evidence for temporal decay in working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1302–1317. doi: 10.1037/a0022933

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Berman, M. G., Jonides, J., and Lewis, R. L. (2009). In search of decay in verbal short-term memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 317–353. doi: 10.1037/a0014873

Bock, J. K., and Cutting, J. C. (1992). Regulating mental energy: performance units in language production. *J. Mem. Lang.* 31, 99–127. doi: 10.1016/0749-596X(92)90007-K

Bock, J. K., and Eberhard, K. M. (1993). Meaning, sound, and syntax in English number agreement. *Lang. Cogn. Process.* 8, 57–99. doi: 10.1080/01690969308406949

Bock, J. K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93.

Dillon, B., Mishler, A., Sloggett, S., and Phillips, C. (2013). Contrasting interference profiles for agreement and anaphora: experimental and modeling evidence. *J. Mem. Lang.* 69, 85–103. doi: 10.1016/j.jml.2013.04.003

Drenhaus, H., Saddy, D., and Frisch, S. (2005). "Processing negative polarity items: when negation comes through the backdoor," in *Linguistic Evidence-Empirical, Theoretical, and Computational Perspectives,* eds S. Kepser and M. Reis (Berlin: Mouton de Gruyter), 145–165. doi: 10.1515/9783110197549.145

Engelmann, F., Jäger, L. A., and Vasishth, S. (2019). *The Effect of Prominence and Cue Association in Retrieval Processes: A Computational Account*. University of Manchester.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain: Papers From the First Mind Articulation. Project Symposium,* eds A. P. Marantz, Y. Miyashita, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1411–1423. doi: 10.1037/0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., and Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *J. Mem. Lang.* 51, 97–114. doi: 10.1016/j.jml.2004.02.003

Gordon, P. C., Hendrick, R., Johnson, M., and Lee, Y. (2006). Similarity-based interference during language comprehension: evidence from eye tracking during reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1304–1321. doi: 10.1037/0278-7393.32.6.1304

Hartsuiker, R. J., Antón-Méndez, I., and van Zee, M. (2001). Object attraction in subject–verb agreement construction. *J. Mem. Lang.* 45, 546–573. doi: 10.1006/jmla.2000.2787

Haskell, T. R., and MacDonald, M. C. (2003). Conflicting cues and competition in subject–verb agreement. *J. Mem. Lang.* 48, 760–778. doi: 10.1016/S0749-596X(03)00010-X

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometr. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Jäger, L. A., Engelmann, F., and Vasishth, S. (2017). Similarity-based interference in sentence comprehension: literature review and Bayesian meta-analysis. *J. Mem. Lang.* 94, 316–339. doi: 10.1016/j.jml.2017.01.004

Kaan, E. (2002). Investigating the effects of distance and number interference in agreement processing: an ERP study. *J. Psycholinguist. Res.* 31, 165–193. doi: 10.1023/A:1014978917769

Kintsch, W. (1970). "Models for free recall and recognition," in *Models of Human Memory*, ed D. A. Norman (New York, NY: Academic Press), 331–373. doi: 10.1016/B978-0-12-521350-9.50016-4

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). LmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Kwon, N., Lee, Y., Gordon, P. C., Kluender, R., and Polinsky, M. (2010).Cognitive and linguistic factors affecting subject/object asymmetry: an eye-tracking study of pre-nominal relative clauses in Korean. *Language* 86, 546–582. doi: 10.1353/lan.2010.0006

Kwon, N., and Sturt, P. (2013). Null pronominal (pro) resolution in Korean, a discourse-oriented language. *Lang. Cogn. Process.* 28, 377–387. doi: 10.1080/01690965.2011.645314

Kwon, N., and Sturt, P. (2014). The use of control information in dependency formation: an eye-tracking study. *J. Mem. Lang.* 73, 59–80. doi: 10.1016/j.jml.2014.02.005

Kwon, N., and Sturt, P. (2016a). Processing control information in a nominal control construction: an eye-tracking study. *J. Psycholinguist. Res.* 45, 779–793. doi: 10.1007/s10936-015-9374-2

Kwon, N., and Sturt, P. (2016b). Attraction effects in honorific agreement in Korean. *Front. Psychol.* 7:1302. doi: 10.3389/fpsyg.2016.01302

Lago, S., Shalom, D., Sigman, M., Lau, E., and Phillips, C. (2015). Agreement processes in Spanish comprehension. *J. Mem. Lang.* 82, 133–149. doi: 10.1016/j.jml.2015.02.002

Lee, Y., Lee, H., and Gordon, P. C. (2006). Linguistic complexity and information structure in Korean: evidence from eye-tracking during reading. *Cognition* 104, 495–534. doi: 10.1016/j.cognition.2006.07.013

Lewandowsky, S., Oberauer, K., and Brown, G. D. (2009). No temporal decay in verbal short-term memory. *Trends Cogn. Sci.* 13. 120–126. doi: 10.1016/j.tics.2008.12.003

Lewis, R. L. (1996). Interference in short-term memory: the magical number two (or three) in sentence processing. *J. Psycholinguist. Res.* 25, 93–115. doi: 10.1007/BF01708421

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 1–45. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., and Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* 10, 447–454. doi: 10.1016/j.tics.2006.08.007

McElree, B. (2000). Sentence comprehension is mediated by content addressable memory. *J. Psycholinguist. Res.* 29, 111–123. doi: 10.1023/A:1005184709695

McElree, B., Foraker, S., and Dyer, L. (2003). Memory structures that subserve sentence comprehension. *J. Mem. Lang.* 48, 67–91. doi: 10.1016/S0749-596X(02)00515-6

Nairne, J. S. (2002). Remembering over the short-term: the case against the standard model. *Annu. Rev. Psychol.* 53, 53–81. doi: 10.1146/annurev.psych.53.100901.135131

Nicol, J. L., Forster, K. I., and Veres, C. (1997). Subject–verb agreement processes in comprehension. *J. Mem. Lang.* 36, 569–587. doi: 10.1006/jmla.1996.2497

Oberauer, K., and Kliegl, R. (2006). A formal model of capacity limits in working memory. *J. Mem. Lang.* 55, 601–626. doi: 10.1016/j.jml.2006.08.009

Parker, D., and Phillips, C. (2017). Reflexive attraction in comprehension is selective. *J. Mem. Lang.* 94, 272–290. doi: 10.1016/j.jml.2017.01.002

Pearlmutter, N. J., Garnsey, S. M., and Bock, K. (1999). Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456. doi: 10.1006/jmla.1999.2653

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: http://www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372

Sohn, H.-M. (1999). *The Korean Language*. Cambridge: Cambridge University Press.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562. doi: 10.1016/S0749-596X(02)00536-3

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition* 105, 477–488. doi: 10.1016/j.cognition.2006.10.009

Sturt, P., and Kwon, N. (2015). The processing of raising and nominal control: an eye-tracking study. *Front. Psychol.* 6:331. doi: 10.3389/fpsyg.2015.00331

Tanner, D., Nicol, J., and Brehm, L. (2014). The time course of feature interference in agreement comprehension: multiple mechanisms and asymmetrical attraction. *J. Mem. Lang.* 76, 95–215. doi: 10.1016/j.jml.2014.07.003

Thornton, R., and MacDonald, M. C. (2003). Plausibility and grammatical agreement. *J. Mem. Lang.* 48, 740–759. doi: 10.1016/S0749-596X(03)00003-2

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 407–430. doi: 10.1037/0278-7393.33.2.407

Van Dyke, J. A., and McElree, B. (2011). Cue-dependent interference in comprehension. *J. Mem. Lang.* 65, 247–263. doi: 10.1016/j.jml.2011.05.002

Vasishth, S., Brüssow, S., Lewis, R., and Drenhaus, H. (2008). Processing polarity: how the ungrammatical intrudes on the grammatical. *Cogn. Sci.* 32, 685–712. doi: 10.1080/03640210802066865

Vigliocco, G., and Nicol, J. (1998). Separating hierarchical relations and word order in language production: is proximity concord syntactic or linear? *Cognition* 68, 13–29. doi: 10.1016/S0010-0277(98)00041-9

von der Malsburg, T., and Angele, B. (2017). False positives and other statistical errors in standard analyses of eye movements in reading. *J. Mem. Lang.* 94, 119–133. doi: 10.1016/j.jml.2016.10.003

Wagers, M., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2

Xiang, M., Dillon, B., and Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain Lang.* 108, 40–55. doi: 10.1016/j.bandl.2008.10.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 June 2019
doi: 10.3389/fpsyg.2019.01152

Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences

Émilie Courteau[1,2†], *Lisa Martignetti*[2,3†], *Phaedra Royle*[1,2] *and Karsten Steinhauer*[2,3*]

[1] Faculty of Medicine, School of Speech Language Pathology and Audiology, University of Montreal, Montreal, QC, Canada, [2] Centre for Research on Brain, Language and Music (CRBLM), Montreal, QC, Canada, [3] Faculty of Medicine, School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada

The present event-related brain potential (ERP) study investigates mechanisms underlying the processing of morphosyntactic information during real-time auditory sentence comprehension in French. Employing an auditory-visual sentence-picture matching paradigm, we investigated two types of anomalies using entirely grammatical auditory stimuli: (i) semantic mismatches between visually presented actions and spoken verbs, and (ii) number mismatches between visually presented agents and corresponding morphosyntactic number markers in the spoken sentences (determiners, pronouns in liaison contexts, and verb-final "inflection"). We varied the type and amount of number cues available in each sentence using two manipulations. First, we manipulated the verb type, by using verbs whose number cue was audible through subject (clitic) pronoun liaison (liaison verbs) as well as verbs whose number cue was audible on the verb ending (consonant-final verbs). Second, we manipulated the pre-verbal context: each sentence was preceded either by a neutral context providing no number cue, or by a subject noun phrase containing a subject number cue on the determiner. Twenty-two French-speaking adults participated in the experiment. While sentence judgment accuracy was high, participants' ERP responses were modulated by the type of mismatch encountered. Lexico-semantic mismatches on the verb elicited the expected N400 and additional negativities. Determiner number mismatches elicited early anterior negativities, N400s and P600s. Verb number mismatches elicited biphasic N400-P600 patterns. However, pronoun + verb liaison mismatches yielded this pattern only in the plural, while consonant-final changes did so in the singular and the plural. Furthermore, an additional sustained frontal negativity was observed in two of the four verb mismatch conditions: plural liaison and singular consonant-final forms. This study highlights the different contributions of number cues in oral language processing and is the first to investigate whether auditory-visual mismatches can elicit errors reminiscent of outright grammatical errors. Our results emphasize that neurocognitive mechanisms underlying number agreement in French are modulated by the type of cue that is used to identify auditory-visual mismatches.

Keywords: subject-verb number agreement, event-related brain potentials (ERPs), auditory-visual sentence-picture matching paradigm, cross-modal number mismatches, French language, online grammaticality judgment, N400 and P600, sustained frontal negativity

OPEN ACCESS

Edited by:
Sendy Caffarra, Basque Center on Cognition, Brain and Language, Spain

Reviewed by:
Patrick Sturt, University of Edinburgh, United Kingdom
Ryan C. Taylor, University of British Columbia, Vancouver, BC, Canada

***Correspondence:**
Karsten Steinhauer
karsten.steinhauer@mcgill.ca

[†] *First authors*

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *22 December 2018*
Accepted: *01 May 2019*
Published: *25 June 2019*

Citation:
Courteau É, Martignetti L, Royle P and Steinhauer K (2019) Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences. Front. Psychol. 10:1152. doi: 10.3389/fpsyg.2019.01152

INTRODUCTION

Few ERP studies have investigated real-time auditory sentence comprehension in French. Importantly, French subject-verb agreement has specific properties (such as clitic-verb liaison, see below) that are relevant to the study of agreement processing but have not received much attention in the ERP literature. Furthermore, many studies of agreement rely on visual word presentation, where morphosyntactic information is presented simultaneously with other lexical information, rather than unfolding over time, as in natural spoken language. These reading studies may not capture temporal aspects typical of spoken language processing, and ERP components may differ across modalities. Moreover, there is increasing interest in ERP methods that do not rely on violation paradigms. Considering these issues, we developed an ERP study where we implemented an auditory-visual sentence-picture matching task to investigate on-line processing of lexico-semantic and morphosyntactic information. Creating mismatches between grammatical auditory sentences and picture stimuli has been shown to elicit ERPs in lexico-semantic noun mismatches (e.g., Willems et al., 2008). To our knowledge, these mismatches between modalities have not been used to study morphosyntactic processing, nor lexico-semantic verb mismatches. Therefore, we examined whether the auditory presentation of a grammatical sentence combined with a picture that doesn't match its morphosyntactic features would elicit the same ERP components as in classic paradigms using ungrammatical sentences. Our innovative approach is motivated by the long-term aim of our research program, which is to study language processing in children with developmental language disorder (previously referred to as specific language impairment, SLI) using ecologically valid stimuli. Combining images and speech resembles other common activities such as shared picture-book reading, or watching documentary or educational videos, where an image is presented concurrently with an oral description. In these cases, people being read or spoken to might make predictions about what the reader will say, and notice any incongruencies, as participants were expected to do during our experiment. Thus, we investigate: (i) lexical-semantic mismatches between visually presented actions and spoken verbs, and (ii) auditory-visual subject number mismatches while varying number-cue types at different positions in the sentence. These manipulations should allow us to better understand how French-speakers handle semantic and grammatical cues online and should also elucidate if cross-modal paradigms elicit similar ERP components as classic within-sentence agreement violations. We will first review relevant ERP findings and then develop our research questions.

In ERPs, lexical-semantic processing is typically reflected by the centro-parietal N400 component between 300–500 ms after word onset (Kutas and Hillyard, 1980). This brain wave can be elicited by lexical-semantic expectancy violations (Steinhauer and Connolly, 2008; Kutas and Federmeier, 2011). Its amplitude may reflect processing effort during lexical retrieval (Lau et al., 2008) and post-lexical integration (Steinhauer et al., 2017), or it can be described as an error signal reflecting the difference between one's lexical-semantic expectations (i.e., the "current model") and the actual word input (Bornkessel-Schlesewsky and Schlesewsky, 2019; henceforth BSS2019). Although most evidence for N400s has come from reading studies, this component has also been observed in bimodal (auditory-visual) lexical-semantic violations where an incongruous image is presented concurrently with an auditory utterance, for instance: *Je vois un !soulier vert sur la table* "I see a green !shoe on the table" with an image of a [HAT on a table] (Royle et al., 2013; see also Friedrich and Friederici, 2004; Willems et al., 2008). The N400 is generally considered a reliable ERP correlate of increased lexico-semantic processing difficulties.

Morphosyntactic agreement-error processing in reading studies is often indexed by one or two components, the left anterior negativity (LAN) and a later positive shift (the P600). The LAN has been reported for a range of morphosyntactic violations, including subject-verb agreement violations (e.g., *As a turtle grows, its shell *grow too* Kutas and Hillyard, 1983), especially in languages with relatively free word order and rich morphological agreement marking (Angrilli et al., 2002; Barber and Carreiras, 2005), but also in languages with less rich paradigms (Osterhout and Mobley, 1995; Hagoort and Brown, 2000). Like the N400, this component typically emerges between 300 and 500 ms after stimulus presentation. Most agreement studies eliciting LANs have been conducted in the written modality, but some auditory studies have also reported LAN-like negativities for a range of morpho-syntactic anomalies (Friederici et al., 1993; Balconi and Pozzoli, 2005; Rossi et al., 2006; Hasting and Kotz, 2008; Morgan-Short et al., 2010; Dube et al., 2016; Haebig et al., 2017). Compared to reading studies, LANs in auditory studies tend to have an earlier onset, and can have a much longer duration (∼100–1,200 ms), and a bilateral frontal distribution (e.g., Hasting and Kotz, 2008). However, several reading studies do not report LANs for agreement violations (Osterhout and Mobley, 1995; Tokowicz and MacWhinney, 2005; Lau et al., 2006; Nevins et al., 2007; Foucart and Frenck-Mestre, 2011, 2012) and report only P600s (see below). Whether or not LANs are reliable reflections of ERP morphosyntactic processes, whether different ERP morphologies reflect distinct processes, and what their functional significance may be, is therefore under debate (Molinaro et al., 2011a; Steinhauer and Drury, 2012; Royle et al., 2013; Tanner, 2015).

The LAN is usually followed by a late parietal positive-going component, the P600, roughly between 500 and 1,000 ms (Osterhout and Holcomb, 1992, 1993; Hahne and Friederici, 1999; Steinhauer et al., 1999). In contrast to the LAN, the P600 is widely viewed as the most consistent ERP signature for a large range of grammatical anomalies. It has been observed for morphosyntactic agreement violations (Frenck-Mestre et al., 2008; Foucart and Frenck-Mestre, 2011, 2012; Molinaro et al., 2011a; Royle et al., 2013), syntactic violations (Friederici, 2002), garden path sentences (Osterhout and Holcomb, 1992; Holcomb, 1993), and has also been elicited by semantic anomalies in conjunction with N400s (Hagoort, 2003; Steinhauer et al., 2010; Royle et al., 2013). While many agree that the P600 is a brain response related to controlled sentence reanalysis and

repair (Hahne and Friederici, 1999), some argue that it is an ERP correlate of implicit syntactic processing (Tokowicz and MacWhinney, 2005). Another interpretation is that the P600 is a member of the parietal P300 (P3b) family of components reflecting stimulus categorization (e.g., in an acceptability judgment task) (Royle et al., 2013; Sassenhagen et al., 2014; BSS2019).

ERP studies have also revealed different patterns for various agreement error types. A majority of studies on agreement are reading tasks, and most use serial word-by-word visual presentation. Molinaro et al. (2011a) present a review of number and gender agreement processing in various languages. Regarding subject-verb number agreement violations, of 17 studies reviewed, all revealed P600s and 13 revealed LANs. The authors correlate the LAN with morphosyntactic error processing and explain the absence of a LAN in certain studies by differences in morphosyntactic saliency. For example, when these are underspecified (i.e., not morphologically expressed on the singular), a LAN may not be triggered. Molinaro et al. (2011b), found that in conditions such as **Il ragazzo e la ragazza corre…* "The boy and the girl run.3rd.SINGULAR," the conjoined noun phrase (NP) does not contain any overt plural marking and in its absence no LAN is triggered. However, Frenck-Mestre et al. (2008) do not observe any negativities resembling a LAN but find a P600 in French native speakers in response to subject-verb agreement violations such as **Le matin je mangez* [mãʒe] "In the morning I eat.2nd.PLURAL." Their data contradict Molinaro's (2011a) interpretation, as the LAN was absent even though subject number properties were clearly expressed by the singular pronoun *je* "I" as well as the verb *mangez*.

In sum, while both the P600 and the LAN can be observed following various agreement-error types, it is still unclear whether they are modulated by the languages, structures, or contexts used to elicit them. The present study attempts to answer the following questions, using entirely grammatical sentences in all conditions. First, whether French speakers will elicit an N400 component for cross-modal (audio-visual) lexico-semantic mismatches realized on actions/verbs—rather than nouns/objects—and whether this violation type will elicit P600s as observed in other cross-modal lexico-semantic mismatch studies. Second, whether cross-modal number mismatches between the pictures' agents and the determiners/pronouns or verb morphology in our sentences elicit biphasic LAN/N400-P600 complexes as in previous morphosyntactic violation studies. To the best of our knowledge, this has not been investigated before. Given that our sentences were grammatical, one could argue that cross-modal number mismatches may cause either (a) conceptual-semantic problems typically associated with N400s instead of LANs, or (b) logical-semantic conflicts related to truth values, which have be found to elicit local N400s or sentence wrap-up effects (Bokhari, 2015) and P600s followed by (but not preceded by) late LANs (L-LANs; cf. Steinhauer et al., 2010). The third question was whether participants, when presented with multiple cues for number mismatch disambiguation, will rely on the first available auditory cue, as indicated by ERP responses.

MATERIALS AND METHODS

Participants

Twenty-eight neurotypical adults aged 18–40 years participated in the experiment. The protocol was approved by Institutional Review Boards at McGill and University of Montreal (UdeM). All participants gave written informed consent in accordance with the Declaration of Helsinki. All were right-handed as assessed using the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision, French as their mother tongue and their everyday language, and did not learn any other language before age 5. None had learning disabilities, neurological damage, or hearing loss. Working memory was assessed orally at session's end. Participants were recruited from Montreal university student populations. Participants were compensated $45 for their time (3.5 h). Six data sets had to be excluded due to excessive eye movement artifacts, such that data from 22 participants were retained for analyses (range: 18–38 years; mean 25; 12 female, 10 male). We consider this sample size as enough to provide a good estimate of the effects of interest, since in Royle et al. (2013) a group of 15 French-speaking adults participating in a similar paradigm (7 in a task-based group and 8 in a no-task one) showed significant ERPs related to adjective agreement errors and noun-image semantic incongruencies in each group.

Materials and Design

As illustrated in **Tables 1–3**, materials consisted of spoken grammatical sentences in French, half of which mismatched with a concurrently-displayed picture, either through the action described or the number of agents (singular/plural mismatch). As we developed the study for younger populations (to be tested after adults), word selection was constrained by age-of-acquisition norms (see **Supplementary Materials** for details). Verbs were presented within sentences containing third person singular or plural subject pronouns (*he/she/they*), and a sentence continuation with a direct object NP, or prepositional phrase (PP, e.g., … *in the public pool*) to avoid sentence-final (or "wrap-up") effects in ERPs time-locked to verbs (Hagoort, 2003; see also Stowe et al., 2018). Verbs were selected based on their number-agreement morphological characteristics, as explained below.

Selected critical verbs were inspired by the fLEX evaluation tool (Pourquie et al., 2016), with their imageability in mind, as they were presented alongside illustrations, and were matched on lemma frequency, age of emergence, and length (syllables and phonemes). Auditory stimulus recording, normalizing, and splicing was supervised by trained research assistants with a background in speech editing (**Supplementary Materials**). For each sentence, one color drawing was created by a professional artist, emphasizing the action being described, and the agent(s) carrying it out. Drawings maintained a constant visual complexity level, avoiding superfluous or distracting details.

In order to enhance the comparability of ERP effects between semantic and number mismatches, we decided to create semantic mismatches on the verb, the main element disambiguating mismatches in our number conditions (see below). Thus, for

TABLE 1 | Experimental sub-conditions for lexico-semantic manipulations and a corresponding visual stimulus.

Visual stimulus

Sample visual stimulus presented concurrently with auditory stimuli for matching lexico-semantic conditions (1a-b) and mismatching ones (2a-b). Note that, in addition to the mismatch at the target verb ("sings" vs. "swims"), conditions 2a-b also include a second semantic mismatch in the prepositional phrase (here: "concert venue" vs. "public pool").

Condition	Context	Sample	auditory stimuli
Congruent semantics	Neutral	(1a)	*Chaque semaine* | *elle chante dans une salle de concert* "Each week | she sings at a concert venue"
	Subject NP	(1b)	*La vedette* | *elle chante dans une salle de concert* "The star | she sings at a concert venue"
Incongruent semantics	Neutral	(2a)	*Chaque semaine* | *elle !nage dans la piscine publique* "Each week | she !swims in the public pool"
	Subject NP	(2b)	*La vedette* | *elle !nage dans la piscine publique* "The star | she !swims in the public pool"

Critical words are underlined. Subj NP, overt subject noun phrase; !, lexico-semantic mismatch; |, cross-splicing point.

semantic mismatches, the spoken verb did not correspond to the depicted action (e.g., the sound file described "she swims…" and the image depicted "she sings…"). Sentences in this condition were created with 60 invariable regular verbs, 30 with a singular and 30 with a plural pronoun ("he/she," "they"). Each pronoun+verb item was then combined with (a) a subject NP context providing a lexical NP with early number information (e.g., "*The*.PLURAL *girls*, they swim"[1] and (b) a neutral context without number information (e.g., "*In the evening*, they swim"), resulting in 120 spoken items. In total, 240 stimuli were created; 120 congruent and 120 in incongruent ones, by splicing the incongruent verb into the sentence (see e.g., **Tables 1**, **2A**).

Number mismatches between the depicted subject and the one presented in the auditory stimulus (e.g., the sound file describes "she swims" and the image depicts "they swim") were realized at different sentence positions using cross-splicing techniques (see **Tables 2**, **3**).

Two verb types were used: 60 liaison (LIAIS) verbs and 60 consonant-final (CONS) verbs. LIAIS verbs had vowel onsets and were regular 1st conjugation verbs, such as *aimer* "to-love," which provide no audible cues or disambiguation between 3rd person singular (*aime* [ɛm]) and plural forms (*aiment* [ɛm]). This allowed us to ensure that the only cue for number disambiguation was located at the junction (liaison) between the subject pronoun and the verb, indexed by the presence or absence of the pronoun's plural marker "*s*" [z] (e.g., *elle aime* [ɛlɛm] "she loves" vs. *elles‿aiment* [ɛl**z**ɛm] "they love"). Unlike LIAIS verbs, CONS verbs were from the 2nd and 3rd conjugation classes, such as *rugir* "to-roar," where number distinctions between singular and plural forms are audible on verb endings (e.g., *il rugit* [ilʁyʒ**i**] "he roars" vs. *ils rugi**ss**ent* [ilʁyʒɪ**s**] "they roar"). This was the only number cue provided by CONS verbs. A total of 120 verbs (60 LIAIS and 60 CONS) were produced in singular and plural sentences, with both NP and neutral contexts. This resulted in 480 audio files and 960 stimuli: 480 in the congruent condition, and 480 in the incongruent one, where there was a mismatch between the spoken sentence and the picture's verb number.

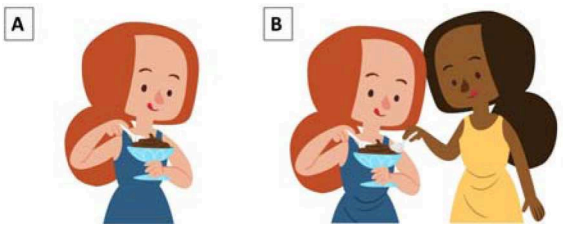
The 1,200 different sentence-picture combinations (240 for conceptual semantics and 960 for agreement) were evenly distributed across four lists (with no sentence repetition within a given list). Three hundred stimuli sentences with accompanying images were presented to participants in each list (60 for conceptual semantics and 240 for morphosyntax) and were pseudo-randomized (see **Supplementary Materials** for details). Item versions for each condition were distributed across lists as follows: For semantics, one version of a given verb was included in each list, such that a participant heard one audio file and saw one image (either congruent or incongruent) for each verb. For each LIAIS and CONS verb type two sentence versions of a given verb were included in each list. These sentences were maximally distinct such that they differed in: (1) number (singular vs. plural), (2) context type (neutral vs. subject NP), and (3) congruency (match vs. mismatch with the image), and were presented in different halves of the experiment. This entailed that each subject be presented the same image twice (one match and one mismatch context), but with two completely different audio files.

Procedure

Experimental sessions took place in a quiet room at the UdeM in the third author's lab. Upon arrival, participants read and signed the consent form, after which they completed

[1]Note that in oral Quebec French, a subject with an overt NP "The girl" followed by a pronoun "she" is grammatical (some say the pronoun is obligatory) contrary to written French.

TABLE 2 | Experimental sub-conditions involving liaison (LIAIS) verbs and corresponding visual stimuli.

Visual stimulus	
	Image A: sample visual stimulus for match conditions (1a-b) and mismatch conditions (2c-d) in the singular. Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d).

Condition	Number	Context	Sample	auditory stimuli
Congruent morphosyntax	Singular	Neutral	(1a)	*Au dessert* | *elle aime la mousse au chocolat* "For desert | she likes chocolate mousse"
		Subject NP	(1b)	*La fille* | *elle aime la mousse au chocolat* "*The* girl | she likes chocolate mousse"
	Plural	Neutral	(2a)	*Au dessert* | *elles⌣ aiment la mousse au chocolat* "For desert | they like chocolate mousse"
		Subject NP	(2b)	*Les filles* | *elles⌣ aiment la mousse au chocolat* "*The* girls | they like chocolate mousse"
Incongruent morphosyntax	Singular	Neutral	(1c)	*Au dessert* | *elle * aime la mousse au chocolat* "For desert | she *likes chocolate mousse"
		Subject NP	(1d)	**La fille* | *elle * aime la mousse au chocolat* "**The* girl | she *likes chocolate mousse"
	Plural	Neutral	(2c)	*Au dessert* | *elles⌣* aiment la mousse au chocolat* "For desert | they *like chocolate mousse"
		Subject NP	(2d)	**Les filles* | *elles⌣* aiment la mousse au chocolat* "**The* girls | they *like chocolate mousse"

*Critical words are underlined. Subj NP, overt subject noun phrase; *, number mismatch; |, cross-splicing point.*

the Edinburgh Handedness Inventory (Oldfield, 1971) and a language background questionnaire. They were then fitted with an EEG cap, and completed three sub-experiments, all of which used an auditory-visual sentence-picture matching paradigm. The first and second study examined gender-agreement processing (Royle et al., 2013) and word order in French noun phrases. Data from the third experiment are reported here. Total session duration was ∼3.5 h, including consent form and other questionnaire completion, WM test administration, preparation, and clean up.

Participants were seated at a desk at a distance of ∼40 cm from a computer monitor. Sentences and images were presented using an "Alien learning paradigm," where an alien visited Quebec and was learning French. A story containing filler sentences, images and animations was created. These were interspersed throughout the experiment to maintain interest and attention. Participants listened to spoken sentences presented binaurally via insert earphones (ER-1 Insert Earphones, Etymotic Research), while images were presented on the computer monitor. A pause was programmed after every three experimental blocks (60 items).

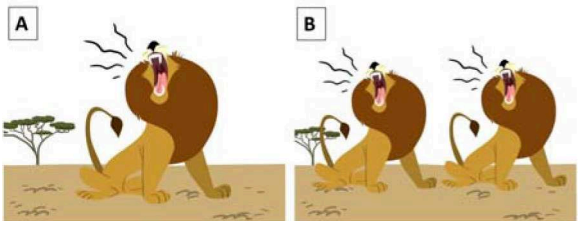
Participants were instructed to listen to each sentence, while attending to all aspects of grammar and meaning, and judge sentence acceptability in relation to the simultaneously presented image by pressing one of two keys on a response pad ("acceptable" or "not acceptable"). In order to avoid laterality effects, the "acceptable" button was programmed on the right side of the pad for half the participants, and the left side for the other half. Participants were instructed to minimize movement and to keep their eyes open during stimuli presentation. Six practice trials were presented at experiment onset and were excluded from subsequent analyses. At least one researcher or assistant was present throughout the session. EEG recording was monitored throughout, and participants were given feedback about eye blinks and other body movements whenever necessary, in order to reduce artifacts.

Each trial began with a fixation cross centered on the screen 1,000 ms before stimulus presentation. The image was presented 500 ms before sentence onset, and stayed on screen until the auditory stimulus ended. Then, a response prompt ("???") appeared on the screen until a response button was pressed, followed by a blank screen for 1,000 ms, during which subjects were instructed to blink their eyes before the next trial began.

Analysis Time-Locking

In order to quantify the time course of number mismatch and lexical-semantic effects, our analyses were time-locked to relevant lexical-semantic and morphophonological cues (Steinhauer and Drury, 2012), using triggers at relevant speech signal positions. **Figure 1** depicts an example waveform for the sentence *Le lion, il rugit dans la savane* "The lion, he roars in the savannah" as well

TABLE 3 | Experimental sub-conditions involving consonant-final (CONS) verbs, and corresponding visual stimuli.

Visual stimulus	
	Image A: sample visual stimulus for match conditions (1a-b) and mismatch conditions (2c-d). Image B: sample visual stimulus for match (2a-b) and mismatch conditions (1c-d)

Condition	Number	Context	Sample	auditory stimuli
Congruent morphosyntax	Singular	Neutral	(1a)	*En soirée* | *il rugit dans la savane* In the evening | he *roars* in the savannah
		Subject NP	(1b)	*Le lion* | *il rugit dans la savane* *The* lion | he *roars* in the savannah
	Plural	Neutral	(2a)	*En soirée* | *ils rugissent dans la savane* In the evening | they *roar* in the savannah
		Subject NP	(2a)	*Les lions* | *ils rugissent dans la savane* *The* lions | they *roar* in the savannah
Incongruent morphosyntax	Singular	Neutral	(1c)	*En soirée* | *il *rugit dans la savane* During evening | he **roars* in the savannah
		Subject NP	(1d)	**Le lion* | *il *rugit dans la savane* *The* lion | he **roars* in the savannah
	Plural	Neutral	(2c)	*En soirée* | *ils *rugissent dans la savane* In the evening | they **roar* in the savannah
		Subject NP	(2d)	**Les lions* | *ils *rugissent dans la savane* *The* lions | they **roar* in the savannah

*Critical words are underlined. Subj NP, overt subject noun phrase; *, number mismatch; |, cross-splicing point.*

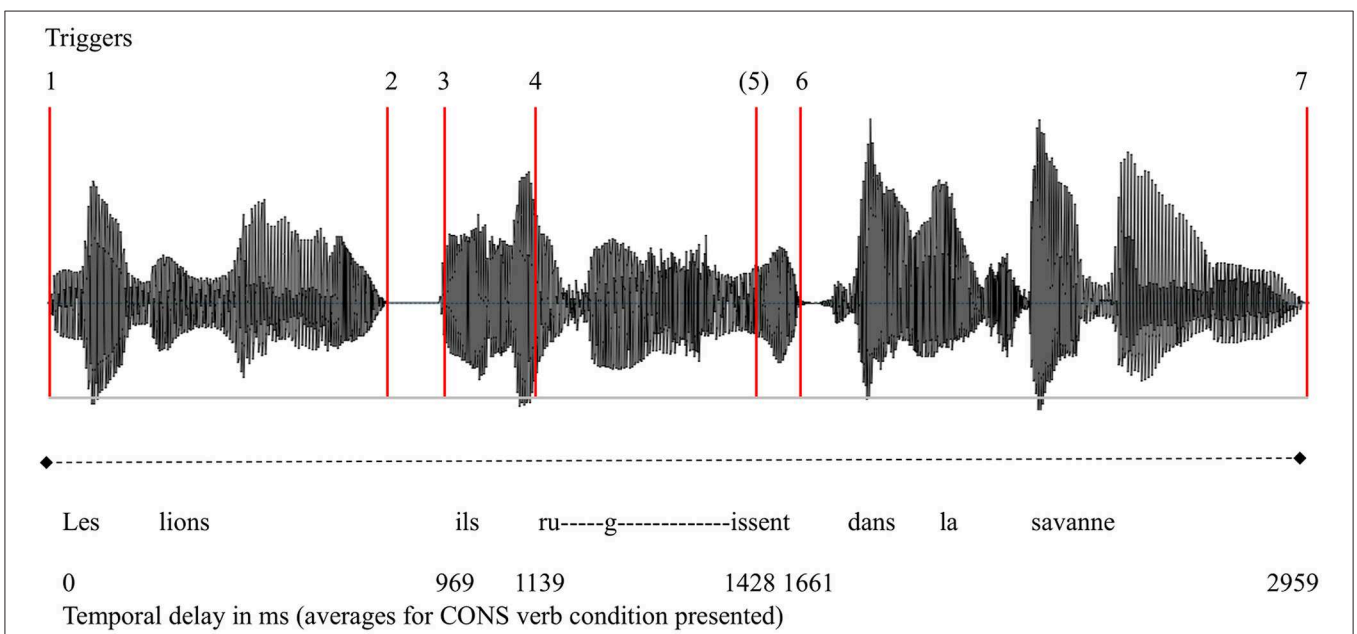


FIGURE 1 | Example waveform of an auditory stimulus for the sentence *Les lions, ils rugissent dans la savane*. The red lines represent the various cue points, called "triggers," measured in the audio file. Trigger 1, sentence onset; Trigger 2, context phrase offset; Trigger 3, pronoun clitic onset; Trigger 4, verb onset; Trigger 5, onset of verb-final consonant (only Type 2 verbs); Trigger 6, verb offset; Trigger 7, sentence offset.

as its trigger points. Analyses presented in this paper use triggers 1 (sentence onset) and 4 (verb onset).

EEG Recording and Data Analysis

The EEG was recorded continuously with a 500 Hz sampling rate from 64 cap-mounted electrodes (WaveGuard caps, ANT; Enschede, NL) placed according to the extended International 10/20 system. The electrodes used for recording covered frontal, central, parietal, temporal and occipital lobes (FP1, FP2, F3, F4, F7, F8, Fz, C3, C4, Cz, P3, P4, Pz, T3, T4, T5, T6, O1, O2, Oz). All impedances were maintained below 5 kΩ and were checked every 45 minutes throughout the experiment. The EEG was amplified using an ANT Neuro Eego^{TM} sports amplifier referenced to the CPz electrode. All subsequent EEG/ERP data processing steps and analyses were carried out using EEProbe software package (ANT; Enschede, The Netherlands) and statistical analyses were performed in R (R Studio Team, 2015), Boston, MA[2] using the Easy analysis and factorial experiments visualization package (Lawrence, MA. 2011, R package version 4.4-0[3]).

Offline, raw data were re-referenced to linked mastoids and filtered using a Gaussian bandpass filter of 0.3 to 40 Hz. Trials contaminated with eye blinks or other artifacts were rejected using a 30 μV criterion. All uncontaminated trials were entered into the final analysis. Using a 600 ms pre-stimulus baseline interval, single-subject EEG waveforms per condition were averaged separately over 2,100 or 3,100 ms epochs (−600 to 1,500 or 2,500 ms), time-locked to the relevant critical word onset (underlined words in **Tables 1**–**3** above) and entered into grand average ERPs. After artifact rejection, an average of 48/60 trials for semantic mismatches and 192/240 trials for number mismatches were analyzed per participant. Based on visual inspection and the previous literature, we identified representative time-windows for statistical analyses of lexical-semantic and number mismatches, during which ERP components were quantified as the mean EEG signal voltage (in μV).

In all analyses, we compare mismatch conditions to their corresponding match conditions presenting the exact same spoken sentence but with a different picture. For example, a number mismatch analysis for singular sentences compares singular spoken sentences with subject NPs, combined with a corresponding picture showing one agent (match condition) or with a similar picture showing two agents (mismatch condition). ERP analyses for midline electrodes and lateral electrodes were performed separately. At midline electrodes, global ANOVAs for the semantic condition included 2 factors: CONDITION (2 levels: mismatch vs. match), and ELECTRODE position (4 levels: Fz, Cz, Pz, and Oz). At lateral electrodes, the global ANOVA included four factors: CONDITION (2 levels: mismatch vs. match), HEMISPHERE (2 levels: right vs. left), ANTERIORITY (3 levels: anterior vs. central vs. posterior), and LATERALITY (2 levels: lateral vs. medial). For the *number* mismatch conditions, two additional factors were included for both analyses: CONTEXT (neutral vs. subject NP) and NUMBER (singular vs. plural). Greenhouse-Geisser corrections were applied in order to address potential violations of sphericity. In these cases, the original degrees of freedom and corrected probability levels are reported. A hierarchically-organized analysis of variance was pursued whereby only theoretically relevant interactions (i.e., CONDITION effects and their interactions with scalp distribution effects) and attendant *post-hoc* analysis results are reported. Given that the ERP effects of interest are generally observed close to the midline rather than at more lateral recording sites, 12 representative electrodes are used to illustrate effects, while head maps for difference waves cover the whole scalp.

Arcsine transformed accuracy data from acceptability judgments were analyzed using repeated-measure ANOVAs, computed separately for semantic and number (mis-)match conditions. The global ANOVA for number mismatches included four factors with 2 levels each: CONDITION, CONTEXT, GENDER, and NUMBER.

RESULTS AND INTERIM DISCUSSIONS

Following a reviewer's suggestion, we first present behavioral data, followed by ERP results and discussion for lexico-semantic mismatches, and finally results and discussion for number mismatches.

Behavioral Data Results

Accuracy for acceptability judgments on *lexical-semantic* conditions were nearly at ceiling for both match and mismatch sentences (see **Table 4**), and a global ANOVA indicated no CONDITION effect ($p < 1$). Global ANOVAs for *number* mismatches on LIAIS verbs revealed significant main effects of CONDITION [$F_{(1,21)} = 6.39$, $p = 0.0196$] in favor of matches, and NUMBER [$F_{(1,21)} = 5.67$, $p = 0.0269$] in favor of the plural (Singular: Mean 93.6, $SD = 0.045$; Plural: Mean = 95.7, $SD = 0.048$), qualified by interactions for CONDITION × NUMBER [$F_{(1,21)} = 8.97$, $p = 0.0069$], CONDITION × CONTEXT [$F_{(1,21)} = 5.90$, $p = 0.0242$], and NUMBER × CONTEXT [$F_{(1,21)} = 9.60$, $p = 0.0054$]. All these interactions are primarily driven by lower rejection rates for singular mismatches in neutral

TABLE 4 | Accuracy means (and standard deviations) for audio-visually matching and mismatching trials in lexico-semantic and number conditions for both liaison and consonant-final inflection conditions.

Conditions	Match	Mismatch
Semantics	93.9 (0.060)	92.5 (0.070)
Number: Liaison verbs	96.3 (0.035)	93.0 (0.066)
Singular: NP context	96.1 (0.069)	94.2 (0.082)
Singular: Neutral context	97.6 (0.036)	**86.5 (0.105)**
Plural: NP context	94.4 (0.074)	94.7 (0.079)
Plural: Neutral context	97.2 (0.046)	96.8 (0.087)
Number: Consonant-final verbs	94.8 (0.039)	91.6 (0.077)

Sub-conditions (for number and context) are listed only where statistical analyses indicated different patterns (i.e., for LIAS verbs).

[2]http://www.rstudio.com/

[3]http://CRAN.R-project.org/package=ez

contexts (in bold, **Table 4**), where number disambiguation was realized by the lack of a plural marker at the liaison. See section ERPs for Number Mismatches on Verbs for further discussion. A global ANOVA for CONS verbs revealed that these differed significantly by CONDITION [$F_{(1,21)} = 4.52$, $p = 0.0455$], but no other significant effects were found. Mismatches were responded to less accurately than matches.

ERPs for Lexico-Semantic Mismatches

As depicted in **Figure 2**, compared with the match condition, the semantic mismatch condition elicited a series of negativities across both context conditions at verb onset. First, we observe a posterior N400-like negativity between roughly 300–700 ms. Secondly a subsequent negative deflection emerges around 1,200 ms and lasts until 2,000 ms. It shows a frontal distribution until 1,700 ms and becomes more posterior afterward. Recall that the verb was always followed by an object noun phrase (NP) or a prepositional phrase (PP) that ended the sentence, and that nouns within these phrases also mismatched with the depicted information (see **Table 1** for an example). On average, verbs ended 550 ms after onset, and participants heard the NP/PP between 600 and 1,800 ms. Based on this time course, we analyzed the negativities in five different time windows: 300–500 ms for the core N400, 500–700 ms for the extended N400, 700–1,100 ms for the interval that did not elicit effects, 1,200–1,700 ms for the negativity related to the NP/PP mismatch, and 1,700–2,000 ms for a presumed sentence-final N400-like negativity.

Statistical analyses for all time windows, separately for lateral and midline electrodes, are summarized in **Table 5**. Significant interactions in the global ANOVA were decomposed to identify scalp electrodes displaying the strongest condition differences. In both the 300–500 ms and 500–700 ms time windows, the most dominant and consistent effects included CONDITION × ANTERIORITY interactions at both lateral and midline electrodes, as well as a CONDITION × LATERALITY interaction at lateral electrodes. Decomposing these interactions confirmed that the N400 reached significance only at posterior electrodes at or near the midline (Pz and Oz, and posterior medial electrodes). As expected, for the 700–1,100 ms time-window, we found no significant main effects or interactions involving CONDITION. As can be seen in **Figure 2** (e.g., at Pz), the absence of an effect in this contrast cannot be attributed to the presence of a P600 that may have canceled out any ongoing negativities due to component overlap. In fact, there is not the slightest indication of a positive dip that could point to a "hidden" P600, including at posterior electrodes where P600s are usually found.

A global ANOVA for time-window 1,200–1,700 ms yielded a significant CONDITION effect at midline and lateral electrodes, as well as CONDITION × ANTERIORITY, CONDITION × LATERALITY, CONDITION × LATERALITY × ANTERIORITY, and CONDITION × LATERALITY × CONTEXT interactions. The first three interactions indicate that this broadly distributed late negativity is most prominent at frontal electrodes and along the entire midline, whereas it gradually decreases at more lateral and posterior sites over both hemispheres (see voltage map). Finally, decomposing the interaction involving CONTEXT, we found that the negativity was more broadly distributed in the NP context, but limited to medial electrodes in the neutral one. Global ANOVAs for the sentence "wrap-up" effect in the 1,700–2,000 ms time-window yielded a CONDITION main effect in the midline with no other interactions.

Discussion for N400 Effects

Lexico-semantic mismatches on verbs were reliably detected by participants and elicited a large N400 component, as expected. Importantly, our study focused on mismatches involving verbs/actions, and not nouns/objects as in Royle et al. (2013) and other previous studies. We have therefore demonstrated that an N400 can be reliably elicited in adult French native speakers in response to verb-action mismatches. We believe that these require more complex cognitive matching processes than noun-object pairings, as they involve syntactic and thematic relations between a verb and its arguments. For example, in order to appropriately illustrate the ditransitive verb *give*, one must include an agent, a patient, and a beneficiary.

After the classic N400 time-window (300–500 ms), the N400 continued until 700 ms post verb-onset. There are various possible interpretations for this finding. First, mismatches involving verbs rather than nouns may require more complex processing. Secondly, in auditory studies, the N400 sometimes shows a longer duration due to word variability across trials (Holcomb and Neville, 1990). Thirdly, extended N400s with durations up to 700 ms have been discussed as reflections of additional post-lexical integration. The relevant discussion concerns the N400's functional interpretation, and whether it simply reflects automatic expectancy-based processing (i.e., lexical access typically between 300 and 500 ms, Kutas et al., 2006; Federmeier, 2007; Lau et al., 2008) or whether it also reflects controlled post-lexical integration (i.e., spoken word integration into a higher-order meaning representation after 500 ms, e.g., Brown and Hagoort, 1993; Holcomb, 1993; Steinhauer et al., 2017). Fourthly, 2/3 of our verbs were immediately followed by a direct object, which, in this condition, also mismatched with the visual stimulus, and may therefore have elicited a second N400. Note that the negativity's scalp distribution between 500 and 700 ms resembled the N400 preceding it, such that it is impossible to rule out any of these explanations without additional analyses beyond the scope of this paper.

