

PERCEPTUAL LINGUISTIC SALIENCE: MODELING CAUSES AND CONSEQUENCES

EDITED BY : Alice Blumenthal-Dramé, Adriana Hanulíková and Bernd Kortmann
PUBLISHED IN: Frontiers in Psychology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-177-7

DOI 10.3389/978-2-88945-177-7

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: researchtopics@frontiersin.org

PERCEPTUAL LINGUISTIC SALIENCE: MODELING CAUSES AND CONSEQUENCES

Topic Editors:

Alice Blumenthal-Dramé, University of Freiburg, Germany

Adriana Hanulíková, University of Freiburg, Germany

Bernd Kortmann, University of Freiburg, Germany



Recent years have seen an upsurge of interest in the notion of salience in linguistics and related disciplines. While in top-down salience, perceivers endogenously direct their attention to a certain stimulus, in bottom-up salience, it is the stimulus itself which attracts attention. In prototypical cases of bottom-up salience, the stimulus stands out because it is incongruous with a given ground by virtue of intrinsic physical characteristics. But a stimulus may also cause surprise by virtue of deviating from a cognitive ground, e.g., when violating social or probabilistic expectations. This has prompted researchers to examine the relationship between expectations and the perceptual salience of linguistic stimuli in new ways.

This e-book features contributions from different scientific frameworks. The reader will find commentaries, reviews, and original research articles on models of sociolinguistic and morphological salience, the role of attention, affect, and predictability, and on how salient items are processed, categorized and learned.

Radu Bercan/Shutterstock.com

Image used under license from Shutterstock.com

ceptual salience of linguistic forms and variants can be theoretically framed and methodologically operationalized in different areas of linguistic processing.

Taken together, the articles in this volume contribute to our understanding of how the perceptual

Citation: Blumenthal-Dramé, A., Hanulíková, A., Kortmann, B., eds. (2017). *Perceptual Linguistic Salience: Modeling Causes and Consequences*. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-177-7

Table of Contents

- 04 Editorial: Perceptual Linguistic Salience: Modeling Causes and Consequences**
Alice Blumenthal-Dramé, Adriana Hanulíková and Bernd Kortmann

SECTION 1: MODELING LINGUISTIC SALIENCE

- 07 Salience and Attention in Surprisal-Based Accounts of Language Processing**
Alessandra Zarcone, Marten van Schijndel, Jorrig Vogels and Vera Demberg
- 24 The Salience of Complex Words and Their Parts: Which Comes First?**
Hélène Giraudo and Serena Dal Maso
- 32 Toward a Unified Socio-Cognitive Framework for Salience in Language**
Hans-Jörg Schmid and Franziska Günther
- 36 What the Heck Is Salience? How Predictive Language Processing Contributes to Sociolinguistic Perception**
T. Florian Jaeger and Kodi Weatherholtz

SECTION 2: SOCIOLINGUISTIC SALIENCE

- 41 Estimating the Relative Sociolinguistic Salience of Segmental Variables in a Dialect Boundary Zone**
Carmen Llamas, Dominic Watt and Andrew E. MacFarlane
- 59 Linking Place and Mind: Localness as a Factor in Socio-Cognitive Salience**
Marie M. Jensen
- 72 The Penefit of Salience: Salient Accented, but not Unaccented Words Reveal Accent Adaptation Effects**
Ann-Kathrin Grohe and Andrea Weber

SECTION 3: SALIENCE AND LANGUAGE ACQUISITION

- 89 Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus**
Myrna C. Cintrón-Valentín and Nick C. Ellis
- 110 Social Salience Discriminates Learnability of Contextual Cues in an Artificial Language**
Péter Rácz, Jennifer B. Hay and Janet B. Pierrehumbert



Editorial: Perceptual Linguistic Salience: Modeling Causes and Consequences

Alice Blumenthal-Dramé^{1,2*}, Adriana Hanulíková^{2,3*} and Bernd Kortmann^{1,2}

¹ Department of English, University of Freiburg, Freiburg, Germany, ² Freiburg Institute for Advanced Studies, University of Freiburg, Freiburg, Germany, ³ Department of German, University of Freiburg, Freiburg, Germany

Keywords: prediction, language learning, morphology, salience, surprisal, social markers, dialects, language variation and change

Editorial on the Research Topic

Perceptual Linguistic Salience: Modeling Causes and Consequences

Recent years have seen an upsurge of interest in the notion of salience in linguistics and related disciplines. The attention literature distinguishes two broad types of perceptual salience (Summerfield and Egner, 2009; Awh et al., 2012). First, a stimulus can be salient—i.e., foremost in one's mind—because it is cognitively preactivated. This type of salience, sometimes referred to as top-down salience, may occur if a stimulus is expected because it is part of a cognitive routine, if it has recently been mentioned, or due to current intentions of the perceiver. Research on salience as a semantic-pragmatic phenomenon has shown that top-down salience can account for systematic preferences in the interpretation of figurative utterances, pronominal antecedents, implicatures, and discursive links (Geeraerts, 2000; Giora, 2003; Chiarcos et al., 2011; Jaszczolt and Allan, 2011).