Discussion for Sustained Frontal and Posterior Negativities

Following N400 effects, we observed late sustained negativities, the first between 1,200 and 1,700 ms with a frontal distribution, and the second between 1,700 and 2,000 ms with a broad distribution, but a central-parietal maximum consistent with an N400. The frontal negativity was elicited while direct objects (NP) or prepositional phrases (PP) were being processed. Both the NP and the noun in the PP also mismatched with the picture (i.e., one sees a woman singing on a stage but hears "*she swims in the public pool*," see **Table 1**). A comparison of this condition and the number mismatch conditions, where no incongruencies were present between the NP/PP in erroneous and correct sentences

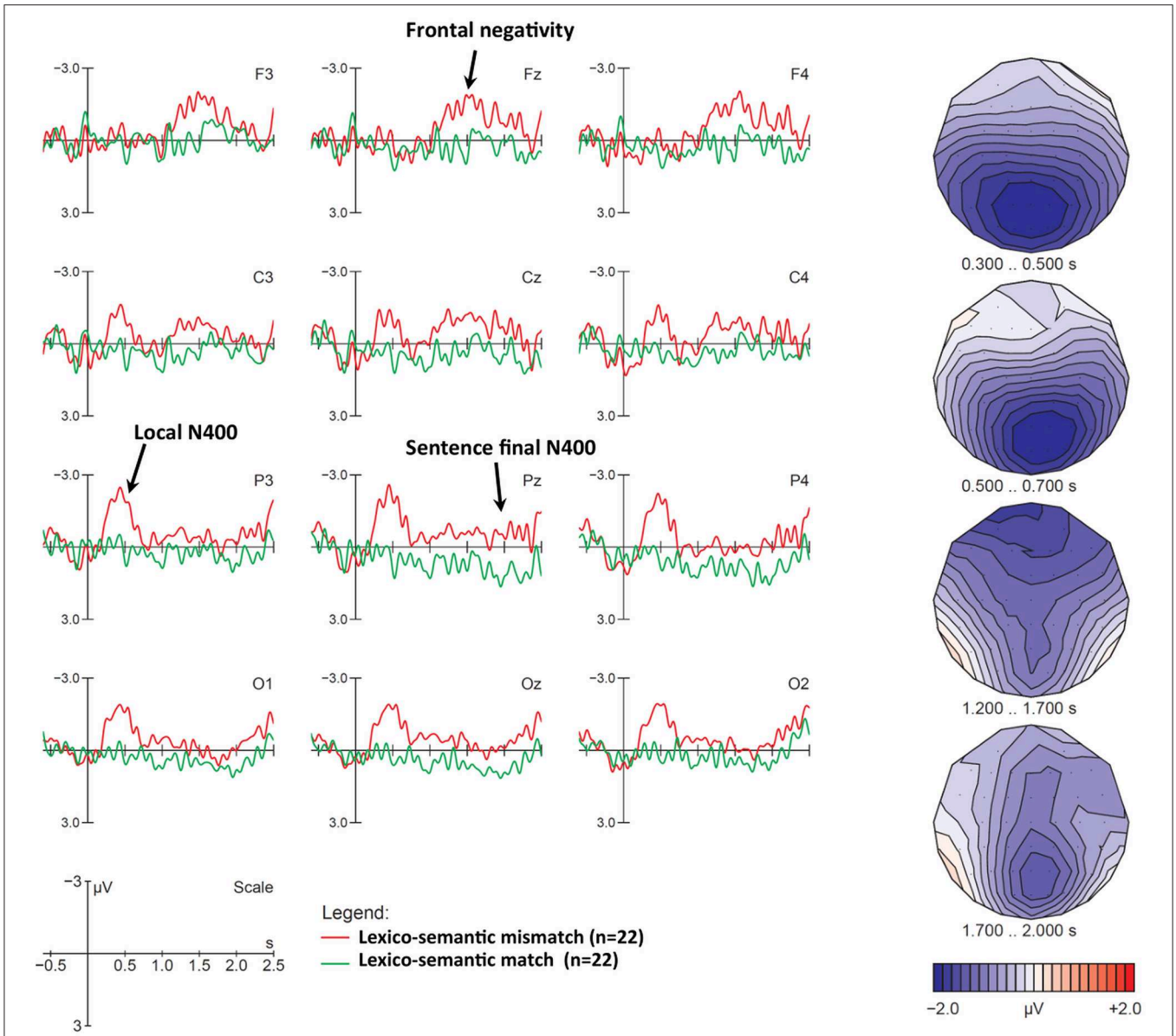


FIGURE 2 | N400 and other negativities elicited by the lexico-semantic mismatch condition, collapsed across NP and neutral contexts. Displayed are grand-average ERPs at midline and eight lateral electrodes, as well as voltage maps illustrating the difference waves, for all participants, time-locked to the onset of the critical verb using a baseline of −600 to 0 ms. The vertical bar marks the onset of the critical verb. On average the verb ended 550 ms after onset; between 600 and 1,800 ms participants heard a NP/PP, which included a second semantic mismatch and ended the sentence. Compared with the correct match condition (green line), the semantic mismatches (red line) elicited a large extended N400 between 300 and 700 ms, followed by a frontal negativity during the NP or PP (1,200–1,700 ms), and a subsequent posterior sentence wrap-up N400 between 1,700–2,000 ms. Negative polarity is plotted upwards. Voltage maps represent difference waves (violation minus control), with negativities in blue and positivities in red. For illustration purposes only, ERP plots have been 10 Hz low-pass filtered.

(see **Figures 5** and **6** below), shows that we observe a sustained negativity between 1,200 and 1,700 ms only in the lexico-semantic mismatch condition, suggesting that it is related to this additional semantic mismatch. However, its frontal distribution is not typical of an N400 and may point to a combination of mismatch effects proper and frontal expectancy effects reflecting anticipation of an additional semantic mismatch. Similar effects have been found for anticipation of a predictable comma likely to render a sentence ungrammatical, and was interpreted as a contingent negative variation (CNV, Steinhauer, 2003). We interpret the late portion of the negativity as a potential "sentence wrap-up effect", which we discuss in the section Sentence-Final Negativities and Wrap-Up Effects below.

Discussion for P600 Effects

Recall that the P600 has sometimes been elicited by semantic anomalies in conjunction with the N400, notably in a cross-modal mismatch paradigm (Royle et al., 2013), but also

TABLE 5 | Global repeated measures ANOVAs for lexico-semantic conditions (Trigger 4) at time-windows of interest.

			(N400)			Late negativity	Wrap up effects
		df	300–500	500–700	700–1,100	1,200–1,700	1,700–2,000
Lateral	Condition	(1, 21)	–	–	–	7.14**	–
	Condition × Anteriority	(2, 42)	5.26***	8.63***	–	6.28**	–
	Frontal: Condition	(1, 21)	–	–	–	9.82**	–
	Central: Condition	(1, 21)	–	–	–	5.02*	–
	Posterior: Condition	(1, 21)	8.08**	8.08**	–	–	–
	Condition × Laterarlity	(1, 21)	9.59**	5.34*	–	4.77*	–
	Medial: Condition	(1, 21)	5.44*	3.60†	–	7.06*	–
	Lateral: Condition	(1, 21)	–	–	–	5.26*	–
	Condition × Ant × Context	(2, 42)	–	5.26*	–	–	–
	NP context: Con × Ant	(2, 42)	–	13.16***	–	–	–
	NP context Ant: Con	(2, 42)	–	8.09**	–	–	–
	Condition × Lat × Ant	(2, 42)	–	–	–	4.51*	–
	Central: Con × Lat	(1, 21)	–	–	–	4.92*	–
	Central, medial: Con	(1, 21)	–	–	–	5.67*	–
	Posterior: Con × Lat	(1, 21)	–	–	–	10.13**	–
	Lateral: Con × Ant	(2, 42)	–	–	–	11.38***	–
	Lateral, frontal: Con	(1, 21)	–	–	–	11.21***	–
	Condition × Lat × Cont	(2, 42)	–	–	–	5.69*	–
	Neutral: Condition × Lat	(1, 21)	–	–	–	7.80**	–
	Neutral, medial: Con	(1, 21)	–	–	–	8.55**	–
Midline	Condition	(1, 21)	5.56*	–	–	10.26***	7.35**
	Condition × Electrode	(3, 63)	6.34*	10.79***	–	–	–
	Pz: Condition	(1, 21)	9.29***	7.35**	–	–	–
	Oz: Condition	(1, 21)	9.21***	12.85***	–	–	–

*Only significant results and trends are presented. Con, Condition; Ant, Anteriority; Lat, Laterality; Cont, Context. †p < 0.10, *p < 0.05, **p < 0.01, and ***p < 0.001.*

in purely auditory ones (Hagoort, 2003), and in reading studies (Steinhauer et al., 2010), and has therefore been argued to reflect mental monitoring and processing load related to language reanalysis (i.e., it is not specific to grammatical processing; Kolk et al., 2003; Steinhauer and Connolly, 2008; van de Meerendonk et al., 2009). Others have argued that these positivities are tightly linked to acceptability judgment tasks, potentially as a linguistic variant of the P300 component (Coulson et al., 1998; Friederici et al., 2001; Sassenhagen et al., 2014). The absence of positivities in the lexico-semantic condition, despite our use of a judgment task, may be explained by our particular mismatches. First, as reflected by the subsequent frontal negativities, participants seemed quite engaged in anticipating and processing additional semantic mismatches in the following NPs and PPs, and may not have categorized the sentence as unacceptable when encountering semantic mismatches on verbs. Another possibility is that semantic mismatches realized on verbs do in fact involve more complex conceptual-semantic processing than those realized on nouns and may draw attention away from whatever processes may elicit positivities found on nouns. As we are not aware of any other ERP studies using verb/action mismatches, this would need to be further investigated. Finally, P600s are certainly not a consistent finding for conceptual mismatches: the motivation for explaining their absence is primarily based on their presence in a recent study from our lab that used a very similar cross-modal paradigm (Royle et al., 2013). Perhaps the most important point is that the absence of a P600 in our semantic mismatch condition contrasts with the P600s observed in other mismatch conditions that we will discuss next.

ERPs for Number Mismatches

Sentence Onset Effects

At sentence onset we observed distinct ERP patterns for neutral contexts (with no disambiguation at this point) and NP contexts, where the NP either matched or not with the picture in number at the determiner (*le/la/les* "the.M.SG/F.SG/PL"). The distinction between LIAIS and CONS verbs does not play a role at this point, such that we can collapse across these conditions, which we did. **Figure 3** displays match and mismatch conditions for both NP and neutral contexts, collapsed across singular and plural sub-conditions. Recall that the mismatch in neutral contexts happens only downstream on the verb and is, therefore, not yet expected to elicit mismatch components. The first 900 ms (−600 to 300 ms) are largely dominated by visual onset components (most prominently at occipital electrodes)

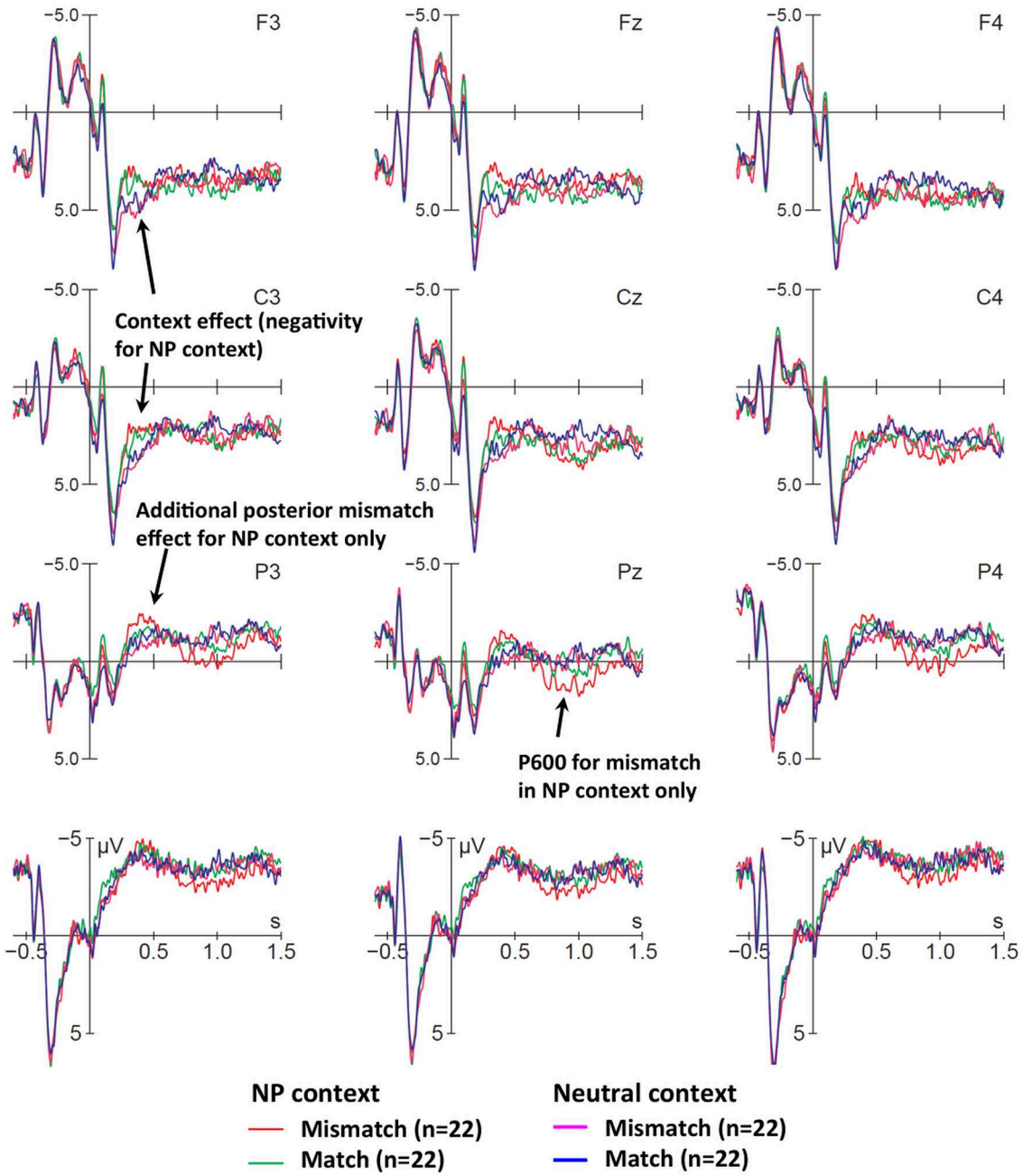


FIGURE 3 | Early ERP effects of context and number mismatches at sentence onset. Displayed are grand-average ERPs at midline and lateral electrodes for all participants, time-locked to the onset of the determiner (vertical bar) with a baseline of −600 to 0 ms. Compared with neutral context correct (blue), and neutral context mismatch (magenta), the NP context correct condition (green), and the NP context mismatch condition (red) elicited an early negativity (300–500 ms). Furthermore, number mismatches with NP context display a small increased negativity (between 300–450 ms) and a large positivity between 700 and 1,200 ms. The two neutral conditions will be disambiguated further downstream at the verb and do not yet show differences at sentence onset.

for pictures (presented at −500 ms) and by auditory onset components (most prominently at fronto-central electrodes) for spoken sentences (starting at 0 ms), respectively. As can be seen, all conditions are virtually indistinguishable up to 300 ms after sentence onset, at which point the first context-effect emerges.

NP contexts, compared to neutral ones, elicited an early slightly left-lateralized fronto-central negativity (300–450 ms) after determiner onset. In the same time-window, we observe an additional enhanced negativity for NP context mismatches, which is followed by a P600 (700–1,200 ms). We will show how singular and plural mismatches in NP contexts contribute to

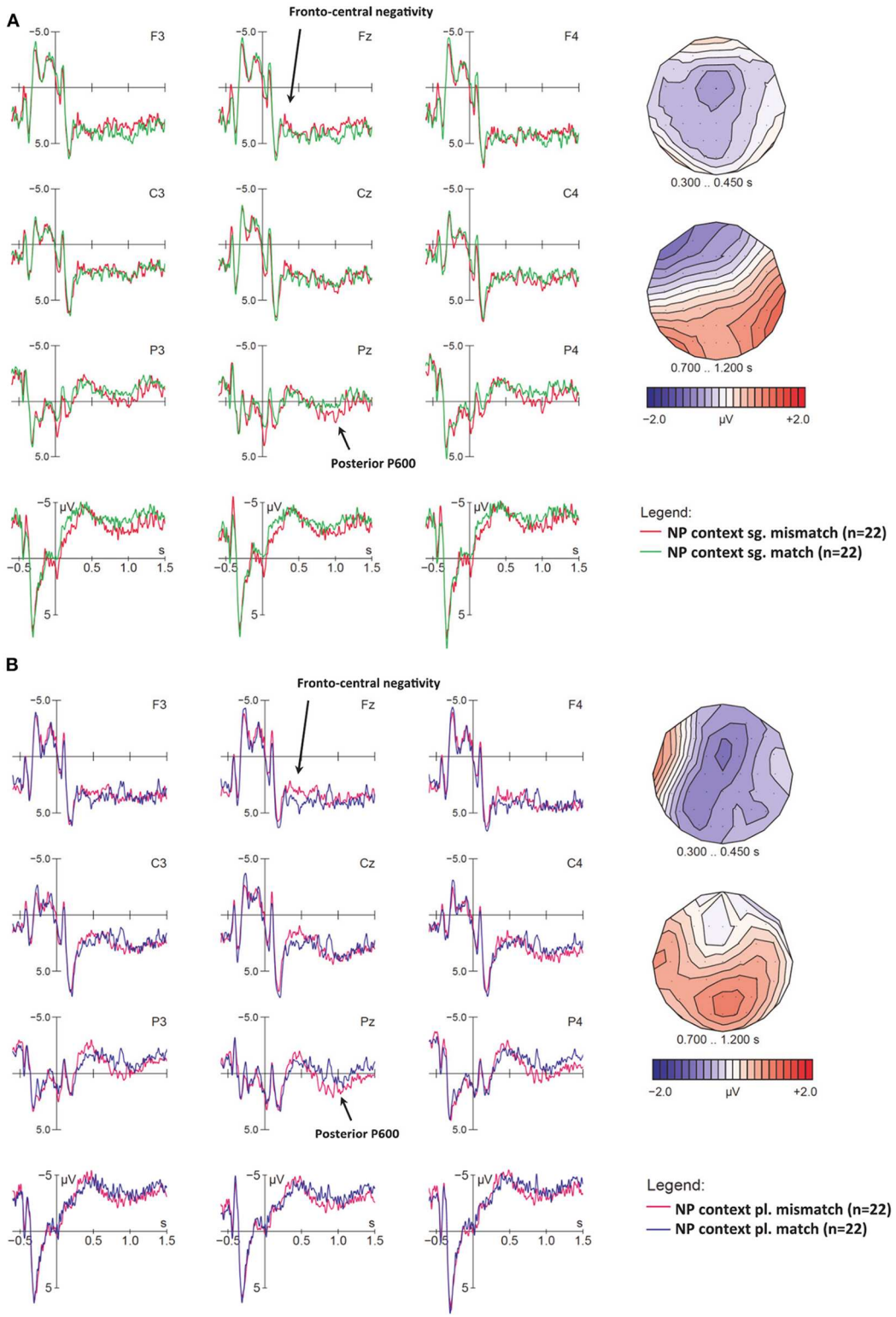


FIGURE 4 | Early effects of cross-modal number mismatches in NP contexts, for **(A)** singular and **(B)** plural NPs at sentence onset. ERPs are time-locked to the onset of the determiner (vertical bar) with a baseline of −600 to 0 ms; voltage maps illustrate the difference waves of relevant effects. **(A)** Singular mismatches (red) elicited a small fronto-central negativity in the N400 time-window relative to singular matches (green), as well as a parietal P600. **(B)** Plural mismatches (magenta) elicited a larger N400 as well as a parietal P600, as compared to plural matches (blue). Voltage maps of these effects (mismatch minus control) show that singular and plural mismatches elicited quite similar components.

TABLE 6 | Global repeated measures ANOVAs for sentence onset effects (Trigger 1) at time-windows of interest.

			(N400)	(P600)
		df	300–450	700–1,200
Lateral	Condition	(1, 21)	–	–
	Context	(1, 21)	29.03***	2.99†
	Condition × Context	(1, 21)	5.32*	–
	Condition × Lat × Cont	(1, 21)	6.70*	–
	Condition × Ant × Cont	(2, 42)	–	7.95**
	NP context: Ant × Cond	(2, 42)	–	9.56***
	NP context, Post: Cond	(1, 21)	–	11.69***
NP context	Condition × Anteriority	(2, 42)	–	7.89**
	Posterior: Condition	(1, 21)	–	10.95*
	Condition × Ant × Hem × Num	(2, 42)	–	4.56*
	Posterior: Condition × Hem × Num	(1, 21)	–	10.95*
	Con × Ant × Hem × Num × Lat	(2, 42)	–	5.81*
	Left Hem: Condition × Ant	(2, 42)	–	7.80**
	Left Hem: Condition × Ant × Num	(2, 42)	–	5.97*
	Left Hem Sg: Condition × Ant	(2, 42)		10.49***
	Left Hem Sg Front: Condition	(1, 21)	–	3.23†
	Left Hem Sg Post: Condition	(1, 21)	–	5.03*
	Left Hem Pl: Condition × Ant	(2, 42)		3.90*
	Left Hem: Condition × Ant × Num × Lat	(2, 42)	–	4.25*
Midline	Condition	(1, 21)	–	–
	Context	(1, 21)	20.56***	9.58**
	Condition × Context	(1, 21)	9.78**	–
	NP context: Condition	(1, 21)	4.43*	–
	Condition × Elec × Context	(3, 63)	–	8.52***
	Pz: Condition	(1, 21)	–	8.44**
	Pz: Condition × Context	(1, 21)	–	6.22*
	Pz, NP: Condition	(1, 21)	–	12.10***
	Oz: Condition × Context	(1, 21)	–	10.92*
	Oz, NP: Condition	(1, 21)	–	9.34**
NP context	Condition	(1, 21)	4.43*	9.14***
	Pz: Condition	(1, 21)	–	11.35***
	Oz: Condition	(1, 21)	–	8.36**

Only significant results and trends are presented. Ant, Anteriority; Con, Condition; Cont, Context; Elec, Electrode; Front, Frontal; Hem, Hemisphere; Lat, Laterality; Num, Number; Pl, Plural; Post, Posterior; Sg, Singular. †$p < 0.10$, *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

this pattern. In neutral context conditions as expected no clear differences are visible, as confirmed by the absence of significant effects in all time-windows discussed below (see also **Table 6**). We return to neutral contexts at later sentence positions—at verb onset—where they are disambiguated.

ERPs for singular and plural mismatches in NP contexts

For sentences with *singular* NPs, we observe a small fronto-central negativity in the N400 time-window, followed by a posterior P600 in the mismatch condition between 700 and 1,200 ms after sentence onset (see **Figure 4A**). In the *plural* contrast (**Figure 4B**), we see a similar biphasic pattern for mismatches. However, the fronto-central negativity appears slightly larger and seems to extend more clearly to left posterior electrodes.

Statistical analyses for sentence-initial positions are summarized in **Table 6**. Global ANOVAs in the 300–450 ms time-window yielded a highly significant CONTEXT main effect. Mismatch effects were reflected by CONDITION × CONTEXT interactions in midline and lateral electrodes, as well as a CONDITION × CONTEXT × LATERALITY interaction in lateral electrodes. These interactions confirmed that the negativity for visuo-auditory number mismatches was limited to disambiguating NP contexts, and was largely limited to medial electrodes. Surprisingly, the absence of significant ANTERIORITY and NUMBER interactions suggested that (a) the apparent frontal focus of the negativity was not reliable across subjects and (b) the apparent differences in size and scalp distribution of negativities between singular and plural conditions (**Figure 4A** vs. **Figure 4B**) were not meaningful. Statistically, there was only

a broadly distributed negativity in both singular and plural mismatches with NP contexts.

In the P600 time window (700–1,200 ms), global ANOVAs yielded a significant CONDITION× ELECTRODE × CONTEXT interaction at midline electrodes, and CONDITION × CONTEXT and CONDITION × LATERALITY × CONTEXT interactions in lateral electrodes (see **Table 6**). Decomposing these interactions confirmed that the P600 had a posterior distribution and was limited to number mismatches in NP contexts. While this P600 was consistent across singular and plural at midline electrodes (significant CONDITION main effect at Pz and Oz), additional interactions with factor NUMBER and topographical factors at lateral electrodes indicated that only for singular mismatches the P600 time-window also showed a (non-significant) frontal negativity over the left hemisphere. Overall, both singular and plural mismatches with NP contexts elicited consistent P600s that lasted until 1,200 ms.

Note that this relatively long P600 duration means that this effect was still present when the verb was presented (average verb onset at 1,140 ms, $SD = 149$ ms) and would have contaminated baselines and ERP analyses time-locked to verb onset (see Steinhauer and Drury, 2012). For these reasons, we refrained from analyzing the NP-context conditions at the verb, even though it would have been interesting to see whether additional disambiguating information elicited more mismatch effects further downstream.

Discussion for sentence-initials effects

Independent of mismatches, context manipulations at sentence onset elicited a larger negativity for NP contexts between 300 and 450 ms after sentence onset: this was likely triggered by the first word. Both NP contexts and neutral contexts started with function words (e.g., *Au dessert* "at-the desert" = "for desert" in neutral contexts, *La/les fille/s* "The girl/s" in NP contexts) for which N400 effects are rather atypical. In addition, the context-driven negativity had a more frontal distribution than a classic N400. We speculate that this context main effect may reflect enhanced alertness once participants had identified that a sentence started with a determiner and could, therefore, provide the first disambiguating task-relevant cue.

Interestingly, determiner *mismatches* elicited an additional, more broadly distributed negativity in virtually the same time-window, which was followed by a posterior P600, for both singular and plural mismatches. The mismatch negativity could be interpreted either as a lexical prediction effect (i.e., an N400, Tanner and Van Hell, 2014; BSS2019) or an effect of reference resolution (i.e., an N-ref component, e.g., Van Berkum et al., 1999). In the first scenario, participants would expect a specific determiner coherent with the number (and gender) of depicted potential subjects, and process a mismatch as a lexical (or phonological) error. In the second scenario, participants might wonder, when there are multiple potential subjects, who *la fille* 'the girl' refers to. However, reference resolution effects only seem to make sense—and have only been reported—for singular nouns where contexts provide multiple potential referents, while we found no statistical differences between our singular and plural conditions and, moreover, we found them at the determiner rather than the noun. For these reasons we believe that this negativity reflects a mismatch for specific predictions. Our finding is reminiscent of that by DeLong et al. (2005) who reported an N400 on determiners for unexpected sentence continuations after a highly constraining context (e.g., *an airplane* rather than *a kite* after "… the boy went outside to fly_"). Whether this effect is primarily lexical or phonological in nature remains unclear.

The following P600-like positivity in our data may either reflect (a) an immediate categorization of the sentence as unacceptable (Sassenhagen et al., 2014) or (b) cross-modal integration of conflicting number information as in previous morphosyntactic (dis-)agreement studies, possibly linked to structural disambiguation or revisions (see e.g., Molinaro et al., 2011a, for a review), or both. In line with our previous work and the literature (e.g., Friederici et al., 2001; Steinhauer and Connolly, 2008; Royle et al., 2013), we maintain the view that the P600 typically reflects multiple cognitive processes and comprises multiple subcomponents. A P600 account involving structural (rather than purely lexical) mismatches or revisions would imply that participants in our study syntactically integrated the determiner with the subsequent noun, which was phonologically compatible with both a singular and a plural form (*fille/s* [fij]). However, a picture of two girls would have suggested (and pre-activated) a plural referent, which then mismatched with the spoken singular determiner (*la* "the.SING.FEM"), thereby resulting in a traditional number agreement violation (i.e., *la* **filles*). Given that these early-disambiguating contexts were followed by additional information disambiguating subject number on the verb, one might expect higher confidence (and thus higher accuracy) in grammaticality ratings compared to sentences with neutral contexts. However, as discussed above (see also **Table 4**), this was not the case, supporting immediate categorization at the first available cue. We anticipate that this pattern may be different in children, especially those with language impairment, who are currently being tested with this same paradigm.

For obvious reasons, number mismatch effects at sentence-*initial* words (as in our study) are absent from the previous literature as they can only be created in relation to a previously presented context (here: a picture). Overall, it is remarkable that this sentence-initial number mismatch elicited an N400-P600 pattern previously found for morpho-syntactic agreement violations. It suggests that non-linguistic visual information from the environment can be immediately used (in < 500 ms) to make strong predictions about appropriate linguistic representations, or that "feature checking" processes are not constrained to linguistic representations. The elicitation of a P600 at this early position in a sentence is clearly compatible with accounts of "conflict monitoring" (Kolk et al., 2003) and "well-formedness categorization" (Sassenhagen et al., 2014), but more difficult to explain in terms of a structural "reanalysis" (Friederici, 2002).

ERPs for Number Mismatches on Verbs

We will now turn to mismatch effects at target verbs in neutral contexts. At sentence onset, LIAIS and CONS verbs did not differ, but at trigger 4 (verb onset) they did, because for

LIAIS verbs, number disambiguation is available at verb onset (e.g., *elles[z]aiment* "they like"), while for CONS verbs, this information is available only at the verb final phoneme (e.g., *ils rugissent [ʁyʒɪ***s***]* "they roar"). We will first focus on LIAIS verbs and then turn to CONS ones and consider only neutral contexts because these are the ones being disambiguated for the first time on the verb.

ERPs for liaison verbs at verb onset at Trigger 4

As with sentence initial effects, we analyzed singular and plural violations separately. **Figure 5A** shows number mismatches time-locked to singular LIAIS verbs. In this comparison we did not observe the expected pattern but rather an apparent early left-anterior positivity between 150 and 450 ms after verb onset, and a posterior right-lateralized late negativity between 1,000 and 1,200 ms. However, global ANOVAs on singular LIAIS verbs in neutral conditions yielded no significant effects involving CONDITION at either the midline or lateral electrodes. (Note that the very early left-frontal positivity was partly driven by one participant's enhanced horizontal eye movements in this condition only, resulting in a polarity inversion of this difference between left-anterior and right-anterior electrodes—especially F7 and F8. Analyses excluding this data set did not change results, however. For consistency, we decided to present ERP data including this data set). Overall, our analyses did not point to any consistent ERP pattern for these number mismatches. Recall that this was also the condition with the lowest overall accuracy rate in our mismatch conditions (**Table 4**).

As illustrated in **Figure 5B,** for plural mismatches we observed an early left-lateralized fronto-central negativity between 100 and 300 ms, followed by a posterior P600-like positivity (500–900 ms), which then seems to be followed by a second late frontal and somewhat left-lateralized negativity from ∼800 to 1,200 ms. In fact, when inspecting the left-anterior electrode F3 alone, the patterns looks like a sustained early negativity, starting around 100 ms and lasting until ∼1,400 ms.

ANOVAs for plural verbs in the 100–300 ms time window yielded a significant CONDITION main effect at midline and lateral electrodes, and a CONDITION × LATERALITY interaction in lateral electrodes (see **Table 7**). This interaction means that the negativity was strong at medial electrodes, but only marginally significant at more lateral electrodes. Given that the early negativity seemed most prominent over left-frontal electrodes (especially F3), the lack of interactions involving factors HEMISPHERE or ANTERIORITY was somewhat surprising. However, this was due to the fact that (a) the negativity was stronger at medial than lateral electrodes over *both* hemispheres, and (b) at posterior electrodes, the negativity was almost equally strong over both hemispheres (suggesting a second and more posterior N400-like negativity near the midline). An ANOVA in the P600 time-window (500–900 ms) yielded significant interactions of CONDITION × ELECTRODE at midline, and CONDITION× ANTERIORITY as well as CONDITION × HEMISPHERE at lateral electrodes. These interactions point to a posterior P600 co-occurring with an ongoing left-frontal negativity that gains strength once the P600 dissipates. In fact, between 800 and 1,100 ms we found a significant CONDITION effect at F3 ($p < 0.02$) and Fz ($p < 0.03$), but not at more posterior electrodes. This pattern of an early frontal negativity and its reoccurrence after an intervening positivity is reminiscent of that previously described for various syntactic violations in auditory ERP studies (Steinhauer and Drury, 2012), suggesting a sustained frontal negativity and a temporarily overlapping P600. We will return to this below.

TABLE 7 | Global repeated measures ANOVAs for liaison verbs (LIAIS) in neutral contexts, for both singular and plural (Trigger 4) at time-windows of interest.

			(LAN)	(P600)
		df	100–300	500–900
NEUTRAL CONTEXTS SINGULAR VERBS				
Lateral	Condition	(2, 42)	–	–
	Condition × Anteriority	(2, 42)	–	–
Midline	Condition	(3, 36)	–	–
	Condition × Electrode	(3, 36)	–	–
NEUTRAL CONTEXTS PLURAL VERBS				
Lateral	Condition	(1, 21)	6.39*	–
	Condition × Laterality	(1, 21)	6.12*	–
	Medial: Condition	(1, 21)	7.22**	–
	Lateral: Condition	(1, 21)	3.63†	–
	Condition × Anteriority	(2, 42)	–	6.66**
	Condition × Hemisphere	(1, 21)	–	4.40*
Midline	Condition	(1, 21)	6.20*	–
	Condition × Electrode	(3, 36)	–	5.23*

Only significant results and trends are presented. †$p < 0.10$, *$p < 0.05$, and **$p < 0.01$.

ERPs for consonant-final verb conditions at Trigger 4

While liaison verbs phonologically disambiguated number at verb onset, consonant verbs provided number information on the verb-final "morpheme" consonant. Due to this difference, one would expect mismatch effects to occur somewhat later than for liaison verbs. As shown in **Figure 6A**, for mismatch CONS *singular* verbs, the most prominent difference between match and mismatch conditions was a broadly distributed, slightly right-lateralized negativity in the N400 time window (400–500 ms after verb onset), which does not seem to be followed by a clear positivity in the P600 time-window. Note however that at anterior electrodes the N400 is both preceded and followed by a negativity starting around 100 ms, which seems to end around 600 ms and re-occur around 1,000 ms. This pattern could, once again, reflect temporary ERP-component overlap, namely an early but sustained negativity with a frontal maximum (from 100 to 1,500 ms), which is superimposed first by a parietal N400 that temporarily results in a more posterior scalp distribution (from 400 to 500 ms) and then by a left-lateralized and posterior positivity (from 800 to 1,000 ms) that temporarily cancels out the negativity at most electrodes (especially over the left hemisphere), until the frontal negativity re-emerges. The assumption that the early (100–300 ms) and late negativity (1,050–1,500 ms) may reflect the same ongoing ERP component is supported by

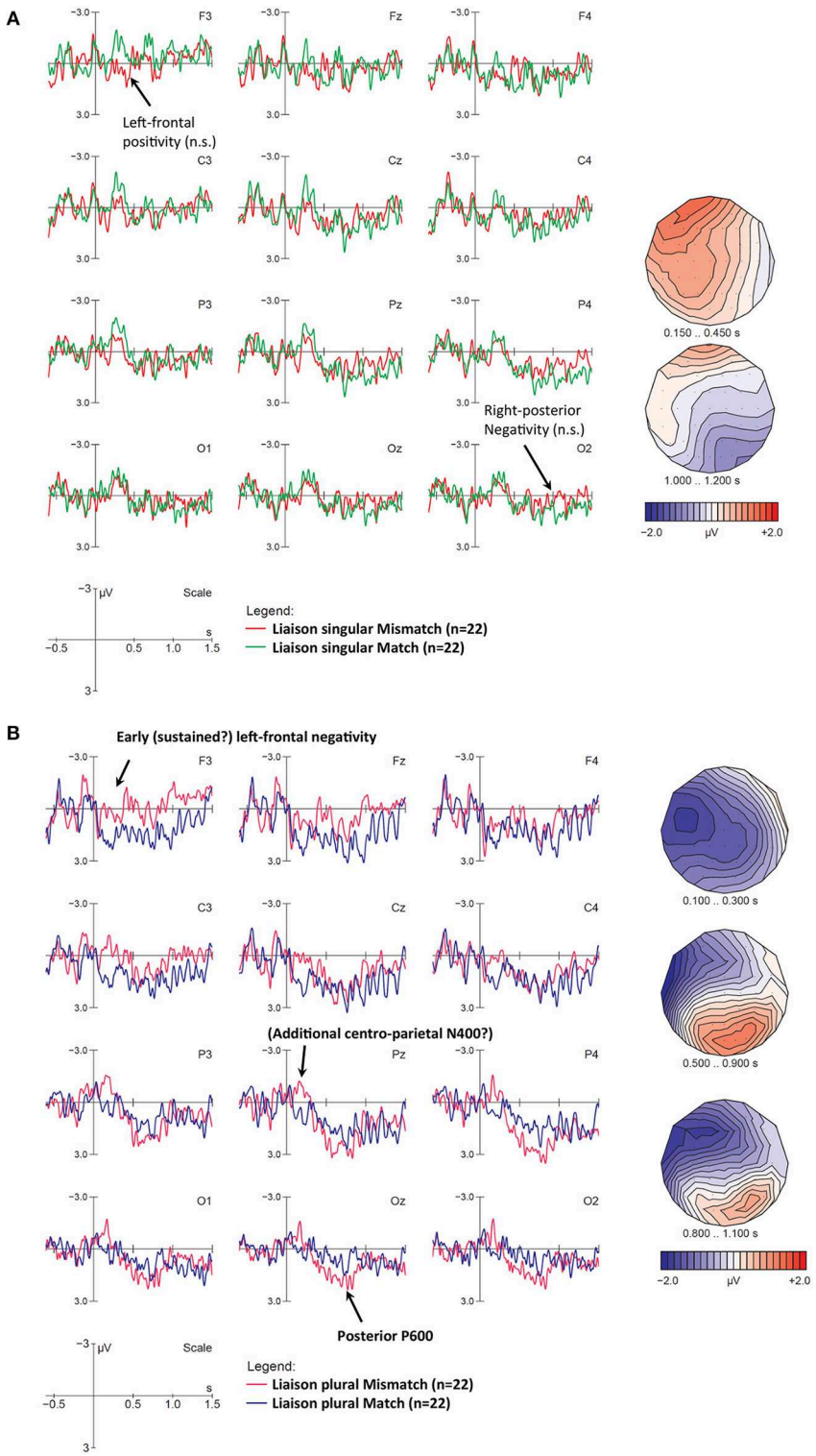


FIGURE 5 | ERP effects for number mismatches at liaison verbs with neutral context, **(A)** for singular and **(B)** for plural verbs. Displayed are grand-average ERPs at midline and lateral electrodes for all participants, time-locked to the onset of the liaison using a baseline of −600 to 0 ms. The vertical bar marks the onset of the liaison. **(A)** For singular verbs, neither the early frontal positivity between 150 and 450 ms nor the posterior negativity (1,000–1,200 ms) reached significance. **(B)** Compared to the correct control condition (blue lines), plural mismatches (magenta lines) show early negativities (100–300 ms), followed by a posterior P600 (500–900 ms). After the end of the P600, a negativity seems to re-emerge at frontal and central electrodes (third voltage map).

their similar scalp distribution (see first and last voltage maps in **Figure 6A**).

To test this assumption statistically, we ran ANOVAs directly comparing the two time windows (i.e., including the additional factor TIMEWINDOW). As expected, all significant effects involving the factor CONDITION were found to display the same scalp distribution in both time windows (100–300 ms and 1,050–1,500 ms, respectively), i.e., they did not interact with TIMEWINDOW. At midline electrodes, we found a CONDITION × ELECTRODE interaction [$F_{(3,63)}$ 3.49, $p = 0.04$], reflecting a frontal negativity [in Fz only, $F_{(1,21)}$ 5.59, $p = 0.03$], whereas lateral electrodes showed a main CONDITION effect [$F_{(1,21)}$ 4.96, $p = 0.04$]. In contrast, for the N400 between 400 and 500 ms, the ANOVA yielded significant CONDITION effects at midline and lateral electrodes, as well as a CONDITION × LATERALITY interaction at lateral electrodes (see **Table 8**). This interaction reflects a main CONDITION effect at medial electrodes. As a whole, this broadly distributed pattern along the midline strongly suggests the presence of a second (more posterior) negativity in addition to the ongoing frontal one. Lastly, in the P600 time window (800–1,050 ms), we observe a significant CONDITION × HEMISPHERE interaction along with higher-order interactions involving CONDITION, HEMISPHERE, ANTERIORITY, and LATERALITY at lateral electrodes, and no effect at the midline. These interactions reflect a right-lateralized (and somewhat anterior) negativity, and a left-lateralized (somewhat posterior) positivity that largely cancel each other out at the midline (see third voltage map in **Figure 6A**).

For CONS *plural* verbs (depicted in **Figure 6B**, statistics in **Table 9**) we observe a number mismatch effect reflected by a more delayed N400 than in singular contrasts (650–800 ms after verb onset), followed by a frontal P3a-like positivity (800–900 ms) and a late posterior one (1,100–1,300 ms). We ran an ANOVA for plural CONS verbs in the later N400 time-window (650–800 ms). This yielded a significant CONDITION main effect at midline and a CONDITION × ANTERIORITY interaction at lateral electrodes. Decomposition of this interaction revealed a main CONDITION effect at both central and posterior electrodes. An ANOVA in the 800–900 ms time-window yielded a CONDITION × ELECTRODE interaction at midline, and a CONDITION × ANTERIORITY interaction at lateral electrodes. These interactions reflect a significant frontal positivity (main effects of CONDITION at Fz), and a corresponding trend at anterior lateral electrodes. Finally, a main effect of CONDITION was found in the late P600 (1,100–1,300 ms) time-window, but only in posterior electrodes. No other main effects or interactions were found.

Discussion for number mismatches on verbs

Whereas cross-modal lexico-semantic mismatches have been shown to elicit N400s in a number of previous studies, number mismatches between visual and auditory input have not been studied so far. Given that our paradigm used grammatical sentences it was unclear whether our number mismatches would elicit ERP profiles typical for morphosyntactic agreement violations, i.e., LAN/N400s and P600s. Number disambiguation in neutral contexts only became available on the verbs. Not unlike mismatch effects at sentence onset, ERPs at verb onset elicited biphasic (N400-P600) profiles in three out of four contrasts. As expected, component latency was influenced by the availability of disambiguating number information (earlier for verb-initial liaisons than for verb-final consonants, and earlier for shorter singular than for longer plural CONS verbs). In addition, two conditions (LIAS plural and CONS singular) displayed sustained anterior negativities, resulting in complex patterns of overlapping ERP components. In contrast, singular LIAIS mismatches did not display any systematic ERP effects at all.

For LIAIS verbs, we first discuss the lack of ERP components for the singular condition before turning to effects found in the plural.

Number mismatches on singular liaison verbs

The absence of ERP effects in the singular LIAIS condition corresponds to relatively poor behavioral performance in that particular condition, i.e., sentences with neutral contexts (e.g., *"For dessert, she likes…"* concurrently with an image illustrating two girls). The different ERP mismatch effects for singular vs. plural sentences with neutral contexts in LIAIS verbs may therefore reflect these difficulties. Note that we cannot explain these effects by appealing to differences between commission and omission, nor plural vs. singular forms (singular being the default), since CONS singular forms *did* elicit ERP components. Similarly rule strength or predictability would promote better perception of differences in liaison, as this process is obligatory in French, and also reliably occurs in determiner-noun contexts. We explore phonological salience, truth-value interpretations assigned to sentences, and sociolinguistic variability as explanations for these results.

Phonological salience (or perceptual salience) refers to the ease with which we can hear or perceive a given structure (Goldschneider and DeKeyser, 2005). Applied to our materials, we can expect that arriving at an accurate sentence interpretation is facilitated by overt phonological cues for number. We used an overt cue for number with LIAIS verbs, which in the plural is arguably more salient—due to the presence of a /z/—than in the singular without a /z/. It seems very unlikely that a participant—after hearing *elles aiment* [ɛl**z**ɛm]—would be willing to deny the cue's presence and assume she may have hallucinated, just because the picture only shows a single potential subject. However, if the same participant sees a picture with two girls and hears singular forms such as *elle aime* [ɛlɛm], it seems possible to conclude to having misperceived liaison. Similar differences between the presence vs. absence of phonological (and visual) evidence have been found for prosodic boundaries and commas (leading to the "Boundary Deletion Hypothesis," cf. Steinhauer and Friederici, 2001; Pauker et al., 2011). Phonological salience thus seems to provide a plausible explanation for the absence of ERP mismatch effects for singular sentences with neutral contexts. However, it does not account for all of our data, as singular CONS mismatches (which were also marked by a non-salient cue) did in fact elicit ERP responses.

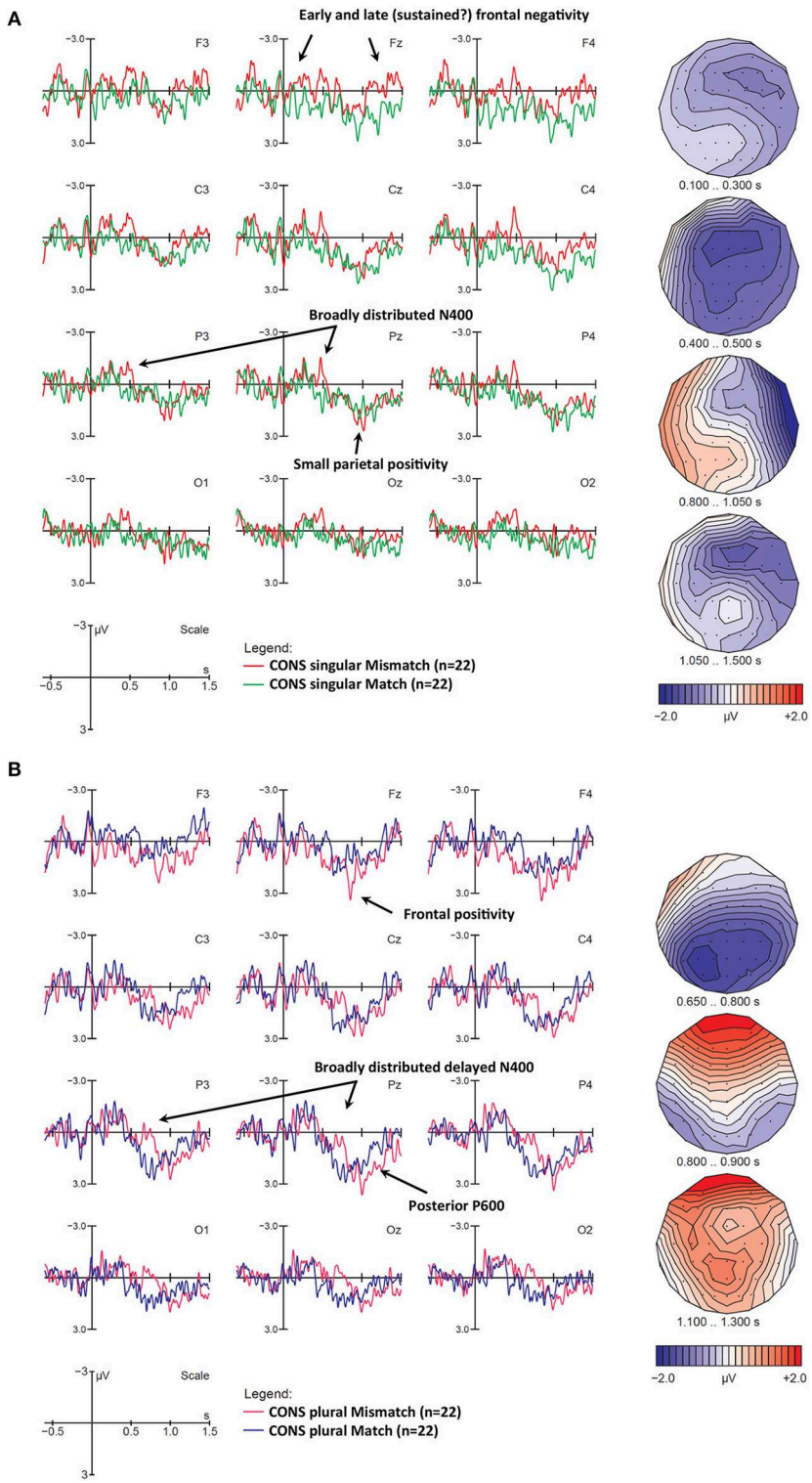


FIGURE 6 | ERP number mismatch effects at consonant-final verbs with neutral contexts, **(A)** for singular and **(B)** for plural verbs. Displayed are grand-average ERPs at midline and lateral electrodes as well as voltage maps illustrating the difference waves, for all participants, time-locked to the onset of the critical verb using a baseline of −600 to 0 ms. The vertical bar marks the onset of the critical verb. **(A)** Compared to the match condition (green lines), singular mismatches (red lines) show an early sustained negativity at frontal electrodes (100–1,500 ms; cf. Voltage maps 1 and 4), an additional N400 (400–500 ms), and an intermediate time window during which a right-anterior negativity and a left-posterior negativity seem to cancel each other out along the midline (800–1,050 ms). **(B)** Compared to the match condition (blue lines), plural mismatches (magenta lines) show an N400-like negativity (650–800 ms), followed by a frontal positivity (800–900 ms) and a posterior P600 (1,100–1,300 ms).

TABLE 8 | Global repeated measures ANOVAs for final consonant singular verbs (CONS) in neutral contexts (Trigger 4) at time-windows of interest.

			(N400)	(P600)	Negativity
		df	400–500	800–1,050	1,050–1,500
NEUTRAL CONTEXTS, SINGULAR VERBS ONLY					
Lateral	Condition	(1, 21)	6.88*	–	4.72*
	Condition × Laterality	(1, 21)	6.36*	–	–
	Medial: Condition	(1, 21)	9.30**	–	–
	Condition × Hemisphere	(1, 21)	–	5.68*	–
	Right Hem: Conditon	(1, 21)	–	6.70*	–
	Condition × Hem × Anteriority	(1, 21)	–	3.56†	–
	Condition × Hem × Laterality	(1, 21)	–	6.55*	–
	Right Hem: Condition	(1, 21)	–	6.67*	–
	Lateral: Condition	(1, 21)	–	5.85*	–
	Condition × Hem × Lat × Ant	(2, 42)	–	6.72**	–
	Left Hem: Con × Lat × Ant	(2, 42)	–	4.08*	–
	Left Hem, front: Con × Lat	(1, 21)	–	5.53*	–
Midline	Condition	(1, 21)	7.78**	–	–
	Condition × Electrode	(3, 36)	–	–	3.66*
	Fz: Condition	(1, 21)	–	–	5.54*

*Only significant results and trends are presented. Ant, Anteriority; Cent, Central; Con, Condition; Front, Frontal; Hem, Hemisphere; Lat, Laterality; Post, Posterior. †$p < 0.10$, *$p < 0.05$, and **$p < 0.01$.*

TABLE 9 | Global repeated measures ANOVAs for final consonant (CONS) plural verbs in neutral contexts (Trigger 4) at time-windows of interest.

			(N400)	(late N400)	(P600, frontal)	(P600, posterior)
		df	400–500	650–800	800–900	1,100–1,300
Lateral	Condition	(1, 21)	–	–	–	–
	Condition × Anteriority	(2, 42)	–	4.36*	5.50*	–
	Anterior: Condition	(1, 21)	–	–	3.25†	–
	Central: Condition	(1, 21)	–	5.45*	–	–
	Posterior: Condition	(1, 21)	–	8.41**	–	10.05**
Midline	Condition	(1, 21)	–	5.47*	–	–
	Condition × Electrode	(3, 36)	–	2.88†	4.62*	–
	Fz: Condition	(1, 21)	–	–	4.50*	–

*Only significant results and trends are presented. †$p < 0.10$, *$p < 0.05$, and **$p < 0.01$.*

Alternatively, the null result for LIAIS singular mismatches might be due to their enhanced acceptability, based on truth-values. Acceptability assigned to our sentences can be either logically or pragmatically motivated. For example, the sentence *Some triangles have three edges* is logically true, but under-informative and pragmatically odd. Similarly, when presented with an image of two girls eating chocolate mousse, describing the picture with "*She likes . . .* " is also logically true, but pragmatically odd. The ERP literature suggests that people differ in their bias toward logical vs. pragmatic processing (e.g., Barbet and Thierry, 2016). If some of our participants were biased toward logical processing, we would expect reduced or absent mismatch effects for neutral singular mismatching sentences. Crucially, however, even though one could argue that a lacking mismatch effect due to logical processing biases should be limited to singular sentences, there is no reason why it should be limited to sentences that are disambiguated by LIAIS verbs. That is, number mismatches disambiguated by CONS verbs would be subject to the same logic, but they did elicit clear ERP mismatch effects.

Yet another way of explaining the absence of ERPs for LIAIS singular mismatches comes from sociolinguistics. According to Prof. Julie Auger at Indiana University (personal communication), *elles* "she.PLUR" does not exist in informal Québec French, due to a process of neutralization (i.e., masculine and feminine plural pronoun clitics have become indistinguishable). Both are pronounced [i] before a consonant and [j] before a vowel (e.g., *les filles/les garcons y'aiment* "The girls/the boys, *they* like" are equally grammatical), although there is some variability between dialects. Two corpora from French monolingual speakers in Quebec City and bilingual speakers in Ottawa-Hull reveal few uses of *elles*, and omission or replacement of *elles* by *ils* "they.MASC" in addition to /l/-deletion (i.e., /il/

or /ilz/ pronounced [i], [iz], or [j], but rarely [ɛl/z] or [ɪl/z] the standard forms for plural) (Poplack and Walker, 1986; Bourget, 1987). The [j], being a semi-vowel, is licit before a vowel-onset verb and no additional liaison is necessary, and could in fact block liaison, since the verb onset is filled. Thus, perception of a subject-verb agreement error in liaison might be less systematic in singular conditions due to loss, or variability, of this grammatical feature, an interpretation that is coherent with our behavioral data where only these forms showed lower accuracy rates. We do not know of a psycholinguistic study that directly investigates liaison processing in Québec French, and so this interesting account remains somewhat speculative. While it appears to best explain our ERP null result for LIAIS singular mismatches (and is not applicable to CONS verbs), we should recall that participants still recognized the mismatches more than 85% of the time. We suggest that the absence of consistent ERP effects with LIAIS singular verbs reflects increased variability in processing strategies across participants, which may very well be influenced by sociolinguistic variability. As reflected by later sentence-wrap-up effects (see **Supplementary Materials**), in some cases error processing might also have been delayed.

Number mismatches on plural liaison verbs

The early-onset and sustained frontal negativity for plural mismatches resembles a classic morphosyntactic (dis-)agreement effect in auditory studies, possibly corresponding to more short-lived LAN-like effects in reading studies (Hasting and Kotz, 2008; Steinhauer and Drury, 2012). The extremely short onset latency of this effect, around 100 ms in our data, may be slightly overestimated due to possible co-articulation prior to the verb onset trigger (Trigger 4 in **Figure 1**) and the presence of the phoneme /z/ indexing a plural pronoun preceding it. As with Hasting and Kotz (2008), this is another illustration that morphosyntax-related processing difficulties that are clearly *not* driven by phrase structure violations can elicit this type of negativity (contra Friederici, 2002, 2011). Another similarity with Hasting and Kotz (2008), as well as many other auditory studies, is our finding of a complex pattern of overlapping ERP components (as discussed in Steinhauer and Drury, 2012). That is, sustained negativities are often superimposed by posterior P600 effects leading to a temporary mutual cancellation of components in at least certain electrodes. In our particular case, the negativity's scalp distribution in the early 100–300 ms time-window points to an even more complex pattern, as the P600 (500–800 ms) seems to be preceded by an additional, more posterior (N400-like) negativity from 100 to 300 ms that also overlaps with the frontal negativity. In our opinion, this is what explains the rather broad distribution of negativities in this time-window as reflected by statistical analyses, whereas the last portion of the "re-emerging" frontal negativity was limited to left-frontal electrode sites.

As with mismatch effects at sentence onset, the N400 effect may primarily indicate a lexical/phonological mismatch with what was predicted based on the picture. That is, participants saw a single person (e.g., one girl eating, thus predicting *elle* [ɛl], i.e., "she") but heard sentences such as *Au dessert, elles aiment …* "For dessert, *they* like …". Importantly, at least initially this mismatch is compatible with a number of interpretations. First, it is possible that the perceived mismatch included both the pronoun and the verb (*elles aiment* "they.FEM.PLUR like" instead of *elle aime* "she.FEM.SING likes"). This implies that the auditorily presented sentence as a whole was processed as a grammatical plural sentence, and the pronoun + verb as a whole mismatched across modalities. The first mismatching cue was provided by the pronoun at verb onset (liaison) and elicited an N400, as with NP contexts at sentence onset. The subsequent P600 was also triggered by the pronoun + verb and either reflected conflict monitoring and mismatch resolution or task-relevant categorization of a mismatching trial, or both. In this scenario, it is also possible that participants considered a *generic* interpretation. That is, "they (i.e., girls) like chocolate mousse" is an assertion that, in principle, could be illustrated with one single girl. As in English, French generic expressions are realized in the plural. However, for a generic (acceptable) interpretation we would predict a higher acceptability rate (which we did not find) and not expect a P600 (which we did find). Secondly, it is possible that the visual presentation of a single person activated a very strong expectation for a singular sentence. Knowing that incoming sentences were always supposed to describe the pictures, all spoken information up to phoneme /z/ at the liaison (including the entire context and most of the pronoun [ɛl]) was compatible with a singular interpretation, and it is conceivable that the longer the ambiguity lasted, the more this singular interpretation was strengthened. This expectation of a singular sentence may have led to two processing strategies that are both distinct from the first one discussed above: One is that only the pronoun, but not the verb, was processed as a plural form. Recall that liaison verbs were phonologically indistinguishable between singular and plural, i.e., *aime/nt* [ɛm]. So hearing *elles aiment'* [ɛlzɛm] could have been interpreted as *elles *aime*, "they likes," a classical morphosyntactic agreement violation. In this scenario, the P600 would reflect some process of reanalysis toward a singular interpretation. The other possibility assumes that the initial expectation of a singular sentence was so strong that it led participants to temporarily mis-parse the incoming speech signal. Instead of interpreting /z/ as the pronoun plural marker (*elles* [ɛlz] + *aime(nt)* [ɛm]) they may have interpreted it as a verb-initial phoneme (i.e., *elle* [ɛl] + *zaime(nt)* [zɛm]). This latter scenario is a possibility, as certain properties of French may have supported this. For instance, pronouns do not normally carry stress and are cliticized with the next content word to form one prosodic word where the content word carries word-final stress. Moreover, according to the "maximal onset principle" (Selkirk, 1981), the plural pronoun marker /z/ is syllabified into the verb's first syllable, as [ɛl.**zɛm**] and not [ɛlz.**ɛm**] (bold font indicates stress). This is the same pattern one would expect for a singular utterance (i.e., *elle zaime*). Importantly, even though the verb *zaimer* does not exist in French, there are a number of French verbs that do start with /z/ (e.g., *zigonner* "to dally," *zigouiller* "to kill," *zigzaguer* "to zigzag," *zézayer* "to lisp," *zyeuter* "to observe intently," *zébrer* "to decorate with stripes"). In other words, given the large number of different verbs used in our study (without any within-subject repetition), it is conceivable that in the LIAIS plural condition participants might have checked their

lexicon for a verb that starts with /z/. We propose that ambiguity complexity in this particular condition may have elicited the additional sustained negativity, possibly reflecting evaluation of multiple options.

Number mismatches on singular consonant-final verbs

We will now turn to number mismatches on consonant-final verbs. The singular CONS mismatch condition with neutral contexts again elicited three components: a sustained anterior negativity (AN), an N400 and a small slightly left-lateralized P600. This pattern resembles that found in plural LIAIS mismatches with, however, a reduced P600. The later onset for the N400 as compared with LIAIS verbs can be straightforwardly explained by the later appearance of disambiguating information in the CONS condition's sound-streams. Interestingly the AN does not differ in distribution between early and late time-windows. According to Steinhauer and Drury (2012), this is one way of demonstrating that two negativities are likely early and late portions of the *same* (ongoing) ERP component. In the intervening time-windows, it is first superimposed by an N400 and then canceled out by a P600, which themselves may have overlapped and canceled each other out to some extent (explaining the absence of either effect between 500 and 800 ms). In contrast to both sentence onset and LIAIS verb conditions, here number ambiguities lasted until the verb-final consonant. That is, when participants saw a picture of two lions roaring and heard *En soirée il rugit* [ilʁyʒ**i**] *dans la savane* "In the evening he roars in the savannah," only the lack of the verb-final consonant [s] (*rugissent* [ilʁyʒɪ**s**]) indicated a mismatch. Importantly, as the singular and plural pronouns *il* and *ils* are homophonous ([il]), we assume that the pronoun was initially processed as a plural (as suggested by the picture). Thus, one interpretation of what happened at the disambiguation point is that participants interpreted the auditory input as *ils* **rugit*, ("he.PLUR roar.SING"), which corresponds to a classical oral-language agreement violation. As before, the N400 would reflect a lexical-phonological mismatch, and the P600 would be associated with both categorization of this sentence as a mismatch and a potential attempt to revise its structure. Recall however, that (a) phonologically, the absence (omission) of a verb-final consonant is not very salient, and (b) participants were strongly biased toward a plural interpretation. Therefore, it is conceivable that participants were not entirely sure if the perceived mismatch was real or if they had simply missed an actually present consonant. Similar temporary confusions based on strong predictions are known from e.g., Itzhak et al. (2010) who demonstrated that listeners perceive a prosodic boundary in absence of any acoustic markers, if both lexical information and syntactic structure strongly predict it. Moreover, and only in the CONS singular condition, it is possible that participants initially parsed the subsequent preposition's word-initial consonant as a verb-final plural marker. In our example (*il(s) rugit* [ilʁyʒ**i**] *dans* …). Misinterpreting the /d/ of *dans* as a plural marker would result in [ilʁyʒ**id**], which could—in principle—be interpreted as a plural verb form (i.e., *ils rugident*). However, in the singular, the stem-final vowel is stressed due to the absence of a word-final coda (compare *ils rugissent* [ilʁyʒ**Is**]), and is a strong cue to word structure. At this point, participants would need to check this verb's stem forms in their mental lexicon and verify which one is legal in the plural. We believe that the complexity involved in this ambiguity is the reason why we find, once again, a sustained frontal negativity, resembling the LIAIS plural condition. As in previous conditions, we interpret the N400 as a reflection of an initial lexical-phonological mismatch, and the P600 as an attempt to resolve its structural consequences. The fact that the frontal negativity lasted beyond the P600 duration (as in LIAIS plurals and previous auditory agreement studies, e.g., Hasting and Kotz, 2008) suggests that the P600 does not always reflect the final stage of evaluation processes. One particularity of the CONS singular mismatch pattern was that the P600 itself did not reach statistical significance. Several previous studies have refrained from interpreting similar findings (e.g., Ye et al., 2006; Hasting and Kotz, 2008), but Steinhauer and Drury (2012) have argued that in the presence of ongoing negativities, the existence of a P600 can be inferred if this negativity is temporarily canceled out during the P600 time window (and at plausible electrode sites) and then re-emerges. We will come back to this point below.

Number mismatches on plural consonant-final verbs

Unlike singular CONS verbs, mismatches with plural CONS verbs elicited only a posterior N400 followed by a large P600, but no AN. As expected (see above), both components emerged slightly later than in the singular condition (due to the longer plural form duration). In many ways the plural condition resembles the singular one, however, the mismatching information is (a) phonologically salient and (b) an unambiguous plural verb marker. Thus, once plural information has been encountered, there can be no doubt that the verb is incompatible with an initial assumption of a singular pronoun (akin to a garden path sentence). In our example, the most likely lexical representation would be *En soirée il* **rugissent* [ilʁyʒɪ**s**]—a classical case of morphosyntactic number disagreement. In fact, we believe that—of all number mismatch conditions in our study—this condition is closest to a traditional oral-language agreement violation. As both the presence and the nature of this mismatch are extremely obvious, both the N400 and the P600 were found to be strong and consistent, while no AN reflecting effortful evaluation of a more ambiguous scenario was elicited.

Sentence-Final Negativities and Wrap-Up Effects

A subset of number-mismatch conditions (see **Supplementary Materials**), as well as the lexico-semantic condition, elicited a late posterior negativity at sentence end (1,700–2,000 ms), which we interpret as potential "sentence wrap-up" effects for both types of error. In contrast to positive waveforms that tend to occur in sentence-final positions of correct sentences, negativities are typically associated with preceding linguistic anomalies and may reflect additional processing load involved in reconsidering the anomaly and integrating the entire sentence (Osterhout and Mobley, 1995). A recent study from our lab on conceptual and logical semantic anomalies also showed that sentence final N400-like "wrap-up" effects are common, irrespective of the type of linguistic violation occurring in mid-sentence positions and of whether these elicited local N400s or P600s (Bokhari, 2015). Recently, Stowe et al.

(2018) have raised the question of whether "sentence wrap-up effect" is an appropriate label for these negativities given the link to anomalies; these authors suspect that task requirements may also play a role in eliciting them. "Anomaly-related sentence-final negativity" may thus be a more neutral term to characterize these ERP effects.

GENERAL DISCUSSION

The present study used ERPs to investigate whether visual-auditory mismatches between a picture and a perfectly grammatical spoken sentence would elicit similar brain responses as typically seen for *within*-sentence linguistic anomalies. We included both cross-modal semantic mismatches, realized on verbs, and number mismatches (singular vs. plural) that occurred at different sentence positions using a range of linguistic number markers in spoken French (determiners, liaison, and verb-final consonants). Analyses also contrasted potential differences between singular and plural mismatches. Overall, our data demonstrate that cross-modal mismatches result in ERP profiles known from the literature for linguistic anomalies, and seem to distinguish between mismatches that can be described as purely conceptual-semantic and those that can be viewed as concerning grammar.

N400s, P600s, and ANs—Evidence for Agreement Violations?

Returning to our initial research questions, our data have demonstrated that (a) cross-modal semantic mismatches realized on verbs elicit typical N400s and that (b) participants use the first available linguistic cues to detect number mismatches between a picture and a spoken sentence. Whether the ERP components found for cross-modal number mismatches are indistinguishable from those typically observed for "purely linguistic" within-sentence agreement violations, is less clear. On the one hand, all components we observed for number mismatches are within the range of ERP effects previously observed for morphosyntactic agreement violations. On the other hand, Molinaro et al. (2011a) reported that previous studies on number agreement violations have typically found LANs and P600s. While most of our negativities preceding the P600s did show a LAN-like frontal distributions, sometimes even with a left-lateralized prominence, statistical evidence usually pointed to a broadly distributed negativity compatible with an N400. Moreover, clearer evidence for left-anterior negativities (i.e., in LIAIS plural verbs) could be attributed to an early-onset sustained negativity at left frontal electrodes (e.g., F3). Overall, we believe our data are more compatible with an N400-P600 profile than with a LAN-P600 one. However, most previous ERP studies on number (dis-)agreement have focused on effects within NPs (determiner-adjective-noun) in the written modality. It is still controversial to what extent LANs (especially in reading studies) result from component overlap between N400s and P600s (e.g., Tanner and Van Hell, 2014). Nevertheless, our data do provide evidence showing that early-onset sustained negativities in mismatch studies can show a clear left-anterior distribution that cannot be explained by component overlap. Since LANs in reading studies tend to have latencies and durations comparable to N400s (i.e., 300–500 ms), we are increasingly less convinced that sustained (left-)anterior negativities in auditory studies (e.g., Brink and Hagoort, 2004; Hasting and Kotz, 2008) are analogous to LAN components in reading studies. For our current data, we suggest that sustained negativities may index a continued evaluation of more complex cases of ambiguity resolution. The N400s we found virtually in all number mismatch conditions are rather difficult to interpret with confidence, as various accounts would predict N400-like components, including for standard morphosyntactic violations involving predictable inflectional morphemes (e.g., Tanner, 2015; BSS2019), truth-value related approaches (Bokhari, 2015), and phonological mismatch accounts (Connolly and Phillips, 1994). Molinaro et al. (2011a) have argued that phonotactics involved in agreement processes might demote grammatical processing (reflected by LANs) toward a lexical one (reflected by N400s). Our CONS verbs had a variety of final consonants (9 different consonants over our 60 verbs). These consonant changes do not follow systematic morphological rules. They are sometimes described as consonant deletion rules from the plural to the singular (Paradis and El Fenne, 1995). However, since singular forms are the default (and are acquired first), Royle (2011) argues against this approach and proposes rather that consonant-alternating forms in French are lexicalized (her research focused on adjectives, but the same logic can also be applied to verbs). This could promote use of lexical rather than grammatical processing when checking agreement, and thus explain N400 effects observed in plural conditions.

CONCLUSION

With the aim of testing whether cross-modal mismatches between pictures and *grammatical* sentences would elicit similar ERP components to those in the literature on linguistic anomalies, we developed an experiment with auditory-visual sentence-picture matching paradigms and an acceptability judgment task in French. We investigated neurocognitive mechanisms underlying lexico-conceptual semantics and grammatical number processing. This is the first study to test three different linguistic cues for number mismatches at different sentence positions. Our results demonstrated that native French speakers reliably exhibit N400 components in response to cross-modal verb-action mismatches, comparable to previous effects found for noun-object mismatches. Auditory-visual number mismatches usually elicited a biphasic N400-P600 (in some cases superimposing a sustained AN), and our context manipulation demonstrated that participants use the first available sentence cue to disambiguate structures. ERP effects at sentence onset and on the verb suggest that participants immediately tracked mismatches between modalities as soon as conflicting information became available, and that these mismatches were processed in a way that is not

fundamentally different from purely linguistic within-sentence agreement violations.

Our paradigm is exciting for a number of reasons, one being that we used grammatical sentences to induce "agreement error" processing, and elicited well-known ERP components. This approach has the advantage of being more ecologically valid than error-based paradigms, as it resembles more closely the mostly error-free speech we are exposed to daily. Having developed this experiment for younger populations, we are confident that our approach will reveal, in children, what types of information are being used at which point in the speech stream to disambiguate information. This type of paradigm also has potential for the study of developmental language disorder as well as second-language learning, as is the visual-world paradigm used in eye-tracking studies (e.g., Hopp and Lemmerth, 2018).

We can anticipate future directions of inquiry from this initial study of verb-based visual-auditory mismatches. As we have seen, not all incongruent number mismatch conditions elicited strong P600s despite the fact that we used a judgment task, which promotes this component. The N400 component seemed to be a more reliable reflection of our mismatch errors. This might in part be due to the fact that we did not use ungrammatical sentences as input, reducing error-detection based strategies that could have been used in most studies that find the LAN or the P600. Our robust N400s instead of LANs (or ANs), and less robust P600s for mismatches, might be the result of our sentences' *grammatical* status.

As we have appealed to sociolinguistics to explain some of our results, it appears interesting to pursue sociolinguistic studies using ERPs. This combination of domains has rarely been explored and we can identify straightforward implementations, as in second language acquisition research, to study variability in grammars within geographically constrained but linguistically diverse speakers of the same language. Paying attention to how a speaker implements a particular linguistic rule has strong potential to help us better understand the neurocognitive underpinnings of within-group variability in language processing.

In conclusion, our study provides a significant contribution to the field of cognitive neuroscience of language by providing high-quality evidence regarding the generalizability of ERP profiles across modalities and languages. This study extends lexico-semantic mismatches to the domain of verbs, provides insight into context effects and early detection of mismatches, establishes ERP patterns for different types of morpho-phonological and morpho-syntactic cues for number mismatch processing, and demonstrates that even grammatical sentences can elicit ERP patterns associated with "error" processing.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of McGill's Institutional Review Board (IRB) and University of Montreal's Comité d'éthique à la recherche en Santé of the Faculty of Medicine, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the McGill's Institutional Review Board (IRB) and University of Montreal's Comité d'éthique à la recherche en Santé of the Faculty of Medicine.

AUTHOR CONTRIBUTIONS

ÉC and LM collected the data, performed the analyses and the statistics, and should equally be viewed as first authors. LM wrote the first draft of the introduction, the methods section, and certain sections of the results as a part of her master's thesis. ÉC, PR, and KS wrote and edited the final version of the paper. KS and PR developed the general idea for the study and oversaw all stages of data analysis. PR, ÉC, and KS developed the theoretical background and issues addressed in the paper, and wrote the results and discussion of the manuscript. All authors designed the experiment and contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This project was funded by SSHRC *Insight* grants to Royle, Steinhauer et al. (435-2015-1280) and to Steinhauer and Royle (435-2013-0583), by an FQRSC team grant to White, Steinhauer, Royle et al. (2016-SE-188196), by Graduate Scholar stipends to ÉC (CRBLM and *Sciences biomédicales, Université de Montréal*), and by a *Premier* internship grant (Faculty of medicine, *Université de Montreal*) to Mélanie Weiss.

ACKNOWLEDGMENTS

We thank Khadidja Meftah and Mélanie Weiss who helped record and process the auditory stimuli, and collect data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01152/full#supplementary-material

REFERENCES

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., et al. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neurosci. Lett.* 322, 5–8. doi: 10.1016/S0304-3940(01)02528-9

Balconi, M., and Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *J. Psycholinguist. Res.* 34, 71–98. doi: 10.1007/s10936-005-3633-6

Barber, H., and Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17, 137–153. doi: 10.1162/0898929052880101

Barbet, C., and Thierry, G. (2016). Some alternatives? *Event-related potential investigation of literal and pragmatic interpretations of some presented in isolation. Front. Psychol.* 7:1479. doi: 10.3389/fpsyg.2016.01479

Bokhari, F. S. (2015). *Investigating the neurocognitive mechanisms underlying truth-conditional and logical semantic aspects of sentence processing: An event-related brain potential study*. Master's thesis, McGill University. Available online at: http://digitool.library.mcgill.ca/R/-?func=dbin-jump-full&object_id=135679&silo_library=GEN01 (accessed May 10, 2019)

Bornkessel-Schlesewsky, I., and Schlesewsky, M. (2019). Towards a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* 10:298. doi: 10.3389/fpsyg.2019.00298

Bourget, M.-J. (1987). "Variation phonétique dans l'emploi des pronoms de troisième personne en français," in *Tendances Actuelles de la Recherche sur la Langue Parlee [Current Trends in Research on Oral Language]*, eds J. Auger, R. Grenier, R. Lapalme, J.-F. Montreuil, and P. Whitmore (Quebec, QC: Université Laval), 129–140.

Brink, D. V. D., and Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *J. Cogn. Neurosci.* 16, 1068–1084. doi: 10.1162/0898929041502670

Brown, C., and Hagoort, P. (1993). The processing nature of the N400: evidence from masked priming. *J. Cogn. Neurosci.* 5, 34–44. doi: 10.1162/jocn.1993.5.1.34

Connolly, J. F., and Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266. doi: 10.1162/jocn.1994.6.3.256

Coulson, S., King, J. W., and Kutas, M. (1998). Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Process.* 13, 21–58. doi: 10.1080/016909698386582

DeLong, K. A., Urbach, T. P., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121. doi: 10.1038/nn1504

Dube, S., Kung, C., Peter, V., Brock, J., and Demuth, K. (2016). Effects of type of agreement violation and utterance position on the auditory processing of subject-verb agreement: an ERP study. *Front. Psychol.* 7:1276. doi: 10.3389/fpsyg.2016.01276

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Foucart, A., and Frenck-Mestre, C. (2011). Grammatical gender processing in L2: electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Biling. Lang. Cogn.* 14, 379–399. doi: 10.1017/S136672891000012X

Foucart, A., and Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? *Evidence from ERPs and eye-tracking. J. Mem. Lang.* 66, 226–248. doi: 10.1016/j.jml.2011.07.007

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., and Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: evidence from ERPs. *Acta Psychol.* 128, 528–536. doi: 10.1016/j.actpsy.2007.12.007

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84. doi: 10.1016/S1364-6613(00)01839-8

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* 91, 1357–1392. doi: 10.1152/physrev.00006.2011

Friederici, A. D., Mecklinger, A., Spencer, K. M., Steinhauer, K., and Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: a spatio-temporal analysis of event-related brain potentials. *Cogn. Brain Res.* 11, 305–323. doi: 10.1016/S0926-6410(00)00065-3

Friederici, A. D., Pfeifer, E., and Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.* 1, 183–192. doi: 10.1016/0926-6410(93)90026-2

Friedrich, M., and Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *J. Cogn. Neurosci.* 16, 1465–1477. doi: 10.1162/0898929042304705

Goldschneider, J. M., and DeKeyser, R. M. (2005). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: a meta-analysis of multiple determinants. *Lang. Learn.* 55, 27–77. doi: 10.1111/j.0023-8333.2005.00295.x

Haebig, E., Weber, C., Leonard, L. B., Deevy, P., and Tomblin, J. B. (2017). Neural patterns elicited by sentence processing uniquely characterize typical development, SLI recovery, and SLI persistence. *J. Neurodev. Disord.* 9:22. doi: 10.1186/s11689-017-9201-1

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* 15, 883–899. doi: 10.1162/089892903322370807

Hagoort, P., and Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia* 38, 1531–1549. doi: 10.1016/S0028-3932(00)00053-1

Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205. doi: 10.1162/089892999563328

Hasting, A. S., and Kotz, S. A. (2008). Speeding up syntax: on the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *J. Cogn. Neurosci.* 20, 1207–1219. doi: 10.1162/jocn.2008.20083

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology* 30, 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x

Holcomb, P. J., and Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Process.* 5, 281–312. doi: 10.1080/01690969008407065

Hopp, H., and Lemmerth, N. (2018). Lexical and syntactic congruency in L2 predictive gender processing. *Stud. Sec. Lang. Acquis.* 40, 171–199. doi: 10.1017/S0272263116000437

Itzhak, I., Pauker, E., Drury, J. E., Baum, S. R., and Steinhauer, K., (2010). Event-related potentials show online influence of lexical biases on prosodic processing. *Neuro Rep.* 21, 8–13. doi: 10.1097/WNR.0b013e328330251d

Kolk, H. H., Chwilla, D. J., Van Herten, M., and Oor, P. J. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain Lang.* 85, 1–36. doi: 10.1016/S0093-934X(02)00548-5

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., and Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Mem. Cognit.* 11, 539–550. doi: 10.3758/BF03196991

Kutas, M., Van Petten, C. K., and Kluender, R. (2006). "Psycholinguistics electrified II (1994–2005)," in *Handbook of Psycholinguistics, 2nd Edn*, eds M. J. Traxler and M. A. Gernsbacher (Boston, MA: Elsevier), 659–724. doi: 10.1016/B978-012369374-7/50018-3

Lau, E., Stroud, C., Plesch, S., and Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain Lang.* 98, 74–88. doi: 10.1016/j.bandl.2006.02.003

Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nat. Rev. Neurosci.* 9, 920–933. doi: 10.1038/nrn2532

Molinaro, N., Barber, H. A., and Carreiras, M. (2011a). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930. doi: 10.1016/j.cortex.2011.02.019

Molinaro, N., Vespignani, F., Zamparelli, R., and Job, R. (2011b). Why brother and sister are not just siblings: repair processes in agreement computation. *J. Mem. Lang.* 64, 211–232. doi: 10.1016/j.jml.2010.12.002

Morgan-Short, K., Sanz, C., Steinhauer, K., and Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: an event-related potential study. *Lang. Learn.* 60, 154–193. doi: 10.1111/j.1467-9922.2009.00554.x

Nevins, A., Dillon, B., Malhotra, S., and Phillips, C. (2007). The role of feature number and feature-type in processing Hindi verb agreement violations. *Brain Res.* 1164, 81–94 doi: 10.1016/j.brainres.2007.05.058

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Osterhout, L., and Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806. doi: 10.1016/0749-596X(92)90039-Z

Osterhout, L., and Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Lang. Cogn. Process.* 8, 413–437. doi: 10.1080/01690969308407584

Osterhout, L., and Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773. doi: 10.1006/jmla.1995.1033

Paradis, C., and El Fenne, F. (1995). French verbal inflection revisited: Constraints, repairs and floating consonants. *Lingua* 95, 169–204. doi: 10.1016/0024-3841(95)90105-1

Pauker, E., Itzhak, I., Baum, S. R., and Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *J. Cogn. Neurosci.* 23, 2731–2751. doi: 10.1162/jocn.2011.21610

Poplack, S., and Walker, D. (1986). "Going through (L) in canadian French," in *Diversity and Diachrony*, ed D. Sankoff (Philadelphia, PA: John Benjamins), 137–198. doi: 10.1075/cilt.53.17pop

Pourquie, M., St. Denis, A., and Royle, P. (2016). *Flexion verbale et structure argumentale dans les troubles primaires du langage.* [Verb inflection and agreement structure in primary language impairment]. Talk presented at the Congrés international sur l'orthophonie et l'audiologie, École d'orthophonie et d'audiologie, Montreal, QC: Université de Montréal.

R Studio Team (2015). *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc. Available online at: http://www.rstudio.com/

Rossi, S., Gugler, M. F., Friederici, A. D., and Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: evidence from event-related potentials. *J. Cogn. Neurosci.* 18, 2030–2048. doi: 10.1162/jocn.2006.18.12.2030

Royle, P. (2011). "On the existence of C/Ø alternations in French adjectives: theoretical and empirical questions," in *Proceedings of the 17th ICPhS* (Honk Kong), 1730–1733. Available online at: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Royle/Royle.pdf (accessed October 26, 2016).

Royle, P., Drury, J. E., and Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *Ment. Lex.* 8, 216–244. doi: 10.1075/ml.8.2.05roy

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* 137, 29–39. doi: 10.1016/j.bandl.2014.07.010

Selkirk, E. O. (1981). "English compounding and the theory of word-structure," in *The Scope of Lexical Rules,* eds M. Moortgat, H. Van der Hulst and T. Hoestra (Dordrecht: Foris), 229–277.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain Lang.* 86, 142–164. doi: 10.1016/S0093-934X(02)00542-4

Steinhauer, K., Alter, K., and Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat. Neurosci.* 2, 191–196. doi: 10.1038/5757

Steinhauer, K., and Connolly, J. F. (2008). "Event-related potentials in the study of language," in *Handbook of the Neuroscience of Language*, ed B. Stemmer (New York, NY: Elsevier Ltd.), 91–104. doi: 10.1016/B978-0-08-045352-1.00009-4

Steinhauer, K., and Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* 120, 135–162. doi: 10.1016/j.bandl.2011.07.001

Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., and Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: evidence from event-related potentials. *Neuropsychologia* 48, 1525–1542. doi: 10.1016/j.neuropsychologia.2010.01.013

Steinhauer, K., and Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *J. Psycholinguist. Res.* 30, 267–295. doi: 10.1023/A:1010443001646

Steinhauer, K., Royle, P., Drury, J. E., and Fromont, L. A. (2017). The priming of priming: evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neurosci. Lett.* 651, 192–197. doi: 10.1016/j.neulet.2017.05.007

Stowe, L. A., Kaan, E., Sabourin, L., and Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition* 176, 232–247. doi: 10.1016/j.cognition.2018.03.011

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: a commentary on "grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex* 66, 149–155. doi: 10.1016/j.cortex.2014.04.007

Tanner, D., and Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia* 56, 289–301. doi: 10.1016/j.neuropsychologia.2014.02.002

Tokowicz, N., and MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: an event-related potential investigation. *Stud. Second Lang. Acquis.* 27, 173–204. doi: 10.1017/S0272263105050102

Van Berkum, J. J., Brown, C. M., and Hagoort, P. (1999). Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182. doi: 10.1006/jmla.1999.2641

van de Meerendonk, N., Kolk, H. H., Chwilla, D. J., and Vissers, C. T. W. M. (2009). Monitoring in language perception. *Lang. Linguist. Compass* 3, 1211–1224. doi: 10.1111/j.1749-818X.2009.00163.x

Willems, R. M., Özyürek, A., and Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.* 20, 1235–1249. doi: 10.1162/jocn.2008.20085

Ye, Z., Luo, Y. J., Friederici, A. D., and Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Res.* 1071, 186–196. doi: 10.1016/j.brainres.2005.11.085

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CORRECTION
published: 05 May 2020
doi: 10.3389/fpsyg.2020.00860

Corrigendum: Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences

Émilie Courteau[1,2†], *Lisa Martignetti*[2,3†], *Phaedra Royle*[1,2] *and Karsten Steinhauer*[2,3*]

[1] Faculty of Medicine, School of Speech Language Pathology and Audiology, University of Montreal, Montreal, QC, Canada, [2] Centre for Research on Brain, Language and Music (CRBLM), Montreal, QC, Canada, [3] Faculty of Medicine, School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada

OPEN ACCESS

Edited and reviewed by:
Sendy Caffarra,
Stanford University, United States

***Correspondence:**
Karsten Steinhauer
karsten.steinhauer@mcgill.ca

† First authors

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: 30 March 2020
Accepted: 07 April 2020
Published: 05 May 2020

Citation:
Courteau É, Martignetti L, Royle P and Steinhauer K (2020) Corrigendum: Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences. Front. Psychol. 11:860. doi: 10.3389/fpsyg.2020.00860

Keywords: subject-verb number agreement, event-related brain potentials (ERPs), auditory-visual sentence-picture matching paradigm, cross-modal number mismatches, French language, online grammaticality judgment, N400 and P600, sustained frontal negativity

A Corrigendum on

Eliciting ERP Components for Morphosyntactic Agreement Mismatches in Perfectly Grammatical Sentences

by Courteau, É., Martignetti, L., Royle, P., and Steinhauer, K. (2019). Front. Psychol. 10:1152. doi: 10.3389/fpsyg.2019.01152

In the original article, there was a mistake in **Figure 4** as published. Instead of correctly describing effects in singular (A) and plural (B) NPs, as in the figure caption, the legend incorrectly describes NP contexts in (A) and neutral contexts in (B). The corrected **Figure 4** appears below.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

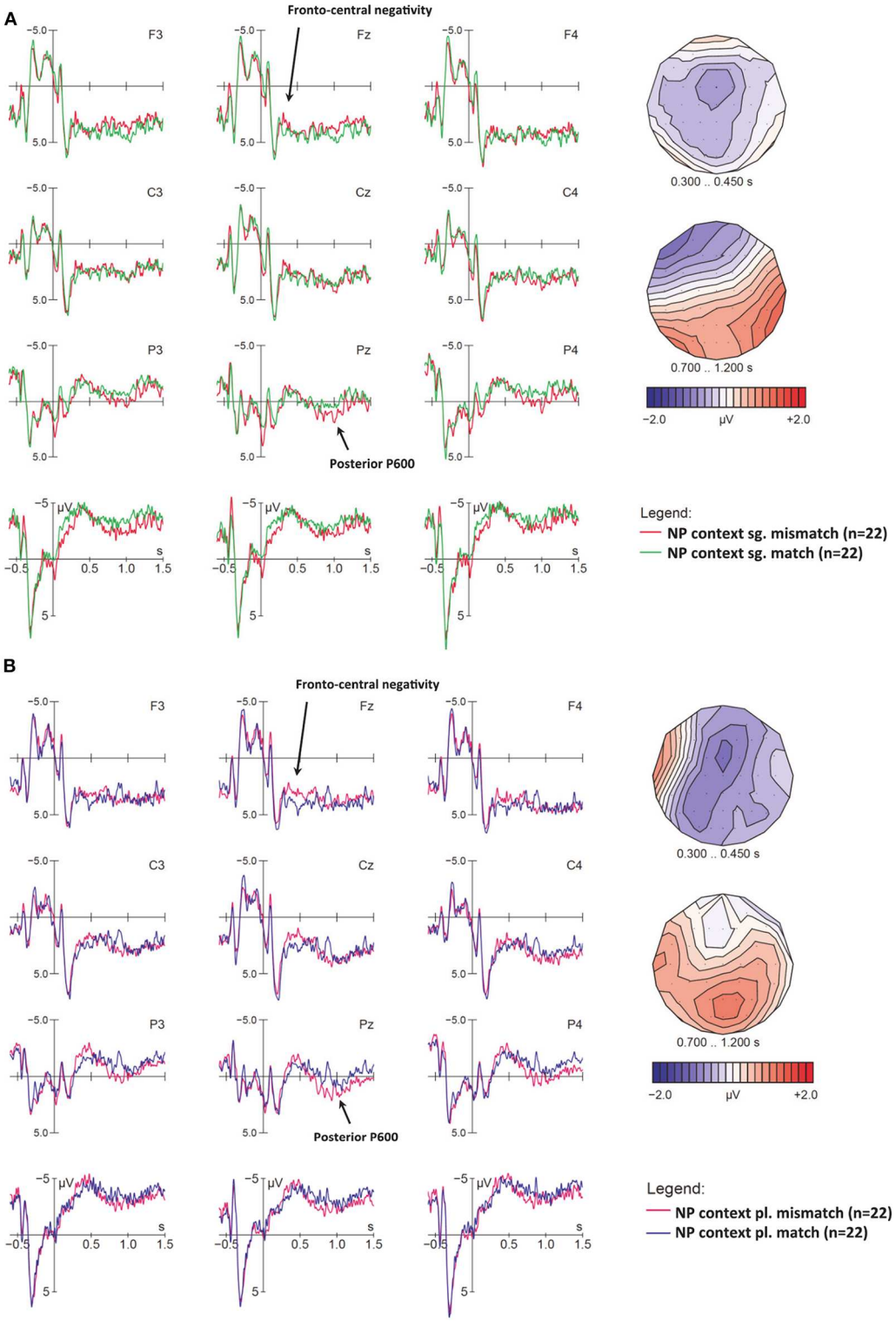


FIGURE 4 | Early effects of cross-modal number mismatches in NP contexts, for **(A)** singular and **(B)** plural NPs at sentence onset. ERPs are time-locked to the onset of the determiner (vertical bar) with a baseline of −600 to 0 ms; voltage maps illustrate the difference waves of relevant effects. **(A)** Singular mismatches (red) elicited a small fronto-central negativity in the N400 time-window relative to singular matches (green), as well as a parietal P600. **(B)** Plural mismatches (magenta) elicited a larger N400 as well as a parietal P600, as compared to plural matches (blue). Voltage maps of these effects (mismatch minus control) show that singular and plural mismatches elicited quite similar components.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 September 2019
doi: 10.3389/fpsyg.2019.02105

Person Features and Lexical Restrictions in Italian Clefts

Cristiano Chesi and Paolo Canal*

Research Center for Neurocognition, Epistemology and Theoretical Syntax (NETS), School of Advanced Studies, Istituto Universitario di Studi Superiori (IUSS), Pavia, Italy

In this paper, we discuss the results of two experiments, one off-line (acceptability judgment) and the other on-line (eye-tracking), targeting Object Cleft (OC) constructions. In both experiments, we used the same materials presenting a manipulation on person features: second person plural pronouns and plural definite determiners alternate in introducing a full NP ("it was $[_{DP1}$ the/you $[_{NP}$ bankers$]]_i$ that $[_{DP2}$ the/you $[_{NP}$ lawyers]] have avoided $_i$ at the party") in a language, Italian, with overt person (and number) subject-verb agreement. As results, we first observed that the advantage of the bare pronominal forms reported in previous experiments (Gordon et al., 2001; Warren and Gibson, 2005, a.o.) is lost when the full NP (the "lexical restriction" in Belletti and Rizzi, 2013) is present. Second, an advantage for the mismatch condition, *Art*$_1$-*Pro*$_2$, in which the focalized subject is introduced by the determiner and the OC subject by the pronoun, as opposed to the matching *Pro*$_1$-*Pro*$_2$ condition, is observed, both off-line (higher acceptability and accuracy in answering comprehension questions after eyetracking) and on-line (e.g., smaller number of regressions from the subject region); third, we found a relevant difference between acceptability and accuracy in comprehension questions: despite similar numerical patterns in both off-line measures, the difference across conditions in accuracy is mostly not significant, while it is significant in acceptability. Moreover, while the matching condition *Pro*$_1$-*Pro*$_2$ is perceived as nearly ungrammatical (far below the mean acceptability across-conditions), the accuracy in comprehension is still high (close to 80%). To account for these facts, we compare different formal competence and processing models that predict difficulties in OC constructions: similarity-based (Gordon et al., 2001, a.o.), memory load (Gibson, 1998), and intervention-based (Friedmann et al., 2009) accounts are compared to processing oriented ACT-R-based predictions (Lewis and Vasishth, 2005) and to *top-down Minimalist* derivations (Chesi, 2015). We conclude that most of these approaches fail in making predictions able to reconcile the competence and the performance perspective in a coherent way to the exception of the *top-down* model that is able to predict correctly both the on-line and the off-line main effects obtained.

Keywords: pronominal determiners, *top-down* derivation, complexity, cue-based retrieval, object cleft, intervention, similarity, memory load

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Ellen F. Lau,
University of Maryland, College Park,
United States
Andrea Santi,
University College London,
United Kingdom

****Correspondence:***
Cristiano Chesi
cristiano.chesi@iusspavia.it

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *09 December 2018*
Accepted: *30 August 2019*
Published: *20 September 2019*

Citation:
Chesi C and Canal P (2019) Person Features and Lexical Restrictions in Italian Clefts. Front. Psychol. 10:2105. doi: 10.3389/fpsyg.2019.02105

INTRODUCTION

A necessary condition for comprehending correctly an Object Relative clause (OR) is to interpret the head of this construction ["the banker" in (1)] as the direct object of the predicate within the relative clause ("praised"):

(1) [**The banker**]$_i$ [that **the barber** praised$_{_i}$] climbed the mountain.

The fact that ORs are generally harder to process than Subject Relative (SR) clauses[1] has been shown systematically using self-paced reading experiments (since King and Just, 1991, a.o.), probe-task paradigms (Wanner and Maratsos, 1978, a.o.), eye movements tracking (Traxler et al., 2002, a.o.), or by monitoring the electrical (Weckerly and Kutas, 1999) or metabolic (Just et al., 1996, a.o.) activity of the brain. Focusing on ORs, Bever (1974) first noticed that their difficulty can be mitigated by varying the type of subject within the relative clause (examples from Gordon et al., 2001[2]):

(2) a. [**The banker**]$_i$ [that **you** praised _$_i$] climbed the mountain
b. [**The banker**]$_i$ [that **Ben** praised _$_i$] climbed the mountain

When pronouns are processed in the subject position within the relative clause, as in (2).a, self-paced reading experiments show that the critical verbal regions ("praised" and "climbed") are read faster than when proper names are present, (2).b; when definite descriptions occupy both the head and the subject position, we obtain the slowest performance on the same critical verbal regions, as in (1)[3].

This effect has been extensively studied both from the theoretical/competence perspective (Friedmann et al., 2009; Belletti and Rizzi, 2013, a.o.) and from the psycholinguistic/performance one (Gordon et al., 2004, a.o.), especially in Object Clefts (OCs), when both the subject and the focalized DP can be definite descriptions, proper names or pronouns, (3) (Warren and Gibson, 2005):

(3) It was [**the banker/Pat/you**]$_i$ that [**the lawyer/Dan/we**] avoided _$_i$ at the party

In this paper, we present a manipulation of person features [3rd (default) vs. 2nd person] in the paradigm in (3) to investigate the role of person agreement in Italian (an overt subject-verb person agreement language) under the presence of a "lexical restriction" (i.e., the NP introduced by the determiner): second person pronouns will be used as determiners and compared to definite articles under the presence of a plural lexical restriction, as exemplified in (4):

(4) Sono/siete [**gli/voi** architetti]$_i$ che [**gli/voi** ingegneri]
are$_{3P_PL}$/are$_{2P_PL}$/ the/you architects that the/you engineers
hanno/avete consultato _$_i$ prima di iniziare il lavoro.
have$_{3P_PL}$/have$_{2P_PL}$/consulted before beginning the work

This study consists of two new experiments (section Materials and Methods): an acceptability judgment (experiment 1) and an eyetracking study (experiment 2). Comparing off-line (acceptability scores in experiment 1 and accuracy in answering comprehension questions after eyetracking in experiment 2) and on-line evidence (all relevant eyetracking measures in experiment 2) we will evaluate the actual fit of some prominent model discussed in literature (section Predicting Processing Difficulties) aiming at explaining the contrasts revealed in (3): from the analysis (section Discussion) of the results of our experiments (section Results) we conclude that none of the models presented readily predicts the behavioral evidence revealed by this study. We will then argue in favor of the left-right, *top-down* derivational minimalist perspective (Chesi, 2015) where the "complexity" of the non-local dependency is computed using a Feature Retrieval and Encoding Cost (FREC) function: this model better integrates the on-line and off-line results gathered here.

In the first part of this paper, we will introduce the structural properties of the Object Cleft sentences under analysis (section The Properties of the Object Clefts (OCs) Under Analysis). A brief state-of-the-art summary on OC processing effects [section Processing Object Clefts (OCs)] will precede the summary of five major models and their predictions on the contrast previously tested: similarity-based (Gordon et al., 2001, section Similarity-Based Accounts), intervention-based (Friedmann et al., 2009, section Intervention-Based Accounts), and Dependency Locality Theory (DLT, Gibson, 1998, 2000, section *memory-load* Accounts) models will be compared with processing-oriented, memory-usage models that make explicit predictions on reading times and are possibly more transparent in terms of brain mechanisms involved: Lewis and Vasishth (2005) model based on ACT-R (section ACT-R-Based Predictions) and a *top-down*, left-right minimalist derivation based on Chesi (2015) (section *top-down* (Left-Right) Minimalist Derivations). Then we will concentrate on person features on the DP triggering overt verbal agreement, especially focusing on pronouns used as determiners (section Pronouns as Determiners and Agreement): this should clarify the rationale behind the proposed manipulation and the semantic/pragmatic impact of this construction on processing.

THE PROPERTIES OF THE OBJECT CLEFTS (OCS) UNDER ANALYSIS

Object Clefts (OCs) are peculiar focalization constructions in which a direct object is displaced in a prominent left-peripheral position (Rizzi, 1997). Following Belletti's (2008) analysis, in these structures the copula selects a truncated CP in which the object was moved into the FocP position as shown below:

(5) … BE [$_{CP}$ ~~Force~~ [$_{FocP}$ ***Object*** [$_{FinP}$ that [$_{TP}$ *Subject* … ***(Object)***]]]]

[1] In SR, the head is related to the subject position within the relative as in "the banker [that _ praised the barber] …".

[2] In this specific experiment, since "Ben"/"you" items are shorter than "the barber," the discussed effects could be biased by the relative NP length. However, the same contrast is reported in other experiments [e.g., (Warren and Gibson, 2002), discussed later] where longer names are used, so we assume here that the relevant contrast is genuine.

[3] Other factors are assumed to induce a facilitation in processing this configuration: for instance animacy of the subject while the object is inanimate (Kidd et al., 2007), but see Belletti and Chesi (2014) for discussion.

According to Belletti, OCs, as opposed to Subject Clefts[4] (SCs), can only convey contrastive/corrective focus (this is the role of FocP), then its realization will be felicitous and perfectly natural only in specific contexts. For instance in correcting/rectifying a statement as below:

(6) X: Ho sentito che Alberto ha salutato qualcuno prima di partire per le vacanze; ha per caso salutato Beatrice prima di partire? (Dopo il litigio che hanno avuto per colpa di Claudia sarebbe stato un segno distensivo)
I heard that A. said goodbye to someone before leaving for holidays; has he said good bye to B. before leaving? (After the fight they had because of C., this would have been a positive sign)
Y: (no, non-era Beatrice, purtroppo) era CLAUDIA che Alberto ha salutato prima di partire!
(no, it wasn't B, unfortunately) it was $C._{focalized}$ that A. said goodbye to before leaving!