While in top-down salience, perceivers endogenously direct their attention to a certain stimulus, in the second type of salience, bottom-up salience, it is the stimulus itself which attracts attention. In prototypical cases of bottom-up salience, the stimulus stands out because it is incongruous with a given ground by virtue of intrinsic physical characteristics. But a stimulus may also cause surprise by virtue of deviating from a cognitive ground, e.g., when violating social or probabilistic expectations (Clark, 2013). This has prompted researchers to examine the relationship between expectations and the perceptual salience of linguistic stimuli in new ways (Hanulíková et al., 2012; Rácz, 2012; Hanulíková and Carreiras, 2015; Blumenthal-Dramé, 2016a,b; Roller, 2016; Blumenthal-Dramé et al., 2017), and inspired us to organize a workshop devoted to this particular area.

In October 2014, the Freiburg Institute of Advanced Studies (FRIAS) hosted the workshop “Perceptual linguistic salience: Modeling causes and consequences”, organized by the editors of this volume. Bringing together researchers from psycholinguistics, sociolinguistics, neurolinguistics, and cognitive linguistics, the workshop sought to explore the notion of perceptual salience and its explanatory potential for the domains of language processing, variation, and change. Several questions arising from the stimulating discussions were listed in the call for papers for this Research Topic and included the following:

- Which cognitive processes underlie the differential treatment of salient vs. non-salient linguistic percepts?
- How can these processes be accommodated within psycholinguistic models?
- How can the perceptual salience of linguistic forms and variants be operationalized?
- To what extent is salience an intrinsic feature of linguistic forms (e.g., dialectal variants), and to what extent does it result from contextual factors or prior experience with language?

OPEN ACCESS

Edited and reviewed by:

Manuel Carreiras,
Basque Center on Cognition, Brain
and Language, Spain

*Correspondence:

Alice Blumenthal-Dramé
alice.blumenthal@anglistik.
uni-freiburg.de
Adriana Hanulíková
adriana.hanulikova@germanistik.
uni-freiburg.de

Specialty section:

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

Received: 28 February 2017

Accepted: 06 March 2017

Published: 22 March 2017

Citation:

Blumenthal-Dramé A, Hanulíková A
and Kortmann B (2017) Editorial:
Perceptual Linguistic Salience:
Modeling Causes and Consequences.
Front. Psychol. 8:411.
doi: 10.3389/fpsyg.2017.00411

This volume features nine contributions including five original research articles, one review, and three commentaries that addressed the above questions in very interesting ways. Several contributions discuss which factors or prior experience with language underlie the differential treatment of salient linguistic percepts, and how can they be operationalized and modeled. Jaeger and Weatherholtz argue that sociolinguistic salience can be quantified using computational psycholinguistics. A distinction is made between the initial salience of a novel variant and the cumulative product of experienced exposures to a variant. A variant's salience may be predicted based on its surprisal and frequency. In support of this view, Schmid and Günther propose a unified framework of salience which aims at reconciling seemingly contradictory uses of this notion in the literature: cues are either categorized as salient because they confirm expectations, or because they violate them. Zarcone et al. suggest that an articulated model of salience should take into account attention, affect, and predictability at different levels of processing, and that these dimensions and their interactions can be straightforwardly accommodated within the Predictive Coding framework. Finally, Giraudo and Del Maso present a critical review of so-called decompositional accounts of morphological processing. They argue that the salience of morphemes cannot be reduced to formal factors, and that semantic factors and relationships between holistically represented complex words should also be integrated into models of morphological processing.

Several contributions address the hypothesis that salient items might function as cognitive reference points that structure and give access to certain cognitive domains (e.g., sociolinguistic stereotypes), thereby influencing the perception and categorization of less salient items of the same domain (Rosch, 1975; Langacker, 1993; Hanulíková and Weber, 2012). On the basis of recent theories of enregisterment and exemplar processing, Jensen investigates percepts resulting from sociolinguistic or socio-cognitive salience, more exactly the salience of various morphosyntactic forms in vernacular Tyneside (Northeast England). This study brings to the fore the role of place as strongly shaping both a community's and an individual's linguistic identity and self-representation.

Llamas et al. present metrics for determining the relative salience of phonetic variables in the Scottish-English border zone. This paper substantiates the fact that the choice of features which ultimately become sociolinguistically salient is largely arbitrary. What matters is sufficient agreement among the members of the relevant speech community as to which structural features are considered to function as signals of group membership. Using eye-tracking, Grohe and Weber show for regional dialects of German that salience clearly has an effect on native accent adaptation, but only if objective criteria for salience apply.