Notice that a presupposition of existence [p.c. Benincà in (Belletti, 2008), (7).a] and uniqueness, as well as exhaustivity [as in Identificational Focus discussed in E. (Kiss, 1998), (8)], are also implied by the cleft constructions [contra standard focalization, both in root, (7).b, or in embedded contexts, (7).c]:

(7) a. *(non) è NESSUNO che ho incontrato (non-tutti)
it is (not) NOBODY that I met (not everybody)
b. NESSUNO ho incontrato (non-tutti)
NOBODY I met (not everybody)
c. ho detto che NESSUNO assumeranno (non-tutti)
I have said that NOBODY they will hire (not everybody)

(8) è UNA MELA che ho mangiato (non-una pera o qualcos'altro)
it is AN APPLE that I have eaten (not a pear or anything else)

Despite their peculiarities, these are perfect configurations for testing non-local crossing dependencies in comprehension: from a processing perspective, the distal argument (the focalized DP) must be retained in memory and retrieved, later on, when the verbal predicate is encountered, crucially after the subject has been interpreted as the agent of the predication. Notice that the absence of an appropriate context does not preclude the possibility of correctly processing and interpreting these constructions: in all the experiments that will be mentioned in the next section [section Processing Object Clefts (OCs)], any context introducing OCs was absent.

Processing Object Clefts (OCs)

The performance contrasts elicited by OCs suggest that the nature of both DPs present in the construction plays a major role (Gordon et al., 2001, 2004; Warren and Gibson, 2005): a definite DP (D), a proper name (N), or a pronoun (P) occupying the two relevant positions produce different effects according to their relative distribution. The full prototypical paradigm (Warren and Gibson, 2005), introduced in (3) and expanded below in (9) for convenience, is used to illustrate these contrasts.

(9)

	$object_{focalized}$	subject	verb	spill-over	condition
a.	It was **the banker**	that **the lawyer**	**avoided** _	at the party	$[D_1\text{-}D_2]$
a'.	It was **the banker**	that **Dan**	**avoided** _	at the party	$[D_1\text{-}N_2]$
a".	It was **the banker**	that **we**	**avoided** _	at the party	$[D_1\text{-}P_2]$
b.	It was **Patricia**	that **the lawyer**	**avoided** _	at the party	$[N_1\text{-}D_2]$
b'.	It was **Patricia**	that **Dan**	**avoided** _	at the party	$[N_1\text{-}N_2]$
b".	It was **Patricia**	that **we**	**avoided** _	at the party	$[N_1\text{-}P_2]$
c.	It was **you**	that **the lawyer**	**avoided** _	at the party	$[P_1\text{-}D_2]$
c'.	It was **you**	that **Dan**	**avoided** _	at the party	$[P_1\text{-}N_2]$
c".	It was **you**	that **we**	**avoided** _	at the party	$[P_1\text{-}P_2]$

Warren and Gibson (2005) evidence [average reading times reported in (10)[5]], based on the paradigm in (9), shows that the greatest slowdown in self-paced reading at the critical verbal segment (*avoided*) is associated to the D_2 condition [i.e., when the subject of the cleft is a definite description, (9).a,b,c]. This correlates with the lowest accuracy rate in comprehension questions. Similar (non-significantly different) reading times are revealed for the N_1-N_2 matching condition (9).b′, while the P_1-P_2 matching condition (always presenting a person mismatch), as well as the other conditions in which *P* is the subject of the cleft, produce the fastest reading times of the critical verbal region.

(10)

Condition	Average RT (SE) ms
$\mathbf{D_1\text{-}D_2}$	365 (19)
$\mathbf{D_1\text{-}N_2}$	319 (12)
$\mathbf{D_1\text{-}P_2}$	306 (14)
$\mathbf{N_1\text{-}D_2}$	348 (18)
$\mathbf{N_1\text{-}N_2}$	347 (21)
$\mathbf{N_1\text{-}P_2}$	291 (14)
$\mathbf{P_1\text{-}D_2}$	348 (18)
$\mathbf{P_1\text{-}N_2}$	311 (15)
$\mathbf{P_1\text{-}P_2}$	291 (13)

The authors reported a reliable effect of subject type with P_2 conditions averaging 30 ms faster than N_2 conditions, which is 28 ms on average faster than D_2 conditions. A marginally significant interaction (mainly driven by the slowest D_1-D_2 and N_1-N_2 conditions) indicates that the matching conditions, overall, are significantly slower than the mismatch conditions.

Also in comprehension, Warren and Gibson reported a main effect on the subject type (D_2 condition is harder than N_2, which is harder than P_2) and an interaction between focalized object type and subject type, with all matching conditions inducing lower accurate results.

These results confirmed and expanded other results discussed in literature (e.g., Gordon et al., 2001).

[4]Belletti (2008) assumes that, in Subject Clefts, the Subject raises to CP to satisfy an EPP feature (hence the CP becomes a small clause in the sense of Stowell, 1983); this would create an intervention context in case the object or the indirect object would move across the subject position. As a consequence, only the subject can realize the new information focus, using the *vP* periphery of the matrix copula while the focalized (indirect) object moves to the CP peripheral FocP, where no EPP is present (exactly as a wh-item). In this position, it can only express contrastive/corrective (and not simply new information) focus.

[5]Thanks to Tessa Warren for sharing the original data. Differences in gray shades indicate statistically significant differences.

Predicting Processing Difficulties

Various models have been proposed to account for the performance data presented in the previous section. Here five models will be discussed, all considering as key factors: (i) the nature of the DPs involved in the dependency (*memory-load* account, Warren and Gibson, 2005, section *memory-load* Accounts), (ii) the similarity between the two DPs (*similarity-based*, Gordon et al., 2001 and section Similarity-Based Accounts *intervention-based* Friedmann et al., 2009, section Intervention-Based Accounts, accounts), (iii) the distance/activation of the focalized object with respect to the predicate (Lewis and Vasishth, 2005, section ACT-R-Based Predictions), or a combination of these factors (*top-down Minimalist* account, Chesi, 2015 section *top-down* (Left-Right) Minimalist Derivations). The predictions these models differ substantially both in terms of general complexity factor (DP types vs. matching/mismatching conditions), relevant features inducing difficulty and the exact moments in which such difficulty can be revealed (encoding at DP vs. retrieval at VP; on-line as slow down at specific regions vs. off-line as comprehension accuracy).

Similarity-Based Accounts

Gordon et al. (2001) explicitly focus on working memory demands in their studies using self-paced reading paradigms. Their proposal is based on the idea that having two DPs "of the same kind" stored in memory makes the OR/OC more complex than SR/SC. This is sufficient to model memory interference during encoding, storage, and retrieval (Crowder, 1976). When similarity between DPs is calculated considering noun type (proper vs. common), gender, number, animacy, case, and person, this theory is sufficient to predict asymmetries for most of the contrasts presented in section Processing Object Clefts (OCs): D_1-D_2 and N_1-N_2 matching condition are expected to be the hardest configurations, while the P_1-P_2 matching configuration might result slightly easier than the other matching conditions because of person features mismatch. In all other mismatch cases, this approach predicts (both on-line and off-line) lighter effects because of the difference in type/features without being able to distinguish between D_1-N_2 [easier, according to (10)] and N_1-D_2 [harder in (10)] conditions or between P_1-D_2 [harder in (10)] and D_1-P_2 [easier in (10)]. This is also expected under the assumption that all features equally contribute to memory confusion.

Memory-Load Accounts

Memory-load accounts (Gibson, 1998; Warren and Gibson, 2002, 2005, a.o.) explain most of the contrasts presented in section Processing Object Clefts (OCs) by postulating an "integration cost" (Gibson, 1998, Syntactic Prediction Locality Theory, SPLT) proportional to new discourse referents[6]: since pronouns do not introduce new discourse referents and names are referentially lighter than definite descriptions (Warren and Gibson, 2005), memory-load accounts predict faster reading times at the cleft verbal region when the subject is a pronoun and slightly longer reading times when it is a proper name. On the other hand, this account incorrectly predicts faster reading times for the N_1-N_2 matching condition ("it was Patricia that Dan *avoided* at the party") than for the D_1-D_2 condition ("it was the lawyer that the businessman *avoided* at the party"), even though no significant differences emerged from this contrast.

Intervention-Based Accounts

The intervention-based accounts (Friedmann et al., 2009; Belletti and Rizzi, 2013, a.o.) can explain the symmetry revealed in the D_1-D_2 and N_1-N_2 matching conditions in terms of similarity of the critical intervening features: Friedmann et al. (2009), building on Rizzi (1990) locality constraint, assume that whenever features are shared between a filler, *X* [e.g., "the banker" in (1)] and a structural intervener, *Z* [e.g., "the barber" in (1)], the relation between *X* and the related selected gap, *Y*, gets disrupted in a way that is proportional to the kind (and number) of features involved. Assuming that lexical restriction, rather than referentiality (c.f. section *memory-load* Accounts), is computed and that features expressing such lexical restriction in definite descriptions, proper names and pronouns are distinct (they assume *N* for common nouns, N_{prop} for proper names, and a null *N* for pronouns), the intervention-based account predicts exactly that the matching conditions in which common nouns and proper names are present are comparable, while pronouns induce easier processing since *N* is absent. A crucial assumption here is that only features triggering movement should cause intervention (Friedmann et al., 2009:83). In this respect, the lexical restriction should not play a significant role, since this "feature" is buried within the DP and does not seem to trigger movement. However, (Belletti and Rizzi, 2013) explicitly consider the lexical restriction as a movement trigger[7], hence rescuing the idea that its presence has an impact in terms of intervention. This model does not predict differences when pronouns are in the focalized object or in the subject position, neither it makes explicit predictions in the D_1-N_2 and N_1-D_2 cases: being *N* and N_{prop} distinct, either we assume that they play a role in triggering movement, hence expecting milder effects than in the matching conditions, or we assume that *N* and N_{prop} are not involved in movement hence these cases should be comparable with respect to the other matching cases. Since verb region in the D_1-N_2 is significantly faster than in the N_1-D_2 condition according to (10), neither assumptions lead to the correct prediction.

[6]Assuming an accessibility hierarchy (Ariel, 1990), the model postulates a discourse referent cost following this referentiality scale: (less accessible) definite description > proper names > referential pronoun (*he/she/him/her*) > deictic pronoun (*I/we* and *you*) (more accessible). The structural integration cost is proportional to the referent cost.

[7]The relevant opposition they propose is the one discussed in Munaro (1999):

(i) a. Con **che tosat** à-tu parlà?
with which boy did you speak?
b. Avé-o parlà de **chi**?
have you spoken of whom?

Notice that the difference between the *wh*-items involved in this construction, *per se*, could be responsible for the different landing sites; the relevant comparison should involve a minimal pair of the kind "that" vs. "that boy" for which distinct positions are targeted according to the presence of the lexical restriction.

ACT-R-Based Predictions

Lewis and Vasishth (2005) present an explicit moment-by-moment model of parsing[8] based on independently motivated working memory principles. Their predictions, both on similarity effects (c.f. section Similarity-Based Accounts) and probability to retrieve the correct, accessible, syntactic chunk over time (c.f. section *memory-load* Accounts), follow from their assumptions based on an implementation of some components of the Adaptive Control of Thought-Rational (ACT-R) architecture (Anderson and Matessa, 1997; Anderson, 2005) within the sentence comprehension perspective. By focusing on working-memory retrieval, this model is able to estimate precisely the integration cost of a non-local constituent relying on its distance from its re-attachment point: the structural chunks are stored in memory and their activation (i.e., a purely numerical value) fades over time; stored chunks receive an activation boost whenever re-accessed. The longer is the time passed after the last re-activation, the longer it will take to retrieve the correct chunk. This plainly explains the difference between retrieving the subject in a SC, which is relatively fast, or the focalized object in OCs, which is relatively slow due to the time spent in attaching the intervening subject. This model can be used to simulate memory decay and difficulty in re-accessing specific constituents. Moreover, since their attempt is to explicitly describe processes and memory structures giving rise to specific linguistic configuration by providing a psychologically motivated theory of processing, this approach has a high explanatory potential. However, unless specific cues pre-activate the object (e.g., agreement as in clitic doubling constructions), this model can hardly predict the relevant asymmetries in the paradigm discussed in (9): we could assume retrieval and attachment of 1st and 2nd person pronouns to be slightly faster than default 3rd person DPs because of their higher saliency in the context; in this sense when P is in the focalized position (P_1) or in the subject position (P_2), this might somehow reduce the general cost paid for retrieving a distal argument, but any prediction in all other cases requires extra assumptions.

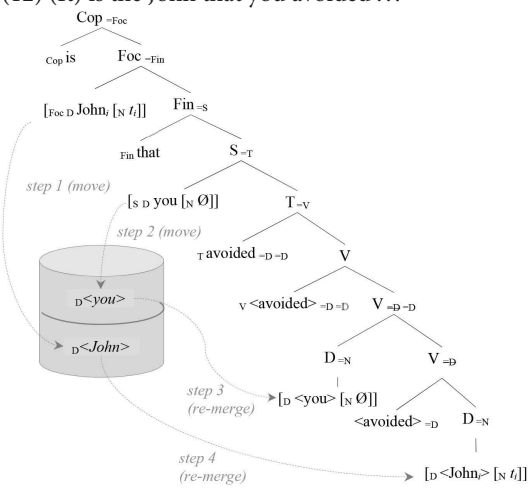
Top-Down (Left-Right) Minimalist Derivations

An alternative way to look at processing, without assuming any specific parsing algorithm or declarative grammatical rule format (as in Lewis and Vasishth, 2005), while maintaining an incremental left-right perspective, is presented in Chesi (2015, 2017): in the *Minimalist top-down* derivation proposed, *merge* is the sole structure building operation and it operates by attaching new (incoming) items always to the right of the phrase structure built so far (c.f. Phillips', 1996, *merge right*). The integration of new items is guided by the expectations triggered by the select feature(s) lexically encoded in the items already merged: for instance, a lexical entry like [$_V$ run $_{=D}$][9] indicates that the verb ("V") "run" selects a determiner ("=D"), namely that a DP is expected next; expectations are always projected after the lexical item is processed. If a category [$_X$] is expected (as result of the expansion of "=X" select feature) either a new [$_X$ *item*] or [$_X$ $_Y$ *item*] can be merged next; in the second case, [$_Y$ <*item*>] must be stored in memory since the "Y" categorial feature was unexpected (i.e., unselected) and must be remerged in the structure as soon as a lexical item with a "=Y" select feature is processed. This is how a non-local, filler-gap dependency (movement) is implemented. In (11), we exemplify the OC derivation using the grammatical knowledge (*Lex*) for the relevant paradigm in (9)[10]:

(11) *Lex* = {
Categories and default selections = {
[$_{Cop}$ =Foc] [$_{Foc}$ =Fin], [$_{Fin}$ =S], [$_S$ =T], [$_T$ =V],
[$_D$ =N], [$_N$], [$_V$]
}
Lexical items = {
[$_{(S)\ (Foc)\ D}$ John$_i$ [$_N$ t_i]], [$_{(S)\ (Foc)\ D}$ you [$_N$ Ø]],
[$_{(S)\ (Foc)\ D}$ the], [$_N$ banker],
[$_{Fin}$ that], [$_{T\ V}$ avoided $_{=D}$ [$_{=D}$]], [$_{Cop}$ is]
}
}

The relevant part of the derivation (equivalent both in generation and in parsing) can be schematized as follows:

(12) (It) is the John that you avoided …



This tree diagram summarizes the history of the derivation, which implements the OC analysis presented in (5) and it is transparent with respect to processing (both parsing and generation).

As we see after step 1 and step 2, both arguments must be stored in memory because of their unexpected (i.e., unselected) "argumental" (D) feature right after they get first merged into the

[8]Context-Free Grammar rules (Chomsky, 1965) and a Left Corner Parsing algorithm (Aho and Ullman, 1972) is assumed as well as various HPSG details (Pollard and Sag, 1994). Refer to **Supplementary Materials** section ACT-R-Based Model for technical details.

[9]We adapt here Stabler's (1996) Minimalist Grammar formalism: features associated to a lexical item are enclosed under squared brackets; categorial features are capital letters to the left of the lexical entry (e.g., "V" in [$_V$ run …]), select features are ordered after a lexical entry and are prefixed by the equal sign (e.g., "=D" in [… run $_{=D}$]).

[10]*S* indicates the subject (topic-related/EPP) functional position (see Bianchi and Chesi, 2014 for a discussion of the subject-related positions in this framework). Round brackets indicate optional features, i.e., *(S)* and *(Foc)* can be associated to *D* or not. The structure of the various DPs implements Longobardi's (1994) raising analysis. Retrieval from *M*(emory) will always preempts lexical insertion; *M* is a last-in-first-out memory. Angled brackets indicate unpronounced copies of an item. See the **Supplementary Materials** section Top-Down Minimalist model for technical details and the full derivation.

structure. Both arguments are retrieved after the verb is merged and its selection requirements are expressed (steps 3 and 4). To predict processing difficulties at retrieval, we associate a cost to the memory buffer access: this cost grows exponentially with respect to the number of items stored (m), linearly with respect to the number of new features to be retrieved from memory (nF), and it is mitigated (linearly, again) by the number of distinct cued features (dF) by x (the region where retrieval is requested, in this case the verbal predicate). This is the core of the Feature Retrieval Cost (FRC)[11] function:

(13) $FRC(x) = \prod_{i=1}^{n} \frac{(1+nFi)^{m_i}}{(1+ dFi)}$

Using the lexicon in (11), we expect to retrieve definite descriptions like [$_D$ *the* [$_N$ *lawyer*]] (namely a D and a N category), proper names like *John* = [$_D$ *John*$_i$ [$_N$ t_i]] (i.e., a contextually salient D index and a coindexed N) and pronouns like *you* = [$_D$ *you* [$_N$ Ø]] (i.e., just a contextually salient D index; for more details on these DP analyses, see section Pronouns as Determiners and Agreement). With these feature structures, we obtain the following FRCs at the verb segment for each conditions in (9). These can be easily compared with the average reading times in (10):

(14)

Condition	Average RT (SE) ms	log(FRC)[12]
D$_1$-D$_2$	365 (19)	1,43
D$_1$-N$_2$	319 (12)	1,08
D$_1$-P$_2$	306 (14)	0,78
N$_1$-D$_2$	348 (18)	1,26
N$_1$-N$_2$	347 (21)	1,26
N$_1$-P$_2$	291 (14)	0,6
P$_1$-D$_2$	348 (18)	1,26
P$_1$-N$_2$	311 (15)	0,9
P$_1$-P$_2$	291 (13)	0,6

Pearson correlation between average reading times and log(FRC) is extremely significant: $r(7) = 0.98$, $p < 0.001$.

Notice that the FRC plainly subsumes (and integrates) Friedmann et al. (2009) account. Moreover, it precisely characterizes the triggers of the filler-gap dependency similarly to the cue-based memory retrieval approach (Van Dyke and McElree, 2006): we expect confusion (higher FRC) when the cued characteristic features are non-unique at retrieval. This also explains the cross-linguistic variation revealed, for instance, in Hebrew vs. Italian with respect to gender vs. number (Belletti et al., 2012): since in Hebrew the verb agrees with its subject also in gender, in Hebrew, but not in Italian, gender mismatch facilitates ORs processing. Under this perspective, this is because the verb uses such cues to retrieve the relevant argument from memory (hence, in FRC terms, gender mismatch increases dF in Hebrew since cued by the verb), while just number mismatch, but not gender mismatch, helps in Italian for the same reason (dF increases when cued number is in a mismatch configuration).

In addition to FRC, an encoding cost must be considered whenever an element is merged into the structure (similarly to Gibson's 1998 new discourse referent cost): this is the Feature Encoding Cost (FEC), a numerical value associated to each new item merged that is proportional to the number of new relevant features integrated in the structure:

(15) $FEC(x) = \sum_{i=1}^{n} eF_i$

eF is the cost of each new relevant feature to be encoded at x. For simplicity $eF = 1$ for a new categorial feature introduced (e.g., 1 for D and 1 for N), 2 for a duplication of the same lexical category still requiring a structural integration as selected argument (i.e., 2 for the second N both in D_1-D_2 and N_1-N_2), 0 otherwise. In the paradigm (9) the FEC predictions are the following ones:

(16)

	object$_{focalized}$	*subject*	*verb*	*spill-over*	*condition*
a.	It was **the banker**	that **the lawyer**	**avoided** _	at the party	[D$_1$-D$_2$]
	1 2	1 3	2	3	
a'.	It was **the banker**	that **Dan**	**avoided** _	at the party	[D$_1$-N$_2$]
	1 2	1 1	2	3	
a".	It was **the banker**	that **we**	**avoided** _	at the party	[D$_1$-P$_2$]
	1 2	1 0	2	3	
b.	It was **Patricia**	that **the lawyer**	**avoided** _	at the party	[N$_1$-D$_2$]
	1 1	1 2	2	3	
b'.	It was **Patricia**	that **Dan**	**avoided** _	at the party	[N$_1$-N$_2$]
	1 1	1 2	2	3	
b".	It was **Patricia**	that **we**	**avoided** _	at the party	[N$_1$-P$_2$]
	1 1	1 0	2	3	
c.	It was **you**	that **the lawyer**	**avoided** _	at the party	[P$_1$-D$_2$]
	1 0	1 2	2	3	
c'.	It was **you**	that **Dan**	**avoided** _	at the party	[P$_1$-N$_2$]
	1 0	1 1	2	3	
c".	It was **you**	that **we**	**avoided** _	at the party	[P$_1$-P$_2$]
	1 0	1 0	2	3	

[11] The intent of this formula (compared to more complex and complete ones, Van Dyke, 2007; Chesi, 2017) is to highlight the factors that are necessary and sufficient to account for the minimal variations in the paradigm under discussion.

[12] Logarithmic function (i.e., log(FRC)) is provided for making this cost directly usable in on-line predictions at a comparable scale with respect to other metrics [e.g., (Gibson, 1998) or (Lewis and Vasishth, 2005)]. This is how the FRC is calculated, condition by condition (see **Supplementary Materials** section Top-Down Minimalist model for more details):

D$_1$-D$_2$ = 9·3 = 27: 9 for retrieving D$_2$, since $nF = 2$ (D and N count as one), $m = 2$ because two DPs are in memory at this time, and $dF = 0$ because no feature is cued by the verb distinguishing one DP from the other; 3 for retrieving D$_1$, since $nF = 2$ (D and N are new), $m = 1$ and $dF = 0$;

D$_1$-N$_2$ = 4·3 = 12: 4 for retrieving N$_2$ ($nF = 1$, that is N, since D is contextually salient as we will see in section 0, $m = 2$, $dF = 0$), 2 for retrieving D$_1$ ($nF = 2$, m =1, $nF = 0$);

D$_1$-P$_2$ = 2·3 = 6: 2 for retrieving P$_2$ ($nF = 1$ even if deictic pronouns are contextually salient, the correct person must be retrieved, $m = 2$, $dF = 1$ since a distinct case on pronouns is cued by the verb), 3 for retrieving D$_1$ ($nF = 2$, $m = 1$, $nF = 0$)

N$_1$-D$_2$ = 9·2 = 18: 9 for retrieving D$_2$ ($nF = 2$, since D and N are new, $m = 2$, $dF = 0$), 2 for retrieving N$_1$ ($nF = 1$, $m = 1$, $nF = 0$);

N$_1$-N$_2$ = 9·2 = 18: 9 for retrieving N$_2$ ($nF = 2$ even though D should be contextually salient, being two proper names presents, the same D, i.e., a co-referential index, cannot be sufficient to distinguish them, then an extra cost must be paid here as in the D–D condition, $m = 2$, and $dF = 0$), 2 for retrieving N$_1$ ($nF = 1$, just N is new since the determiner is now contextually salient and unique, $m = 1$ and $dF = 0$);

N$_1$-P$_2$ = 2·2 = 4: 2 for retrieving P$_2$ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving N$_1$ ($nF = 1$, $m = 1$, $dF = 0$);

P$_1$-D$_2$ = 9·2 = 18: 9 for retrieving D$_2$ ($nF = 2$, $m = 2$, $dF = 0$); 2 for retrieving P$_1$ ($nF = 1$, $m = 1$, $dF = 0$);

P$_1$-N$_2$ = 4·2 = 8: 4 for retrieving N$_2$ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving P$_1$ ($nF = 1$, $m = 1$, $dF = 0$);

P$_1$-P$_2$ = 2·2 = 4: 2 for retrieving P$_2$ ($nF = 1$, $m = 2$, $dF = 1$); 2 for retrieving P$_1$ ($nF = 1$, $m = 1$, $dF = 0$);

More precisely, in accordance with the structural assumptions expressed in (11), definite descriptions generally require the encoding of two critical new features (a determiner and a nominal restriction), proper names one (the contextually salient determiner is "free" and the proper name nominal restriction costs 1), while deictic pronouns have no encoding cost since contextually already present in the context (hence already pre-activated).

The absence of cost and the extra cost associated, respectively, to an already introduced feature and to the duplication of a category is coherent with a conception of memory as a pattern associator: if the pattern p_f, encoding feature *f*, has been just activated, re-activating it should have a minor cost (priming effect), while forcing a differentiation in a fully-overlapping pattern should induce the recruitment of extra memory units, hence an extra cost.

Summary

Summarizing, **Table 1** reports the predictions made by the models just described and the average reading times revealed at the verb segment in (9) [data from (10)]:

Theories based on the referentiality hierarchy (Ariel, 1990; Gibson, 1998; Warren and Gibson, 2005, a.o.; *memory-load* prediction in the table) fail to predict that also N_1-N_2 matching condition induces a low performance comparable to the D_1-D_2 matching condition. Similarity-based accounts (Gordon et al., 2004, a.o.) capture this fact, but fail in distinguishing any order permutation in mismatching conditions (e.g., D_1-P_2 vs. P_1-D_2); Intervention-based accounts (e.g., Belletti and Rizzi, 2013, a.o.) correctly predict harder times with both D_1-D_2 and N_1-N_2 matching condition, also expecting better performances with pro-intervening conditions (i.e., D_1-P_2, N_1-P_2, P_1-P_2), but fail in predicting any distinction among other conditions (e.g., P_1-D_2 vs. D_1-P_2 or D_1-N_2 vs. D_1-N_2). Notice moreover that the processing costs at the verb segment do not follow from this perspective (as in any other bottom-to-top, movement-based approach). The ACT-R-based models, only relying on distance and pre-activation of the relevant argument, can predict easier retrieval only when *P* is present either at the subject or at the focalized object position. Also in these cases, P_1-D_2 (and more marginally P_1-N_2) would be predicted to be easier than it actually is. In the end, the *top-down* model (Chesi, 2015) correctly predicts more efforts in processing the D_1-D_2 and N_1-N_2 matching conditions (with D_1-D_2 being the hardest configuration), medium difficulty when D_2 and N_2 are integrated in the subject position (different encoding costs) and lighter effects when P_2 is present because of case (nominative "we" vs. "us" morphology in English, while 1st/2nd vs. 3rd person asymmetry can not be used as a cue because of past tense of the cleft predicate).

In conclusion, memory-load and *top-down* predictions present the highest level of correlation with respect to the revealed average reading times, but they make quite different predictions: in *memory-load* theories interference is irrelevant, while the *top-down* prediction crucially relies on the fact that retrieval is at issue, especially when features overlapping among items to be re-merged occurs. For the *top-down* model, also an encoding cost is considered, but the prediction is that the items already present (salient) in the discourse environment and, more generally, those features already merged in the structure, pay a minor cost at encoding (providing a precise characterization of the "more accessible" referents in memory-load accounts), with the exception of the re-introduction of a categorial feature (*N* in this case) and, possibly, its saliency specification (D), whenever they must be kept distinct in memory.

Pronouns as Determiners and Agreement

One way to dig further into the predictions emerging from these different assumptions is to keep all peculiar factors of OCs constant while investigating the specific contribution of single cued features using an overt subject-verb agreement language: focusing on person features, 1st and 2nd person, unlike 3rd person, are anchored to the speech event, being always present in a speech act (and in a left-peripheral structural dedicated position, Bianchi, 2003, 2006; Sigurdhsson, 2004). Because of their saliency (and dedicated structural position), we might expect 1st and 2nd person features to facilitate the integration of an argument better than default 3rd person (a non-person, in Sigurdhsson, 2004 terms). This could have an impact in terms of encoding: "highly accessible" deictic pronouns are lighter both for the *memory-load* (higher position in the accessibility hierarchy, section *memory-load* Accounts) and for the *top-down* models (being already present in phrase structure, they do not pay an extra FEC, section *top-down* (Left-Right) Minimalist Derivations).

On the other hand, at retrieval, different hypotheses can be formulated: considering person mismatch as a general facilitation (default hypothesis, H1), both the *similarity-based* (under any condition) and the *top-down* model (only when the relevant person mismatch is cued by the verbal agreement morphology) predict a facilitation[13]; an alternative hypothesis (H2), considering the salience of 1st and 2nd person features,

[13]Intervention-based approach would predict a specific facilitation for the 1st/2nd vs. 3rd person mismatching condition only under the further assumption that these features are relevant as movement triggers.

TABLE 1 | Summary of the predictions for the paradigm in 10 (data from Warren and Gibson, 2005).

Condition	D_1-D_2	D_1-N_2	D_1-P_2	N_1-D_2	N_1-N_2	N_1-P_2	P_1-D_2	P_1-N_2	P_1-P_2
Read. time	**365**	**319**	**306**	**348**	**347**	**291**	**348**	**311**	**291**
(SE) ms	**(19)**	**(12)**	**(14)**	**(18)**	**(21)**	**(14)**	**(18)**	**(15)**	**(13)**
Memory-load prediction	Hard	Medium	Easy	Hard	Medium	Easy	Hard	Medium	Easy
Similarity-based prediction	Hard	Medium	Easy	Medium	Hard	Easy	Easy	Easy	Medium
Intervention-based prediction	Hard	?	Easy	?	Hard	Easy	Easy	Easy	Easy
ACT-R-based prediction	Hard	Hard	Medium	Hard	Hard	Medium	Medium	Medium	Easy
Top-down prediction	Hardest	Medium	Medium–Easy	Hard	Hard	Easy	Hard	Medium	Easy

should predict a facilitation only for 1st and 2nd person and not for 3rd person. Under this second hypothesis, the only model making different predictions, based on the arguments involved, is the *top-down* model: only the subject and only when verb agreement is overt, 1st/2nd vs. 3rd person mismatch should produce a facilitation [this is because only 1st/2nd person feature mismatch would be considered as a *dF* facilitation for the FRC in (13)]. A neurophysiological evidence supporting the idea that (1st/2nd vs. 3rd) person features are peculiar in terms of subject-verb agreement (vs. number) is discussed in Mancini et al. (2011).

Notice that the introduction of a lexical restriction after the pronoun (e.g., "[$_D$ you [$_N$ bankers]]") would remove any advantage of the bare pronominal condition according to the Intervention-based model (section Intervention-Based Accounts) and, for different reasons (increased number of features to be retrieved and compared), for the *top-down Minimalist* model (section *top-down* (Left-Right) Minimalist Derivations), but not for the other models. According to Belletti and Rizzi (2013) pronouns are to a lesser extent interveners because of their lack of a lexical restriction. In fact, pronouns, given an appropriate context, can function as determiners (Postal, 1966, a.o.) and, unlike determiners, bear person features other than default 3rd person.

On the usage of pronouns as determiners[14], we refer to Elbourne (2005) analysis and we consider them as (empty) definite determiners taking an index and an NP predicate as arguments. According to Elbourne (2005), this is the structure shared by all DPs referring to individuals, namely proper names, pronouns and definite descriptions. In this sense, the "lexical restriction" would be an NP predicate which is, semantically speaking, denoted by type $<e,t>$, while the denotation of the pronominal determiner ("you") is expressed as follows[15]:

(17) $[\![you_i]\!]^{g,a} = \lambda f : f \in D_{<e,t>} \ \& \ a \leq_i g(j) \ \& f\left(g\left(j\right)\right) = 1.g(j)$

This means that "you," when used as a determiner in the construction "you bankers," takes the NP ["banker(s)"] with denotation *f* and returns, as the denotation of the full DP, some contextually salient plural individual *j,* such that the addressees *a* (deictic use of "you") must be part of *j* and *j* must be *f* (i.e., a banker).

The contextual salience of the relevant individuals is necessary for the sentence to be acceptable and it must be postulated in out of the blue sentences; this means that if a relevant context is not provided to the reader, s/he must infer by her/himself that a salient group of individuals is presupposed by the sentence even if s/he does not share this information with the speaker at the utterance time. On the one hand, we might expect this missing contextual information, related to unexpected saliency, to produce some slowdown in processing, forcing the reader to update his knowledge of the common ground in order to accept this specific utterance; on the other, this is a perfectly grammatical construction and it should be correctly interpreted even when an appropriate context is missing. As far as we can tell, the presence/absence of an appropriate context licensing this usage of second person pronouns as determiners has never been tested before.

Considering the cleft sentences under analysis, this could even happen twice:

(18) It was [**you/we** bankers]$_i$ that [**you/we** lawyers] avoided _ $_i$ at the party

Both the focalized DP and the cleft subject require that both a group of bankers and a group of lawyers be salient in the context and that the speaker and/or the addressee be part of one specific group. If the context is provided and the two groups of bankers and lawyers are in the common ground, the sentence should sound perfectly acceptable, if not, the reader should postulate the presence of the two groups after she/he realizes that none of them was accessible in her/his contextual knowledge. To our knowledge, this again has never been tested before.

Assuming a given cost for definite descriptions, a pronominally restricted DP ("you bankers") would pay either the same cost (default assumption), a minor cost as pronouns (coherently with their implicit referentiality in the *memory-load* model) or an extra cost (whenever they are non-salient in the context or they get re-introduced twice, as predicted in the *top-down* model).

Testing the Different Predictions

The paradigm expanded in (19) will be used to test the specific contribution of 2nd vs. 3rd person (default) in an overt subject verb agreement language, Italian, under the assumptions previously discussed:

(19) a. Sono [**gli** architetti]$_i$ che [**gli** ingegneri] **hanno** consultato _$_i$ prima di iniziare i lavori.
*are$_{3P_PL}$ **the** architects that **the** engineers **have$_{3P_PL}$** consulted before beginning the works*
b. Sono [**gli** architetti]$_i$ che [**voi** ingegneri] **avete** consultato _$_i$ prima di iniziare i lavori.
*are$_{3P_PL}$ **the** architects that **you** engineers **have$_{2P_PL}$** consulted before beginning the works*
c. Siete [**voi** architetti]$_i$ che [**gli** ingegneri] **hanno** consultato _$_i$ prima di iniziare i lavori.
*are$_{2P_PL}$ **you** architects that **the** engineers **have$_{3P_PL}$** consulted before beginning the works*
d. Siete [**voi** architetti]$_i$ che [**voi** ingegneri] **avete** consultato _$_i$ prima di iniziare i lavori.
*are$_{2P_PL}$ **you** architects that **you** engineers **have$_{2P_PL}$** consulted before beginning the works*

Hereafter, we will refer to (19).a as the Art_1-Art_2 matching condition, (19).b as Art_1-Pro_2 mismatch condition, (19).c as Pro_1-Art_2 mismatch condition and (19).d as Pro_1-Pro_2 matching condition. DP_1 is the focalized object, DP_2 the OC subject.

[14] We do not have space here to discuss the restriction that seems to force pronouns used as determiners to be either 1st or 2nd (and marginally 3rd) person plural in most languages. The interested reader should refer to Postal (1966), but also check the footnote in (Elbourne, 2005):60 reporting usage of first person singular in Early Modern English.

[15] *g* is a variable assignment, *a* is the addressee, $\leq_i$ the *individual part-of* relation Link (1983).

TABLE 2 | Theory by theory overall (off-line) predictions on the paradigm (19).

Condition	Art_1-Art_2	Pro_1-Pro_2	Art_1-Pro_2	Pro_1-Art_2
Similarity-based prediction	Hard	Hard	Medium	Medium
Intervention-based prediction	Hard	Hard	Medium	Medium
Top-down prediction (FRC)—H1	Hard	Hard	Medium	Medium
Top-down prediction (FRC)—H2	Hard	Hardest	Medium	Hard
Memory-load prediction—A1	Hard	Hard	Hard	Hard
Memory-load prediction—A2	Harder	Hard	Hard	Harder
Memory-load prediction—A3	Hard	Harder	Harder	Hard
ACT-R-based prediction	Hard	Hard	Hard	Hard

TABLE 3 | On-line predictions on the paradigm (19); at the verb segment (encoding+retrieval).

	Condition	DP_1	DP_2	Verb	PP
Art_1-Art_2	Memory-load prediction	2	2	4 (1+3)	4
	Top-down prediction	2	3	3.43 [2+log(27)]	5
Art_1-Pro_2	Memory-load prediction	2	3	5 (1+4)	4
	Top-down prediction	2	4	3.38 [2+log(24)]	5
Pro_1-Art_2	Memory-load prediction	3	2	4 (1+3)	4
	Top-down prediction	3	3	3.43 [2+log(27)]	5
Pro_1-Pro_2	Memory-load prediction	3	3	5 (1+4)	4
	Top-down prediction	3	5	3.68 [2+log(48)]	5

Overall, in off-line terms, similarity-based (section Similarity-Based Accounts), intervention-based (section Intervention-Based Accounts) and *top-down Minimalist* (based on FRC, section *top-down* (Left-Right) Minimalist Derivations) models would all predict that matching conditions (Art_1-Art_2 and Pro_1-Pro_2) should be more difficult than the mismatching conditions (Pro_1-Art_2 and Art_1-Pro_2). The *top-down* model, in particular, would predict a facilitation under the mismatch condition due to the distinct cued features at the verbal predicate [higher *dF* coefficient in the FRC, as expressed in (13)]. By default (hypothesis H1 in **Table 2**), as discussed in section Pronouns as Determiners and Agreement, any feature mismatch could help, hence the facilitation should be similar for both Pro_1-Art_2 and Art_1-Pro_2 conditions. However, according to the 1st and 2nd person anchoring hypothesis (H2 in section Pronouns as Determiners and Agreement), a global facilitation for the Art_1-Pro_2 condition compared with the Pro_1-Art_2 is expected by the *top-down* model: retrieval of Pro_2 due to the cued 2nd person should be favored (hypothesis H2 in **Table 2**). Similarly, a retrieval penalty due to 2nd person feature matching is expected under the Pro_1-Pro_2 condition. Notice that H2 does not change in any relevant sense the predictions of any other model except for the *top-down* one.

As for *memory-load* hypotheses (section *memory-load* Accounts), *Pro* condition has never been discussed previously, so the model here needs further assumptions: on the one hand, we might expect *Pro* and *Art* conditions to be referentially similar; under this assumption (A1 in **Table 2**) no difference would be predicted whatsoever in the paradigm. If *Pro* and *Art* are assumed to be referentially different, either *Art* turns out to be less accessible than *Pro* (A2 assumption in **Table 2**), then we should expect an extra encoding cost for *Art* and a related complexity signature for retrieving the focalized object in the Art_2 condition, or *Pro* is less accessible than *Art* (A3 assumption in **Table 2**), then a greater effort should be paid in the Pro_2 condition.

From a different perspective, ACT-R based approaches (section ACT-R-Based Predictions) would predict no difference across conditions since the distance between the focalized DP_2 and the OC predicate is always the same and no cue can help in retrieving/reactivating the focalized object.

In the end the retrieval/intervention cost predictions for all the models can be summarized in the table below [consider "hard" to be the baseline, based on the evidence discussed in section Processing Object Clefts (OCs)]:

In terms of on-line predictions, costs at the VERB regions should be proportional to the predictions expressed in **Table 2**. Moreover, *memory-load* and *top-down Minimalist* models also predict specific encoding costs that should have an effect in terms of on-line measures at the related DP_1, DP_2, predicate and PP final regions: according to Gibson (2000), the verb introduces an event referent (+1), while the integration cost must be calculated as crossing the verbal event (+1) and the cost of the intervening nominal referents (+1 in Gibson, 2000). According to the *top-down* approach, the predicate introduces two relevant features, a temporal index T and V predicate, in conformity with the DP encoding hypothesis: *Art* should have an encoding cost of 2, an index *D* and a *N* predicate (section Pronouns as Determiners and Agreement). We expect *Pro* to pay an extra encoding cost due to the out of the blue inclusion of a speech act participant into the N predicate (+1) as suggested in section Pronouns as Determiners and Agreement; as result, *Pro* condition would cost globally 3 units. Taking $Art = 2$ and $Pro = 3$ as encoding cost baseline, both for *memory-load* and *top-down* models, only the *top-down* model predicts an extra cost for any duplicated category (+1 for *N* in DP_2) and for a referential mismatch (+1 for II person matching at DP_2 in Pro_1-Pro_2 condition).

Under these assumptions, we can compare on-line predictions at the relevant segments[16] (**Table 3**).

[16]The integration cost is calculated as crossing the verbal event (+1) and 2 or 3 nominal referents (+2/+3 respectively for *Art* and *Pro*). The cost of +2 for

Crucially, the predicted cost at DP_2 is different since the *top-down* model, and not the *memory-load* one, predicts an extra encoding cost which is proportional to the number of matching features (and consequently to the necessity of updating an expectation, also in terms of speech act participants, otherwise salient, hence "free"). Moreover, at the verb segment, the predictions of the two models differ: the *memory-load* model predicts major efforts in the Pro_2condition, while the *top-down* model expects milder differences and essentially a penalty to be paid for the Pro_1-Pro_2 condition. Under H2 (i.e., only 1st/2nd person features are cued by the verb), also a mild facilitation for the Art_1-Pro_2 would be expected by the *top-down* model as compared to the Art_1-Art_2 and Pro_1-Art_2 conditions. Under H1 (i.e., all person features count as distinct cues), also Pro_1-Art_2 would benefit by the cued feature mismatch.

MATERIALS AND METHODS

Ethics Statement

The experiment was approved by the Ethics Committee of the "Dipartimento di Scienze del Sistema Nervoso e del Comportamento" of the University of Pavia. A written informed consent was obtained from the participants of this study.

Participants

Fifty-three participants (age range 20–52, $M = 34.57$, SD = 8.09, 28 female; all speakers of center-north Italian variety) voluntarily signed up for the acceptability study (experiment 1).

A different sample of 33 Italian native speakers of the same Italian variety (age range = 19–35; 18 female) took part in the eye-tracking study (experiment 2). After the end of each experimental session, in experiment 2, we assessed the participants' Verbal Working Memory Capacity (WM) using a test (in Lewandowsky et al., 2010) that is a variant of the Sentence Span test originally designed by Daneman and Carpenter (1980). Participants were asked to carry out a dual task: they were presented with series of 3 to 8 statements, each followed by a consonant. Participants had 4 sec to judge whether the statement was True or False then they were asked to remember the series of consonants that was presented after each sentence. At the end of the series a question mark appeared, signaling participants to type in all the consonants presented. A score ranging from 0 to 1 is obtained, indicating the individuals verbal working memory capacity. The scoring procedure takes into account the length of the series and the accuracy to the Ture/False judgment (see Lewandowsky et al., 2010, for more details).

Stimuli

We created 32 paradigms expressing the four possible conditions presented in (19), for a total of 128 items. DPs were introduced by articles and second person pronouns while keeping number (plural, in order to make all the oppositions sound) and gender (masculine) constant. The nouns within each sentence were balanced for (i) number of letters ($DP_1 = 8.86$, SD = 1.46; $DP_2 = 8.96$, SD = 1.95; $t < 1$), (ii) logarithmic frequency (based on Repubblica corpus, Baroni et al., 2004) of nominal items ($DP_1 = 7{,}66$, SD = 1.68; $DP_2 = 7.58$, SD = 1.95; $t < 1$), and (iii) concreteness and imageability (all were concrete nouns referring to professions). The plural masculine article in Italian is sensitive to the beginning of the following noun: when a vowel is present "gli" is used ("gli architetti," the architects vs. "i giorni," the days). We used "gli" to maximize the length of the determiner and match the length of the pronoun ("voi," you) whenever possible, but in most cases (22/32), due to semantic congruity and to keep DP frequency comparable within sentences we used "i" determiner for nouns beginning with a consonant. Experimental lists included 32 critical items (eight per condition) and 112 fillers. As in previous experiments, we did not include any relevant context. Fillers included 64 declarative sentences, half of them perfectly grammatical and the other half presenting a subject-verb number agreement error; other 48 fillers were questions with various degrees of acceptability, ranging from perfect grammatical *wh*- long distance question ("what do you think that John saw?") to violation of locality ("what do you wonder who see?"). Both experiments used these materials.

A Latin square design was used to counterbalance conditions across experimental lists, in a way that in each experiment participants were exposed to one only version of each paradigm, and each item within paradigms was presented to an almost equal number of subjects across lists.

Procedure

For experiment 1 (acceptability judgment), a web-based questionnaire was created using Osucre open source software (Van Acker, 2007). The experimental lists were presented one at a time on a single line and participants were asked to judge each for acceptability on a 7-points Likert-scale.

For experiment 2 (eye-tracking), participants were individually tested in a dimly lit room. They sat in front of a 17 inches computer screen and kept their head on a chin rest so that the distance between the display and their eyes was 56 cm. They were instructed to read the sentences carefully, as they would have to answer one question following each sentence. Each trial (presented in pseudo-randomized order) consisted of a fixation cross appearing at the center of the screen for 1,500 ms, and was followed by the sentence, displayed on one single line. Participants could press the space bar to signal the reading completion, or the sentence display timed out after 20 sec. After the sentence display a second fixation cross appeared for 1,500 ms, just before the presentation of the question, that remained on screen until the Yes/No response. Half of the questions concerned the subject (e.g., "did the architects consult someone?") the other half the object of the cleft (e.g., "did someone consult the engineers?"); half of the questions included a relevant PP (e.g., "did the architects consult someone after the meeting?"); 50% of the questions required a positive answer, 50% a negative one.

Art is considered for uniformity with respect the *top-down* model. These are the assumptions leading to the best possible predictions for the *memory-load* model. As for the *top-down* model, the FRC predicted is: Art_1-$Art_2 = 9{\cdot}3 = 27$; Art_1-$Pro_2 = 8{\cdot}3 = 24$; Pro_1-$Art_2 = 9{\cdot}3 = 27$; Pro_1-$Pro_2 = 16{\cdot}3 = 48$; See **Supplementary Materials,** section ACT-R-Based Model and section Top-Down Minimalist model for details.

Analysis

Data Acquisition and Pre-processing of Eye-Tracking Data

Eye-movements in experiment 2 were recorded with an Eyelink® 1000 system (SR-Research, Ottawa, CA), tracking the dominant eye and using the desktop mount configuration. Eye gaze was sampled at 1,000 Hz frequency. Consecutive fixations between 50 and 80 ms occurring at one character distance were grouped into one single fixation (1.35% of data). Fixations that were (a) shorter than 80 ms, (b) longer than 1,200 ms, (c) occurring within 20 ms from blink onset/offset, and (d) occurring outside sentence boundaries were excluded from the analyses (overall rejection rate 3.16%). Four participant were excluded on the basis of their performance on the comprehension questions (<60%), leaving 29 participants in the analyzed dataset.

Four canonical reading time measures (Rayner, 1998) were computed. *First Fixation* (FF) and *Gaze Duration* (GD) were defined as the time (ms) spent on each region when participants entered it (from the left side) for the first time: FF was the duration of the very first fixation only, whereas GD was defined as the time spent from the first time entering the region to the first time leaving it, to the right or to the left. Words in the sentences could be fixated after a regression (i.e., entering the region from the right) and the time spent in a region entering it from the right was defined as *Second Pass* (SP) reading time. Total duration Time (TT) was the total time spent on a given region.

We also analyzed regression data: a regression event occurred when participants looked back in the sentence. For each regression event, we determined the region from where the eye left (*R-from*), and the region where the gaze landed (*R-in*).

Sentences consisted of six different regions: BE, DP_1, C, DP_2, VERB, SPILLOVER. Regression analyses were carried out for each region, when the number of observations was sufficient for carrying out statistics. In particular, only Total Duration and Second Pass could be computed for BE and C regions.

Statistics

Results were analyzed using linear mixed-effects regression models (Baayen et al., 2008), using lme4 package (version 1.1.21) in R environment (R Core Team, 2018). Mixed models are widely used in eye-movements research (e.g., Staub, 2010; Kuperman and Van Dyke, 2011) as they conveniently handle imbalanced designs and missing values, typically occurring with eye-tracking data.

Linear mixed models were used for acceptability judgments of Experiment 1 and Reading Times in Experiment 2. The dependent variables for the linear mixed models were the single trials acceptability scores (ranging from 1 to 7) for Experiment 1, and the log transformed Reading Times for Experiment 2. Log transformation was needed because the distribution of times and the models' residuals were far from the normal distribution. For the analysis of regressions in Experiment 2 the dependent variable was binary (0,1) depending on whether a regression event was or was not recorded, and data were analyzed by fitting generalized mixed models using the logit response function (e.g., Jaeger, 2008)—as it was done for the Accuracy data on the comprehension question of Experiment 2.