The notion of perceptual salience is inextricably linked to issues concerning language acquisition. Cintrón-Valentín and Ellis examine effects of physical salience and attentional biases in the visual and auditory modalities in second language acquisition. Chinese and English native speakers were trained on Latin tense morphology under different types of explicit form-focused instructions, some of which successfully increased learners' attention to less salient morphological features. Rácz et al. use artificial language learning and show that the social-cognitive salience of non-linguistic contexts influences learning of morphological features. Learning is easier with a coherent and interpretable social context (such as gender of the speaker) as opposed to accidental links between the speaker and the construction (such as front-facing vs. side-facing).

Taken together, the papers featured in this volume contribute to our understanding of how the perceptual salience of linguistic forms and variants can be theoretically framed and methodologically operationalized in different areas of linguistic processing.

AUTHOR CONTRIBUTIONS

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

FUNDING

Funding for the workshop was provided by the Freiburg Institute for Advanced Studies (FRIAS) at Albert-Ludwigs University in Freiburg, Germany.

REFERENCES

- Awh, E., Belopolsky, A. V., and Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends Cogn. Sci.* 6, 437–443. doi: 10.1016/j.tics.2012.06.010
- Blumenthal-Dramé, A. (2016a). "Entrenchment from a psycholinguistic and neurolinguistic perspective," in *Entrenchment and the Psychology of Language Learning. How We Reorganize and Adapt Linguistic Knowledge*, ed H.-J. Schmid (Boston, MA: APA and Walter de Gruyter), 129–152.
- Blumenthal-Dramé, A. (2016b). What corpus-based cognitive linguistics can and cannot expect from neurolinguistics. *Cogn. Linguist.* 27, 493–505. doi: 10.1515/cog-2016-0062
- Blumenthal-Dramé, A., Glauche, V., Bormann, T., Weiller, C., Musso, M., and Kortmann, B. (2017). Frequency and chunking in derived words: a parametric fMRI study. *J. Cogn. Neurosci.* doi: 10.1162/jocn_a_01120. [Epub ahead of print].
- Chiarcos, C., Claus, B., and Grabski, M. (2011). *Salience: Multidisciplinary Perspectives on its Function in Discourse*. Berlin: Walter de Gruyter. doi: 10.1515/9783110241020
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Geeraerts, D. (2000). "Salience phenomena in the lexicon: a typology," in *Meaning and Cognition: A Multidisciplinary Approach*, ed L. Albertazzi (Amsterdam: Benjamins), 79–101. doi: 10.1075/celcr.2.05gee
- Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780195136166.001.0001
- Hanulíková, A., and Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by speaker's gender. *Front. Psychol. Cogn.* 6:1396. doi: 10.3389/fpsyg.2015.01396
- Hanulíková, A., van Alphen, P. M., van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent

- on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103
- Hanulíková, A., and Weber, A. (2012). Sink positive: linguistic experience with th substitutions influences nonnative word recognition. *Atten. Percept. Psychophys.* 74, 613–629. doi: 10.3758/s13414-011-0259-7
- Jaszczolt, K. M., and Allan, K. (2011). *Saliency and Defaults in Utterance Processing*. Berlin: Walter de Gruyter. doi: 10.1515/9783110270679
- Langacker, R. W. (1993). Reference-point constructions. *Cogn. Linguist.* 4, 1–38. doi: 10.1515/cogl.1993.4.1.1
- Rác, P. (2012). Operationalising saliency: definite article reduction in the North of England. *English Lang. Linguist.* 16, 57–79. doi: 10.1017/S1360674311000281
- Roller, K. (2016). *Saliency in Welsh English Grammar: A Usage-Based Approach*. Freiburg: University Library Press. doi: 10.1017/S1360674311000281
- Rosch, E. (1975). Cognitive reference points. *Cogn. Psychol.* 7, 532–547. doi: 10.1016/0010-0285(75)90021-3
- Summerfield, C., and Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409. doi: 10.1016/j.tics.2009.06.003
- 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.

Copyright © 2017 Blumenthal-Dramé, Hanulíková and Kortmann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Salience and Attention in Surprisal-Based Accounts of Language Processing

Alessandra Zarcone^{1*}, Marten van Schijndel², Jorrig Vogels¹ and Vera Demberg¹

¹ Computational Linguistics and Phonetics, Universität des Saarlandes, Saarbrücken, Germany, ² Department of Linguistics, The Ohio State University, Columbus, OH, USA

OPEN ACCESS

Edited by:

Alice Julie Blumenthal-Dramé,
Albert-Ludwigs-Universität Freiburg,
Germany

Reviewed by:

LouAnn Gerken,
The University of Arizona, USA
Stefan Frank,
Radboud University Nijmegen,
Netherlands

*Correspondence:

Alessandra Zarcone
zarcone@coli.uni-saarland.de

Specialty section:

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

Received: 29 February 2016

Accepted: 20 May 2016

Published: 06 June 2016

Citation:

Zarcone A, van Schijndel M, Vogels J
and Demberg V (2016) Salience and
Attention in Surprisal-Based Accounts
of Language Processing.
Front. Psychol. 7:844.
doi: 10.3389/fpsyg.2016.00844