In Experiment 1 we used the "maximal" random structure allowing for by-subject and by-item intercept adjustments and by-subject slopes adjustments for DP_1 by DP_2 interaction. However, the maximal random structure of the models for Experiment 2 (where the number of participants was lower) often caused convergence issues. We therefore adopted a random structure chosen on grounds of feasibility (Matuschek et al., 2017). For reading measures (FF, GD, TT, and SP) the random structure was allowed by-subject and by-item intercept adjustments, and by-subject slope adjustments of DP_1 and DP_2. For regression measures the random structure was initially "minimal," but, when possible, the model used to report the final estimates and the contrasts between conditions had the same structure used for the reading measures.

To evaluate the presence of significant main effects and interactions we used likelihood ratio tests (LRT) comparing the fit of (nested) models of increasing complexity (e.g., [factor A + B] compared to [factor A + B + AxB]). Tables in the Results section describe the outcome of the LRTs, reporting the value of chi square and the level of significance. In the analysis of Experiment 2 the models were further specified: the Null model consisted of the effect of Trial Order alone, and Trial Order was kept in all subsequent models. The contribution to the model fit was assessed for the factors DP_1 (Art, Pro), DP_2 (Art, Pro), and WM (continuous) and their interactions. To make the main effects interpretable as in standard ANOVAs we adopted contrast coding for categorical factors, while continuous predictors were z-centered around the mean (e.g., Levy et al., 2013). The effects of the factors of interest and their interactions are further described in the text and in the figures providing the size of the effect in the response measure (i.e., back transforming log(ms), in ms, and log odds into probability) using functions in the emmeans package (Lenth et al., 2019).

RESULTS

Off-Line Results (Acceptability Judgment and Accuracy in Comprehension Questions)

We compare here (**Figure 1**) the results of the offline data gathered from both experiments: the acceptability rate of experiment 1 and the accuracy in answering comprehension questions after eyetracking (experiment 2).

Summarizing the acceptability data collected in Experiment 1, we obtained the following pattern[17]:

Art_1-Pro_2 ($M = 5.31$, SE $= 0.19$) $\geq$ Art_1-Art_2 ($M = 5.22$, SE $= 0.19$) $>$ Pro_1-Art_2 ($M = 4.54$, SE $= 0.22$) $>$ Pro_1-Pro_2 ($M = 2.50$, SE $= 0.18$).

The analysis of Experiment 1 LRTs showed that judgments were influenced by the interaction between DP_1 and DP_2

[17]"$\geq$" indicates a numerical, though non-significant difference; ">" indicates a significant difference.

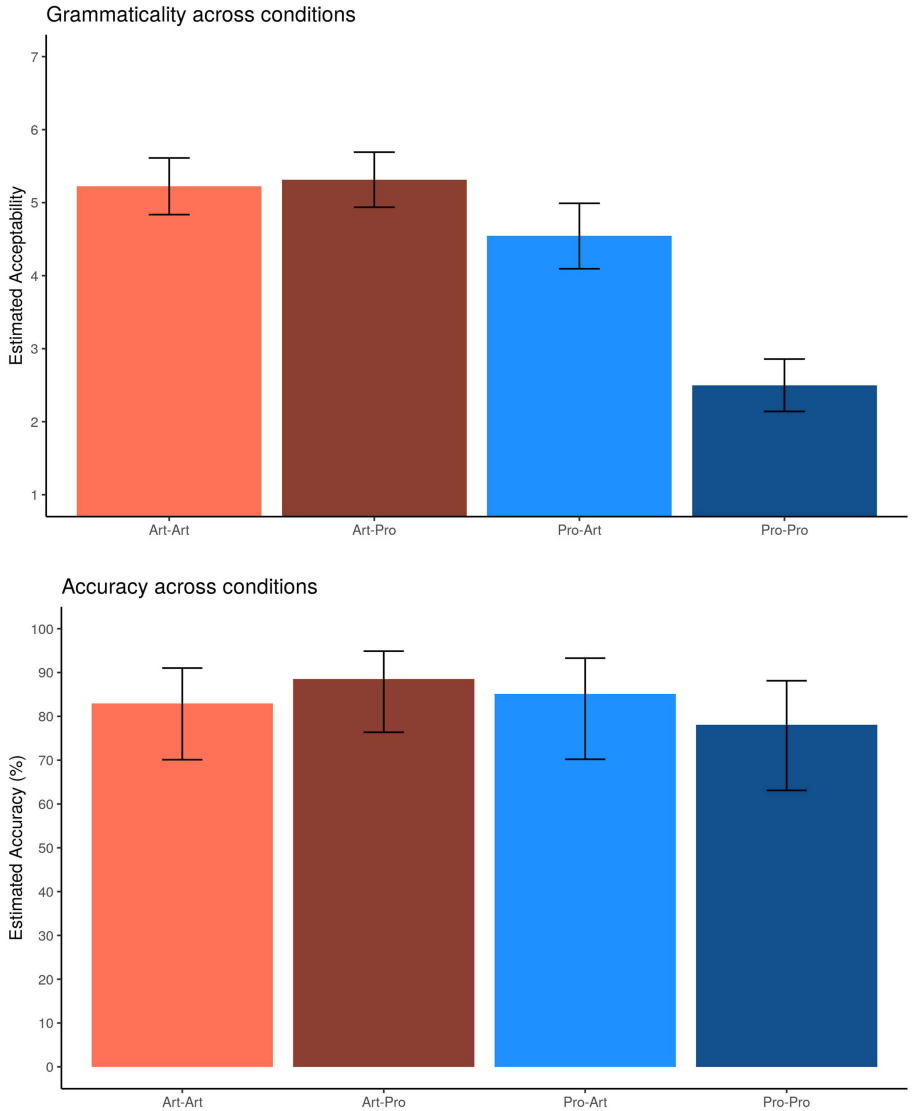


FIGURE 1 | Offline results: estimated acceptability (7-point Likert scale) judgments (experiment 1) and accuracy (%) in answering comprehension questions (experiment 2). Error bars represent 95% confidence intervals around the models' estimates.

TABLE 4 | Acceptability and Accuracy effects depending on DP_1-DP_2 types.

	χ^2	*p*
Acceptability		
DP_1 type	**22.09**	**<0.001**
+ DP_2 type	**5.49**	**<0.01**
+ DP_1:DP_2	**51.85**	**<0.001**
Accuracy		
DP_1 type	0.69	
+ DP_2 type	0.00	
+ DP_1:DP_2	3.76	<0.1

Bold values are the significant ones.

(**Table 4**) showing that when DP_1 was *Art* the subject type had no effect on acceptability (Art_1-Art_2 = 5.22; Art_1-Pro_2 = 5.31; $t = -1.04$), whereas when DP_1 was introduced by a pronoun, the effect of subject type was sensible being the matching condition much less acceptable (Pro_1-Art_2 = 4.54; Pro_1-Pro_2 = 2.50; $t = -9.28$, $p < 0.001$).

A similar numerical pattern indicating that Art_1-Pro_2 is better than Pro_1-Pro_2 is revealed in comprehension (while Pro_1-$Art_2 \geq Art_1$-Art_2 is non-significant):

Art_1-Pro_2 ($M = 88.6\%$) $\geq Pro_1$-Art_2 ($M = 85.1\%$) $\geq Art_1$-Art_2 (83%) $\geq Pro_1$-Pro_2 (78.1%).

The marginally significant interaction pointed to the effect of DP_1 when DP_2 is *Pro*, which describes the marginally significant

difference between Art_1-Pro_2 and Pro_1-Pro_2 (+10.5%, $z = 1.70$, $p < 0.1$).

On-Line Results (Eye Tracking Data, Experiment 2)

Reading Times

First fixation (FF)

Due to the small number of observations for FFs in BE ($n = 115$) and in C ($n = 307$), models did not converge and the results from these regions were omitted (**Figure 2**). Analyses (**Table 5**) revealed a main effect of DP_2 type, indicating longer FFs for Pro_2, while reading DP_2 (+10 ms, $t = 2.06$, $p < 0.05$) and VERB (+10 ms, $t = 2.39$, $p < 0.05$) regions. Upon reading SPILL region, FFs were slightly longer with Pro_1 (+9 ms, $t = 1.98$, $p < 0.1$).

Considering the effect of WM, it was significant across the last three regions: FFs were generally shorter for Higher WM participants (β in DP_2: −0.038, $t = -1.95$, $p < 0.1$; β in VERB: −0.070, $t = -3.44$, $p < 0.01$; β in SPILL: −0.056, $t = -2.51$, $p < 0.05$).

Gaze duration (GD)

Due to the small number of observations for GD in BE and in C, models did not converge and the results from these regions were omitted (**Figure 3**). In the analysis of GD (**Table 6**), a main effect of DP_1 type in DP_1, and a main effect of DP_2 type in DP_2, are significant. In DP_1 region, Pro_1 triggered longer GD compared to Art_1 [+64 ms, $t = 6.20$, $p < 0.001$]. In DP_2 region longer GD occurred with DP_2 [DP_2 region: +108 ms, $t = 7.28$, $p < 0.001$]. A marginal indication of a significant main effect of DP_2 emerged also in the VERB region, showing a slow down to Pro_2 [+26 ms, $t = 1.86$, $p < 0.1$].

The effect of WM, indicating shorter GD for participants with higher WM, was robust across regions: in DP_1 (−0.115, $t = -3.45$, $p < 0.001$), DP_2 (β = −0.097, $t = -2.77$, $p < 0.01$), VERB (β = −0.179, $t = -4.37$, $p < 0.001$), and SPILL (β = −0.181, $t = -3.66$, $p < 0.001$).

Total time duration (TT)

As for TT (**Figure 4**), effects of the experimental factors were found on all Regions except for SPILL (**Table 7**).

In BE the significant interaction between DP types was due to the slow down (+43 ms, $t = 1.84$, $p < 0.1$) to Pro_1 compared to Art_1 that resulted marginally significant on the pairwise contrasts ($t = 1.84$, $p < 0.1$) in the Pro_2 condition and not in Art_2 (+2 ms, $t < 1$). In DP_1, reading Pro_1 took longer than Art_1 (+168 ms, $t = +6.71$, $p < 0.001$), and TT in this region is further modulated by the interaction between DP_2 type and WM: the facilitatory effect of WM was stronger ($\Delta\beta = -0.085$, $t = -1.90$, $p < 0.1$) when DP_2 was *Art* (β = −0.07) rather than *Pro* (β = 0.01). In DP_2 region, the effect of DP_2 (longer Pro_2 compared to Art_2: +183 ms, $t = +5.41$, $p < 0.001$) was further modulated by DP_1 type: when DP_2 was *Art*, no effect of DP_1 emerged (−20 ms, $t < 1$), while when DP_2 was *Pro*, longer total times were observed for Pro_1–Pro_2 compared to Art_1–Pro_2 (+137 ms, $t = +3.31$, $p < 0.01$). In the VERB region, as in the previous region, the interaction between DP types revealed that the effect of DP_1 was present in the Pro_2 condition–longer TT for Pro_1–Pro_2 compared to Art_1–Pro_2 (+83 ms, $t = +2.35$, $p < 0.05$)–and not in the Art_2 condition (−17 ms, $t < 1$).

Total times spent in BE and C were globally influenced by WM [in BE: β = −0.095, $t = -1.99$, $p < 0.1$; in C: β = −0.064, $t = -1.73$, $p < 0.1$].

Second pass duration (SP)

The presence of a pronominal restriction in DP_1 and DP_2 was sufficient to cause longer SP reading times in both DP_1 and DP_2 regions (**Figure 5**, **Table 8**): SP in DP_1 were longer for Pro_1 (+73 ms, $t = +2.89$, $p < 0.01$), and, similarly, in DP_2 region SP were longer for Pro_2 (+111 ms, $t = +3.21$, $p < 0.01$). In DP_2, however, participants show longer SP in the Pro_1 condition compared to Art_1 condition (+45 ms, $t = +1.75$, $p < 0.1$).

In the VERB region, a three-ways interaction emerged, showing that *Pro* generally causes a slowdown in SP (Pro_2 condition: +52 ms, $t = 1.78$, $p < 0.1$; Pro_1condition: +57 ms, $t = 2.01$, $p < 0.05$), and that the effect of WM (faster reading times for higher WM) was much stronger in Pro_1-Art_2 (β = −0.214), compared to the other three conditions [Pro_1-Pro_2 β = 0.025, $\Delta\beta = -0.239$, $t = -2.75$, $p < 0.05$; Art_1-Art_2 β = 0.036, $\Delta\beta = -0.250$, $t = -3.04$, $p < 0.05$; Art_1-Pro_2 β = 0.018, $\Delta\beta = -0.232$, $t = -2.45$, $p < 0.1$].

In the end, SP in BE showed a significant DP_1 by DP_2 interaction. The time spent re-reading BE was not affected by DP_1 type when DP_2 was *Art* (18 ms, $t < 1$), while for Pro_2 the difference between Pro_1 and Art_1 was consistent (+49 ms, $t = 1.99$, $p = 0.05$).

Regressions

We first assessed the likelihood of performing a regression From each region, independently of the experimental factors, but distinguishing between regressions in first pass and overall regressions. The probability of performing a regression was largest from the rightmost region SPILL (80%), followed by DP_2 (46.8%, SPILL vs. DP_2: $z = +13.48^{***}$), VERB (29.9%, DP_2 vs. VERB: $z = +7.00^{***}$), DP_1 (23.3%, VERB vs. DP_1: $z = +3.15^{*}$), and C (19.8%, VERB vs. C: $z = +4.26^{***}$).

Notably, the probability of performing a regression during first pass was considerably reduced for the VERB region (9.42%).

VERB first pass regressions were less likely than those occurring in DP_1 (13.75%, $z = -2.79^{*}$) or DP_2 (19.30%, $z = -6.02^{***}$), suggesting that processing difficulties at the VERB may be different from those at DP regions. In particular, the integration efforts at DP_1 or DP_2 trigger immediate regressions, while encoding/retrieval costs do not immediately lead to a regressive saccade, but rather require more reading time.

All in all, the overall probability in regressions pattern suggests that the DP_2 region is particularly hard to process: upon reading this region, being it the first time or not, participants need to go back in the sentence to retrieve additional information.

Regressions from regions (R-from)

We first evaluated the probability of making a R-from each ROI on the total number of fixations on each region (**Figure 6**, **Table 9**). A main effect of DP_1 type in DP_1 region, showed that

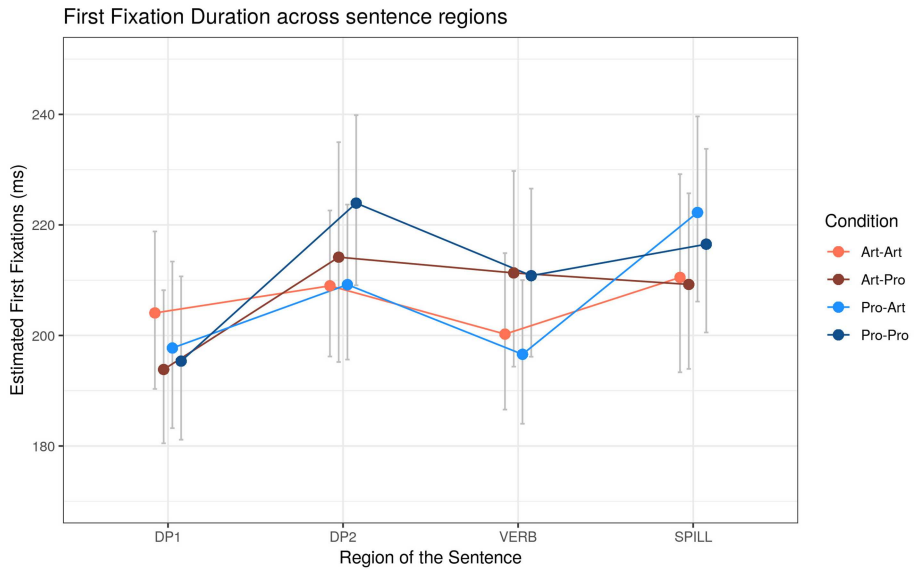


FIGURE 2 | Estimated First Fixation (FF) reading times (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 5 | First Fixation (FF) effects depending on the DP_1-DP_2 types and Working Memory (WM).

First fixation times	Regions											
	BE		DP_1		C		DP_2		Verb		Spill	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
Null (trial order)												
+ DP_1 type							1.21				**3.92**	**<0.05**
+ DP_2 type			2.14				**4.98**	**<0.05**	**6.55**	**<0.05**		
+ DP_1:DP_2												
+ WM			1.29				**4.37**	**<0.05**	**8.58**	**<0.01**	**6.41**	**<0.05**
+ DP_1:WM												
+ DP_2:WM									2.08			
+ DP_1:DP_2:WM									1.32			

$\chi^2 < 1$ are omitted.
Bold values are the significant ones.

regressions from DP_1 were more likely for DP_1 *Pro* (*Pro*$_1$ vs. *Art*$_1$: +0.11, $z = 2.84$, $p < 0.01$). In the pattern of Regressions from VERB and DP_2 regions significant interactions between DP_1 and DP_2 emerged. When DP_1 was *Pro*, more regressions out of VERB were associated to *Pro*$_2$ (*Pro*$_2$ vs. *Art*$_2$: +0.11, $z = 2.06$, $p < 0.05$), while when DP_1 was *Art* the pattern was numerically opposite (*Pro*$_2$vs. *Art*$_2$: −0.06, $z = -1.30$, ns). Considering a *post hoc* mismatching vs. matching pairwise comparison, in addition to DP type comparison, we observe that mismatching conditions are associated to a much smaller regression probability from this region (*mismatching* vs. *matching*: −0.10, $z = -2.754$, $p = 0.005$). As for Regressions from DP_2 the interaction was explained by a similar pattern: here, though, when DP_1 was *Pro*, the higher proportion of regressions out for *Pro*$_2$was not significant (*Pro*$_2$vs. *Art*$_2$: +0.07, $z = 1.45$, ns), while when DP_1 was *Art* a higher likelihood of making a regression was observed for DP_2 *Art* (*Art*$_2$ vs. *Pro*$_2$: +0.13, $z = 2.25$, $p < 0.05$).

The effect of WM, occurring in VERB ($\beta = +0.403$, $t = 2.32$, $p < 0.05$) and SPILL ($\beta = +0.723$, $t = 2.58$, $p < 0.01$) shows that participants with higher WM were more likely to make a regression out of these regions, whereas the marginally significant effect of WM in DP_1 had the opposite direction ($\beta = -0.375$, -1.82, $p < 0.1$), suggesting that low WM participants performed a regression much earlier on. In DP_2, the interaction between DP_2 type and WM was due to the more positive ($\Delta\beta = +0.31$, $z = +1.97$, $p < 0.05$) slope of WM for *Pro*$_2$ ($\beta = +.35$) compared to *Art*$_2$ ($\beta = +0.04$).

Then we evaluated Regressions in first pass, considering the proportion of regressions from each region on the number of fixations made in each region during first pass only. Regressions

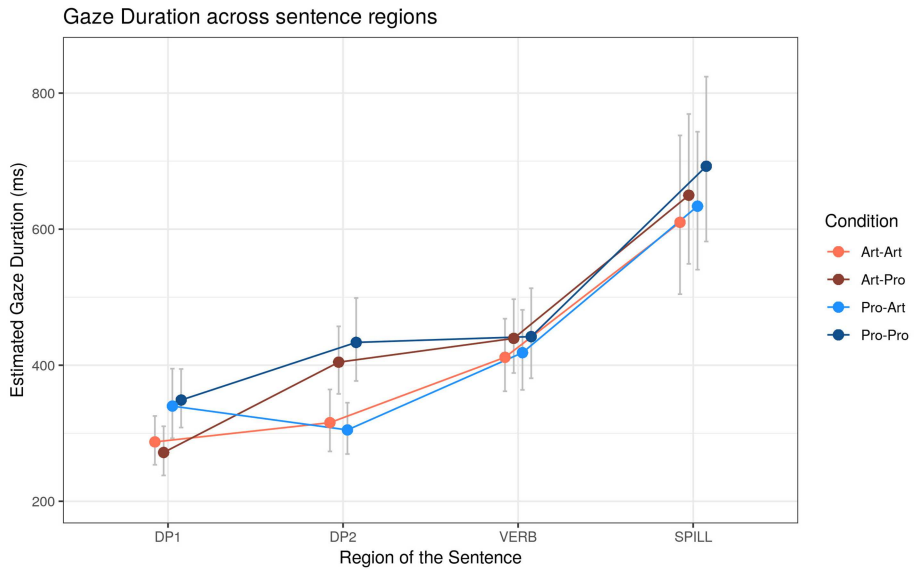


FIGURE 3 | Estimated Gaze Duration (GD) reading times (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 6 | Gaze Duration (GD) effects depending on DP_1-DP_2 type and Working Memory (WM).

Gaze duration times	Regions											
	BE		DP_1		C		DP_2		Verb		Spill	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
Null (trial order)												
+ DP_1 type			**30.1**	**<0.001**							1.6	
+ DP_2 type							**32.09**	**<0.001**	3.42	<0.1	2.93	<0.1
+ DP_1:DP_2			1.11				2.42					
+ WM			**11.80**	**<0.001**			**8.12**	**<0.01**	17.1	**<0.001**	12.59	**<0.001**
+ DP_1:WM												
+ DP_2:WM							1.93		1.01			
+ DP_1:DP_2:WM							1.13					

Bold values are the significant ones.

out of C were more likely for Art_1 (+0.08, $z = 1.93$, $p = 0.05$). First pass Regressions from DP_2 were more likely for Art_2 (+0.12, $z = 3.04$, $p < 0.05$). At the VERB, the DP_1 by DP_2 interaction was due to a larger number of regressions for Art_1 compared to Pro_1 (+0.07, $z = 2.08$, $p < 0.05$) when DP_2 was *Art*, which was absent when DP_2 was *Pro* (−0.02, $z < 1$, ns).

Regressions into regions (R-in)

We first evaluated the likelihood of making a R-in for each ROI on the total number of Regressions events (**Figure 6**, **Table 10**). Regressions in C and DP_2 were affected by the main effect of DP_2 type: in C regressions were more likely for Art_2 (+0.05, $z = 2.84$, $p < 0.01$), while in DP_2 regressions were more likely for Pro_2 (+0.06, $z = 3.00$, $p < 0.01$). The effect of WM had a negative slope in BE ($\beta = -0.444$, $z = -4.01$, $p < 0.001$) and in C ($\beta = -0.257$, $z = -2.32$, $p < 0.05$), suggesting that regressions in these regions were more likely for low WM participants, while the effect of WM had more positive slopes in DP_1 ($\beta = 0.135$, $z = 2.09$, $p < 0.05$), DP_2 ($\beta = 0.212$, $z = 2.55$, $p < 0.05$) and VERB ($\beta = +0.222$, $z = 2.46$, $p < 0.05$), showing that higher WM participants directed their regressions toward these sentence regions.

Summary

Acceptability judgments showed that matching conditions are significantly different (Art_1-Art_2 better than Pro_1-Pro_2). Art_1-Art_2 matching condition results slightly less grammatical than Art_1-Pro_2 mismatching condition. The other Pro_1-Art_2 mismatch condition ranking below them and above Pro_1-Pro_2 matching condition, which is unquestionably considered rather ungrammatical.

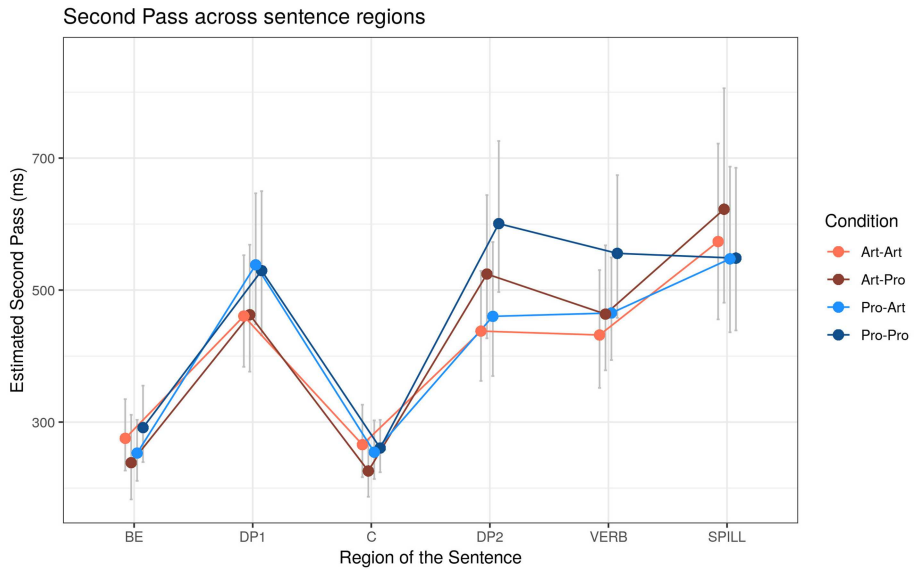


FIGURE 4 | Estimated Total reading Times (TT) (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 7 | Total Time (TT) effects depending on the DP_1-DP_2 type and Working Memory (WM).

	Regions											
Total reading times	**BE**		**DP_1**		**C**		**DP_2**		**Verb**		**Spill**	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
Null (trial order)												
+ DP_1 type	1.52		**28.22**	**<0.001**	2.84	<0.1	1.04		2.34			
+ DP_2 type							**19.39**	**<0.001**	2.77	<0.1		
+ DP_1:DP_2	**4.37**	**<0.05**			1.58		**8.95**	**<0.01**	**6.33**	**<0.05**	1.32	
+ WM	3.53	<0.1			**3.86**	**<0.05**			1.21			
+ DP_1:WM												
+ DP_2:WM			3.60	<0.1			1.90					
+ DP_1:DP_2:WM												

Bold values are the significant ones.

Comprehension questions in the eye-tracking experiment, revealed only an interaction between DP types, showing a difference in accuracy between Art_1-Pro_2 and Pro_1-Pro_2, but overall, the experimental sample correctly (>85%) understood the sentences.

Concerning reading times, and looking at the global results we can summarize:

i. Art_1-Art_2 matching condition constitute the processing baseline in all measures;
ii. Art_1-Pro_2 mismatch condition caused some slow-down, mainly at DP_2, where *Pro* was present (GD, TT, and SP), and marginally at VERB (FF only);
iii. Pro_1-Art_2 mismatch condition caused some slow-down, but only on DP_1 region (TT and SP);
iv. Pro_1-Pro_2 matching condition is numerically the most time consuming condition in all measures, and the numeric differences emerge as statistically consistent with the DP_1 DP_2 interaction in TT for DP_2 and VERB regions.

About factors interaction:

a. the interactions between DP_1 and DP_2 was found in DP_2 and VERB regions for TT; in BE and C for SP; in the regressions from DP_2 and VERB.
b. the effect of DP_2 type is overwhelming: Pro_2 is problematic as revealed in FF, GD, TT, and SP both in DP_2 and VERB regions;
c. mismatching conditions (Art_1-$Pro_2$and Pro_1-Art_2), overall, are associated to a reduced probability to trigger regressions from DP2 and VERB regions.
d. the effect of WM is very strong, and, interestingly, the slope of WM is not always negative (higher WM associated with faster reading or fewer regressions). It has a negative slope in reading times measures FF, GD, TT, and SP. However,

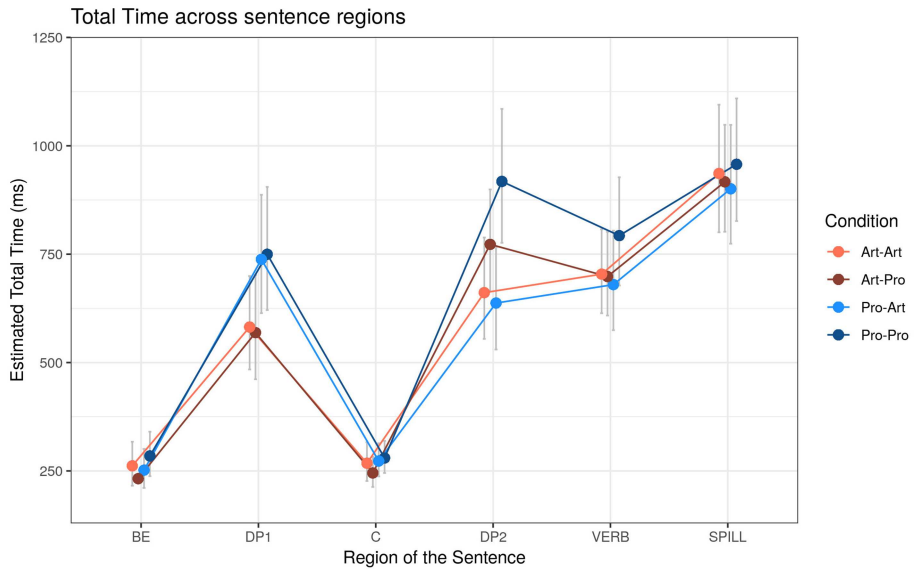


FIGURE 5 | Estimated Second Pass (SP) reading times (ms) across sentence regions. Error bars represent 95% confidence intervals around the models' estimates.

TABLE 8 | Second Pass (SP) reading times effects depending on the DP_1-DP_2 type and Working Memory (WM).

Second pass reading times	Regions											
	BE		DP_1		C		DP_2		Verb		Spill	
	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*	χ^2	*p*
Null (trial order)												
+ DP_1 type			**6.68**	**<0.01**			**3.88**	**<0.05**	**6.33**	**<0.05**	1.64	
+ DP_2 type							**9.70**	**<0.01**	**3.89**	**<0.05**		
+ DP_1:DP_2	**5.06**	**<0.05**			2.73	<0.1						
+ WM	1.89				2.21				2.66			
+ DP_1:WM			1.51				2.13		2.37			
+ DP_2:WM			2.08				1.27		2.47			
+ DP_1:DP_2:WM							2.13		**3.82**	**=0.05**		

$\chi^2 < 1$ are omitted.
Bold values are the significant ones.

it shows different effects on regression probability. Higher WM is associated to a larger proportion of regressions out of VERB and SPILL regions, but to fewer regressions out (during first pass) of earlier sentence regions like DP1. No major interaction between WM * DP_1 * DP_2 type is revealed.

In **Table 11** a summary of the main results.

DISCUSSION

Starting with off-line considerations, the results of both experiments consistently show that a lexically restricted second person pronoun ("you linguists") is at least as hard as a restricted definite article ("the linguist") across all conditions, hence any advantage of the bare pronominal DPs ("you"), revealed in previous experiments (Warren and Gibson, 2005, a.o.), is lost when a lexical restriction is present. This main result is consistent both with the intervention-based (Friedmann et al., 2009; Belletti and Rizzi, 2013) and with the similarity based prediction (Gordon et al., 2001). These models, however, fail to capture contrasts both in matching conditions (with Art_1-Art_2 condition "easier than" Pro_1-Pro_2 condition) and in mismatching ones (with Pro_1-Art_2 condition less acceptable than Art_1-Pro_2 condition). Both contrasts are predicted under the *memory-load* and *top-down* perspectives: *memory-load* approach (Gibson, 1998) can predict an extra cost in processing *Pro* conditions by relying on the absence of an appropriate context (assumption A3 in section Testing the Different Predictions), but it fails to predict the striking asymmetry found between Art_1-Pro_2 and Pro_1-Pro_2. This contrast is only correctly predicted

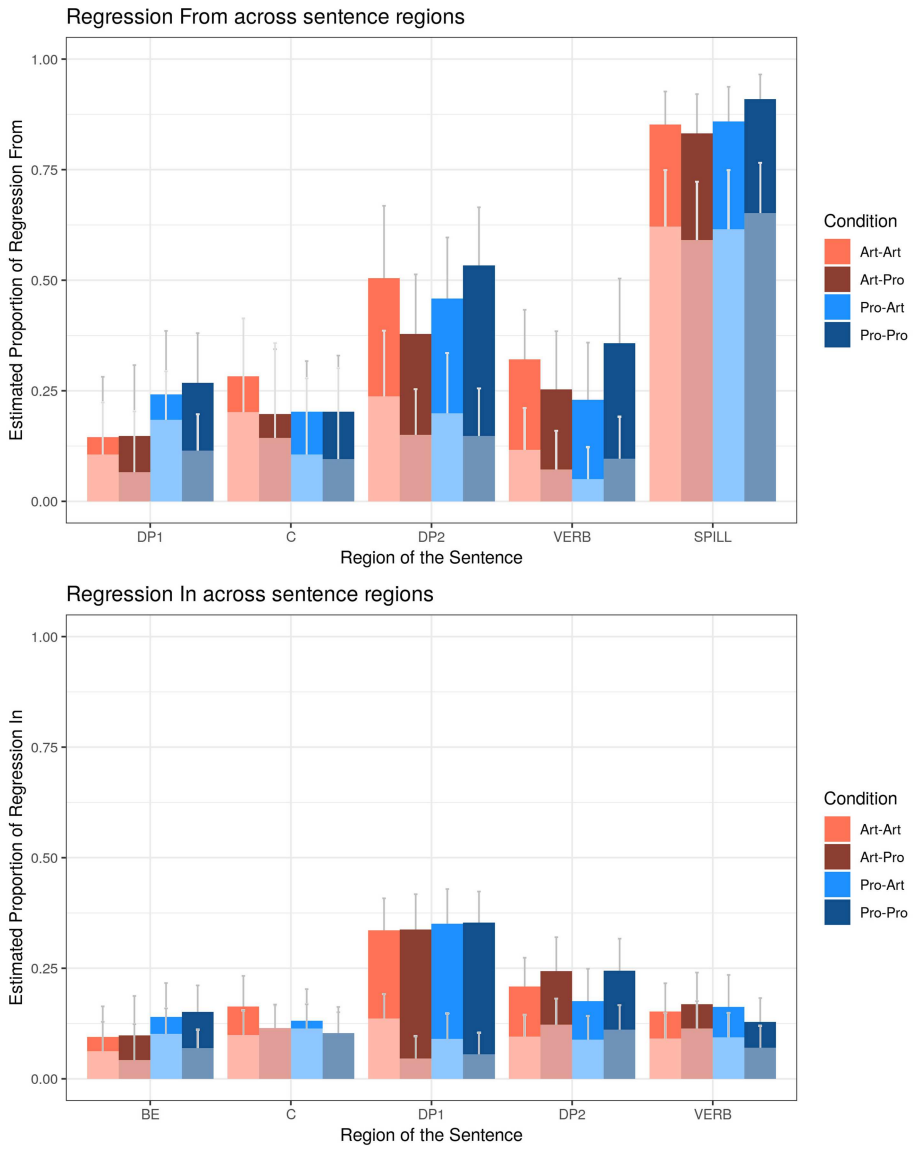


FIGURE 6 | Estimated first (light shades) and total (full colors) regression probabilities (%) In and From the regions of interest. Error bars represent 95% confidence intervals around the models' estimates.

by the *top-down* model: the acceptability pattern revealed a significant difference in these conditions (Art_1-Pro_2 better than both Pro_1-Art_2 and, even more robustly, Pro_1-Pro_2) is predicted by this model under the H2 hypothesis: 2nd person on the subject, in a mismatching condition, induces a facilitation better than 3rd person (section Testing the Different Predictions). Similarly, under the same hypothesis, only the *top-down* model predicts a major effort in processing the Pro_1-Pro_2 with respect to the Art_1-Art_2 condition due to 2nd pronominal matching feature. This model also predicts a milder advantage of the Art_1-Pro_2 condition with respect to the Art_1-Art_2 condition which is numerically present in accuracy but not significant. Generally the pattern across conditions is similar both in acceptability and in accuracy in comprehension questions after eye-tracking. However, the first pattern, but not the second (with the exception of Pro_1-Pro_2 vs. Art_1-Pro_2, again coherently with H2), results in statistically significant contrasts. Notice that the Pro_1-Pro_2 condition is considered nearly ungrammatical by the subjects. This clearly differentiate this condition from the others. Also in comprehension questions this condition leads to the worst performance, but such performance is still surprisingly high (78.1%). This indicates that the subjects correctly answer to questions posed on sentences that they consider unacceptable, suggesting a milder discriminative power of the accuracy measure with respect to acceptability judgments.

Due to a constant set of factors (i.e., distance between the focalized object and the predicate, same context and same

TABLE 9 | Regression from (R-from) depending on the DP_1-DP_2 types and Working Memory (WM).

Regression from	DP1 ($N = 198$)		DP2 ($N = 396$)		Verb ($N = 300$)		Spill ($N = 708$)	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Null (trial order)								
+ DP_1 type	**6.16**	**<0.05**	2.19		<1		1.64	
+ DP_2 type	*2.63*		*8.67*	*<0.01*	1.29			
+ DP_1:DP_2	*1.93*		**7.68**	**<0.05**	**7.80**	**<0.01**		
					4.72	*<0.05*		
+ WM	3.31	<0.1			**4.87**	**< 0.05**	*1.31*	
	6.72	*<0.01*						
+ DP_1:WM								
+ DP_2:WM	3.48	<0.1	**4.10**	**<0.05**				
+ DP_1:DP_2:WM	1.81		2.38				*2.89*	*<0.1*

First Regression from are indicated ($_{fR-from}$) under the total R-From estimate. $\chi^2 < 1$ are omitted.
Bold values are the significant ones.
Italics indicate the fR–from measures.

TABLE 10 | Regression In (R_{in}) depending on the DP_1-DP_2 types and Working Memory (WM).

Regression in	BE ($N = 286$)		DP1 ($N = 534$)		C ($N = 243$)		DP2 ($N = 336$)		C ($N = 228$)	
	χ^2	p	χ^2	p	χ^2	p	χ^2	p	χ^2	p
Null (trial order)										
+ DP_1 type	3.00	<0.1								
+ DP_2 type					**7.01**	**<0.01**	**8.35**	**<0.01**		
+ DP_1:DP_2					2.43				1.09	
+ WM	**13.42**	**<0.001**	**4.69**	**<0.05**	**5.90**	**<0.05**	**6.39**	**<0.05**	**5.60**	**<0.05**
+ DP_1:WM			1.32		3.02	<0.1				
+ DP_2:WM	1.90				1.20					
+ DP_1:DP_2:WM					2.02					

$\chi^2 < 1$ are omitted.
Bold values are the significant ones.

cued-features for the focalized DP), ACT-R-based processing model (Lewis and Vasishth, 2005) flatly predicts no difference among any of the tested configurations suggesting that this model needs extra assumptions to account for the revealed asymmetries.

Considering the online predictions (**Figure 7**), assuming an encoding cost penalty for the *Pro* conditions (hypothesis A3 in section Testing the Different Predictions), *memory-load* model becomes competitive in predicting GD reading times (r (14) = 0.65, $p = 0.006$) and (less robustly) and TT (r (14) = 0.53, $p = 0.034$).

Feature Retrieval and Encoding Cost (FREC) (FRC+FEC) however, with the very same assumptions on *Pro* encoding penalty and under the hypothesis that only 2nd person features on the subject are cued by the verb (H2 in section Testing the Different Predictions), correlates much more precisely both with GD (r (14) = 0.90, $p < 0.001$) and (even better) with TT (r (14) = 0.96, $p < 0.001$). Considering the nature of retrieval, the higher correlation with respect to the most comprehensive, latest, measure (i.e., TT) is expected.

As predicted under Bianchi (2003, 2006), Sigurdhsson (2004) and Elbourne (2005) analyses, the difficulty associated to the processing of the *Pro* condition is mostly due to the out of the blue presentation of the second person feature restricted by a *N* predicate. Therefore, the processing cost revealed at the DP regions where *Pro* occurred is likely due to encoding (need of postulating the salience of the relevant referents and update the common ground accordingly): this is revealed by longer reading times in late measures (GD, TT, and SP) comparable at both DPs regions under the *Pro* condition. In DP_2 region, also an effect at the early FF measure was observed for Pro_1-Pro_2, suggesting an element of surprise as soon as the second restricted pronoun is encountered in the subject cleft position. A similar early effect, dependent on the presence of Pro_2, is revealed also at the verb segment possibly indicating a "spillover effect" of the previous region processing or a retrieval difficulty. The spill-over interpretation only is supported by the fact that the DP_2 effect disappears in later measures at the VERB segment (non-significant effect in GD and TT). On the other hand a retrieval problem

TABLE 11 | Main results summarized.

Region	Condition Measure	Art_1-Art_2	Pro_1-Pro_2	Art_1-Pro_2	Pro_1-Art_2
DP1 (focalized object)	Acceptability	Good	Bad	Good	Medium
	Comprehension	Good	Good	Good	Good
	First fixation	Baseline	Baseline	Baseline	Baseline
	Gaze	Baseline	Slower	Baseline	Slower
	Total	Baseline	Slower	Baseline	Slower
	Second pass	Baseline	Slower	Baseline	Slower
	Regressions from	Baseline	More	Baseline	More
	Regressions in	Baseline	Baseline	Baseline	Baseline
D2 (subject)	First fixation	Baseline	Slower	Slower	Baseline
	Gaze	Baseline	Slower	Slower	Baseline
	Total	Baseline	Slower	Mildly slower	Baseline
	Second pass	Baseline	Slower	Mildly slower	Baseline
	Regressions from	More	More	Baseline	Baseline
	Regressions in	Baseline	More	Baseline	Baseline
Verb	First fixation	Baseline	Slower	Slower	Baseline
	Gaze	Baseline	Slower	Baseline	Baseline
	Total	Baseline	Slower	Baseline	Baseline
	Second pass	Baseline	Slower	Baseline	Baseline
	Regressions from	Slightly more	More	Baseline	Baseline

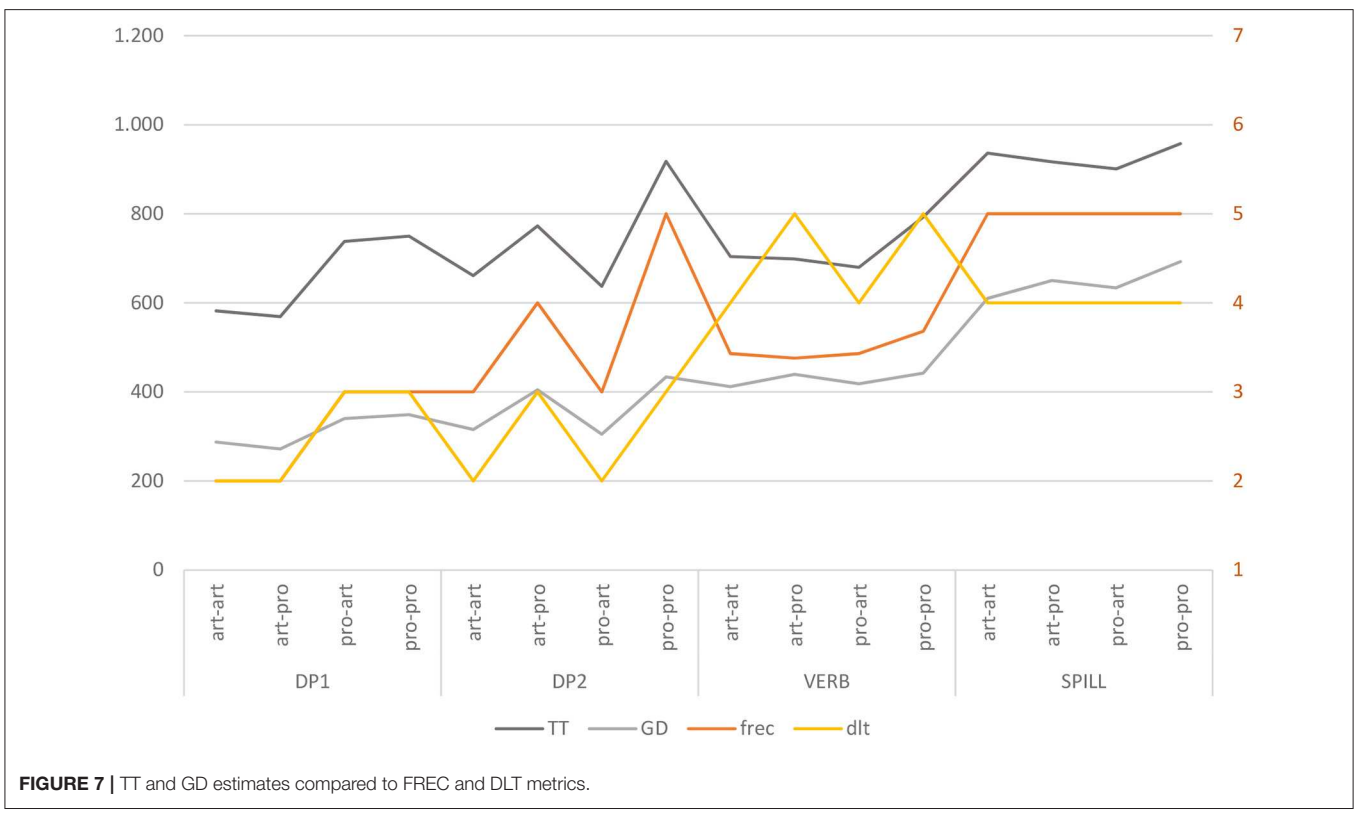


FIGURE 7 | TT and GD estimates compared to FREC and DLT metrics.

in matching conditions is suggested by the first and total regression patterns leaving the VERB region: the significant *post hoc* pairwise comparison indicates less regressions from the verb in the mismatching conditions, possibly revealing a difficulty in retrieving the correct argument in the matching cases (especially with second person feature matching). Considering

that the same effect is observed also at DP_2, we should conclude that the retrieval interpretation is not the sole possible analysis.

This difficulty in the matching conditions is in line with the *top-down*, *similarity*-based and *intervention*-based predictions, but not with the *memory-load* ones.

Under the H2 hypothesis (2nd person feature on the subject cued by verb agreement morphology should facilitate the integration of the subject, while 3rd person default agreement should represent just a minor facilitation) the top-down model would predict an asymmetry also in the mismatching condition. However, no significant on-line evidence of a facilitation in retrieval is encountered for the Art_1-Pro_2 vs. Pro_1-Art_2 neither in terms of reading times at VERB, or regression probability. The on-line results then support hypothesis H1 (section Testing the Different Predictions), namely that both 2nd and 3rd feature mismatch on the VERB might facilitate retrieval (quicker FF), but Pro_2 encoding penalty makes this effect hardly detectable. We agree with the reviewers suggesting that one way to tease apart the actual impact of the encoding penalty with respect to the retrieval effect would be to include in this experimental design the Subject Cleft condition where only encoding (and no retrieval penalty due to matching features to be re-merged) will be present. Another way to disentangle the two components would be to introduce a proper context, then removing the encoding penalty of *Pro* predicted under the assumption A3.

We can only suppose here that the encoding cost, responsible for the on-line slowdown revealed in the *Pro* conditions, could have been partially mitigated by the (mild) facilitation at retrieval in the mismatch Art_1-Pro_2 case, in the end producing an acceptability equivalence between the Art_1-Art_2 matching condition and the Art_1-Pro_2 mismatch condition. Everything being equal, removing the encoding cost, i.e., providing an appropriate context, we expect a difference Art_1-Art_2 vs. (worse than) Art_1-Pro_2 to appear, hence confirming the facilitation related to the usage of a deictic, mismatching, 2nd person feature in the intervening DP whose morphology is cued by the selecting verb (hence confirming the explanatory superiority of H2 over H1 also in on-line measures). Coherently with Staub (2010, p. 77-78), the complexity signatures at DP_2 revealed (also) by a generally higher probability of making a regression from this region, especially in the matching conditions case (and especially in the Pro_1-Pro_2 case) could be interpreted as an indication of "something is wrong" (as in E–Z reader 10 model, Reichle et al., 2009) or, more precisely, an integration failure due to time out (c.f. Staub, 2010, p. 83) because of a context-update request.

To conclude, a final crucial intent of this study was to provide some new evidence for disentangling the (complex) relation between off-line and on-line performance measures. Given the off-line results gathered, first, we observed that acceptability judgments are more discriminative than accuracy in comprehension questions (though both generally correlates on the numerical patterns), second, FRC metrics, based on the *top-down* model, is the one making the closest predictions with respect to the pattern revealed across conditions. This suggests that the retrieval effort, at least in this context, is the best predictor of the overall acceptability and that, despite heavy encoding efforts (revealed by on-line measures), readers are fully rewarded by an adequate comprehension, revealed by accurate answers in all conditions.

As for the on-line data, again FRC+FEC (FREC) shows the best correlation with respect to the revealed "late" measures (GD and TT). Unexpected referents, introducing features that force a revision of the common ground assumptions are correctly predicted to affect performance by the FEC component at specific regions. These predictions crucially rely on a precise linguistic theory that takes into consideration the nature of the OC dependency and the relation between D, N types and person features. It is important to emphasize that no significant interaction between WM and our syntactic manipulation has been revealed: high WM participants simply show faster reading times and more regressive patterns compared to the low WM population across all conditions.