The notion of *salience* has been singled out as the explanatory factor for a diverse range of linguistic phenomena. In particular, perceptual salience (e.g., visual salience of objects in the world, acoustic prominence of linguistic sounds) and semantic-pragmatic salience (e.g., prominence of recently mentioned or topical referents) have been shown to influence language comprehension and production. A different line of research has sought to account for behavioral correlates of cognitive load during comprehension as well as for certain patterns in language usage using information-theoretic notions, such as *surprisal*. Surprisal and salience both affect language processing at different levels, but the relationship between the two has not been adequately elucidated, and the question of whether salience can be reduced to surprisal / predictability is still open. Our review identifies two main challenges in addressing this question: terminological inconsistency and lack of integration between high and low levels of representations in salience-based accounts and surprisal-based accounts. We capitalize upon work in visual cognition in order to orient ourselves in surveying the different facets of the notion of salience in linguistics and their relation with models of surprisal. We find that work on salience highlights aspects of linguistic communication that models of surprisal tend to overlook, namely the role of attention and relevance to current goals, and we argue that the Predictive Coding framework provides a unified view which can account for the role played by attention and predictability at different levels of processing and which can clarify the interplay between low and high levels of processes and between predictability-driven expectation and attention-driven focus.

Keywords: attention, goals, language, predictive coding, predictability, relevance, salience, surprisal

1. INTRODUCTION: THE ATTENTIVE BRAIN AND THE ANTICIPATING BRAIN

The perceptual experience we are continuously subjected to while awake is an “embarrassment of riches” (Wolfe and Horowitz, 2004): for example, when we process a visual scene, we need to focus our maximum visual acuity (the fovea) on the most useful or interesting parts of the scene (Mackworth and Morandi, 1967). In doing so, we are guided by attention: the “attentive brain” filters out the *relevant* information, prioritizing between stimuli and giving certain stimuli a special status, thus easing the processing burden. The stimuli attracting attention are said to be *salient* (literally, “standing out from the ground”, Chiarcos et al., 2011). The notion of *salience* has been widely used in linguistics as the explanatory factor for a diverse range of phenomena:

to indicate a property of a sociolinguistic variable that makes it cognitively prominent and thus noticeable (Trudgill, 1986; Kerswill and Williams, 2002; Rácz, 2013), or a property of discourse entities exploited in anaphoric binding (Grosz et al., 1995; Osgood and Bock, 1977; Prat-Sala and Branigan, 2000), but also, according to a simulation view of language comprehension, the property of prominent entities in the described situation (Claus, 2011).

The *predictability* of the stimulus also affects our perceptual experience. Our brain's ability to anticipate new stimuli is key to its adaptive success (Bar, 2011; Clark, 2013): the "anticipating brain" keeps track of what it has experienced (and how often), adapts to regularities, predicts upcoming stimuli based on recent context, but also detects *surprising* stimuli and reacts to unexpected ones if the predictions go wrong (Ranganath and Rainer, 2003). For example, when looking at a series of static pictures implying motion, people mentally simulate implicit motion, going beyond what they see in the pictures and preparing for what is coming next (Freyd, 1983; Hubbard, 2005). Language is no exception: the linguistic units we process (at different levels: phonemes, words, syntactic constituents) may be expected or unexpected, depending on preceding context. The difference between expected and unexpected stimuli is determined by their frequency and *conditional probability* given preceding context. *Surprisal* is a function of the input's conditional probability given preceding context, corresponding to how predictable the input is, and has been shown to influence processing costs as well as production choices (Hale, 2001; Levy, 2008).

Saliency has been identified with (e.g., Rácz, 2013) or at least related to surprisal / predictability (e.g., Blumenthal-Dramé et al., 2014), and given the success of information-theoretic models of language it would be tempting (and theoretically elegant) to reduce saliency to surprisal. While it is clear that both predictability and saliency(s) affect language processing, the relationship between the two has not been adequately elucidated, leaving the question open of whether saliency can be reduced to surprisal. The main goal of this review is to address this question by disentangling the notions of saliency and predictability and the role they both play during linguistic processing, distinguishing between their cognitive correlates and identifying their interplay.

The first challenge to face is undoubtedly a lack of terminological consistency among linguists: while in visual cognition the term *saliency* refers to bottom-up stimulus-driven perceptual saliency, linguists use the term to refer either to bottom-up, perceptual properties of incongruous stimuli (low-predictability stimuli, expected to require additional processing effort, Hanulíková et al., 2012; Blumenthal-Dramé et al., 2014), or to top-down, discourse-driven properties of accessible, congruous or recently accessed entities (high-predictability stimuli, expected to facilitate processing, Claus, 2011). This inconsistency leads to potentially contradictory hypotheses on the relationship between predictability and saliency (saliency corresponds to low-predictability vs. saliency corresponds to high-predictability).

The second challenge pertains to the interaction between high- and low-level representations involved in language processing. Predictability-based approaches to language

comprehension have shown that high-level information (e.g., what we know about the speaker or the situation) might influence lower-level predictions, at a phoneme or word level. For example, because of our world knowledge including the information that *men do not get pregnant*, when we listen to a man's voice we don't expect him to say he's *pregnant* (van Berkum, 2009). However, the interplay between low- and high-levels of processing and representation has not been explicitly modeled. This interplay becomes more clear if we factor in the role played by attention. For example, people can overlook very unexpected events if they are paying attention to other aspects of the scene: if people are asked to count passes in a basketball video, they will not notice a person in a gorilla costume walking across the scene (inattention blindness effect, Simons and Chabris, 1999). Similarly, if asked *How many animals of each kind did Moses put on the Ark?* (Van Oostendorp and De Mul, 1990) people might be too focused on the high-level task of answering the question to notice that, at the word-level, *Noah* should be in the place of *Moses* (see Sanford and Sturt, 2002, for a review of similar phenomena).

We will argue that the comprehender's attentional focus weights surprisal effects from one level or another, depending on the current goals and on perceived rewards. The Predictive Coding framework (Rao and Ballard, 1999; Friston, 2010; Clark, 2013) provides a unified view which can clarify the interplay between low- and high-levels of processing and between bottom-up, stimulus-driven saliency and top-down, goal-directed attentional control, and has the potential to reconcile low-level computations of surprisal, high-level representations, and goal-mediated attentional control.

We first give a brief overview of studies providing evidence for predictability-driven language comprehension, with a particular focus on recent results from information-theoretic approaches (Section 2). We then address the notion of saliency (Section 3), first by drawing from work in visual cognition and then surveying the different facets of this notion in linguistics, seeking for parallels with visual cognition. We look at visual cognition because predictability and saliency are arguably relevant to many cognitive domains (such as vision and language) and reflect very basic properties of cognition, but also because the field of visual cognition provides us with tools and categories which have been extensively modeled and discussed and have the potential to bring some clarity in the rather contradictory terminology employed in linguistics. We find that work on saliency uncovers aspects of linguistic processing that models of surprisal tend to overlook, namely the role of attention, mediated by the perceiver's category system, by relevance to current goals and by affect. We then focus on recent work in the Predictive Coding framework, and on how surprisal and attention can be understood within this framework (Section 4). Finally we discuss how surprisal models can be extended to account for the role of saliency and attention (Section 5).

2. PREDICTABILITY AND LANGUAGE

Every linguistic stimulus we process comes with a context: for example a visual scene, or a previously processed language

input, or the situation we are in. Depending on previously processed contextual information, a stimulus can be more or less expected. Decades of experimental work in expectation-based approaches to language processing (e.g., Altmann and Kamide, 1999; Trueswell et al., 1994; Elman et al., 2005) have shown that comprehenders draw context-based expectations about upcoming linguistic input at different levels: they build expectations for the next word (Morris, 1994; Ehrlich and Rayner, 1981; McDonald and Shillcock, 2003), but also for their phonological form (DeLong et al., 2005) and gender inflection (van Berkum et al., 2005), for syntactic parses (Spivey-Knowlton et al., 1993; MacDonald et al., 1994; Demberg and Keller, 2008), for discourse relations (Köhne and Demberg, 2013; Drenhaus et al., 2014; Rohde and Horton, 2014), for semantic categories (Federmeier and Kutas, 1999), for typical event participants (Bicknell et al., 2010; Matsuki et al., 2011), for the next referent to be mentioned (Altmann and Kamide, 1999), for the next event to happen in a sequence (Chwilla and Kolk, 2005; van der Meer et al., 2005; Khalkhali et al., 2012), and for typical implicit events (Zarcone et al., 2014). The effects of predictability are measurable, as expectation-matching input facilitates processing, and deviation from expectations produces an increase in processing costs. Predictable words are read faster: they are fixated for less time and are more likely to be skipped than unpredictable words (Ehrlich and Rayner, 1981; Balota et al., 1985; McDonald and Shillcock, 2003; Frisson et al., 2005; Demberg and Keller, 2008); also, the amplitude of the N400 event-related potential increases in a graded way as a function of a word's predictability (Kutas and Hillyard, 1984; Federmeier and Kutas, 1999; Kutas and Federmeier, 2011; Frank et al., 2013).