The actual usage of working memory during the processing of these specific constructions is still to be explored precisely. Nevertheless, we believe that the intuition that identical features that must be (re)merged within the active workspace are lighter to be processed than new ones or "similar" ones that must be kept distinct in memory (like an extra nominal restriction or an extra second person index) is worth further investigation: this idea is coherent with a "primed" active storage in which the "memory units" encoding a specific feature, being just activated would be more accessible than other units (on the line of ACT-R Lewis and Vasishth, 2005 intuition), while forcing a minimal diversification of a new pattern with respect to a pre-activated overlapping one has a considerably high cost.

ETHICS STATEMENT

The experiment was approved by the Ethics Committee of the Dipartimento di Scienze del Sistema Nervoso e del Comportamento of the University of Pavia.

AUTHOR CONTRIBUTIONS

CC designed and directed the project, developed the theoretical top-down framework, the FR(E)C complexity measure, and performed the acceptability judgment experiment. PC performed the eye-tracking experiment and analyzed the results of both experiments. Both authors discussed the results and contributed to the final manuscript.

FUNDING

This work was supported by IUSS as an internal research project (ProGraM-PC–A Processing-friendly Grammatical Model for Parsing and predicting on-line Complexity, PI: CC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02105/full#supplementary-material

REFERENCES

Aho, A. V., and Ullman, J. D. (1972). The theory of parsing, translation, and compiling, Vol. 1, *Parsing*. Englewood Cliffs, NJ: Prentice Hall.

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cogn. Sci.* 29, 313–341. doi: 10.1207/s15516709cog0000_22

Anderson, J. R., and Matessa, M. P. (1997). A production system theory of serial memory. *Psychol. Rev.* 104, 728–748. doi: 10.1037/0033-295X.104.4.728

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G., et al. (2004). Introducing the La Repubblica corpus: a large, annotated, TEI (XML)-compliant corpus of newspaper Italian. *Issues* 2, 5–163. Available online at: https://www.aclweb.org/anthology/L04-1128/

Belletti, A. (2008). The CP of clefts. *Rivista di Grammatica Generativa*. (2008) 33, 191–204. Available online at: http://lear.unive.it/jspui/handle/11707/434

Belletti, A., and Chesi, C. (2014). A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production. *Rivista di Grammatica Generativa*. (2014) 36: 1–28. Available online at: http://lear.unive.it/jspui/handle/11707/5239

Belletti, A., Friedmann, N., Brunato, D., and Rizzi, L. (2012). Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian. *Lingua* 122, 1053–1069. doi: 10.1016/j.lingua.2012.02.007

Belletti, A., and Rizzi, L. (2013). "Intervention in grammar and processing," in *From Grammar to Meaning: The Spontaneous Logicality of Language*, eds I. Caponigro and C. Cecchetto (Cambridge: Cambridge University Press), 293–311.

Bever, T. G. (1974). "The ascent of the specious, or there's a lot we don't know about mirrors," in *Explaining Linguistic Phenomena* ed D. Cohen (Washington, DC: Hemisphere), 173–200.

Bianchi, V. (2003). "On finiteness as logophoric anchoring," in *Temps et Point de Vue/Tense and Point of View*, eds J. Guéron, L. Tasmovski (Paris, Univ. Paris X Nanterre), 213–246.

Bianchi, V. (2006). On the syntax of personal arguments. *Lingua* 116, 2023–2067. doi: 10.1016/j.lingua.2005.05.002

Bianchi, V., and Chesi, C. (2014). Subject islands, reconstruction, and the flow of the computation. *Linguist Inq.* 45, 525–569. doi: 10.1162/LING_a_00166

Chesi, C. (2015). On directionality of phrase structure building. *J. Psychol. Res.* 44, 65–89. doi: 10.1007/s10936-014-9330-6

Chesi, C. (2017). Phase-based Minimalist Parsing and complexity in non-local dependencies. *Proceedings of CLiC-it 2017*. CEUR Workshop Proceedings, Roma:CEUR Workshop Proceedings.

Chomsky, N. (1965). *Aspects of The Theory of Syntax*. Cambridge, MA: MIT Press.

Crowder, R. G. (1976). *Principles of Learning and Memory*. Hillsdale, NJ: Erlbaum.

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Elbourne, P. D. (2005). *Situations and Individuals*. Cambridge, MA: MIT Press.

Friedmann, N., Belletti, A., and Rizzi, L. (2009). Relativized relatives: types of intervention in the acquisition of A-bar dependencies. *Lingua* 119, 67–88. doi: 10.1016/j.lingua.2008.09.002

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). "The dependency locality theory: a distance-based theory of linguistic complexity," in *Image, Language, Brain*, eds Y. Miyashita, A. P. Marantz, and W. O'Neil (Cambridge, MA: MIT Press), 95–126.

Gordon, P. C., Hendrick, R., and Johnson, M. (2001). Memory interference during language processing. *J. Exp. Psychol.* 27:1411. doi: 10.1037//0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., and Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *J. Mem. Lang.* 51, 97–114. doi: 10.1016/j.jml.2004.02.003

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., and Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science* 274, 114–116. doi: 10.1126/science.274.5284.114

Kidd, E., Brandt, S., Lieven, E., and Tomasello, M. (2007). Object relatives made easy: a cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Lang. Cogn. Proc.* 22, 860–897. doi: 10.1080/01690960601155284

King, J., and Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *J. Mem. Lang.* 30:580. doi: 10.1016/0749-596X(91)90027-H

Kiss, K. E. (1998). Identificational focus versus information focus. *Language* 74, 2, 245–273. doi: 10.1353/lan.1998.0211

Kuperman, V., and Van Dyke, J. A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *J. Mem. Lang.* 65, 42–73. doi: 10.1016/j.jml.2011.03.002

Lenth, R., Singmann H., Love J., Buerkner P., and Herve M. (2019). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. R package version 1.4. Available online at: https://github.com/rvlenth/emmeans

Levy, R., Fedorenko, E., and Gibson, E. (2013). The syntactic complexity of Russian relative clauses. *J. Mem. Lang.* 69, 461–495. doi: 10.1016/j.jml.2012.10.005

Lewandowsky, S., Oberauer, K., Yang, L. X., and Ecker, U. K. (2010). A working memory test battery for MATLAB. *Behav. Res. Methods* 42, 571–585. doi: 10.3758/BRM.42.2.571

Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25

Link, G. (1983). "The logical analysis of plurals and mass terms: a lattice-theoretical approach," in *Formal Semantics: The Essential Readings*, eds P. Portner and B. Partee (Oxford: Blackwell Publishing), 127–146.

Longobardi, G. (1994). Reference and proper names: a theory of N-movement in syntax and logical form. *Ling. Inq.* 609–665.

Mancini, S., Molinaro, N., Rizzi, L., and Carreiras, M. (2011). A person is not a number: discourse involvement in subject–verb agreement computation. *Brain Res.* 1410, 64–76. doi: 10.1016/j.brainres.2011.06.055

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

Munaro, N. (1999). *Sintagmi Interrogativi nei Dialetti Italiani Settentrionali*. Padova: Unipress.

Phillips, C. (1996). *Order and structure*, (Doctoral dissertation), Massachusetts Institute of Technology, Cambridge, MA, United States.

Pollard, C., and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Postal, P. (1966). "On so-called pronouns in English," in *Modern Studies in English*, eds D. A. Reibel and S. A. Schane (Englewood Cliffs, NJ: Prentice-Hall), 177–206.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.

Reichle, E. D., Warren, T., and McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychon. Bullet. Rev.* 16, 1–21. doi: 10.3758/PBR.16.1.1

Rizzi, L. (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, L. (1997). "The fine structure of the left periphery," in *Elements of Grammar*, ed L. Haegeman (Dordrecht: Springer), 281–337.

Sigurdhsson, H. Á. (2004). "The syntax of person and speech features," in *The Syntax and Interpretation of Person Features,* eds V. Bianchi, K. Safir (Italian Journal of Linguistics), 16.1.

Stabler, E. (1996). *Derivational minimalism. In International Conference on Logical Aspects of Computational Linguistics*. Berlin; Heidelberg: Springer. 68–95.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition* 116, 71–86. doi: 10.1016/j.cognition.2010.04.002

Stowell, T. (1983). Subjects across categories. *Ling. Rev.* 2, 285–312.

Traxler, M. J., Morris, R. K., and Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *J. Mem. Lang.* 47, 69–90. doi: 10.1006/jmla.2001.2836

Van Acker, F. (2007). *Online Survey Creator 7.4 for Windows. Ms.*

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol.* 33:407. doi: 10.1037/0278-7393.33.2.407

Van Dyke, J. A., and McElree, B. (2006). Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166. doi: 10.1016/j.jml.2006.03.007

Wanner, E., and Maratsos, M. (1978). "An ATN approach to comprehension," in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan, and G. A. Miller (Cambridge, MA: MIT Press).

Warren, T., and Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition* 85, 79–112. doi: 10.1016/S0010-0277(02)00087-2

Warren, T., and Gibson, E. (2005). Effects of NP type in reading cleft sentences in English. *Lang. Cogn. Proc.* 20, 751–767. doi: 10.1080/01690960500051055

Weckerly, J., and Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology* 36, 559–570. doi: 10.1111/1469-8986.3650559

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fpsyg.2019.02994

(Morpho)syntactic Variation in Agreement: Specificational Copular Clauses Across Germanic

Jutta M. Hartmann[1] *and Caroline Heycock*[2]*

[1] General Linguistics, Faculty for Linguistics and Literary Science, Bielefeld University, Bielefeld, Germany, [2] Linguistics & English Language, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, United Kingdom

In this paper we bring together the results of our research into agreement in copular clauses in four different Germanic languages—Dutch, German, Faroese, and Icelandic—in order to provide an overview of the results. These cases present a particularly interesting window into how verbal agreement operates, since there are two potential controllers of agreement, which may disagree in person and/or number (*The source of the rumor* BE *the neighbors/you*-sg/*you*-pl). We will show that there is variation at all levels in which nominal controls agreement: cross-linguistic, inter-speaker within a single language, and intra-speaker. We argue that our data support the following claims: (1) "Downward" agreement for person, as well as number, with a nominal that is not in the canonical subject position is possible and in some cases preferred; (2) The agreement patterns observed in Icelandic and Faroese support the hypothesis that in these languages there are distinct Number and Person heads; (3) "Downward" agreement from a high position in the left-periphery is a grammatically distinct phenomenon from agreement when the verb remains in a lower position in the clause; (4) In some languages and some configurations, speakers show a significant degree of indeterminacy in their judgments and production, suggesting that speakers use more than one grammar. We relate our findings to current discussions in the generative literature on subject agreement and in particular differences between number and person agreement, and possible connections to restrictions on object clitics; we also discuss questions that remain open, and invite new, cross-disciplinary research.

Keywords: Germanic, copular clauses, agreement, downwards agree, number-only agreement

OPEN ACCESS

Edited by:
Andrew Nevins,
University College London,
United Kingdom

Reviewed by:
Carson T. Schütze,
University of California, Los Angeles,
United States
Marcel Den Dikken,
MTA Research Institute for Linguistics, Hungary
Sendy Caffarra,
Basque Center on Cognition, Brain and Language, Spain

***Correspondence:**
Caroline Heycock
caroline.heycock@ed.ac.uk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *20 December 2019*
Accepted: *17 December 2019*
Published: *11 February 2020*

Citation:
Hartmann JM and Heycock C (2020) (Morpho)syntactic Variation in Agreement: Specificational Copular Clauses Across Germanic. Front. Psychol. 10:2994. doi: 10.3389/fpsyg.2019.02994

1. INTRODUCTION

In this paper we bring together results from a series of experiments that we have conducted investigating agreement in a particular type of clause, across four Germanic languages: Dutch, German, Faroese, and Icelandic. Our investigation focusses on SPECIFICATIONAL COPULAR CLAUSES (SCCs henceforth), which feature minimally the copular verb (*be* in English) and two noun phrases (DPs). The definition of these clauses will be gone into in more detail below; (1) gives examples from English.

(1) a. The cause of the riot {was/*were} the pictures of the wall.
b. The cause of the riot {was/*were you}.
c. My favorite authors {*is/are} Heller and Austen.
d. The winning candidates {*is/are} you two.

This type of clause is of interest for the syntax of agreement for various reasons. Notably, languages differ as to which of the two nominals the verb agrees with. As is suggested by the examples above, in English agreement is, to a high degree of consistency, with the leftmost/first DP (DP1); conversely, as discussed in Moro (1991, 1997), in Italian agreement is consistently with the rightmost/second DP (DP2):

(2) a. La causa della rivolta *è/sono le foto
the cause of.the riot *be.3.SG/be.3.PL the pictures
del muro. ITALIAN
of.the wall
'The cause of the riot is the pictures of the wall.'

b. La causa della rivolta *è/sono io.
the cause of.the riot *be.3.SG/be.1.SG I
'The cause of the riot is me.'

While there is general consensus in the literature that English and Italian are consistently "DP1 agreement" and "DP2 agreement" languages, respectively, in the syntax of these copular clauses, in this article we show that in other languages—even those closely related to English—there is a richer and more complex pattern of variation. We give an initial illustration in (3):

(3) a. … weil das grösste Problem deine Eltern
because the biggest problem your parents
sind/***ist**. GERMAN
be.PRES.3PL/*be.PRES.3SG
'…because the biggest problem is your parents.'

b. … dat de oorzaak van het ongeluk kapotte
that the cause of the accident broken
remmen %**waren**/%**was**. DUTCH
brakes be.PST.PL/be.PST.SG
'…that the cause of the accident was broken brakes.'

c. … um orsøkin til eldin %**vóru**/%**var**
if cause-DEF to fire-DEF be.PST.PL/be.PST.SG
tey brennandi kertiljósini. FAROESE
the burning candles.DEF
'…whether the cause of the fire was the burning candles.'

d. … hvort aðalvandamálið %**væri**/
if main problem.DEF be.SBJ.3.SG/
%**væruð**/ %**væru** þið. ICELANDIC
be.SBJ.2.PL/ be.SBJ.3.PL you.PL
'…whether the main problem is you.PL'

First, although in non-copular clauses all of these Germanic languages typically show a pattern very like English, in which the finite verb consistently agrees with a clause-initial subject[1], here we find three different agreement patterns:

1. agreement in number and person with the precopular noun phrase (DP1 agreement), as in English;
2. agreement in number and person with the post-copular noun phrase (DP2 agreement), as in Italian;
3. agreement with the post-copular noun phrase in number only (number-only DP2 agreement)—see the Icelandic example in (3d).

Second, all of the four languages that we investigated allowed at least two of these patterns, but to different extents: Icelandic and to a lesser degree Faroese show all three patterns; Dutch only shows DP1 and DP2 agreement; and German almost categorically requires full DP2 agreement in all but one context. Third, all four languages—even German, which as just stated is almost categorical in the preference for DP2 agreement—show a notable shift toward DP1 agreement in one particular syntactic context, when the copula precedes both DPs, as in (4):

(4) a. Meiner Meinung nach ?war/?? waren das
my opinion after was/were the.SG
Schlimmste am Urlaub die vielen Mücken.
worst at.the holiday the.PL many mosquitos
'In my opinion, the worst part of the holiday was the many mosquitos.' GERMAN

b. Misschien was/?*waren het ergste van de
possibly was/were the.SG worst of the
vakantie de vele muggen. DUTCH
holiday the.PL many mosquitos
'Possibly, the worst part of the holiday was the many mosquitos.'

SCCs in these languages thus provide an interesting testbed for theories of agreement; in particular, for theories which predict severe restrictions on agreement with "low" nominative arguments, i.e., nominative arguments that appear in a position lower than the canonical subject position. They also present a new, relatively unstudied set of cases of agreement variability. In this paper we bring together the results from a series of experimental studies to give an overview of the generalizations that have emerged, and to relate these results to current theories of agreement. While the details of the goals and results of the individual experiments are available in a number of different papers, our aim here is to summarize the results, show the emerging overall picture and relate our findings to current issues discussed in the syntax of copular clauses and agreement. Our hope is that this will facilitate interdisciplinary discussion of the issues raised. Throughout, we will provide the references to papers where more detailed descriptions of experiments have been reported.

In section 2, we will outline some current issues in the syntax of agreement that are relevant to, and we hope illuminated by, our results. In section 3 we give some background on Specificational Copular Clauses and outline an argument that the agreement facts support an "inversion" analysis of these clauses. With this background, in section 4 we discuss patterns that are common to all four languages, and then in section 5 we turn to the variation we find, focussing in particular on person agreement. In section 6 we briefly discuss some of the new questions that have opened up in the course of this investigation, before concluding in section 7.

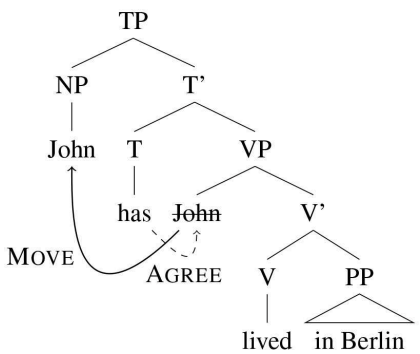
[1]In German and Icelandic in particular there are exceptions to this generalization, as will be discussed later.

2. THE SYNTAX OF AGREEMENT: SOME BACKGROUND

There is a range of theories on how the sharing of features that constitutes agreement can be modeled. In current generative grammar, it is generally assumed that morphological agreement is one possible reflex of a more general syntactic relation, AGREE, that is established between a "probe" (the agreeing element, typically a head) and a "goal" (the agreement controller). There are a number of different proposals concerning the configurational relationship between the probe and the goal. (i) A longstanding position, going back at least to Chomsky (1981), but more recently championed in Koopman (2006), is that agreement holds between a head and an agreement controller in its specifier. (ii) A less constrained alternative is that Agree can be established between a probe and a c-commanding goal that may be more remote than the local specifier: see among others Wurmbrand (2012), Zeijlstra (2012), Bjorkman and Zeijlstra (2019). This is termed either "Upward Agree" or "Reverse Agree" because it reverses the hierarchical relations between probe and goal in the more widely adopted proposal of Chomsky (2000), namely that (iii) the probe must c-command the goal ("Downward Agree"). Depending on the framework and language considered, there is also work that argues for allowing both upwards and downward Agree (with upwards Agree often reducing to specifier-head agreement), see for example Béjar and Rezac (2003), Baker (2008), and Ackema and Neeleman (2018); note that for Béjar and Rezac (2003) the two types do not have equal status: upward Agree obtains only where downward Agree fails[2].

In this paper we will be assuming downward Agree, for reasons that will become clearer when we have introduced the structure of copular clauses. (5) illustrates a simple case of how a downward Agree analysis handles subject-verb agreement in a non-copular sentence like *John has lived in Berlin*.

(5)



In this representation it is assumed that there is a single probe that has unvalued features for both person and number that will be valued by the first set of features that it encounters on a downward search of its c-command domain. Considerable work has been done on the idea that probes may be more or less specified in the features that they are searching for: e.g., a probe might be specified to match not against any person feature, but only, say, 1st or 2nd person, as proposed for Persian in Béjar and Kahnemuyipour (2017); again, we find cross-linguistic variation in this domain, see section 3.2 for further discussion.

There is now in fact a significant body of work establishing that agreement for person and agreement for number do not always behave in the same way; in some analyses it is argued that there are distinct syntactic probes, and in some cases in fact distinct heads associated with person and with number agreement. An argument for this last position is made in Sigurðsson and Holmberg (2008), based on "dative-nominative" configurations in Icelandic where there is a dative subject and a nominative argument lower in the structure. In such cases the verb may agree in number with a 3rd person nominative, as illustrated in (6a), but it cannot agree at all with a non-3rd person nominative, as illustrated in (6b)[3].

(6) a. Honum mundu virðast þeir (vera)
him.DAT would.3.PL seem they.M.NOM (be.INF)
hæfir. ICELANDIC
competent.M.PL
'They would seem competent to him.'
(Sigurðsson and Holmberg, 2008, p. 255)

b. *Henni virtumst við vera duglegar.
she.DAT seem.1.PL we.NOM be industrious
Intended: 'We seemed to her to be industrious.'
(Sigurðsson, 1996, p.76b)

Differences between number and person agreement will be discussed in more detail in section 5.

Two requirements for a successful agreement relation to be established are thus that the probe and the goal must be in the appropriate hierarchical relation to each other, and that the goal must carry the features searched for by the probe. A third requirement is that there can be no "intervening" goal: Agree must establish a match with the first appropriate set of features in its search path (assuming downward Agree, this means that it will seek to match with the highest potential goal in its c-command domain).

In the case of morphological agreement, there also seems to be a further requirement: whether or not a DP with the relevant features can in fact control agreement depends on its morphological case. At least in the Germanic languages, there is a generalization that only nominative DPs can control agreement (see Bobaljik, 2008 for discussion, but also Jónsson, 2009; Ussery,

[2]It should also be noted that in the proposal of Bjorkman and Zeijlstra (2019) that argues for Agree to be uniformly "upward" (with the goal c-commanding the probe), the CHECKING relation established by Agree is followed by a second step of VALUATION, and in certain circumstances this can have the result that agreement can obtain between a probe and a c-commanded element, for example between finite T and a lower nominative argument. This possibility only arises if there is a higher argument that is featurally defective in some way. Bjorkman and Zeijlstra do not discuss SCCs, but it seems that the most natural application of their proposal to account for the possibility of DP2 agreement would require that DP1 be analyzed as featurally deficient. Such an analysis has been advanced in Béjar and Kahnemuyipour (2017, 2018); see Hartmann and Heycock (2018a,c) for arguments against the claim of featural deficiency for DP1.

[3]For these cases where the "low" nominative argument is the subject of a non-finite clause, default (3rd person singular) agreement in the matrix is grammatical for most, possibly all speakers, while this type of default agreement is unacceptable when the nominative is a co-argument of the dative. See Sigurðsson and Holmberg (2008) for details, including interspeaker variation, and the further data and discussion in Thráinsson et al. (2015).

2017 for potential counterexamples). In most configurations, the nominative argument is the structurally highest argument, so in order to see the relevance of case, we need a configuration in which the two are separated. We find such a configuration in German with a number of psych verbs that select for a dative experiencer argument and a nominative theme argument. The dative argument has been shown to be the structurally higher argument (higher before any movement has taken place, and at the point that T is merged) with such verbs like *gefallen* in (7) (see Lenerz, 1977; Sternefeld, 2009, p. 563), subject-verb agreement is nevertheless with the nominative argument. This shows that nominative case is a precondition for agreement in German.

(7) … dass mir diese Bücher gefallen — GERMAN
… that me.DAT these.NOM books please.PL
'… that I like these books'

The Icelandic dative-nominative construction illustrated above in (6) shows a similar effect, but in these cases it has been argued that while the dative argument does not control agreement, it does interact with the agreement probe in some way (a phenomenon referred to in the literature as "defective intervention" see among many others, Holmberg and Hróarsdóttir, 2004; Sigurðsson and Holmberg, 2008; Thráinsson et al., 2015; Ussery, 2017; Hartmann and Heycock, 2018d); we will come back to this briefly in section 5.1.

In all the discussion so far we have been considering an agreement probe associated with finite T[ense], the goal of which is a nominative noun phrase (typically the subject), which results in morphological agreement on the finite verb. While this is the most familiar instance of agreement in Germanic, it has also been observed that in some Germanic languages, agreement with the subject of a clause can additionally be related to the C[omplementizer]-position. One version of this C-agreement is that there is agreement marking on the complementizer in a number of varieties of Dutch, as illustrated in (8), from van Koppen (2005), p. 33.

(8) …datt-e we naar Leie gaan. — KATWIJK DUTCH
…that-PL we to Leiden go
'…that we are going to Leiden.'

We will argue in section 4.3 that C-agreement is the basis for the agreement exemplified in (4b) above. But first we need to look also at the type of copular clauses that are the focus of our investigation.

3. SPECIFICATIONAL COPULAR CLAUSES AND AGREEMENT

3.1. Specificational Copular Clauses: Background

Copular clauses may have various syntactic types of phrase in nonsubject position, including Adjective Phrases, as in (9a), Prepositional Phrases, as in (9b), among others.

(9) a. Alexis is very tall.
b. Alexis is in a very weak position.

However, the case that is of interest to us here is that of "binominal" copular clauses, where both of the phrases that accompany the copular verb are nominals. Such binominal copular clauses have been further subclassified, the most influential classification being the four-way scheme set out in Higgins (1979) and illustrated in (10).

(10) a. Sarah is a genius / the winner. [predicational]
b. The man you saw yesterday is the man Jessie was talking about today. [equative/equational]
c. This is Sarah / the woman who I was telling you about. [identificational]
d. The winner is Sarah. [specificational]

There is a substantial literature on copular clauses: for recent discussion and extensive references to other work, we refer the reader to den Dikken (2006b), Mikkelsen (2011), and Heycock (in press). Here we simply present a brief summary of some relevant distinctions from that literature.

The hallmark of predicational copular clauses like (10a) is that the pre-copular noun phrase, *Sarah* in (10a), is assigned the property described by the post-copular noun phrase, *a genius/the winner* in (10a). The post-copular noun phrase does not introduce a referent, even when it is definite (see Coppock and Beaver, 2015 for a recent discussion of the use of definite nominals as predicates). A syntactic diagnostic for predicative copular clauses in English that is often appealed to is that the same predication is felicitous in a small clause, without any instance of the copula:

(11) a. I consider [Sarah a contender / the winner]
b. With [Sarah a contender / the winner], the Jones family are feeling rather pleased with themselves.

In equative/equational binominal sentences like (10b), two individuals are "equated"; put differently, the two descriptions are asserted to pick out the same referent. Such cases generally cannot appear in small clauses:

(12) a. *I consider [the man you saw yesterday the man Jessie was talking about today].
b. *With [the man you saw yesterday the man Jessie was talking about today], our suspicions were raised.

The ungrammaticality of examples like (12a), while frequently cited as following from the status of the small clause as an equative, is however already predicted by the fact that *consider* is a verb that requires its argument proposition to be open to subjective assessment (Saebø, 2009), and presumably being identical to another entity is not even coercible into a subjective predicate. Absolute adjuncts introduced by *with* are not subject to the same restriction, so that the ungrammaticality of (12b) does not suffer from the same confound[4].

[4] As a referee points out, further evidence in the same direction is that the inclusion of the copula improves (12b), but not (12a):

(i) a. ?*I consider [the man you saw yesterday to be the man Jessie was talking about today].

The third class, identificational copular clauses, illustrated in (10c), have a deictic expression as the first nominal and some referring expression (whether a name or a definite) as the second. There is some discussion in the literature as to whether such sentences should rather be subsumed into one of the other classes; see Partee (1986), Huber (2002), Mikkelsen (2005), Heller and Wolter (2008), and Moltmann (2013) for discussion.

The fourth class are specificational copular clauses (SCCs). This is the type that is our primary focus in this paper. One example was already given in (10d) above, some more are given in (13):

(13) a. The best candidate was Jo.
b. The cause of the riot was the leaked memo.
c. The source of the rumor was probably you.

SCCs typically have a definite description as the first nominal, and some referring expression as the second. Since many definite nominals are ambiguous between a predicational and a referential reading, many sentences are ambiguous between a predicational and a specificational reading. Such sentences can give a sense of the kind of interpretation associated with "specification." Consider for example (14):

(14) My favorite horse is the winner.

On the predicational reading, the sentence is a natural answer to the question *Has your favorite horse just won that race, or has it lost?* On the specificational reading, it is a natural answer to the question *Which horse do you like best, the one that won or the one that lost?* The predicational reading can be forced by adding a proper name as an apposition to the first DP:

(15) My favorite horse, Ardbeg, is the winner.

Equally, the specificational reading can be forced by a proper name in apposition to the second DP

(16) My favorite horse is the winner, Ardbeg.

It is important to observe that in SCCs, at least in Germanic, the first DP occupies the canonical subject position, rather than some topic position high in the left periphery. Thus, for example, the subject of an SCC can immediately follow the auxiliary in a root polar interrogative in English:

(17) Is the best candidate really Jo?

This distinguishes SCCs from cases that have been described in the literature as A′ predicate fronting, of the kind discussed in Birner (1992) and illustrated in the second sentence in (18) (the introductory sentence is included just to provide a favoring environment), where the same diagnostic indicates that the initial phrase does not occupy the canonical subject position. For extended discussion of the contrast between SCCs and A′ predicate fronting, see Heycock and Kroch (1998).

(18) Bad housing is a threat to social cohesion in this area. {An equally serious threat/Equally threatening} are factory closings.

(19) *Are {an equally serious threat/equally threatening} factory closings?

Although, as just discussed, the first DP in an SCC does not occupy some peripheral "topic" position, one of the best-known characteristics of SCCs is that they nevertheless have a fixed information structure. In particular, the second DP has to be in focus (Heggie, 1988). The following exemplification is from Heycock (1994). First, we see that the same predicative copular sentence can be used felicitously in both (20) and (21), where the questions set up either the first DP or the second as the focus in the answer:

(20) A: Who was the culprit? (John or Bill?)
B: JOHN was the culprit.

(21) A: What was John? (Was John the culprit or the victim?)
or
A′ Tell me something about my cousin John and his role in the crime.
B: John/he was the CULPRIT.

In contrast, the specificational sentence is good in only one of these two contexts, where the focus is on the postcopular constituent.

(22) A: Who was the culprit? (John or Bill?)
B: The culprit was JOHN.

(23) A: What was John? (Was John the culprit or the victim?)
or
A′ Tell me something about my cousin John and his role in the crime.
B: *The CULPRIT was John/him.

For experimental evidence of this restriction that makes use of the prosodic contours associated with focus, see Hartmann (2019).

A typical characterization of specificational sentences is that the first nominal, although in the canonical subject position, does not have a simple referential reading (of type *e*). In cases where this nominal could in principle pick out an animate entity, this can be seen by the pronoun used to refer back to it. Thus, while normally *the best candidate* would have to be referred back to by a gendered pronoun if it picks out a human candidate, this is not the case when it appears as the subject of a specificational copular sentence (Mikkelsen, 2005; Heycock, 2012):

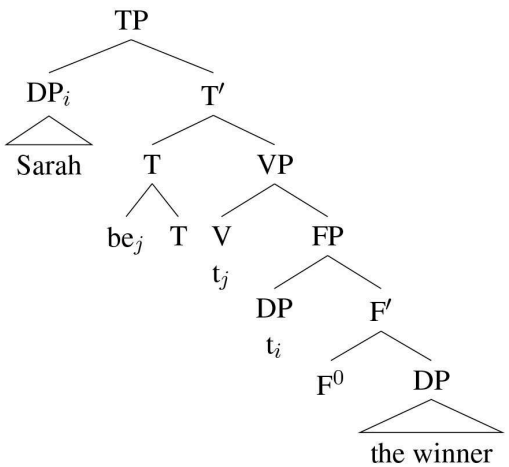
(24) a. #The best candidate was very well-spoken, wasn't it?
b. The best candidate was Jo, wasn't it?

While there is general agreement in the literature that the second nominal in an SCC denotes an individual (in contrast to a predicational copular sentence), and that the first nominal does not, there is less consensus concerning the denotation of the first, and this relates closely to the analyses that different

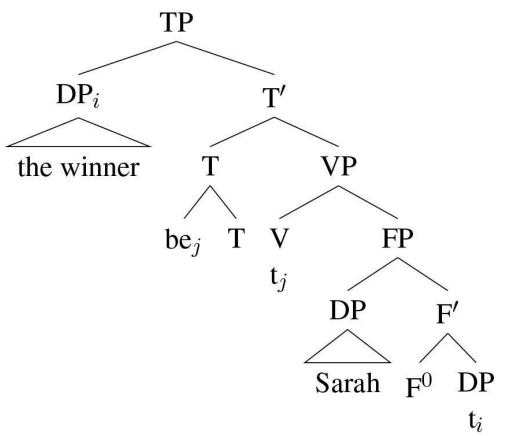
b. With [the man you saw yesterday being the man Jessie was talking about today], our suspicions were raised.

researchers have put forward. One widely-adopted proposal (see among others, Heggie, 1988; Moro, 1991, 1997; Mikkelsen, 2005; den Dikken, 2006a) is that the initial nominal is in fact a predicate (type <e,t>), and that specificational sentences are derived when the predicate, rather than the subject, of a small clause complement to the copula moves into the matrix subject position, as schematized in (25b). "F" is whatever functional head is taken to project the small clause; for most of the writers above it can be taken to be something akin to Bowers' (1993) Pr[edicative]P.

(25) a.



b.



An alternative proposal for the interpretation of the initial nominal is that it is a concealed question, an interpretation available to definite descriptions in cases like (26) (Romero, 2005; Heycock, 2012):

(26) They guessed/announced the best candidate.

This proposal is still compatible with the syntactic "inversion" analysis schematized in (25), as a concealed question denotation can be shifted into a predicative interpretation (in the sense that it can combine with an argument of type e to yield a proposition) just like other definite descriptions, as discussed in Heycock (2012). The possibility of inversion will be important in the discussion to follow, but the precise nature of the semantic contribution of the first nominal will not be important here, so we will not be discussing it further.

Note that all inversion accounts have to explain why the higher DP within the small clause (DP2) does not "intervene" to block movement of the lower, preventing the inversion. There have been a number of proposals for how this problem could be circumvented. The essence of the proposal in den Dikken (2006a) is that the head of the small clause moves to adjoin to the copula *be*, and that this head-movement has the effect of making the two DPs within the small clause "equidistant" from a probe above. For Mikkelsen (2005), Shlonsky and Rizzi (2018), and Hartmann (2016) what is crucial is an informational asymmetry between the two DPs in a specificational sentence. As mentioned briefly above, specificational sentences are unusual in that they have a restricted type of information structure. We have followed the characterization of this as being a requirement that DP2 is in focus; this is the characterization that Shlonsky and Rizzi assume as well. An alternative characterization, adopted in Mikkelsen (2005), is that DP1 has to be a topic. Mikkelsen capitalizes on the informational asymmetry by proposing that the agreement probe on T may optionally carry a [+Topic] feature. If it does, and if in addition the lower of the two DPs (DP1) carries such a feature, then it may move past the higher DP (DP2), simply because that DP cannot match the probe. Shlonsky and Rizzi (2018), on the other hand, argue that DP2 in a specificational sentence moves to a low Focus position at the edge of the VP. In their terms, this is a "criterial position," from which further movement is impossible (a case of "criterial freezing"). The remnant small clause may then move, stranding the focus to its right, and "smuggling" with it the lower DP, which subsequently moves out of it. In this paper we will assume that it is indeed the information structural asymmetry that is crucial in allowing the lower DP within the small clause to cross the higher, we will not discuss further the exact mechanism, but see Hartmann (2016, 2019) for a proposal.

There are a number of criteria that have been used as diagnostics for SCCs: as well as the distinctive pronominalization pattern for apparently animate nominals in initial position, illustrated in (24) above; these include restrictions on A′-extraction and obligatory focus on the second nominal (see Higgins, 1979; den Dikken, 2006b; Moro, 2006 for overviews and references). For the purposes of our studies, we operationalized the category of specificational copular clause as follows:

(27)
I. the clause contains a copula and two nominals;
II. the first nominal is a definite description, headed by a noun that either
   i. denotes a role (like *winner* or *candidate* in (10d), (13a) above; or
   ii. is an abstract noun (like *cause* or *source* in (13b), (13c);
III. the second nominal is either
   i. a name;
   ii. a definite description denoting a human; or
   iii. a pronoun

Clearly this operationalization would include some sentences that are at least ambiguous between specificational and predicational readings, as (14) above was. See section 3.3 for some detail concerning our strategies for avoiding such potential ambiguities in our materials.

3.2. Agreement in SCCs

The important work on specificational sentences in Moro (1991, 1997) showed that SCCs have different agreement properties in English and in Italian, as mentioned in the introduction. Essentially, in English agreement in SCCs is with the linearly leftmost/first/precopular overt nominal (DP1 henceforth), as illustrated in (1) above, repeated here as (28)[5].

(28) a. The cause of the riot {was/*were} the pictures of the wall.
b. The cause of the riot {was/*were you}.
c. My favorite authors {*is/are} Heller and Austen.
d. The winning candidates {*is/are} you two.

Note that, whether DP1s in specificational sentences are concealed questions or predicates, in either case they are predicted to be limited to 3rd person. They may be singular or plural, as just illustrated. In a specificational clause, if DP1 is singular, DP2 can be either singular or plural, and of any person. However, if DP1 is plural, the linearly rightmost/last/postcopular nominal (DP2) is again free to be of any person, but can typically only be plural, as in (28c,d)[6]. Evidently this restricts the types of potential agreement "mismatch" that can be constructed with this kind of specificational sentence. It should be noted, nevertheless, that the ungrammaticality of (28c,d) with singular agreement suggests strongly that the 3rd singular agreement in (28a,b) is controlled by DP1; if it were simply default agreement, the same 3rd singular agreement would be predicted to be acceptable in (28c,d), contrary to fact. We will return to this point in section 4.2.

The agreement pattern in Italian is different, as Moro argued: it is with DP2. This holds true both of number and person agreement. We repeat here the examples given earlier, which are adapted from those in Moro (1997), Ch. 1.

(29) a. La causa della rivolta *è/sono le
the cause of.the riot *be.3.SG/be.3.PL the
foto del muro. ITALIAN
pictures of.the wall
'The cause of the riot was the pictures of the wall.'
b. La causa della rivolta *è/sono io.
the cause of.the riot *be.3.SG/be.1.SG I

Moro (1997) derived this difference between English and Italian from the pro-drop character of the latter language. However, it has been known for some time that this cannot be the whole story. As pointed out in den Dikken (1998), Dutch allows DP2 agreement despite being a non-pro-drop language, and the same is true of German, which allows DP2 agreement even more freely (examples discussed for German go back at least as far as Berg, 1998).

If Dutch and German were invariant DP2 agreement languages (the characterization that Moro assumes for Italian), and English an invariant DP1 agreement language, one might pursue the idea that the difference in agreement is determined by the case properties of the languages. In Italian, Dutch, and German, both DPs in a finite specificational clause are nominative (in Dutch and Italian this is only evident when DP2 is a pronoun, since there is no overt morphological case marking on non-pronominal DPs in these languages). In Present Day English, on the other hand, where DP2 is a pronoun that is not syncretic for case, it is evident that it has to be accusative[7]:

(30) a. The cause of the riot is me/*I.
b. Die Ursache ist der kaputte
the cause be.3.SG the.M.SG.NOM broken
Wasserhahn. GERMAN
tap
'The cause is the broken tap'
c. Die Ursache bin ich.
the cause be.1.SG I.NOM
'The cause is me.'

As discussed in the last section, in all the Germanic languages, only nominative DPs can control morphological agreement on the finite verb. Given that postcopular DPs in English specificational sentences are accusative (for whatever reason), this precludes the possibility of the verb agreeing with them. It might then be possible to set up a system that makes DP2 the first candidate for controlling agreement for some structural reason. In Italian, Dutch, and German the search for an agreement controller would stop there, yielding DP2 agreement; in English

[5]There are limited/sporadic exceptions to this generalization. Typically they involve cases where the initial nominal is ambiguous/underspecified for number, as with light-headed relatives like (i)a or cases that might involve NP ellipsis like (i)b:

(i) a. All I could see {was/were} two staring eyes.
b. The best of the candidates {was Alex / were Alex and Jo}.

For some further discussion see Heycock (2012), p. 213.

[6]One exception to this is if DP1 can be a plurale tantum nominal, semantically singular but formally plural. Thus for example the Icelandic word *upptök* 'cause(s)' is formally plural but semantically singular, and can appear as the first DP in SCCs like (i)

(i) Þau spurðu hvort eldsupptökin væru ekki þurrkurinn.
they asked whether fire.causes.def be.SBJ.3PL not drought.DEF
'They asked if the cause of the fire wasn't the drought.'

Such cases, and the agreement patterns that are found there, are discussed in Hartmann and Heycock (2018a).

A reviewer points out that for speakers of varieties of English where collective nouns like *council* are unambiguously singular, other exceptions would be examples like (ii):

(ii) The defendants were the town council.

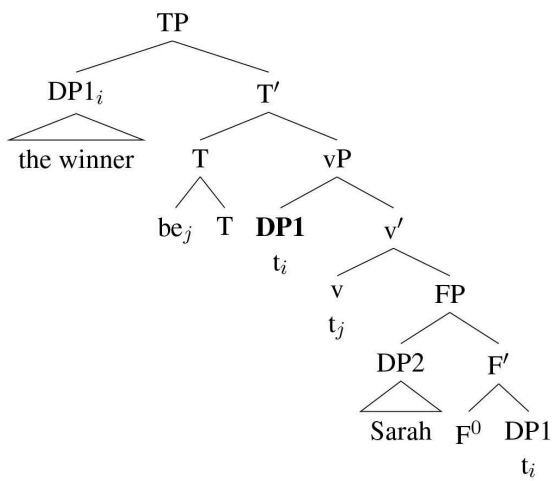
[7]We take it that *It is I* to be a frozen form in Modern Day English, not part of the productive syntax.

however, since DP2 agreement would be precluded by the accusative case, some mechanism could allow the search to continue, to find the nominative DP1 and agree with that.

However, Fischer (2003) established already that Dutch at least is not an "invariant DP2 agreement language." Rather, there is significant inter- and intra-speaker variation between DP1 and DP2 agreement in this language even in SCCs, where DP2 is invariantly nominative. One of the goals of our work on agreement in SCCs in Germanic, then, has been to look in detail at four languages that all have nominative DP2 in SCCs: first to establish what the agreement patterns are in a number of configurations, and then to work toward an analysis that could explain the patterns observed. The languages that we chose to investigate are German, Dutch, Icelandic, and Faroese. All four are Verb Second (V2) languages, as will be discussed further below. German and Dutch have SOV order in subordinate clauses, while Icelandic and Faroese, like the other Scandinavian languages have SVO. All four languages show morphological agreement on the copula (unlike a number of other Germanic languages, including Afrikaans and the standard varieties of all the other Scandinavian languages), but German and Icelandic have "richer" (less syncretic) agreement morphology than Dutch and Faroese.

One possible line of analysis for the difference between DP1 and DP2 agreement in specificational sentences is developed in Béjar and Kahnemuyipour (2017, 2018). These authors adopt the kind of inversion analysis discussed above, according to which DP1 in a specificational sentence originates in the lower position within a small clause. However, rather than assuming that DP1 moves directly from this position to Spec,TP [as sketched in (25) above], they adopt the proposal that DP1 in a specificational sentence always moves initially to a position below T, which they take to be the locus of the agreement probe. They do not discuss the specifics of this position, but it seems that for the purpose of discussion we can identify it with Spec,vP:

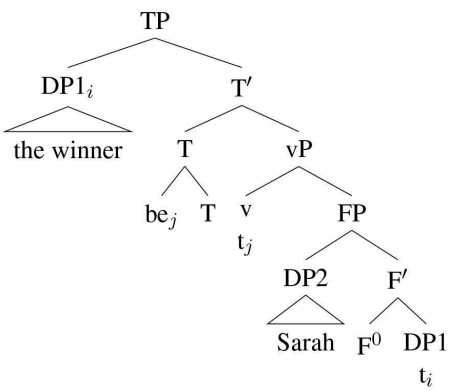
(31)



Given this derivation, DP1 in a specificational sentence, just like DP1 in a predicational sentence, will always be the first DP found by the agreement probe on T. The crucial extra assumption that Béjar and Kahnemuyipour (2017) make is that in a specificational sentence DP1 is deficient in ϕ-features. DP2 agreement then arises if a language has a probe that is searching for the feature(s) that DP1 lacks; such a probe will "skip" DP1 and hence be able to find and Agree with DP2. On the other hand, if a language has a sufficiently underspecified, and hence "undiscriminating" probe, it will match against DP1 and so agreement with DP2 will be blocked. In Béjar and Kahnemuyipour (2018) it is proposed that DP1 in a specificational sentence is deficient in that it lacks person features. Note that this lack of person cannot be common to all non-pronominal DPs, as it must distinguish between specificational subjects (which by hypothesis are skipped by a probe that is searching for person) and "ordinary" DPs occurring, say, as the subjects of predicational sentences (which are not skipped). Under this kind of analysis, variation between DP1 and DP2 agreement in a single language presumably reflects multiple options for the type of probe available (assuming, as is surely the case, that postulating variation in the ϕ-features of DP1 would be highly undesirable).

An alternative approach, which we have outlined in Hartmann and Heycock (2016, 2017) and follow-up work, is to propose that DP1 may fail to be agreed with in a specificational structure not because it is ϕ-deficient, but because in some languages it is possible for DP1 to reach a position above the agreement probe directly from its position within the small clause. Thus, while a derivation such as that in (31) will result in DP1 agreement, in some languages the derivation illustrated in (32) is possible:

(32)



Assuming downward Agree, the highest DP within the c-command domain of T is DP2 (*Sarah)*: this then predicts that agreement will hold between T and DP2. In DP1's base position it will not be found by the agreement probe because DP2 is closer to that probe; and in its derived position it is above the probe, and hence not in a position for Downward Agree to reach it.

Under this view, rather than a difference in the ϕ-sensitivity of the probe, variation between DP1 and DP2 agreement reflects a difference in the initial landing site of DP1, possibly reflecting

the presence of two distinct grammars.[8] We will return to this issue after we have discussed how we have gone about trying to establish the essential facts about agreement in the languages we have been considering.

In order to begin to address the theoretical questions arising from agreement in these copular clauses, we have sought to address the following questions, which we will discuss in more detail in sections 4, 5.

(33)
a. Do all the four languages we study here allow DP2 agreement both for number **and** for person?
b. Do all the four languages allow DP1 agreement?
c. How much variation in agreement is there within each language, and to the extent that there is any, is it inter-speaker, intra-speaker, or both?
d. Do we find evidence for default agreement, that is, 3rd singular agreement that cannot be analyzed as agreement with either DP1 or DP2?
e. Is there evidence for Number and Person being distinct probes and heads in any/all of the languages in question? This question is particularly relevant for Icelandic, as the existence of distinct probes and heads for Number and Person in this language has previously been argued for on the basis of the agreement pattern in dative-nominative configurations in cases like (6) above (see Sigurðsson and Holmberg, 2008)
f. It was observed in Heycock (2012) that in Faroese, DP1 agreement is strongly favored if the finite verb precedes both DPs. Is this pattern replicated in other V2 languages?

The questions in (33) aim at providing the overall picture of variation with respect to the availability of DP1 and DP2 agreement in different syntactic contexts. More specifically, potential differences in DP2 agreement with respect to number vs. person are interesting in the light of recent agreement theories, where it has been argued that downwards agreement with person is more restricted than number agreement, or—depending on whether or not we are dealing with a multiple agree configuration—possibly subject to syncretism effects. For recent analyses of agreement, the availability of DP1 agreement and the lack of default agreement is relevant for a distinction between a configurational approach to agreement in SCC (as we have proposed) and an approach such as Béjar and Kahnemuyipour (2017) in which DP2 agreement is claimed to arise just when DP1 does not have any ϕ-features accessible to the probe[9].

Additionally, one important point of dispute in agreement theories is whether or not person and number should be taken to be different probes or in fact distinct heads in at least some of our languages, see (33e).

3.3. Methodological Issues and Strategies

As discussed above, while earlier work on agreement in SCCs assumed that each language was of a particular "type" (requiring either DP1 or DP2 agreement), more recent work on Dutch and Faroese (Fischer, 2003; Heycock, 2009) suggested that in at least some cases there is intra-language variation of various kinds. Some of this variation is conditioned by syntactic environment (see in particular section 4.3), but some is not (or at least, not evidently). Given the possibility of inter-speaker variation, in order to understand the status of the different agreement options in a language it is essential not to rely on data from a single consultant.

In order to investigate the available patterns of number and person agreement with two nominative DPs in SCCs, we therefore conducted several experimental studies on number and person agreement in Dutch, Faroese, German and Icelandic, combining production studies (fill-in-the-blank) and rating studies (thermometer rating, following Featherston, 2008, which is a variant of the magnitude estimation technique, see Bard et al., 1996). We chose to investigate the issue using both production and rating tasks, as both have their advantages and disadvantages. A production task, such as the fill-in-the-blanks paradigm that we used, allows but does not force speakers to reflect on their own production. This method has been used in previous studies (see Berg, 1998; Fischer, 2003; Heycock, 2009). A further benefit of such a production task is that it allows for speakers to produce forms that the investigators were not previously aware of. However, the production task is to some extent a forced-choice task, in that participants are presented with a sentence in a particular order and can only choose some form to fit a single blank. Hence for example a 50/50 distribution might reflect two fully acceptable options (potentially, completely free variation between two grammatical variants) or two equally degraded options. The rating task can reveal such distinctions.