These and more studies have shown that during language processing comprehenders do not just rely on transitional probabilities between words (McDonald and Shillcock, 2003; Frisson et al., 2005) but exploit various sources of information to narrow down predictions for upcoming input, such as verb subcategorization biases and thematic fit (Trueswell et al., 1993, 1994; Hare et al., 2003, 2009; van Schijndel et al., 2014), verb aspect (Ferretti et al., 2007), but also visual context (Kamide et al., 2003), generalized knowledge about typical events and their participants (Ferretti et al., 2001; Bicknell et al., 2010), knowledge about scenarios (van der Meer et al., 2002, 2005; Khalkhali et al., 2012), discourse markers (Köhne and Demberg, 2013; Drenhaus et al., 2014; Xiang and Kuperberg, 2015), and pragmatic inferences about the speaker's identity and status (van Berkum et al., 2008). These different types of information are drawn upon by language comprehenders at multiple levels of representation (syntactic, lexical, semantic, and pragmatic) at each point in processing to reach a provisional analysis and build expectations at multiple levels based on this provisional analysis (van Berkum, 2010; Kutas et al., 2011; Kuperberg, 2016; Kuperberg and Jaeger, 2016). The flow of information goes both ways: the encountered input activates high-level representations in a bottom-up fashion (e.g., triggering expectations for new syntactic structures, event knowledge, scenarios), and, depending on contextual information, high-level representations influence low-level predictions (Kuperberg, 2016). For example, knowledge about events and their participants cued by previous context (*The*

day was breezy so the boy went outside to fly a...) determines a prediction for a word (... *kite*) but also triggers expectations for a phonological realization of the article against another (*a kite* / *an airplane*, DeLong et al., 2005).

2.1. Models of Surprisal

Information-theoretic notions, such as *surprisal* (Hale, 2001; Levy, 2008), have been proposed to account for the relationship between predictability and processing costs. *Surprisal* is a function of the input's conditional probability given preceding context, corresponding to how predictable the input is and how much information it carries (highly predictable input conveys little information):

$$\text{Surprisal}(\text{linguistic_unit}) = -\log P(\text{linguistic_unit}|\text{context})$$

The surprisal of a word is equivalent to the difference between the probability distributions of possible utterances before and after encountering that word (Kullback-Leibler divergence), quantifying the amount of information conveyed by that word (Levy, 2008). Surprisal Theory has sought to account for certain patterns in language usage as well as for behavioral correlates of cognitive load during comprehension, with the underlying linking hypotheses that cognitive load is proportional to the amount of information conveyed by the input (its surprisal) given preceding context, and that the speakers' production choices tend to keep the amount of information constant (*Uniform Information Density Hypothesis*, Jaeger and Levy, 2007, see also Jurafsky et al., 2001; Gahl and Garnsey, 2004). Surprisal can be modeled at different levels (phonemes, phrases, words) and is often estimated using relatively simple statistical models such as *n*-gram language models or Probabilistic Context-Free Grammars (Hale, 2001; Demberg and Keller, 2008; Frank, 2009; Roark et al., 2009). A word's surprisal has been shown to correlate with its reading time (Hale, 2001; Demberg and Keller, 2008; Levy, 2008; Fossum and Levy, 2012; Smith and Levy, 2013; van Schijndel and Schuler, 2015) and with the amplitude of the N400 at the word (Frank et al., 2013).

2.2. Limitations of Models of Surprisal

A surprisal-based model is typically defined by the linguistic units it takes into consideration and by what level it can condition on. Typically, surprisal-based models do not tackle the problem of how different levels of representation interact with each other, as the probability of a linguistic unit (e.g., a phoneme, a phrase, a word, a situation model) is conditioned on the preceding units at the same level (e.g., preceding phonemes, phrases, words, situation models). Comprehenders, though, exploit information at different levels to build expectations for upcoming input. There have been some attempts at integrating surprisal estimates with a model of semantic surprisal (Mitchell et al., 2010; Frank and Vigliocco, 2011; Sayeed et al., 2015), but not a unified account showing how the probability of lower-level units (e.g., perceptual features) can be conditioned on higher-level units (e.g., situation, world knowledge) to predict processing costs, or how to exploit higher-level information to predictively pre-activate information at lower levels of representation (Kutas et al., 2011; Kuperberg,

2016). We will argue that such an account should include the role played by attention in shifting the focus between different levels to determine at what level surprisal influences processing costs.

Surprisal-based models rely on the linking hypothesis that high surprisal corresponds to high processing costs. But does this relationship between surprisal and processing cost always hold? Kidd et al. (2012) have shown that infants focus their visual attention to sequences whose complexity (surprisal) is neither too low nor too high, but *just right*, that is, it falls within certain optimal complexity margins (this effect is known as the *Goldilocks effect*). Arguably, some sort of Goldilocks effect also affects the attention of adult comprehenders, who react to extreme values of the complexity/predictability spectrum by diverting their attention from extremely complex stimuli that is too demanding or unpredictable (for example, when they are pushed beyond their memory capacity, see Nicenboim et al., 2015, or when they hear a foreign language), or from extremely predictable stimuli. For example, utterances about very predictable events (“*John went shopping. He paid the cashier*”) may trigger pragmatic inferences (*John is a shoplifter*, Kravtchenko and Demberg, 2015), simply because we expect our interlocutors to be informative (if they think it’s worth mentioning that *John paid the cashier*, it must be an exceptional event). Also, as noted by van Berkum (2010), “predictions are even useful when they are wrong”: less expected (marked) combinations (e.g., a cleft sentence construction) may be a way of marking the delivery of a message as worthy of extra attention, thus easing the processing burden on an otherwise surprising stimulus. Previous context may also lead the hearer to expect surprise, e.g., *You’ll never believe it! The thing John was brushing his teeth with was a knife the day before yesterday.* (Futrell, 2012).

A third point concerns the relationship between the model we use to estimate surprisal, and the input’s probability of occurrence in the world. As observed by Pierrehumbert (2006), (log-)frequencies of occurrences, while going a long way in explaining processing costs, do not tell us the whole story: between the frequencies of events and the frequency of memories, “lies a process of attention, recognition, and coding which is not crudely reflective of frequency.” What we store in our memory, and then exploit in expectation-based processing, depends on where our attention is focused, on what stimuli we consider relevant but also on what valence we associate with them. We will argue in Section 4 that we need to factor in the role played by the affect system, that is the neural circuitry that processes valence in the brain, to fill the gap between probability distributions of events in the world and our memory’s probability distributions.

2.3. Bayesian Surprise and the Snow-Screen Paradox

Surprisal does not quantify how useful or relevant the stimulus is, but solely how predictable it is. Itti and Baldi (2009) introduced a Bayesian theory of *surprise*, which weights the predictability of a stimulus by its usefulness or relevance, determining how unexpected we perceive the stimulus to be. The observer’s background beliefs (for example, the probability of seeing *CNN* or *BBC* when turning on the TV) are represented as a prior

probability distribution, which is updated using Bayes’ theorem as new observations are made (e.g., *CNN is on*). *Bayesian surprise* is the difference (Kullback-Leibler divergence) in the belief distribution before and after an observation, indicating how much the observation changed our beliefs about the world. If *CNN* is the most expected outcome given our prior beliefs, when we turn on the TV and see *CNN* the surprise will be minimal. If *BBC* is shown instead, there will be a small amount of surprise and a subsequent belief update. Every subsequent change on the screen (a newscaster’s mouth moving, a commercial break) will also update our beliefs and thus our predictions about upcoming TV content accordingly.

Itti and Baldi (2009) illustrate the difference between *surprisal* and *surprise* using the so-called “snow-screen paradox”: if a random pixel pattern (known as *snow* or *static*) appears when we turn on the TV or while we are watching it, we will be highly surprised, because this outcome is extremely unexpected. At a high level, our belief that the snow would appear was very low (high surprise). At a low level, the pixel configuration before the snow would not have helped us predict the random black-and-white pixel configuration when it first appeared (high surprisal). Also, the snow is interesting at a high-level, because it signals a malfunction, so, after observing it, we will experience a large shift between prior and posterior distributions, strongly favoring the snow against other channels. But if the snow persists after the belief update, it is no longer interesting, because it is now the most expected outcome based on our updated belief (low surprise). At a pixel level, though, the snow frames are still continuously changing at random, making it impossible to predict the status of any pixel at any moment (high surprisal). In Itti and Baldi’s words (2009, p. 1297), “random snow, although in the long term the most boring of all television programs, carries the largest amount of Shannon information” (that is, surprisal). Bayesian surprise differs from surprisal in that it quantifies the *belief update* of the model given the observation, whereas surprisal quantifies how much information the observation conveys (how predictable it is) given a current model, without taking into account a model update.

Griffiths and Tenenbaum (2007) also argue that surprisingness / interestingness rather than mere low probability determines the difference between a simply unlikely event and what we consider to be a coincidence: a coincidence (e.g., many coin flips, all turning out to be heads) is not only an unlikely event, but it is an event which is *less* likely under our currently adopted explanation for the observed state of things than under an alternative explanation (*the coin is unfair*, or *the person flipping the coin can magically control it*), which nevertheless does not have enough support to be adopted through a belief update. If interesting coincidences continue to occur, and if we pay attention to them, then the coincidence can turn into evidence and the alternative hypothesis can be supported via a belief update.

The snow-screen paradox shows that the level of representation that is most relevant to us determines how affected we are by one outcome or the other, and so does our category system: the snow is only interesting at its onset insofar as it signals a malfunction, but its random pixel changes have

no relevance for us. If the observer neither understands English nor knows about different English-speaking channels, both *CNN* and *BBC* are categorized as *TV channels I don't understand*, and it makes very little difference in her belief update which one is showing. Similarly, language learners initially filter the L2-input (and try to build predictions about it) using the categories in their L1, which in turn determine what is surprising in the L2-input and what is not. Also, they rely heavily on L1-L2 similarities, for example by exploiting overlapping categories in the lexical aspect domain or in the grammatical aspect domain (depending on what dimension is marked in their L1) in learning the tense-aspect system of the new language (Izquierdo and Collins, 2008; Shirai, 2009). Learners do not pay attention to the *snow* in L2, that is to stimuli that are highly unpredictable to them because they are beyond their level, but focus on stimuli which they have a meaningful category for (see also Palm, 2012).

In a similar vein, Relevance Theory (Sperber and Wilson, 1986; Wilson and Sperber, 2004) argues that comprehenders are driven by a search for *relevance*, under a *presumption of optimal relevance*. As the goal of comprehension is to construct a plausible hypothesis about the speaker's meaning, stimuli are optimally relevant if and only if (1) they are compatible with what we know of the communicator's abilities and preferences and (2) they are worth the audience's processing effort, because they contribute to confirming or correcting our hypotheses about the speaker's meaning (Wilson and Sperber, 2004). Stimuli that are not relevant enough or that do not yield any cognitive effect (that is, do not confirm a hypothesis or correct a mistaken assumption about the speaker's meaning) are disregarded as not worth the processing effort. *Snow* stimuli are not worth the processing effort as they do not have any effect in confirming or correcting our hypotheses.

Summary

Predictability-based models have been very successful in accounting for processing costs during language comprehension, but (at least in their current implementations) they seem to have overlooked some aspects of linguistic processing, which suggest that the unexpectedness of a stimulus may not be the only factor determining how useful, interesting or difficult the stimulus is. In the next section, we will pinpoint these aspects in terms of salience and attention. In order to do so, we will first clarify some terminological issues related to salience in linguistics and its relation with predictability.

3. SALIENCE IN VISION AND SALIENCE IN LANGUAGE

Salience is a widely used term in linguistics, often referring to very different aspects of language comprehension and production (Chiaros et al., 2011; Blumenthal-Dramé et al., 2014), such as the *acoustic salience* of the linguistic input (Rácz, 2013) or of the *visual salience* of a scene during language-relevant tasks (Kelleher, 2011), but also the *discourse salience* of referents (Osgood and Bock, 1977) or the salience of entities in the described situation (*simulation-based* or *situation-based salience*

Claus, 2011). As with visual cognition, language understanding also seems to be influenced by low-level properties (of the visual scene or of the linguistic stimulus) and by high-level conceptual representations and goals. While in visual cognition salience is mainly used to refer to perceptual salience driven by low-level visual properties, in linguistics the same term is used to refer to two potentially contrasting properties of the stimulus (Blumenthal-Dramé et al., 2014): for example, *acoustic salience* is typically meant to be a low-level perceptual property of the signal (depending on its transitional probabilities), attracting attention in a bottom-up fashion as *visual salience* does, whereas *discourse* and *simulation-based salience* typically exert a top-down influence which makes certain upcoming input more expected.

This terminological inconsistency is not completely unmotivated, as we will see in Section 3.3, but it leads to an apparent paradox when it comes to linking these models to measures of processing cost and to relating salience to predictability. Bottom-up salience, being a property of low-predictability stimuli, is expected to require additional processing effort (Hanulíková et al., 2012), whereas top-down salience, being a property of accessible, high-predictability or recently accessed entities, is argued to facilitate processing (Claus, 2011). We will now address this inconsistency by capitalizing on work on visual search in order to clarify the relationship between predictability and salience.

3.1. Salience in Visual Cognition

Attention is a cognitive necessity: the amount of information our optic nerve receives¹ far exceeds what our brain can process and transform into conscious experience. Attention filters out the relevant information, easing the processing burden (Wolfe and Horowitz, 2004; Awh et al., 2012). Attention is also an evolutionarily beneficial trait: our survival depends on our ability to filter and prioritize useful or interesting parts of our perceptual experience (attention-capturing or *salient* parts) over overtly predictable or uninteresting ones, in order to quickly identify and react to potentially dangerous or rewarding stimuli. Research in visual cognition has long focussed on pinning down factors that drive attention (Mackworth and Morandi, 1967; Loftus and Mackworth, 1978), and has identified two main components of attentional deployment (see Itti and Koch, 2000, for a review): a bottom-up, fast mechanism based on the stimulus *salience* and a slower, top-down mechanism based on goals and tasks.

Salience or *saliency* is defined by early features of the visual stimulus, such as color, intensity and orientation, which are claimed to drive preattentive selection (Koch and Ullman, 1985; Itti and Koch, 2000), determining effects such as the *pop-out* effect (observed when a target stimulus differs from its background distractors on at least one feature dimension). Itti and Koch (2000) describe a computational model of preattentive selection based on *saliency maps*, where each unit is activated based on low-level perceptual features and the competition among active units determines a single, winning location (the most salient one), predicting the location of gaze; the winning

¹On the order of 10⁸ bits per second, (Itti and Koch, 2000)

location is then promptly inhibited and a new winning location is chosen, predicting gaze at the next step, so that the map is able to scan the visual input by visiting different parts in a sequential fashion. Bruce and Tsotsos (2009) move from the idea that efficient sampling should focus on the areas maximizing information, and define salience in information-theoretic terms, as local information (how informative / unexpected the content of a region is, based on surrounding context). Salient parts of the stimulus are outliers (Tatler et al., 2011), deviating from the surrounding area, and are prioritized by efficient sampling strategies as they carry the most information.

Salience is a good predictor of gaze during free visual search, but top-down