It has to be acknowledged that these experiments can only be viewed as a preliminary exploration, as there has been no prior work on this topic on Faroese and Icelandic, and little on Dutch or German. In particular, our experiments were not designed to easily reveal the extent of intra-speaker variability, since each participant saw at most 3 examples of each condition. We have made some preliminary attempts to look at individual speakers, and to establish to what extent it is possible to identify dialect splits: for this we refer the reader to Hartmann and Heycock (2017, 2018a). Further, for logistic reasons we had to conduct most of our experiments on-line, so that they had to be of

[8]Clearly, for this kind of movement to be possible, it must be the case that this longer movement is not in violation of the PIC or any equivalent locality condition. There are already arguments in the literature that any impenetrability induced by v has to be modulated in some way to avoid constraints on locality that are empirically too strict; for example ruling out agreement with "low" nominatives in Icelandic of the type illustrated in (6) above. The proposal in Chomsky (2001) is to weaken the PIC so that the complement of a phase is only spelled out when the next phase head (C, in this case) is merged. An alternative, defended for example in Keine (2017) on the basis of evidence from Hindi-Urdu, is to reject the hypothesis that vP defines a phase at all. Note that also under an account along the lines of Béjar and Kahnemuyipour (2017), an agreement probe on T has to be able to reach a DP within the small clause in order to account for DP2 agreement.

[9]Note that Béjar and Kahnemuyipour (2018) have adjusted their approach to allow for DP1 to have accessible number features in the light of the results in Hartmann and Heycock (2018a); we will come back to why this is still not enough in section 5.3, see also Hartmann and Heycock (2018c).

limited length, and it was not possible for us to have the same participants do both the production and rating tasks. As pointed out by a referee, this is worth bearing in mind in the context of discrepancies that we found in some cases between production and rating data, discussed below.

The experiments were designed and run as parallel as possible to allow for cross-linguistic comparison of the overall patterns/effects, even though direct comparison of individual ratings and productions is not possible[10]. All studies were set-up online using the OnExp online software package[11]. Test sentences and fillers were presented one per screen in randomized orders per participants.

Participants were recruited via personal contacts for Faroese, Dutch, and Icelandic and we additionally used the mailing list "Onze Taal" for Dutch. Experiments for these three languages were run fully online. Participants could sign up to take part in a lottery for a gift voucher after having finished the online study. The studies on German were all run on-site at the University of Tübingen, with individual payment for participation. All participants had to state their mother tongue(s), we only included the data of the participants who declared themselves to be native speakers of the language we were investigating (none of the participants reported themselves as bilinguals). As the test sentences were distributed across various lists, we analyzed the data with roughly an equal number per list per experiment. Per study we had between 8 and 15 participants per list, which adds up to between 50 and 90 participants per study.

For the production studies, participants were presented with sentences with a blank in one position of the sentence, as in (34), which participants were asked to fill with a single word of their own choosing.

(34) Der Psychologe fragte, ob das Problem die Eltern ___.
the psychologist asked if the problem the parents ___
'The psychologist asked if the problem ___ the parents' GERMAN

Before the actual study started, participants went through a short practice phase. All studies included fillers, between 1.5 and 2 times as many as the actual test sentences.

In analysing the data, we excluded all cases where participants used a verb other than the copula. All included cases were coded for number and person agreement on the copula, and then as DP1 and DP2 agreement (plus number-only DP2 agreement in Icelandic, see below). For the statistical analyses, we calculated relative frequency of DP1 agreement. These values (f) were transformed as usual, i.e., arcsine(square-root(f))—and we calculated planned contrasts with participant (F1) or item (F2) as random factors. Where appropriate, we also looked at the variation within and between speakers in more detail.

The rating studies followed the Thermometer Rating task model described in Featherston (2005), a variant of the Magnitude Estimation technique (Bard et al., 1996). Participants are asked to rate the naturalness of a sentence in relation to two reference sentences. The reference sentences are provided with a fixed score: one, a rather natural sentence, is assigned the value 30, one, a less natural sentence, is assigned the value 20. Reference sentences were kept on the screen throughout the experiment; stimulus clauses were presented one at a time, with participants advancing to the next by button press, with no possibility to return to earlier screens. Participants were asked to rate the naturalness of individual examples by providing numerical scores (all positive numbers) for individual sentences. As with the Magnitude Estimation technique, this allows participants to make finer grained judgments and to make distinctions between more or less unacceptable sentences. Before presenting the study, participants went through two short practice phases: the first one gave participants practice in assigning a value to the length of a line in reference to two standard lines assigned the values 30 and 20. Then they practiced rating naturalness with a set of sentences that varied in naturalness, so that they could familiarize themselves with the task.

The resulting scores for the rating experiments were all z-transformed (including fillers) per participant in order to normalize for the different scales participants might still have used. Z-scores were aggregated within conditions for each participant (F1) or item (F2). Where possible and useful, we computed the difference between DP1 and DP2 agreement for participants or items by subtracting DP1 z-scores from DP2 z-scores (positive values indicate that DP2 agreement is overall rated higher; negative values indicate that DP1 agreement is overall rated higher). This procedure allows us to investigate the same contrasts for the production and the rating studies. Depending on the design and more specific goals of each study, we also analyzed the rating data independently from the production data using ANOVA and mixed effect models (see the respective papers for details), especially when considering potential correlations with other factors, or speaker-groups.

All rating studies included a range of different filler sentences, including a set of standard-setting sentences, which help to put the overall acceptability into perspective (along the lines of the ideas presented in Gerbrich et al., 2019).

To test for number agreement with either DP1 or DP2 we used singular DP1 and plural DP2, corresponding to (35a)[12]; to test for person agreement we used singular DP1 and non-3rd person DP2, corresponding to the English example in (35b) (note that in all the

[10]There are two reasons why direct comparison in one model is not easily possible: first, absolute ratings might differ because the reference sentences used for the different languages might not be completely equivalent in their acceptability, so that scales might differ between languages. Second, there are morphological differences between the languages which make it impossible to test all the same conditions for all four languages.

[11]This package was developed by E. Onea at the Göttingen Courant Research Centre "Text Structures" at Göttingen University, see https://onexp.textstrukturen.uni-goettingen.de.

[12]We indicate in the tables whether DP2 was a plural definite, or a plural pronoun. As mentioned in section 3.2, when DP1 and DP2 differ in number, it is always DP1 that has to be singular. This might lead to the conclusion that what appears to be DP1 agreement is actually "default" singular agreement. We address this issue in section 4.2

Germanic languages that we investigated other than English, the pronouns for 2nd person singular and plural are distinct):

(35) a. The problem is your parents/they.
b. The problem is you.SG/you.PL.

While keeping the studies in the different languages as similar as possible, we needed to make adjustments in the design of the experiments due to language specific differences in the morphology of the copula; e.g., Dutch and Faroese do not make person distinctions in the plural, while Icelandic and German do; Icelandic and German on the other hand have syncretic forms for 1/3sg (in both present and past tenses in Icelandic, only in the past tense in German). Additionally, the pronominal forms in Dutch and German have a syncretic form for 3sg feminine and 3pl (*zij* and *sie* in Dutch and German, respectively). In order to avoid ambiguity in production and rating, where we needed a 3rd person plural we could not use a pronoun in these languages but rather had to use a nonpronominal DP.

It was mentioned earlier that individual examples of binominal copular sentences—particularly taken out of context—may be ambiguous, or simply indeterminate. We note here the principal ways we sought to avoid this in our materials:

- In the experiments reported here, DP1 was usually headed by a singular non-animate abstract noun [option II.ii in (27) above] like *reason/cause*, *problem*, *hope*, *inspiration* etc.: e.g., *The reason for the delay* BE *their friends*, except when DP1 needed to be plural (which is not possible with all of these abstract nouns). In the right context it is certainly possible for speakers to use even such abstract nouns to refer to individuals in a kind of metonymy (e.g., *The reason for the delay just walked into the room*), so in principle a copular clause like *The reason for the delay is my husband* is ambiguous between a specificational reading (≈ My husband caused the delay) and a predicational one (≈ The reason for the delay is related to me by marriage), but to our ears the predicational reading is much less readily available.
- Additionally, we did carry out experiments that included conditions where DP1 was headed by a noun denoting some kind of role [the option described in II.i in (27) above], like *the most likely winner(s)*, *the only witness(es)*, *her favorite drinking companion(s)*. This is the type of DP1 that seems most likely to create ambiguities, as DPs like *the winner* can more easily be used to refer to individuals than DPs like *the problem*. This "role" type of DP1 was used almost exclusively when we wanted to test the possibilities for agreement when DP1 is plural, since they are more natural in the plural than the DPs headed by many abstract nouns. For example, we take (36b) to be more natural than (37b):

(36) a. The most likely winner is Marta.
b. The most likely winners are Marta and Nina.

(37) a. Our only hope is Marta.
b. #Our only hopes are Marta and Nina.

In most cases where we used these "role" type DP1s, however, DP2 was a pronoun (and hence unlikely to be given a predicative interpretation, as mentioned above). Further, in our production experiments in Icelandic and German we did a direct comparison between a condition where DP1 was headed by a "role" type noun (e.g., *The most likely winner* ___ *you*.SG), and a condition where DP1 was headed by an abstract noun (e.g., *The main problem* ___ *you*.SG), and we found no difference in participants' choice of agreement on the copula in the two conditions (Hartmann and Heycock, 2017, pp. 249–261).

- In all conditions where we were testing for the distribution/acceptability of **person** agreement in SCCs, DP2 was a pronoun (*The reason for the delay* BE *you*). 1st and 2nd person pronouns have a predicative use only in very restricted circumstances (see e.g., Percus and Sharvit, 2014), so all such cases are highly unlikely to get a predicational construal.
- When we were testing for the distribution/acceptability of **number** agreement in SCCs, DP1 was always singular and DP2 plural. This also strongly disfavors a predicational reading (and, clearly, also an equative one). Consider the examples in (38):

(38) a. The source of the rumor was my favorite violinist.
b. The source of the rumor was my favorite violinists.

(38a), where both DPs are singular, can with some considerable effort get a predicational interpretation, where "the source of the rumor" is taken to be a very indirect way to describe an individual (*the source of the rumor—my friend Michael—was my favorite violinist*). But (38b) clearly cannot get such an interpretation: its only interpretation is as an SCC.

All four languages that we discuss are Verb Second (V2) in root clauses: thus in a root clause, the initial position is not reserved for subjects. This creates possible confounds that do not arise in English, where topicalization of a predicative NP is not string-identical to the specificational order, as show in (39)[13]:

(39) English
a. [$_{TP}$ The culprit is John]
SPECIFICATIONAL SENTENCE
b. [$_{CP}$ The culprit$_i$, [$_{TP}$ John is t$_i$]
PREDICATIONAL SENTENCE + TOPICALIZATION

We discuss in section 4.1 how we avoided this confound.

The production and rating studies on number included root clauses as well as embedded clauses in order to evaluate the possible effect of embedding (that is, of non-V2 vs. V2 structures), the experiments considering person only included embedded contexts in Icelandic, Faroese, and Dutch. In the German production study for person, we tested root clauses.

[13]Even in English there can be ambiguities in root clauses due to the other A′ predicate fronting construction illustrated in (18) above.

Based on our work on production of number agreement in Faroese and Icelandic (Heycock, 2009, 2012; Hartmann and Heycock, 2016, 2017), rating of person agreement in Icelandic (Hartmann and Heycock, 2018d), rating and production of person agreement in Faroese (Hartmann and Heycock, 2018e), rating and production of number agreement in Dutch and German (Hartmann and Heycock, 2018b) and rating and production of person agreement in Dutch (Hartmann and Heycock, 2019), we have arrived at the following answers to the questions raised in (33):

(40) a. All four languages allow DP2 agreement both for number and for person. We see this in the production data where varying numbers of participants provide DP2 agreement forms. In rating, we see an overall advantage for DP2 agreement in Dutch, Faroese and German (with the exception of XP-initial orders, see (40f) below) In Icelandic DP2 agreement is produced more frequently than DP1 agreement, but overall it is rated lower; nevertheless there are a small number of speakers who consistently rate DP2 agreement higher than DP1 agreement.

b. For a subset of speakers of Dutch, Icelandic and Faroese, DP1 agreement is also an available option. This can be seen in the production data, both for number and person. In the rating data, only a small number of speakers prefer DP1 agreement in Dutch and Faroese, while in Icelandic, DP1 agreement is available and preferred by a larger number of speakers.

c. We find considerable variation within languages, including intra-speaker variation. German shows the least variation, with speakers showing an overall higher score for / higher number of productions of DP2 agreement, to a high degree of consistency.

d. We find no evidence that default agreement is ever possible.

e. A subset of Icelandic and Faroese speakers show evidence that Person and Number are not only separate probes, but distinct heads, as argued already for Icelandic, for different reasons, by Sigurðsson and Holmberg (2008). In Icelandic we see this directly in production, in a pattern of "Number-only" DP2 agreement. In Faroese, we provide indirect evidence from a comparison between production and rating data.

f. In XP-initial V2 orders DP1 agreement is the most commonly produced option; in fact used almost to the exclusion of DP2 agreement in this context in all but German.

We will discuss the patterns (40a), (40d), and (40f) that all four languages share in section 4, and the other patterns, which we find only in a subset of the four languages, in section 5.

4. SHARED PATTERNS OF AGREEMENT ACROSS DUTCH, FAROESE, GERMAN, ICELANDIC

Our investigations of copular clauses in Dutch, Faroese, German and Icelandic show that all four languages share two patterns of agreement. First, all four languages allow agreement with DP2 both in V2 clauses, but also—more significantly, given the issues of ambiguity in V2 structures outlined in the last section—in non-V2 contexts. None of the languages we investigated show evidence for default agreement. Additionally, we find that all four languages show high levels of use of DP1 agreement in adjunct/modifier-initial V2 structures (that is, root clauses that have the order XP–*be*–DP1–DP2). We discuss the three patterns in the following subsections.

4.1. DP2 Agreement

In all four languages we found a high level of production of DP2 agreement in root (V2) clauses. When DP1 was 3rd singular and DP2 3rd plural, in German root clauses DP2 agreement was virtually categorical (92%); the lowest rate in root clauses was 62% (Dutch). DP2 agreement was also robustly attested in all four languages in embedded clauses, although at lower rates for all but Dutch, ranging from 88% for German to 46% for Faroese[14].

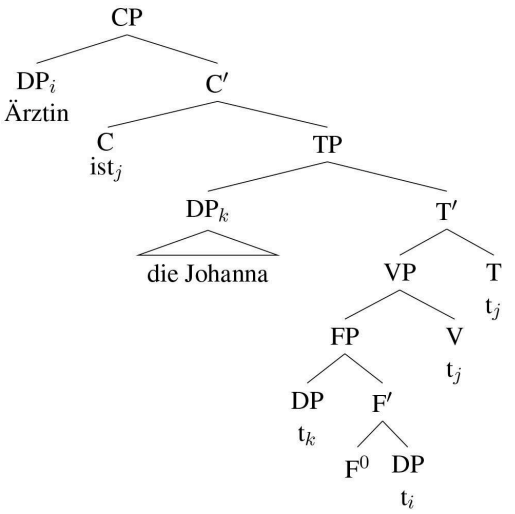
The possibility of DP2 agreement is especially interesting from a theoretical point of view when it comes to person agreement: as outlined above in section 2, person agreement has been claimed to be universally restricted to be impossible downwards (see most prominently Baker's *Structural Condition on Person Agreement*, Baker, 2008, p. 52, discussed below in section 5.1). DP2 agreement in person in SCCs would therefore constitute an important counterexample to the universality of this claim.

In order to be able to show that agreement with DP2 is indeed downward agreement, however, we need to make sure that DP2 is indeed in a position that is below the agreement probe. This issue comes up in different environments in the languages we tested. As all four are V2 languages, the position of DP2 in root V2 clauses is not necessarily a position below the agreement probe: given den Besten's widely adopted analysis that V2 orders involve the verb in a high position in the left periphery (e.g., the Complementizer position) and the initial XP in the specifier of that position, one derivation for a clause like German (41) would have the initial nominal topicalising from a low predicate position, and the second nominal occupying a position above the agreement probe in T.

(41) a. Ärztin ist (nur) die Johanna. GERMAN
doctor is only the.NOM Johanna
'Only Johanna is a doctor.'

[14]We tested agreement with a 3pl DP2 in all four languages. Agreement with 1/2 person DP2 in root clauses was tested only in German, see below.

b.



In the example in (41) it is clear that this is not a specificational sentence, rather the initial nominal is actually a predicate that has presumably reached the initial position in the clause by A′-movement, given that it is a bare (determinerless) nominal of a type that typically cannot appear in subject position (compare for example the ungrammatical case in (42), which is unacceptable because the bare NP *Ärztin* cannot function as the subject of *reich* '*rich*', just like e.g., *mayor* or indeed *doctor* in English):

(42) *Ärztin ist nie reich. GERMAN
doctor is never rich
'Doctor is never rich.'

The issue is potentially more complicated with the specificational structures we are dealing with, however. Because, as discussed above, definites can also get a predicative interpretation (as in e.g., *Joan is the best-paid psychiatrist in Europe* and its translation equivalents), root sentences corresponding to English (39b) on the one hand, which involve predicate topicalization (A′ movement) and to (39a) on the other, which instantiate a specificational structure, are string-identical in the languages we looked at.

It is however possible to establish that the DP2 agreement that we find in specificational sentences in Faroese, Icelandic, Dutch and German is not simply reducible to the result of the kind of A′ predicate fronting + V2 illustrated in (41a), although the situation is most difficult in German. The type of A′ predicate fronting just discussed is generally taken to be a root phenomenon (Heycock and Kroch, 1998; Heycock, 2012), and in a V2 language is thus expected to pattern together with V2 order. In the SOV languages Dutch and German, the embedded clauses in our materials were all verb-final, and hence unambiguously had no possible parse as embedded V2. In the SVO languages Faroese and Icelandic, embedded V2 has been shown to be possible in environments that allow "embedded root phenomena," but it is not freely available across all clause types. In order to avoid the confound of a parse as embedded V2, our materials therefore had the copular clause as an embedded interrogative (introduced by the equivalent of *whether*), as interrogatives are known to be the least favorable environment for embedded V2 (see e.g., Thráinsson, 2007 p. 44 for Icelandic, Heycock et al., 2010 for Faroese and Icelandic). In some of our experiments/conditions for Faroese and Icelandic we added a further control, namely the inclusion of sentential negation. As discussed extensively in Mikkelsen (2002) for Danish, negation in Faroese and Icelandic occupies a position somewhere at or just above the left edge of the VP. Hence, if DP2 follows negation, it must be in a low position, not in the specifier of TP. We tested the influence of negation in the Faroese and Dutch production studies, with the result that negation did not affect agreement patterns (see Hartmann and Heycock, 2018e, 2019 for details), supporting the assumption that in these embedded clauses the first DP below the complementizer is indeed parsed as the subject, and DP2 as occupying a lower position[15]. We did not include negation in all of our experiments/conditions, since the examples were already fairly complex and adding negation to the interrogative adds further complexity.

While, as mentioned above, it is straightforward in the SOV languages Dutch and German to construct clauses that are unambiguously non-V2, these languages present a different issues in that they both, in some circumstances, allow "scrambling" to front a constituent all the way to the left of a subject in an embedded clause, and it has been argued that such scrambling is also an instance of A′ movement (see Neeleman and van de Koot, 2002, 2008; Grewendorf, 2005; Frey, 2010 for discussion and references). For these two languages, then, such a derivation might potentially be another confound. In Dutch this is not a serious concern. Such scrambling is much more restricted than in German: this kind of A′-scrambling requires a very specific context and is not even accepted by all speakers (see Neeleman and van de Koot, 2002, 2008). Additionally, in some of the conditions/experiments we added negation (which marks the edge of the VP) to further support the parse in which DP2 is in a low position. In German, scrambling of an object across a subject is more freely available than in Dutch, making a parse in which DP2 occupies the subject position (above T) with DP1 in a higher A′-position possible. In such a structure, DP2 agreement is not necessarily downwards agreement. However, A′-scrambling of a non-referential DP1 across DP2 in a copular clause is also an information-structurally restricted option even in German, i.e., it only occurs in contrastive/focus/emphasis contexts (see Frey, 2010 for discussion and references). As the sentences were presented out of the blue without such a context, it is unlikely that participants ended up analysing our test sentences as cases where a predicate has undergone A′-movement to a high position in the left periphery. Therefore, while acknowledging that we cannot completely exclude the possibility of an alternative parse, we think that DP2 agreement

[15] A referee notes however that since the examples with and without negation were presented within the same experiment so that the presence/absence of negation could be treated as a variable, there could be an effect of structural priming from the negative sentences to the non-negative ones. To the extent that this kind of priming may have occurred, we cannot be certain that the non-negative sentences would be parsed in the same way in the absence of this hypothesized effect.

TABLE 1 | Production of DP2 agreement in % (DP2 is a non-pronominal, full DP).

Context	DP2		Dutch	German	Faroese	Icelandic
Main Clause	3pl	DP	62%	92%	64%	74%
Embedded Clause	3pl	DP	70%	88%	46%	66%

TABLE 2 | Production of DP2 agreement in % (DP2 is a pronoun).

Context	DP2		Dutch	German	Faroese	Icelandic
Embedded Clause	2sg	Pronoun	97%	(99%)*	12%	48%
Embedded Clause	2pl	Pronoun	98%	(98%)*	66%	68%**

German data is based on main clauses.

***This includes cases of 3pl (number-only agreement) and 2pl (full agreement in both number and person) marking: see section 5.3 for this distinction.*

TABLE 3 | Rating advantage of DP2 agreement (z-scores).

Context	DP2		Dutch	German	Faroese	Icelandic
Embedded Clause	3pl	DP	0.80	1.11	(0.22)*	–
Embedded Clause	2pl	Pronoun	0.35	0.94	0.33	−0.48

**Faroese materials used 3pl pronoun rather than full DP.*

in German SCCs is also most plausibly taken to be an instance of downwards agreement.

In sum, we think that the materials we tested across the four languages indeed represent structures in which DP2 occurs in a position below the agreement probe. In this light let us now turn to the results.

In all four languages, in non-V2 contexts, DP2 agreement occurred (in production) and was judged relatively acceptable in the rating tasks, although as we have documented, the rates of production and the degree of acceptability was not the same across languages. We will discuss the differences in section 5 and concentrate on the shared production results here.

As mentioned at the beginning of this section, speakers of all four languages produced DP2 agreement when the two DPs were mismatched for number in a V2 clause like (43a). Crucially, this was also the case in embedded contexts, (43b). The rates of production for examples like (43) are given in **Table 1**. Note that in all the tables, where we give information about DP2, "3pl DP" means that the nominal in question is a "full" (non-pronominal) plural DP.

(43) a. Das Problem ___ die Eltern. GERMAN
the problem ___ the parents
'The problem ___ the parents.'

b. Der Psychologe fragte, ob das Problem die
the psychologist asked if the problem the
Eltern ___.
parents ___
'The psychologist asked if the problem ___ the parents'

While there are differences between number and person (see section 5) all four languages also show the production of DP2 agreement when DP2 is non-third person as in (44), see **Table 2**[16].

(44) a. …dat het echte probleem jij ___ DUTCH
…that the real problem you.2.SG ___

b. …dat het echte probleem jullie ___
…dat the real problem you.2.PL ___

So from the production data, we conclude that DP2 agreement is a viable option for at least some speakers, in all four languages. The production data show, however, that there is significant variation with respect to the extent to which DP2 agreement is a possible or preferred option. This is also reflected in the rating data that we obtained for DP1 and DP2 agreement both for number agreement (with DP1 and DP2 differing in number only) and person agreement (where DP2 is a 1st or 2nd person pronoun). See **Table 3**[17,18]. In this table we see that there are also differences between languages in the rating advantage of DP2 agreement. So while the difference in German is 0.94 in embedded clauses with 2pl pronouns, in Faroese it only reaches 0.33 for the same case. In Icelandic on the other hand, overall, speakers rated DP1 agreement higher than full agreement (in both person and number) with DP2[19]. We nevertheless find a small group of speakers who rate DP2 agreement over DP1 agreement (see Hartmann and Heycock, 2018d), supporting the conclusion that DP2 agreement for number and person is possible for at least some varieties of Icelandic (for differences in the extent to which DP1/DP2 agreement is possible in all four languages see section 5).

Thus the rating data also illustrates our point that the languages under consideration differ in how much variation they exhibit[20].

[16]Note that in this case we tested V2 clauses in German; embedded clauses for the other three languages.

[17]A referee suggests that the variability in agreement within the results for a single language might be due to speakers' varying in their interpretation of the sentences in the materials as SCCs or as other types of copular clause (predicational or equative). We have outlined in section 3.3 some ways in which we tried to eliminate or at least reduce this confound in our examples. We can now see from these results that the extent of variability is different in the different languages (for example, even if we consider only the data from root clauses in **Table 1**, German is much less variable than any of the other languages). As the items we used in our material are as similar as possible across the experiments in the different languages, this between-language difference in the extent of variation suggests that at most a small amount of this variation could be due to individual variation in construal.

[18] In this table we give the differences in z-scores for each language, rather than absolute values, in order to abstract away from other differences between the languages (see Hartmann and Heycock, 2018b for detailed discussion). So here the "advantage" of DP2 agreement is the figure that results from subtracting z-scores for DP1 agreement from z-scores for DP2 agreement.

[19]Note that we did not run a comparable experiment for differences with number mismatch in Icelandic. Observe also that in this judgment task in Icelandic, the ratings for DP2 agreement with a 2nd plural pronoun that contribute to the figure here were for examples showing full Number+Person agreement; this may be part or all of the reason why DP2 agreement here appears to be disfavoured in comparison with its frequency as shown in **Table 2**.

[20]Again, our point here is that variation exists, the distribution of this variation between and within speakers is the next step to take.

4.2. Lack of Default Agreement

As discussed above, in a specificational sentence DP1 is always 3rd person. Because in many cases mismatches in ϕ-features can only be tested with DP1 being singular and DP2 being plural or non-3rd person, one might be tempted to analyse what we have been calling DP1 agreement rather as "default" 3rd person singular, or lack of agreement with any DP. However, there are configurations where it is possible to tease these possibilities apart. We set up such cases and found that, wherever we were able to test, default is not in fact an option in SCCs. This seems clear in English: if DP1 agreement in (45a) were default agreement, we would expect a 3sg copula also with two plural DPs, but this is sharply ungrammatical (see already Heycock, 2009 for English and Faroese).

(45) a. The cause of the riot {was/*were you}
b. Her favorite authors {*is/are} Heller and Fielding.

Initial informal evidence suggested already that default is not grammatical in SCCs in all the four languages in our studies. We included this configuration in our rating studies to get a value for a clearly ungrammatical agreement pattern with SCC (a baseline), see the examples in (46). The results showed indeed that default is ungrammatical: in Icelandic and Dutch default is significantly worse than any of the other conditions tested; in Faroese default is numerically but not significantly worse than DP1 agreement; in German, which is a consistent DP2 agreement language, default agreement is as bad as DP1 agreement, see **Table 4**, illustrated with respective examples in (46).

(46) a. *Þau voru að velta fyrir sér hvort líklegustu
they were wondering whether likeliest
sigurvegararnir væri ekki þið.
winners be.3.SG not you.PL
'They were wondering whether the most likely winners wasn't you.PL' ICELANDIC
b. *Tey ivaðust í, um teir trúligastu sigursharrarnir
they wondered if the most likely winners
ikki er tit.
not be.3.SG you.PL
'They wondered whether the most likely winners isn't you.PL' FAROESE
c. *De leraar zegt dat de huidige problemen niet
the teacher says that the current problems not
de ouders is.
the parents be.3.SG
'The teacher says that the current problems is not the parents.' DUTCH
d. *Die Nachbarin fragte, ob die Auslöser
the neighbor.F.SG asked, whether the triggers.PL
des Streits Ihr war.
for the dispute you.PL be.3.SG
'The neighbor asked whether the triggers for the dispute was you.PL' GERMAN

This further supports the conclusion from English that 3rd singular agreement with a 3rd singular DP1 is DP1 agreement, not default. Thus, any analysis in which DP1 does not have ϕ-features to be agreed with, see e.g., Béjar and Kahnemuyipour (2017), cannot be generally upheld.

TABLE 4 | Ratings in z-scores for default agreement in contrast to DP1 and DP2 agreement.

	DP1	DP2	Verb	Dutch	German	Faroese	Icelandic
Default	3pl	2pl	3sg	−0.78	−1.13	−0.79	−0.66
DP1	3sg	2pl	3sg	−0.67	−1.18	−0.73	0.34
DP2	3sg	2pl	2pl	−0.32	−0.24	−0.40	−0.14

TABLE 5 | Production of DP2 agreement in % in different V2 (root) contexts.

Context	DP2		Dutch	German	Faroese	Icelandic
DP1-initial V2 Clauses	3pl	DP	62%	92%	64%	74%
XP-initial V2 Clauses	3pl	DP	8%	29%	4%	2%

4.3. DP1 Agreement in XP-Initial V2 Contexts

A further pattern that all four languages share is a specific effect in V2 contexts which we relate to C[omplementizer]-agreement. When considering adjunct-initial root clauses like (47) (which we will refer to as XP-initial V2 clauses in the tables below) in the four languages we discuss, we see that the production rate of DP2 agreement drops significantly in all four languages. This drop is especially striking for German, where we otherwise found a rather stable and strong preference in production for DP2 agreement across all other contexts we tested.

(47) Meiner Meinung nach **?ist/?sind** das eigentliche
my opinion after is/are the real
Problem deine Eltern. GERMAN
problem your parents
'In my opinion the real problem is your parents'

This difference also shows up in acceptability ratings, although the effect seems less dramatic, for reasons that we do not yet understand. As can be seen in **Table 5**, in the production data the adjunct-initial V2 order results in a "flip" from DP2 agreement to DP1 agreement being the most frequently produced order in all four languages (for all but German in fact the production of DP1 agreement in this order is close to categorical). **Table 6** shows the extent to which DP2 agreement is rated higher than DP1 agreement in the judgment task in three different environments, including the adjunct-initial V2 order, In the rating data for Dutch, the "flip" in production corresponds to slightly—but significantly—higher ratings for DP1 agreement over DP2 agreement in the adjunct-initial order only. In the German rating data, the advantage of DP2 agreement in adjunct-initial root clauses is significantly reduced compared to the advantage of that agreement in embedded clauses and DP1-initial main clauses, so that in adjunct-initial V2 clauses there

TABLE 6 | Rating advantage of DP2 agreement (z-scores) for Dutch and German.

Context	DP2		Dutch	German
Embedded clauses	3pl	DP	0.48	0.57
DP1-initial V2 Clauses (Root)	3pl	DP	0.29	0.90
XP-initial V2 Clauses (Root)	3pl	DP	−0.30	0.12

is no significant difference between the ratings for DP1 and DP2 agreement[21].

Evidently, in these V2 languages one effect of the XP-initial order is that the order of the finite verb (in the cases we tested, always the copula itself) and the DPs becomes V_{fin} < DP1 < DP2. An initial hypothesis might therefore be that the increased advantage for DP1 agreement here is some kind of performance effect tied to the linear order. However, whether or not there is a performance effect contributing to the increased production/acceptability of DP1 agreement, it does not seem likely that this is an effect which produces DP1 agreement *in a system in which DP1 agreement is ruled out by the syntax*. This is particularly relevant for German, where our results based on other configurations suggest that DP1 agreement is essentially ungrammatical. Consider, for example, that given the relatively free word order of German, it is possible for the DP immediately following the finite verb in an XP-initial non-copular sentence to be the object, even though the "default" order would be for this position to be occupied by the subject:

(48) a. Heute holt der Vater die Tochter ab.
today pick.3.SG the.NOM father the daughter up
Today the father picks up his daughter.
b. Heute holt die Tochter
today pick.3.SG the daughter.NOM/ACC.SG
der Vater ab.
the.NOM father up
Today the father picks up his daughter.

As there is syncretism between feminine singular nominative and accusative, in (48b) it is only when the unambiguously nominative singular DP *der Vater* is reached that it becomes evident that *die Tochter* cannot be the subject. One can reproduce such a structure with plural noun phrases, which are equally syncretic for nominative and accusative. If there is a performance effect that induces agreement with the first DP following the finite verb in second position, we would expect to find that examples like (49) are both produced and judged grammatical:

(49) *Heute holen die Töchter
today pick.3.PL the daughters.NOM/ACC.PL
der Stefan ab.
the.M.NOM Stefan up
Intended: Today Stefan is picking the daughters up.

In the judgment of the German-speaking author of this paper, (48b) rather has the effect of a garden-path sentence, and (49) is simply unacceptable. This seems quite different to what is observed with the XP-initial copular clauses with DP1 agreement.

We have therefore pursued a different approach to explaining agreement in the adjunct-initial order. Namely, we have argued that what we observe here is not regular subject-verb agreement, associated with an agreement probe in T; instead we propose that here the agreement on the verb is in fact the exponent of a probe on C, which agrees with the closest DP in its c-command domain. As mentioned above in section 2, it has been known for some time that complementizers in Germanic sometimes also carry agreement features. The case mentioned above was so-called complementizer agreement, which is particularly associated with West Germanic varieties, see Bayer (1984), Ackema and Neeleman (2004), van Koppen (2005), and van Koppen (2017) among many others. An example from a Dutch variety was given above as (8), (50) is a further example, this time from Flemish, where the complementizer *dat* in (50) is inflected for number and person in agreement with the subject that immediately follows it.

(50) K peinzen dat-n die studenten nen buot gekocht
I think dat-3.PL those students a boat bought
ee-n FLEMISH
have-3.PL
'I think that those students have bought a boat.' (van Koppen, 2017, p. 2)

Note that this type of complementizer agreement only obtains when the subject immediately follows the complementizer in the linear order.

A second type of C-related agreement occurs in cases of so-called "inversion agreement" in Dutch. The distinct marking of the 2nd person singular in Standard Dutch is obligatorily omitted when the 2nd person subject immediately follows the finite verb in exactly the kind of adjunct-initial V2 structures where we find the unexpected high rates of DP1 agreement in our copular clauses (for a discussion of inversion agreement in other varieties of Dutch see Don et al., 2013 and references therein):

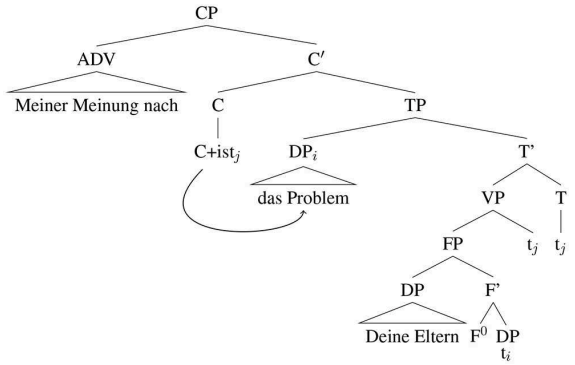
(51) a. dat jij dagelijks met een hondje over straat
that you daily with a doggy over street
loopt DUTCH
walk.2.SG
'that you walk with a doggy in the street every day'
b. Jij **loopt** dagelijks met een hondje over
you walk.2.SG daily with a doggy over
straat.
street
'You walk in the street with a doggy every day.'
c. Dagelijks **loop** jij met een hondje over straat.
daily walk you with a doggy over street
'Daily you walk in the street with a doggy.'

As just noted, both types of C-related agreement discussed in the literature only obtain when the DP immediately follows the C position (see the proposals in the literature referenced in van Koppen, 2017 on how this adjacency requirement can be implemented). This is exactly the configuration in which

[21] So far, we only have acceptability ratings for Dutch and German on this phenomenon.

we found increased production/acceptability of DP1 agreement in the specificational sentences. Given this parallel behavior, we analyse the significant increase in the use/rating of DP1 agreement in SCCs when the copula has moved to C and is immediately followed by DP1 as the result of a type of inversion agreement. The agreement probe on C—to which the finite verb has moved—probes downwards and finds the closest available target, which is DP1, as in (52)[22].

(52)



Thus, the agreement on the verb is the exposition of agreement of a person/number probe in the C-domain, whereas agreement in the T-domain is not expressed; here we have to assume that when both C-agreement and T-agreement conflict, but have to be realized on a single head (the verb), the conflict is resolved in favor of C agreement for most speakers[23]. In most of our languages this C-agreement is not usually visible, because in most sentences no differences between agreement in the T-domain and C-domain can arise with usually only a single nominative argument being present, see Hartmann and Heycock (2018b) for details.

5. FINE-GRAINED DIFFERENCES: NUMBER VS. PERSON

5.1. Background on Person vs. Number Agreement

Work on agreement especially in the last 15 years has drawn attention to the fact that agreement for person and agreement for number are not fully parallel. This topic is explored in depth in Baker (2008), Preminger (2011), Preminger (2014), and Ackema and Neeleman (2018) among many others (see in particular the references in Ackema and Neeleman, 2018). In general, person agreement is more restricted (in the terms of Baker, 2008 more "fragile") than number agreement. In the Germanic family that we are concerned with here, the most prominent case comes from Icelandic. As mentioned earlier, Icelandic has a number of verbs whose subject has to appear with dative case-marking[24]. Some of these verbs are transitive, and have their lower argument appear in the nominative case; another class consists of verbs that select for non-finite clauses of one type or another, with the subject of the embedded non-finite clause again appearing in the nominative. Strikingly, the finite verb may agree in number with the low nominative argument (as seen in the (a) examples), but not in person (as seen in the (b) examples).

(53) a. Honum virtust þær (vera) duglegar. ICELANDIC
him.DAT seemed.3.PL they.F.PL.NOM (be.INF) industrious.F.PL.NOM
'They seemed industrious to him.'

b. *Honum virtumst við (vera) duglegar.
him.DAT seemed.1.PL we.NOM (be.INF) industrious.F.PL.NOM
Intended: 'We seemed industrious to him.'

(54) a. Henni líkaðu þeir. ICELANDIC
her.DAT liked.3.PL they.M.NOM
'She liked them.'

b. *Henni líkaðir þú.
her.DAT liked.2.SG you.NOM
'She liked you.'

As noted earlier with respect to (6), an additional complexity here is that "default" 3rd singular agreement rescues examples like (53b), where the nominative argument is not an argument of the higher clause, but does not have the same effect—at least for many speakers—on examples like (54b):

(55) a. Honum virtist við (vera) duglegar.
him.DAT seemed.3.SG we.NOM (be.INF) industrious.F.PL.NOM
Intended: 'We seemed industrious to him.'

b. *Henni líkaði þú.
her.DAT liked.3.SG you.NOM
'She liked you.'

There are a range of suggestions as to how to account for such restrictions in general, and the Icelandic case in particular. One prominent approach is that of Baker (2008), where it is claimed that while number agreement can obtain "at a distance," this is ruled out for person agreement, which can only be established

[22]In this tree structure we assume that (i) German has a TP and (ii) that DP1 in SCCs moves to Spec,TP, neither of which is crucial for our analysis. At the same time, both assumptions are under discussion: e.g., Haider (1997) and Sternefeld (2009) have argued against the presence of T in German. If T is absent, the agreement probe remains on the highest verbal projection. In both approaches, with or without T, the crucial aspect for our analysis is that DP1 moves directly above the agreement probe. Second, given that there is a TP in German, it has been extensively discussed whether subjects move into Spec,TP (see among others, Abraham, 1993; Haider, 1993; Alexiadou and Anagnostopoulou, 1998; Biberauer, 2004) or can just simply A-scramble to adjoin to TP. Again the two options are both compatible with our analysis as long as DP1 A-moves to a position above the agreement probe.

[23]In German, based on the data we obtained, there seems to be a small group of speakers who consistently prefer DP2 agreement in XP-initial clauses, however, there are many speakers whose ratings on both DP1 and DP2 agreement in this condition vary considerably.

[24]For evidence that the dative is indeed the subject in examples like these, see e.g., Zaenen et al. (1985) and Sigurðsson (1989).

via a specifier-head relation, as expressed more formally in the SCOPA:

(56) *Structural Condition on Person Agreement (SCOPA)*
A functional category F can bear the features +1 or +2 if and only if a projection of F merges with a phrase that has that feature, and F is taken as the label for the resulting phrase.
(Baker, 2008, p. 52)

An alternative family of proposals relates the "person effect" seen in this configuration in Icelandic to a constraint observed in combinations of direct and indirect object clitics in a number of languages, the Person Case Constraint. This type of proposal is built on two core ideas. First, it is argued that a significant set of cases where there is a "low" 1st/2nd person argument that gives rise to ungrammaticality [including Icelandic examples like (54) and (53)], a higher argument intervenes between the agreement probe and the 1st/2nd person argument, preventing agreement from being established with that lower argument. In the case of the Icelandic dative-nominative constructions, this intervening argument is the dative DP. Second, 1st and 2nd person pronouns have the special property that they need to be licensed via agreement with a relevant probe, see Béjar (2003) and Béjar and Rezac (2003) for key proposals, and Preminger (2014) for a recent discussion of Icelandic cases like the ones just presented. This special property of 1st/2nd person pronouns is summed up in the Person Licensing Condition (PLC) of Béjar and Rezac (2003).

(57) Person Licensing Condition (PLC)
Interpretable 1st/2nd-person features must be licensed by entering into an *Agree* relation with an appropriate functional category. (Béjar and Rezac, 2003)

This formulation is subsequently amended in Preminger (2011) to exempt person features in clauses—even small clauses—without person ϕ-probes, precisely to account for the grammaticality of examples like (55a):

(58) Person Licensing Condition (PLC)—Revised version
A 1st/2nd-person pronoun in the same clause as a person ϕ-probe must be agreed with by that ϕ-probe. (Preminger, 2011)

The net effect is that when some nominal intervenes between an agreement probe on some functional category and a 1st or 2nd person pronoun, the probe will fail to "reach" the pronoun (intervention) and the resultant lack of agreement will be fatal (PLC). More has to be said about why agreement for number with a "low" nominal (3rd person pronoun or nonpronominal DP) is possible even in the presence of an apparent intervenor; we leave this aside here, but see the cited works for details.

A third option for deriving the restrictions on person agreement also relies on the fact that such cases involve an "intervening" nominal, but assumes that agreement can be established with both DPs in such cases ("multiple agreement"). Ungrammaticality arises if there is no possible morphological exponent that is consistent with both of the agreement features that are copied onto the agreeing head. Thus it is argued that in (54b) and (53b) the dative argument triggers default (3rd person) agreement, and the nominative triggers 2nd or 1st person, respectively. The resulting conflict in feature values can however be resolved if there is a morphological form that happens to be syncretic for the two distinct values. Thus for example (59) is argued to be grammatical in Sigurðsson and Holmberg (2008) because for the verb *virðast* 'seem' there is syncretism in the plural between 2nd and 3rd person:

(59) Henni virtust þið eitthvað
her.DAT seemed.2/3.PL you.NOM.PL somewhat
einkennilegir. ICELANDIC
strange
'You.PL seemed somewhat strange to her.'

See Schütze (2003) (based on data from Sigurðsson, 1996) and Sigurðsson and Holmberg (2008) and Ackema and Neeleman (2018) for this kind of proposal for Icelandic Dative-Nominative structures[25].

All the proposals just listed have been argued to be general restrictions on person agreement. It then becomes relevant to ask whether their effects are evident also in SCCs—and not only in Icelandic. That is, we might expect that agreement with DP2 should be possible only for number, and not person; and that failure to agree with a 1st or 2nd person DP2 should result in ungrammaticality. And indeed such a claim is made for Dutch in den Dikken (2019), on the basis of his own judgments. However, our data suggest that other Dutch speakers show a different effect, as will be made clear in the next sections.

5.2. Person Agreement in SCCs

First, as reported above in **Table 2**, DP2 agreement in person is produced in all four languages, though to varying degrees.

German is in general quite consistent in having agreement with DP2, as shown in **Table 7**[26]. In addition, the rating data show a consistent overall higher rating for DP2 agreement in this language[27].

[25] In Hartmann and Heycock (2018d) we provide experimental data that supports the existence of an effect of syncretism in Dative–Nominative structures in Icelandic, and argue on that basis that the multiple agreement approach is the most plausible. The syncretism effect is however far from categorical, and we argue that syncretic forms are a repair mechanism which is not available to all speakers in the same degree.

[26] Please be aware that the data in the language specific summaries in **Tables 7–10** are based on three/four different experiments: the data in each box bounded by lines on top/bottom and left/right belong together. There are differences in the ratings across experiments, and absolute z-score values cannot be compared across experiments. We do find differences between experiments, which may result from the fact that we had different participants for the different experiments within a single language and/or that we used different fillers in the respective experiments. We have to leave the work of teasing apart these possibilities to future research.

[27] Note: the person agreement production data reported here on German is from root clauses, whereas in the rating study we used embedded clauses. Note also that in the first two rows of all the following tables, which show the production and rating of different agreement options when DP1 and DP2 differ only in number, DP2 was a full (lexically headed) DP, while in the other rows, where the mismatch is for person (plus number in some cases), it was necessarily a pronoun for 1/2 person. For third person, we indicate in the tables whether or not we had a full (non-pronominal) DP or a pronoun. As mentioned in section 3.3, we were not

TABLE 7 | Agreement patterns in SCCs in German.

Context	DP2		Production			Rating (z-scores)		
			DP1	DP2	%DP2	DP1	DP2	DP2 advantage
Main clause	3pl	DP	10	129	92%	−0.48	0.42	0.90
Embedded clause	3pl	DP	16	117	88%	−0.54	0.03	0.57
	2sg	Pronoun	1	158	99%	–	0.03	–
	2pl	Pronoun	3	131	98%	−1.18	−0.24	0.94
	3pl	DP	–	–	–	−0.73	0.48	1.11
			(Main clauses)			(Embedded clauses)		

TABLE 8 | Agreement patterns in SCCs in Dutch.

Context	DP2		Production			Rating (z-scores)		
			DP1	DP2	%DP2	DP1	DP2	DP2 advantage
Main clause	3pl	DP	69	113	62%	−0.54	−0.25	0.29
Embedded clause	3pl	DP	53	127	70%	−0.76	−0.28	0.48
Embedded clause	2sg	Pronoun	6	211	97%	−0.71	−0.30	0.41
Embedded clause	2pl	Pronoun	2	146	98%	−0.67	−0.32	0.35
Embedded clause	3pl	DP	–	–	–	−0.52	0.22	0.80
Embedded clause	3sg	Pronoun	–	–	–	−0.38*		–

**The final row shows the rating where DP2 is a 3rd singular pronoun, and agreement is 3rd singular.*

Dutch is clearly different from German in a number of respects: see **Table 8**. Observe that more Dutch speakers than German produced DP1 agreement in number (where DP2 was a plural non-pronominal DP, Dutch speakers produced DP1 agreement in 38% of root clauses, and 30% of embedded clauses; the corresponding figures for German are 8% and 12%). However, in both languages we tested embedded clauses with 2nd person pronouns (both singular and plural) as DP2: in these cases the rate of DP2 agreement in Dutch rises to match that of German. The rating studies reveal though that in Dutch ratings drop in general when DP2 is a personal pronoun, independent of whether there is syncretism with 3rd person in the verbal agreement, or indeed whether the pronoun is 2nd or 3rd person. Thus there is no difference in the ratings between the conditions in (60a) and (60b) and the very small difference to (60c) is not significant. On the other hand, there is a significant difference between the cases where DP2 is a pronoun (whether 2sg, 2pl, or 3sg) and those where it is a full DP, as in (60d). In German we also see a difference between the ratings for cases where DP2 is a full DP, and those where it is a pronoun (2sg or 2pl), with the full DP condition rated more acceptable overall. However, we do not have a direct comparison within a single experiment that compares a full DP with a third person pronoun as DP2. Thus, our data cannot be used to argue for a pronoun effect in German. Informal discussions with native speakers of German and Dutch seem to suggest that there is indeed a difference between Dutch and German in that focused pronouns in SCCs are problematic in Dutch, but not in German: this clearly requires further investigation.

(60) a. …dat het echte probleem niet jij
…that the real problem not you.SG
bent DUTCH
be.PRES.2.SG
z-score: -0.30

b. …dat het echte probleem niet jullie zijn
…dat the real problem not you.PL be.PRES.PL
z-score: -0.32

c. …dat het echte probleem niet hij is
…dat the real problem not he be.PRES.3.SG
z-score: -0.38

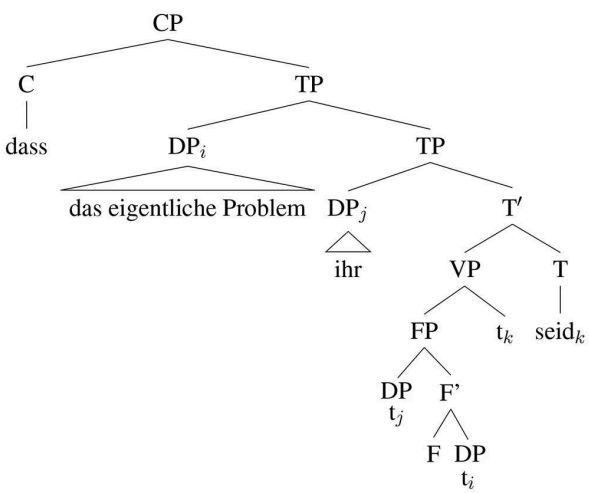
d. …dat het echte probleem niet de ouders
…that the real problem not the parents
zijn.
be.PRES.3.PL
z-score: +0.22

Thus, in our data, Dutch DP2 agreement in SCCs exhibits a pronoun effect and there is no evidence for a person effect.

From the perspective of whether or not downwards person agreement with a low nominative is possible in these languages (contra SCOPA), we need to look at the data quite carefully, and take into consideration independently known facts about these languages. German shows DP2 agreement both with number and person. On the face of it, this looks like a clear case of downwards person agreement into the VP. However, of the languages we are considering, German has the most "free" word order within TP; most relevantly here, object pronouns generally move out of the VP to the left edge of the "middle field", which could be—depending on the analysis of scrambling—outside the c-command domain of the agreement probe (presumably T, but see footnote 22), as in (62):

(61) …dass das eigentliche Problem ihr seid
…that the real problem you.PL be.PRES.2.PL

(62)



able to make an entirely minimal comparison in German because the 3rd person plural pronoun is homophonous with 3rd singular feminine (*sie* in both cases) and so could not be used.

If speakers parse our copular sentences with pronominal DP2 as involving such movement, these examples would not necessarily involve agreement into the VP, and could not be used as an argument against theories that treat downwards agreement for person as impossible (as e.g., Baker's SCOPA) However, the alternative parse with the nominative DP2 in a low position is at least equally plausible: leftward movement is not obligatory, and focused pronouns in particular tend to remain within the VP. In an SCC the second DP is obligatorily focussed, so it is at the least possible that the pronominal DP2 has indeed remained in a low position. The usual way to force such a parse is to include negation, which would precede an unmoved, "low" pronominal. While we did not include negation in our materials for German, informally elicited judgments from native speakers informants suggest that the presence of negation does not affect the preference for DP2 agreement in any way.

In Dutch, the production data in **Table 8** show that when presented with SCCs with a 2nd person pronoun as DP2, participants overwhelmingly chose to agree with the pronoun, regardless of whether this agreement was syncretic with 3rd person (97%–98%). Evidently, this is consistent with downwards person agreement being grammatical in this language. It is the case that, as we just saw, in Dutch, pronouns in general are less acceptable in the low position following negation and it could be objected that this production task does not allow us to determine whether speakers simply found this order unacceptable, and that they were making agreement choices for sentences that were ungrammatical for them. If we inspect the ratings data in the same table, however, we can see that while the ratings for the examples with pronouns were low, they were not at floor. Further, we find no additional effect of 1st/2nd person. That is, if there was a "person effect" on top of the pronoun effect, the ratings should be worse for 1/2 person pronouns than for 3rd person pronouns, but that is not what we see. Thus, we find no "person effect" in Dutch either, and we have evidence that person agreement with low nominatives is possible in this language.

We can strengthen this point by looking at the data in Faroese and Icelandic, in **Tables 9**, **10**, respectively. Both languages are VO, so DP2 is clearly in a low position in the embedded interrogatives that we tested, since it follows the verb. In production in Faroese, we see that native speakers produce DP2 agreement to a significant extent (see section 5.3 below, and also Hartmann and Heycock, 2018e, for discussion of why DP2 agreement appears to be produced at an unusually low rate just when DP2 is the 2nd singular pronoun). In the rating data, we see that they in general prefer DP2 agreement over DP1 agreement, though ratings in general are rather low for SCCs with pronominal DP2. In Faroese there is some evidence that the rather low ratings when DP2 is a pronominal is not due to a person effect, since the ratings when DP2 is a 3rd person pronoun are not significantly higher than the ratings when it is 1st or 2nd person. It could be that there is a "pronoun effect", as in Dutch, but to establish this would take further research, as the Faroese rating experiment did not include a condition with a non-pronominal DP2.

In Icelandic, we see that overall, DP2 agreement is preferred with number agreement, but not person agreement, which is also reflected in the rating data, where DP1 agreement is preferred over DP2 agreement. Despite this fact, a more detailed investigation into the data reveals is that there are still some speakers in Icelandic (though few in our sample) who consistently prefer DP2 person agreement over DP1 agreement: for this see Hartmann and Heycock (2018d).

In summary, we do not find any clear evidence in SCCs of the kind of "person effect" (ungrammaticality of "low" 1st/2nd person nominatives) that is present in Icelandic dative-nominative constructions. To the extent that these languages allow for pronouns to appear as DP2, agreement is possible regardless of person.

It is also relevant to consider the production and rating of DP1 agreement. In Icelandic, the production and grammaticality (at least for some speakers) of DP1 agreement when DP2 is a 1st or 2nd person pronoun constitutes an argument against the general applicability of the requirement for 1st/2nd person pronouns to be agreed with (the Person Licensing Condition (PLC) described in section 5.1 above). It is clear from the Icelandic data that DP1 agreement is a viable option and the preferred option for many speakers in our sample.

Finally in our data we did not find any evidence for a syncretism effect that could be taken as evidence for multiple agreement in SCCs (recall the discussion in section 5.1 of multiple agreement as an account of the person effect in dative-nominative constructions in Icelandic). In all languages we tested syncretic forms (German: 1/3 plural, Dutch 1/2/3 plural, Icelandic: 1/3 singular, Faroese: 1/2/3plural) as potentially providing evidence for multiple agree, but either we found that syncretism did not have a significant effect (German, Dutch, Icelandic) or that what looks like a syncretism effect in Faroese, akin to what is found in dative-nominative constructions in

TABLE 9 | Agreement patterns in SCCs in Faroese.

Context	DP2		Production			Rating (z-scores)		
			DP1	DP2	%DP2	DP1	DP2	DP2 advantage
Main clause	3pl	DP	18	32	64%	–	–	–
Embedded clause	3pl	DP	20	17	46%	–	–	–
Embedded clause	2sg	Pronoun	100	14	12%	–	–	–
Embedded clause	1sg	Pronoun	–	–	–	−0.56	−0.40	0.16
Embedded clause	2pl	Pronoun	39	76	66%	−0.73	−0.40	0.33
Embedded clause	3pl	Pronoun	51	54	51%	−0.66	−0.44	0.22

TABLE 10 | Agreement patterns in SCCs in Icelandic.

Context	DP2		Production			Rating (z-scores)		
			DP1	DP2	%DP2	DP1	DP2	DP2 advantage
Main clause	3pl	DP	50	139	74%	–	–	–
Embedded clause	3pl	DP	63	123	66%	–	–	–
Embedded clause	2sg	Pronoun	109	99	48%	–	–	–
Embedded clause	2pl	Pronoun	see **Table 11**			0.34	−0.14	−0.48
Embedded clause	3pl	Pronoun	74	143	66%	–	–	–
Embedded clause	3sg	Pronoun	–	–	–	0.43		

Icelandic, in fact has a ratings profile that requires a different explanation (see Hartmann and Heycock, 2018e, and below). Thus, multiple agreement does not arise where there are two nominative arguments in a single clause (though it is a viable analysis for the dative-nominative construction, as we argue in Hartmann and Heycock, 2018d)[28].

5.3. Person and Number Are Separate Probes in Icelandic and Faroese

There is one further aspect in the agreement domain in which we find variation in the languages under consideration that we wish to present here: Icelandic and Faroese show evidence that person and number are actually distinct heads. In Icelandic—where the distinction between the two probes has been made previously based on the pattern of agreement in dative-nominative cases, see Sigurðsson and Holmberg (2008)—we find direct evidence for this. In the production test in Icelandic we presented speakers with sentences where DP2 could differ from DP1 in number, in person, or in both, as illustrated in (63)[29]:

(63) Hann var að velta fyrir sér hvort … ICELANDIC
he was wondering if
'He was wondering whether …

a. aðalvandamálið ___ þeir.
main problem.DEF ___ they
'the main problem is them.'

b. aðalvandamálið ___ þú.
main problem.DEF ___ you
' the main problem is you.SG.'

c. aðalvandamálið ___ þið.
main problem.DEF ___ you.PL
'the main problem is you.PL.'

d. líklegustu sigurvegararnir ___ þið.
most likely winners.DEF ___ you.PL
'the most likely winners are you.PL.'

The choices made by the participants are tabulated in **Table 11**.

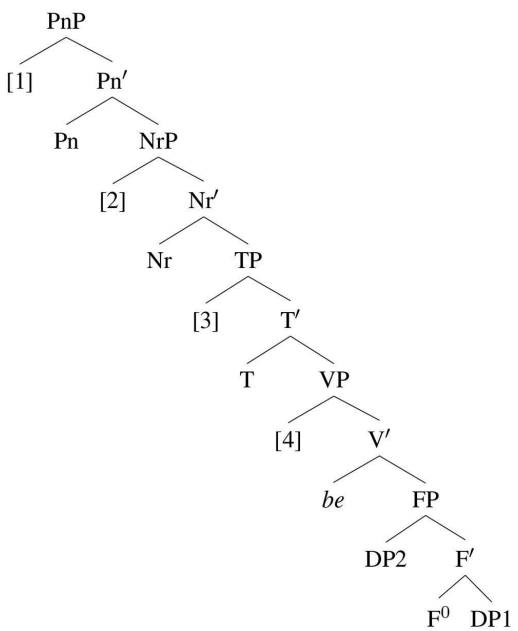
The interesting case is condition C. In German, the other of the four languages where 2nd person is distinctively marked in the plural, nearly all responses (98%) had 2nd person plural agreement in the corresponding condition (full DP2 agreement), with just a few choices of DP1 agreement, 3rd singular (2%)—see again **Table 7**. In this condition in Icelandic, however, just under a third of the responses consisted of the 3rd plural form instead of either 3rd singular (DP1 agreement)[30] or 2nd plural (full DP2 agreement). Thus in these Icelandic responses we see agreement with DP2 in number (plural), but not person.

TABLE 11 | DP1 vs. DP2 Agreement per condition in Icelandic (irrelevant cases excluded).

	DP ϕ-features		Copula agrees with			
Cond	**DP1**	**DP2**	**DP1**	**DP2 (all)**	**DP2 (Nr only)**	**Total**
A	3sg	3pl	74 (34%)	143 (66%)	n.a.	217
B	3sg	2sg	109 (52%)	99 (48%)	n.a.	208
C	3sg	2pl	68 (32%)	80 (38%)	63 (30%)	211
D	3pl	2pl	118 (56%)	91 (44%)	n.a.	209

It was argued in Sigurðsson and Holmberg (2008) that Number and Person are distinct heads in Icelandic, with Person higher than Number (consistent with the morphology in Icelandic, where Person morphology on the verb is consistently outside Number morphology). Sigurðsson and Holmberg (2008) argue, on the basis of the variation of patterns in agreement in the kind of dative-nominative constructions discussed earlier, that with these additional heads come additional landing sites for movement. The spine that they propose for the clause in Icelandic can be schematized as follows, where we have numbered the potential landing sites for ease of reference[31]:

(64)



[28] In this paper, we argue that the crucial difference between the dative-nominative construction and the cases discussed here is that the dative is an intervener: it agrees with the probe but does not halt it. Thus the probe can enter a agree relationship with both the dative and the nominative, giving rise to syncretism effects. The nominatives in SCCs halt the probe, so there is no multiple agreement and no true syncretism effect.

[29] There was in fact one further condition, which was like the (b) condition except that DP1 matched the type used in the (d) condition (e.g., *winner* rather than *problem*), this was mentioned briefly in section 3.3 above. As the type of DP1 was shown not to have a significant effect, we do not discuss it further here; for detailed discussion see Hartmann and Heycock (2017, pp. 249–261).

[30] As DP1 in SCCs cannot be 1st or 2nd person, as discussed earlier, there is no way to test in this specific case whether or not this is DP1 agreement or default. However, as we have argued above, where we were able to test for "default" agreement in SCCs, it was always judged ungrammatical, so we exclude it as option here, too.

[31] Note that in Preminger (2011, 2014) the order of the two heads has number higher than person in contrast to the proposal in Sigurðsson and Holmberg (2008). As will be shown below, our data and analysis support the order proposed in Sigurðsson and Holmberg (2008), which is consistent with the morphology on the verb in Icelandic.

Importantly, there is a potential landing site for DP1 below Person but above Number [Position [2] in (64)]. If DP1 moves directly to position [1] above all agreement probes, the first DP that will be encountered by the agreement probes for both number and person will be DP2. This should result in full DP2 agreement. If DP1 moves rather to position [2] it will be accessible to the person probe, but DP2 will be found by the number probe. This should result in person agreement with DP1 (which in SCCs is always 3rd person) but number agreement with DP2. Thus this derives the Number-only DP2 agreement just described. Positions [3] and [4] as landing sites for DP1 will result in DP1 agreement in both person and number as in this case, DP1 is the closest target for both agreement probes. At present we do not see how this pattern of Number-only DP2 agreement could be derived by appealing only to the possibility of differential ϕ-sensitivity of the probe (Béjar and Kahnemuyipour, 2017).

It does have to be recognized that despite the robust production of this type of agreement in our experiment, in the ratings task it was rated rather low (z-score: -0.40), though it was significantly above the rating for default agreement (which was -0.66, see **Table 4**). There are two options why this might be the case. First, we only have a small number of speakers (4 to be precise) who rate the condition consistently (i.e., in all 3 occasions they rate it) above their average. Alternatively, we think that it is possible here that speakers are more aware of prescriptive pressures in the ratings task than in the production task[32]. Recall that in the case of number-only agreement, the morphology of the verb is neither a full match for DP1 or DP2. Clearly this is a rare configuration in the language, and although as far as we know there are no articulated prescriptions about agreement in specificational sentences in Icelandic, any speaker who is hesitant about the "correctness" of their response is unlikely to conclude that number-only agreement is the prescribed form. In the production experiment, speakers were never presented with forms of the copula: they generated these forms to fill the blanks. On the other hand, in the ratings task speakers were presented with examples in other conditions where the copula can be interpreted as agreeing fully with DP1 and/or DP2. We therefore suggest tentatively that speakers may be more conscious in the ratings task of alternative forms that seem more "standard" and that this may account at least in part for the rating of number-only agreement being lower than would be expected from its frequency in production.

In Hartmann and Heycock (2018e), we argue that Faroese also shows evidence that the person and number probes are on distinct heads in this language though the evidence is much less direct than in Icelandic, as number-only agreement in Faroese can conflate either with DP1 or full DP2 agreement. In the paper we show how the frequency differences and rating results are best understood in terms of two probes; below we will concentrate on the production data; readers are referred to the paper for full details.

[32]It is also the case that in the rating study—but not the production study—the embedded clauses all contained negation. This makes the sentences more complex and is thus expected to depress the ratings to some extent.

TABLE 12 | Conditions and results of the Faroese production study on Person.

	DP ϕ-features		Copula agrees with			
Condition	**DP1**	**DP2**	**DP1**	**DP2**	**Total**	**%DP2 agreement**
A	3sg	3pl	51	54	105	51%
B	3sg	2sg	100	14	114	12%
C	3sg	2pl	39	76	115	66%

In Faroese we tested agreement in the following three conditions[33]:

(65) Hann ivaðist í, um … FAROESE
he wondered if
'He wondered if …
A: høvuðstrupulleikin _____ tey
the main problem they
'the main problem _____ them'
DP.SG _____ Pronoun.3.PL
B: høvuðstrupulleikin _____ tú
the main problem you.SG
'the main problem _____ you.SG'
DP.SG _____ Pronoun.2.SG
C: høvuðstrupulleikin _____ tit
the main problem you.PL
'the main problem _____ you.PL'
DP.SG _____ Pronoun.2.PL

The results are tabulated in **Table 12**. It is important to bear in mind that Faroese has no distinct person marking in the plural on the copula, but 2nd person (and 1st in the present tense only) is marked distinctively in the singular, as illustrated in (66).

(66)

	Present		Past	
Person	Singular	Plural	Singular	Plural
1	*eri*	*eru*	*var*	*vóru*
2	*ert*	*eru*	*vart*	*vóru*
3	*er*	*eru*	*var*	*vóru*

The observation of interest here is that the amount of DP2 agreement for person drops to 12% with 2nd person singular DP2, while it is much higher in the other two conditions[34].

Once we take into consideration that person and number are separate heads, and as a result that number-only agreement is a viable option in Faroese, we can see why there is such a difference between production of apparent DP2 agreement in B on the one hand, and A and C on the other. As set out in **Table 13**, in both conditions A and C, number-only agreement [the agreement

[33]We had two more conditions including negation for condition A and B; as there was no difference with or without negation (as expected), we do not discuss this here any further.

[34]The figures in column 4 (ϕ-features on the copula matching DP1) are the number of 3rd singular verb forms; the figures in column 5 are the number of verb forms showing plural agreement (in A and C) and 2nd person singular agreement in B.

TABLE 13 | Agreement features, verb form and coding per DP1 position for Faroese.

Condition	DP2 features		Feature realization per DP1 position [1]	[2]	[3]	[4]
A	3pl	V-features	3.pl	3.pl	3.sg	3.sg
		verb form	eru/vóru	eru/vóru	er/var	er/var
		coding	DP2	DP2	DP1	DP1
B	2sg	V-features	2.sg	3.sg	3.sg	3.sg
		verb form	ert/vart	er/var	er/var	er/var
		coding	DP2	DP1	DP1	DP1
C	2pl	V-features	2.pl	3.pl	3.sg	3.sg
		verb form	eru/vóru	eru/vóru	er/var	er/var
		coding	DP2	DP2	DP1	DP1

pattern associated with DP1 occupying position [2] in the tree in (64)] conflates with DP2 agreement (the agreement pattern associated with position [1]); in condition B, on the other hand, the agreement morphology associated with these two positions is distinct. Thus the apparent lower production of "DP2" agreement in condition B can be explained because it is the realization of only one possible configuration (DP1 occupying position [1]) while in conditions A and C, apparent DP2 agreement can be the realization of two configurations (DP1 occupying either position [1] *or* position [2])[35].

Thus, we conclude that both Faroese and Icelandic have person and number as separate heads, which provides an additional landing site for DP1. As a result, number-only DP2 agreement is a possible option, even though it might not be overtly marked in all cases.

5.4. Summary

Summarizing our findings and relating them to other works on agreement patterns in specificational copular clauses, we found that these patterns are due to general properties of the agreement system of each language and properties of SCCs. The relevant factors that we isolated are: (i) case (ii) structural configuration (iii) number of agreement probes; (iv) type of agreement probe.

(i) Reviewing previous literature, we pointed to one first relevant aspect for agreement, namely the case of the two DPs, and as a result their availability for agreement. In English, DP2 appears in accusative case, which makes it inaccessible as a controller of agreement in English. In the languages that we discussed, this is not an issue. In all four languages we looked at, both DP1 and DP2 are nominative and as such potential controllers of agreement.

(ii) We have argued that the crucial source of variation in the Germanic languages arises from the SCCs being inversion structures, which creates a configuration in which the initially lower DP1 can become accessible to a higher agreement probe, because it moves above DP2 to become the highest DP below a yet higher probe or probes, see (64) above. This sets SCCs apart from predicational copular clauses, which do not show significant variation in agreement patterns in Germanic.

(iii) A third relevant factor is the number and structure of agreement probes. This is relevant in the discussed languages for two effects we saw in the data. First, the separation of the number and person probe in the T-domain in Icelandic leads to a third possible pattern of agreement: number-only DP2 agreement. Second, it provides an explanation for an apparent increase in DP2 agreement in Faroese where this is indistinguishable from number-only DP2 agreement due to morphological syncretism.

(iv) Additionally, we take the increase in the production and acceptability of DP1 agreement in all four languages[36] to be due to an agreement probe in the C-domain. The effects of this probe are usually not visible, as there is typically just one target for agreement for both probes; they are manifest in specificational copular clauses because there are two.

6. NEWLY OPENED QUESTIONS

Overall, we intend our work to contribute a new range of data relevant both to specific questions concerning copular clauses, and to more general questions about how agreement goals are "chosen" when there is more than one, and how apparent restrictions on person agreement might be explained. We have tried to highlight throughout how these new data bear on existing theoretical questions about agreement. At this point we would like to add the perspective of what new questions are opened up by this data that are relevant for future research in this domain.

First of all, we have shown for Dutch, Faroese, and Icelandic, that there are a range of agreement options that native speakers choose from. Considering the patterns that we have presented above, an important question that immediately arises is what independent factors determine the agreement options available in the different languages. Clearly the hypotheses that can be entertained depend on prior decisions regarding the most promising analyses of these different options. We have outlined two alternative types of analysis above. In one, initially set out in Béjar and Kahnemuyipour (2017, 2018), DP1 agreement arises in specificational sentences because DP1 in such sentences is ϕ-deficient and the agreement probe on T may be sensitive to exactly the feature or features that DP1 lacks, so that DP1 may be "skipped." In the other, initially set out in Hartmann and Heycock (2016, 2017, 2018d,e, 2019), we have proposed that DP1 may "evade" agreement by moving directly to a position above the agreement probe (or, where Person and Number are not only distinct probes but distinct heads, above Number but below Person).

[35]The production data alone are consistent with an alternative analysis, namely that the lower proportion of apparent DP2 agreement when DP2 is 2nd person singular could be due to a morphological conflict that does not arise in the plural due to syncretism (an explanation given for the"person effect" in dative-nominative constructions in Icelandic, see section 2 above). However, this alternative can be ruled out by taking into consideration also the corresponding rating examples. We do not discuss this further here, see Hartmann and Heycock (2018e) for details.

[36]As noted in section 4.3 we have production data for this phenomenon from all four languages, but rating data only from Dutch and German.

We take as a baseline assumption that the ϕ-features on DP1 in a specificational sentence do not vary between languages (*a fortiori* they do not vary between idiolects within a single language). Under the first type of analysis, then, the kind of variation between DP1 and DP2 agreement that we have documented here would have to reflect variation in the ϕ-sensitivity of the probe—intra-speaker variation, for many speakers of Dutch, Faroese, and Icelandic[37].

Under the alternative approach that we have outlined, the observed variation has to be due to the possibility of DP1 making use of different landing sites. In Heycock (2012) it was proposed that direct movement above the agreement probe (taken in that paper to reside in T) may occur if the copula, rather than uniformly instantiating *v*, may instead instantiate T, which is then the lowest functional head above the small clause. This is evidently only a possible analysis for the finite copula, given that in all the languages under consideration here, and also in English, the copula may appear in non-finite contexts, for example below a modal. It was noted in that paper that initial results from Faroese indicated that exactly in such contexts, the production of DP2 agreement dropped sharply, to the point where a possible conclusion was that it was in fact ungrammatical. In our subsequent studies reported on here, we have not been able to include conditions testing for this kind of locality effect across the languages at issue: evidently this is a question that demands further research, if we are to be able to answer questions about locality effects in these structures and what they (or their absence) can tell us about the right analysis. As noted already in Heycock (2012), however, at least in German our impression, based on the German-speaking author's judgment and informally gathered judgment from other German-speakers, is that agreement with DP2 is possible even when the copula is embedded below e.g., a modal:

(67) Sie sagte, dass der Ursprung des Gerüchts ihr
She said that the source the.GEN rumor.GEN you
beide sein sollt
two be should.2.PL GERMAN
'She said that the source of the rumor ought to be you two.'

Taken together with the virtually categorical DP2 agreement (setting aside what we have analyzed as C-agreement) in German, this suggests that in this language there is never a landing site for DP1 below the agreement probe. This would be in line with the discussion in the literature showing that German provides hardly any evidence for a T projection independent of verbal projections including projections for auxiliaries and modals (see Haider, 1997; Sternefeld, 2009). If this is indeed the case, in German the agreement probe might in fact be on *v*-related heads. If these did not provide specifier positions—for a reason that would have to be determined—any landing site for DP1 would be above it.

[37] Some additional proposal would have to be made for this type of analysis to be extended to account for the agreement pattern in the XP–V_{fin}–DP1–DP2 order discussed in section 4.3. Further, as just discussed, it is not clear that the number-only DP2 agreement pattern attested in Icelandic could be accounted for under this analysis.

The difference between German on the one hand and Dutch, Faroese and Icelandic then could be that the latter have a separate T projection. This has the effect that the edge of vP is a possible landing site for DP1, below the agreement probe. For many speakers of Faroese, Dutch, and Icelandic, of course, although DP1 agreement is possible, it is not the only option, and in this they differ from English. That is, the movement to the edge of vP is a possible option, but direct movement to the position higher than the agreement probe, directly to the subject position is also possible.

A second, closely related question, is what independent evidence a learner has for these differences between the languages. One potential factor in this is the role of the morphological exponence of agreement features. For example, we presented above evidence that Icelandic has distinct heads for the number and person probes. An obvious hypothesis is that this might correlate with a transparent morphological distinction between number and person morphology on the verb (as pointed out in Sigurðsson and Holmberg, 2008, Icelandic verbal agreement morphology systematically has distinct number morphology close to the verb stem, and then person morphology). However, we found evidence for a "split" probe also in Faroese, where there is no similar clear split between person and number morphology. Further, German has a morphological paradigm that is comparable to the one in Icelandic, but there is no evidence for split probes in this system. This suggests that any relation between the nature of the morphological expression of agreement and the range of options will not be a simple one, but more work is needed here.

A third question that arises in the context of our work concerns the question of what determines native speakers' choice of form when they have more than one option available. That is, what factors influence/determine the choice of variant when there is (at least the possibility for) intra-speaker variation? These choices might be influenced by factors such as formal/informal context, processing constraints, and the like. Speakers might also develop preferences based on features that are known to affect agreement in other languages: pronoun vs. full noun phrase, definiteness, information structure, animacy or even the task/goal for the expression used (see below) etc. In general, it seems to us that specificational copular clauses in Faroese, Icelandic, and Dutch provide an interesting new testbed for the study of syntactic/morphological variation/optionality.

A fourth question that arises in all cases of agreement is the role of linear order and the difference between "true" grammatical agreement and processing effects associated with linear order, for example the much-discussed case of "agreement attraction" (see among many other, Bock and Miller, 1991; Franck et al., 2002; Wagers et al., 2009; Patson and Husband, 2016 and references therein). For example, we found an effect in our production experiments concerning number that could potentially be a processing effect of distance: for all our languages, DP2 agreement decreases in V2 clauses when an adverbial intervenes between DP1 and DP2 compared to the same structures without an adverb, see **Table 14**.

This might in principle be an interaction of optionality and processing preference: when two options are grammatical, speakers might tend to choose the option which allows agreement

TABLE 14 | Production of DP2 agreement with and without intervening ADV % in V2 contexts.

	Dutch	German	Faroese	Icelandic
DP1 V DP2	62%	92%	64%	74%
DP1 V ADV DP2	37%	82%	42%	45%

with a more local DP. This is a possible explanation for Dutch, Faroese and Icelandic (where we have argued both DP1 and DP2 agreement are grammatical options), however, it leaves the effect in German unexplained (where only DP2 agreement is possible in the T-domain). Even though the increased use of DP1 agreement in German is smaller compared to the other languages, it is still significant, see the details in Hartmann and Heycock (2018b). It remains to be established whether this effect should be considered an effect of C-agreement or a processing effect with DP2 being linearly more distant. Establishing this would require cross-disciplinary work to tease apart the two types of effects.

Finally, our research also raises the question of the kind of evidence obtained from production and rating studies. As discussed in section 3.3, we used both methods in order to combine the merits of both (and control for the limitations of both). We did, however, find some mismatches between the data from the two types of study that we do not yet understand. For example, considering the production study in Icelandic, with a 2nd person plural DP2 (the third row of data in **Table 11**), the production of the three types of agreement (DP1, DP2, and person-only DP2) is roughly equal. In the rating data, however, DP1 agreement is rated significantly higher than either type of DP2 agreement. Part of this seems to be due to the fact that some speakers seem to prefer one or the other agreement pattern generally, and in the rating data, we had more speakers who prefer DP1 agreement over DP2 agreement. But further investigation into the effect of the task on the results seems necessary. The production task, which is effectively a forced-choice task as far as agreement is concerned, may be less sensitive to small differences in the materials, as the possible options for the participants are very limited. In the rating task, especially with such fine-grained methods such as the magnitude estimation and thermometer ratings, ratings might be more directly affected by small differences in the materials. Systematic consideration of these methodological effects therefore seems in need of further study.

7. CONCLUSION

In this paper we have provided an overview of agreement patterns in Specificational Copular Clauses in Germanic. Based on experimental work summarized here and reported in more detail in Heycock (2009, 2012), Hartmann and Heycock (2016, 2017, 2018b,d,e, 2019), we have attempted to synthesize the main insights and generalizations, and we have proposed an analysis of specificational clauses along the lines given in (64), but also discussed alternatives.

According to the analysis we have outlined, depending on the landing site of inversion of DP1 in a position below, above or between agreement probes (where they are split), different agreement arises. For the four languages under discussion we see the following patterns. First, all four languages under investigation show DP2 agreement to a greater or lesser degree, i.e., all four languages have [1] as a landing site for DP2. German is the one language that shows exclusively such agreement, potentially indicating the lack of the T-domain. Second, we find variation with respect to the other positions: Icelandic and Faroese have number and person split, so both languages have position [2] as one available option for speakers. This results in number-only DP2 agreement. Additionally, Icelandic, Faroese and Dutch allow for DP1 agreement (again to varying degrees), i.e., DP1 can land in position [3]/[4] below these agreement probes. Finally, we have isolated a further pattern of agreement which is located in the C-domain, so independent of the positions in the tree in (67). This C-related agreement appears in XP-initial V2 clauses in all four languages.

Overall we consider that agreement patterns in SCCs in Germanic help to understand SCCs as inversion structures, and provide further insight in factors that play a role for agreement within and across languages, namely the number of agreement probes and their location (in the C- or T-domain), the syntactic configuration, and the option of downwards agreement with a low nominative.

DATA AVAILABILITY STATEMENT

The overview provided in this paper is based on a series of experiments reported in Hartmann and Heycock (2016, 2017, 2018b,d,e, 2019). The data sets can be accessed via the University of Edinburgh data share repository here: http://dx.doi.org/10.7488/ds/2329.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Linguistics & English Language Ethics Committee of the University of Edinburgh, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Linguistics & English Language Ethics Committee.

AUTHOR CONTRIBUTIONS

JH and CH contributed equally to the reported research and the writing of the article.

FUNDING

The research reported on here was partly supported by a British Academy/Leverhulme Trust Small Research Grant awarded to the two authors (SG142530: 2015-2017). Some of the reported research is partially based on materials supported by the German Research Foundation (DFG) via the grant to the SFB 833, project A7 (PIs: JH and Susanne Winkler; 2013–2017).

ACKNOWLEDGMENTS

We would like to thank again all the participants who contributed their time and linguistic expertise to our project, also Sigríður Mjöll Björnsdóttir, Zakaris Svabo Hansen, Marloes Oomen, Uni Johannesen, and Höskuldur Thráinsson for their extensive help with the materials, Julia Restle for help with OnExp and the processing of the results, and Robin Hörnig for assistance with statistical analysis. We would also like to express our thanks to the three reviewers and the editor, Andrew Nevins, for their help in improving the paper. All errors remain our responsibility.

REFERENCES

Abraham, W. (1993). Null subjects in the history of German: from IP to CP. *Lingua* 89, 117–142. doi: 10.1016/0024-3841(93)90050-7

Ackema, P., and Neeleman, A. (2004). *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: Oxford University Press.

Ackema, P., and Neeleman, A. (2018). *Features of Person*. Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Alexiadou, A., and Anagnostopoulou, E. (1998). Parametrizing Agr: word order, V-movement and EPP-checking. *Nat. Lang. Linguist. Theory* 16, 491–539. doi: 10.1023/A:1006090432389

Baker, M. (2008). *The Syntax of Agreement and Concord*. Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

Bard, E., Robertson, D., and Sorace, A. (1996). Magnitude Estimation of linguistic acceptablity. *Language* 72, 32–68. doi: 10.2307/416793

Bayer, J. (1984). COMP in Bavarian syntax. *Linguist. Rev.* 3, 209–274. doi: 10.1515/tlir.1984.3.3.209

Béjar, S. (2003). *Phi-syntax: a theory of agreement*. (Ph.D. thesis). University of Toronto, Toronto, ON, Canada.

Béjar, S., and Kahnemuyipour, A. (2017). Non-canonical agreement in copular sentences. *J Linguist.* 53, 463–499. doi: 10.1017/S0022226717000010X

Béjar, S., and Kahnemuyipour, A. (2018). Not all phi-features are created equal: a reply to Hartmann and Heycock. *J. Linguist.* 54, 629–635. doi: 10.1017/S0022226718000154

Béjar, S., and Rezac, M. (2003). "Person licensing and the derivation of PCC effects," in *Romance Linguistics: Theory and Acquisition*, eds A. T. Perez-Leroux and Y. Roberge (Amsterdam: John Benjamins), 49–62.

Berg, T. (1998). The resolution of number conflicts in English and German agreement patterns. *Linguistics* 36, 41–70. doi: 10.1515/ling.1998.36.1.41

Biberauer, T. (2004). Reconsidering the EPP and Spec-TP in Germanic. *Cambridge Occasional Papers Linguist. COPiL* 1, 15–40.

Birner, B. (1992). *The discourse function of inversion in English*. (Ph.D. thesis). Northwestern University, Evanston, IL, United States.

Bjorkman, B. M., and Zeijlstra, H. (2019). Checking up on (ϕ-)agree. *Linguist. Inq.* 50, 527–569. doi: 10.1162/ling_a_00319

Bobaljik, J. D. (2008). "Where's Phi? Agreement as a postsyntactic operation," in *Phi Features Across Interfaces and Modules*, eds D. Harbour, D. Adger, and S. Béjar (Oxford: Oxford University Press), 295–328.

Bock, K., and Miller, C. A. (1991). Broken agreement. *Cogn. Psychol.* 23, 45–93. doi: 10.1016/0010-0285(91)90003-7

Bowers, J. (1993). The syntax of predication. *Linguist. Inq.* 24, 591–656.

Chomsky, N. (1981). *Lectures on Government and Binding*. Studies in Generative Grammar 9. Dordrecht: Foris.

Chomsky, N. (2000). "Minimalist inquiries: the framework," in *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, eds R. Martin, D. Michaels, and J. Uriagereka (Cambridge, MA: MIT), 89–155.

Chomsky, N. (2001). "Derivation by phase," in *Ken Hale: A Life in Language*, ed M. Kenstowicz (Cambridge, MA: MIT Press), 1–52.

Coppock, E., and Beaver, D. (2015). Definiteness and determinacy. *Linguist. Philos.* 38, 377–435. doi: 10.1007/s10988-015-9178-8

den Dikken, M. (1998). Appraising 'The raising of predicates'. *Linguist. Berichte* 174, 246–263.

den Dikken, M. (2006a). *Relators and Linkers: The Syntax of Predication, Predicate Inversion, and Copulas*. Cambridge, MA: MIT.

den Dikken, M. (2006b). "Specificational copular sentences and pseudoclefts," in *The Blackwell Companion to Syntax*, Vol. IV, eds M. Everaert and H. van Riemsdijk (Malden, MA; Oxford: Blackwell), 292–409.

den Dikken, M. (2019). The attractions of agreement: why person is different. *Front. Psychol.* 10:978. doi: 10.3389/fpsyg.2019.00978

Don, J., Fenger, P., and Koeneman, O. (2013). Micro-variation as a tool for linguistic analysis. lingbuzz/001954.

Featherston, S. (2005). Magnitude estimation and what it can do for your syntax: some wh-constraints in German. *Lingua* 115, 1525–1550. doi: 10.1016/j.lingua.2004.07.003

Featherston, S. (2008). "Thermometer judgements as linguistic evidence," in *Was ist Linguistische Evidenz?*, eds C. M. Riehl and A. Rothe (Aachen: Shaker Verlag), 69–90.

Fischer, G. (2003). *The problem is/are your parents: resolving number conflicts in equative sentences in Dutch and German*. (Undergraduate Honours Dissertation in Linguistics). Edinburgh: University of Edinburgh.

Franck, J., Vigliocco, G., and Nicol, J. (2002). Subject-verb agreement errors in French and English: the role of syntactic hierarchy. *Lang. Cogn. Process.* 17, 371–404. doi: 10.1080/01690960143000254

Frey, W. (2010). A′-Movement and conventional implicatures: about the grammatical encoding of emphasis in German. *Lingua* 120, 1416–1435. doi: 10.1016/j.lingua.2008.09.016

Gerbrich, H., Schreier, V., and Featherston, S. (2019). "Standard items for English judgement studies: syntax and semantics," in *Experiments in Focus: Information Structure and Semantic Processing*, eds S. Featherston, R. Hörnig, S. von Wietersheim, and S. Winkler (Berlin: De Gruyter Mouton), 305–327.

Grewendorf, G. (2005). "The discourse configurationality of scrambling," in *The Free Word Order Phenomenon: Its Syntactic Sources and Diversity*, eds J. Sabel and M. Saito (Berlin; New York, NY: Mouton de Gruyter), 75–135.

Haider, H. (1993). *Deutsche Syntax Generativ: Vorstudien zur Theorie einer Projektiven Grammatik*. Tübinger Beiträge zur Linguistik. Tübingen: Gunther Narr.

Haider, H. (1997). "Projective economy: on the minimal functional structure of the German clause," in *German: Syntactic Problems - Problematic Syntax*, eds W. Abraham and E. van Gelderen (Tübingen: Niemeyer), 83–103.

Hartmann, J. M. (2016). *The Syntax and Focus Structure of Specificational Copular Clauses and Clefts*. Tübingen: Habilitationsschrift, Tübingen University.

Hartmann, J. M. (2019). "Focus in nominal copula clauses in English," in *Experiments in Focus: Information Structure and Semantic Processing*, Linguistische Arbeiten, eds S. Featherston, R. Hörnig, S. von Wietersheim, and S. Winkler (Berlin: Mouton de Gruyter), 71–103.

Hartmann, J. M., and Heycock, C. (2016). "Evading agreement: a new perspective on low nominative agreement in Icelandic," in *Proceedings of the Forty-Sixth Annual Meeting of the North East Linguistic Society (NELS)*, Vol. 2, eds C. Hammerly and B. Prickett (Amherst, MA: GLSA Publications), 67–80.

Hartmann, J. M., and Heycock, C. (2017). "Variation in copular agreement in Insular Scandinavian," in *Syntactic Variation in Insular Scandinavian, Volume 1 of* Studies in Germanic Linguistics (SIGL), eds Thráinsson, H., Heycock, C., Petersen, H. P., and Hansen, Z. S. (Amsterdam: John Benjamins), 233–275.

Hartmann, J. M., and Heycock, C. (2018a). A remark on Béjar & Kahnemuyipour 2017: specificational subjects do have phi-features. *J. Linguist.* 54, 611–627. doi: 10.1017/S0022226718000117

Hartmann, J. M., and Heycock, C. (2018b). *Agreement in Copula Clauses: Evidence for a Dual Mechanism of Agreement*. Mannheim; Edinburgh: IDS Mannheim, University of Edinburgh Manuscript.

Hartmann, J. M., and Heycock, C. (2018c). More on phi-features in and out of copular sentences: a reply to B&K 2018. *J. Linguist.* 54, 637–646. doi: 10.1017/S0022226718000233

Hartmann, J. M., and Heycock, C. (2018d). *Person Effects in Low Nominative Agreement in Icelandic: Bringing New Data to Bear*. Mannheim; Edinburgh: IDS Mannheim, University of Edinburgh Manuscript.

Hartmann, J. M., and Heycock, C. (2018e). Postverbal subjects and agreement: specificational Copular Clauses in Faroese. *Ital. J. Linguist.* 30, 187–212. doi: 10.26346/1120-2726-129

Hartmann, J. M., and Heycock, C. (2019). Restrictions on "Low" Person agreement in Dutch Copular Constructions. *Linguist. Netherlands* 36, 131–147. doi: 10.1075/avt.00028.har

Heggie, L. (1988). *The syntax of copular constructions*. (Ph.D. thesis). University of Southern California, Los Angeles, CA, United States.

Heller, D., and Wolter, L. (2008). "That is Rosa: identificational sentences as intensional predication," in *Proceedings of Sinn und Bedeutung*, Vol. 12 (Oslo), 226–240.

Heycock, C. (1994). "The internal structure of small clauses," in *Proceedings of NELS 25*, Vol. 1, ed J. Beckman (Amherst, MA: GLSA), 223–238.

Heycock, C. (2009). Agreement in specificational sentences in Faroese. *Nordlyd (Tromsø Working Papers Lang. Linguist.)* 36, 57–77.

Heycock, C. (2012). Specification, equation, and agreement in copular sentences. *Can. J. Linguist.* 57, 209–240. doi: 10.1017/S0008413100004758

Heycock, C. (in press). "Copular sentences," in *The Wiley Blackwell Companion to Semantics*, eds L. Matthewson, C. Meier, H. Rullmann, T. E. Zimmerman, and D. Gutzmann (London: Wiley-Blackwell).

Heycock, C., and Kroch, A. (1998). "Inversion and equation in copular sentences," in *Papers in Linguistics*, Vol. 10, eds A. Alexiadou, N. Fuhrhop, U. Kleinhenz, and P. Law (Berlin: Zentrum für Allgemeine Sprachwissenschaft, Sprachtypologie und Universalienforschung (ZAS)), 71–87.

Heycock, C., Sorace, A., and Hansen, Z. S. (2010). V-to-I and V2 in subordinate clauses: an investigation of Faroese in relation to Icelandic and Danish. *J. Compar. Germanic Linguist.* 13, 61–97. doi: 10.1007/s10828-010-9035-7

Higgins, F. R. (1979). *The Pseudo-Cleft Construction in English*. New York, NY: Garland.

Holmberg, A., and Hróarsdóttir, T. (2004). Agreement and movement in Icelandic raising constructions. *Lingua* 114, 651–673. doi: 10.1016/j.lingua.2004.01.002

Huber, S. (2002). *Es-Clefts und det-Clefts: zur syntax, semantik und informationsstruktur von Spaltsätzen im Deutschen und Schwedischen*. (Ph.D. thesis). Lund University, Lund, Sweden.

Jónsson, J. G. (2009). "Covert nominative and dative subjects in Faroese," in *Nordlyd 36.2: NORMS Papers on Faroese*, eds P. Svenonius, K. Bentzen, C. Heycock, J. í. L. Jacobsen, J. B. Johannessen, J. K. Parrott, T. E. Strahan, and Ø. A. Vangsnes (Tromsø: CASTL, University of Tromsø), 142–164.

Keine, S. (2017). "Agreement and vP phases," in *A Schrift to Fest Kyle Johnson*, Vol. 1, eds N. LaCara, K. Moulton, and A.-M. Tessier (Amherst, MA: Linguistics Open Access Publications), 177–185.

Koopman, H. (2006). "Agreement configurations: in defense of Spec-head," in *Agreement Systems*, Linguistik Aktuell, ed C. Boeckx (Amsterdam: Benjamins), 159–199.

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Gunter Narr.

Mikkelsen, L. (2002). "Specification is not inverted predication," in *Proceedings of NELS 32*, ed M. Hirotani (Amherst, MA: GLSA, University of Massachusetts), 403–422.

Mikkelsen, L. (2005). *Copular Clauses: Specification, Predication and Equation*. Amsterdam; Philadelphia, PA: Benjamins.

Mikkelsen, L. (2011). "Copular clauses," in *Semantics: An International Handbook of Natural Language Meaning*, Vol. 2, eds K. von Heusinger, C. Maienborn, and P. Portner (Berlin: De Gruyter), 1805–1829.

Moltmann, F. (2013). Identificational sentences. *Nat. Lang. Semant.* 21, 43–77. doi: 10.1007/s11050-012-9091-4

Moro, A. (1991). "The raising of predicates: copula, expletives, and existence," in *More Papers on Wh-movement*, MIT Working Papers in Linguistics, eds L. Cheng and H. Demirdache (Cambridge, MA: MIT), 119–181.

Moro, A. (1997). *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge; New York, NY: Cambridge University Press.

Moro, A. (2006). "Copular sentences," in *The Blackwell Companion to Syntax*, Vol. 2, eds M. Everaert and H. van Riemsdijk (Malden, MA; Oxford: Blackwell), 1–23.

Neeleman, A., and van de Koot, H. (2002). The configurational matrix. *Linguist. Inq.* 22, 529–574. doi: 10.1162/002438902762731763

Neeleman, A., and van de Koot, H. (2008). Dutch scrambling and the nature of discourse templates. *J. Compar. Germanic Linguist.* 11, 137–189. doi: 10.1007/s10828-008-9018-0

Partee, B. (1986). "Ambiguous pseudoclefts with unambiguous *be*," in *Proceedings of NELS 16*, eds S. Berman, J.-W. Choe, and J. McDonough (Amherst, MA: GLSA), 354–366.

Patson, N. D., and Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *Q. J. Exp. Psychol.* 69, 950–971. doi: 10.1080/17470218.2014.992445

Percus, O., and Sharvit, Y. (2014). "Copular asymmetries in belief reports," in *Poster presented at SALT 24*. New York, NY.

Preminger, O. (2011). Asymmetries between person and number in syntax: a commentary on Baker's SCOPA. *Nat. Lang. Linguist. Theory* 29, 917–937. doi: 10.1007/s11049-011-9155-z

Preminger, O. (2014). *Agreement and Its Failures*. Number 68 in Linguistic Inquiry Monographs. Cambridge, MA: MIT Press.

Romero, M. (2005). Concealed questions and specificational subjects. *Linguist. Philos.* 28, 687–737. doi: 10.1007/s10988-005-2654-9

Saebø, K. J. (2009). Judgment ascriptions. *Linguist. Philos.* 32, 327–352. doi: 10.1007/s10988-009-9063-4

Schütze, C. (2003). "Syncretism and double agreement with Icelandic Nominative objects," in *Grammar in Focus: Festschrift for Christer Platzack*, eds L.-O. Delsing, C. Falk, G. Josefsson, and H. Á. Sigurðsson (Lund: Department of Scandinavian Languages, Lund University), 295–303.

Shlonsky, U., and Rizzi, L. (2018). "Criterial Freezing in small clauses and the cartography of copular constructions," in *Freezing*, Studies in Generative Grammar, eds J. M. Hartmann, M. Jäger, A. Kehl, A. Konietzko, and S. Winkler (Berlin; Boston, MA: De Gruyter Mouton), 29–65.

Sigurðsson, H. Á. (1989). *Verbal syntax and case in Icelandic in a comparative GB approach*. (Ph.D. thesis). University of Lund.

Sigurðsson, H. Á. (1996). Icelandic finite verb agreement. *Work. Pap. Scand. Syntax* 57, 1–46.

Sigurðsson, H. Á., and Holmberg, A. (2008). "Icelandic dative intervention: Person and Number are separate probes," in *Agreement Restrictions*, eds R. D'Alessandro, S. Fischer, and G. H. Hrafnbjargarson (Berlin: Mouton de Gruyter), 251–279.

Sternefeld, W. (2009). *Syntax*, Vol. 2. Tübingen: Stauffenburg.

Thráinsson, H. (2007). *The Syntax of Icelandic*. Cambridge: Cambridge University Press.

Thráinsson, H., Sigurðsson, E. F., and Jónsson, J. G. (2015). "Samræmi (agreement)," in *Tilbrigði í íslenskri setningagerð*, volume II, Chapter 12 (Reykjavík: Málvísindastofnun Háskóla Íslands), 203–231.

Ussery, C. (2017). "Dimensions of variation: agreement with nominative objects in Icelandic," in *Syntactic Variation in Insular Scandinavian, Volume 1 of* Studies in Germanic Linguistics (SIGL), eds Thráinsson, H., Heycock, C., Petersen, H. P., and Hansen, Z. S. (Amsterdam: John Benjamins), 165–197.

van Koppen, M. (2005). *One probe - two goals: aspects of agreement in Dutch dialects*. (Ph.D. thesis). Universiteit Leiden, Leiden, Netherlands.

van Koppen, M. (2017). "Complementizer agreement," in *The Wiley Blackwell companion to syntax*, The Wiley Blackwell Companions to Linguistics, eds M. Everaert and H. C. van Riemsdijk (Hoboken, NJ: John Wiley & Sons, Inc.), 1–40.

Wagers, M. W., Lau, E. F., and Phillips, C. (2009). Agreement attraction in comprehension: representations and processes. *J. Mem. Lang.* 61, 206–237. doi: 10.1016/j.jml.2009.04.002

Wurmbrand, S. (2012). Parasitic participles in Germanic: evidence for the theory of verb clusters. *Taal Tongval* 64, 129–156. doi: 10.5117/TET2012.1.WURM

Zaenen, A., Maling, J., and Thráinsson, H. (1985). Case and grammatical functions: the Icelandic passive. *Nat. Lang. Linguist. Theory*, 3.4, 441–483. doi: 10.1007/BF00133285

Zeijlstra, H. (2012). There is only one way to Agree. *Linguistic Rev.* 29, 491–539. doi: 10.1515/tlr-2012-0017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership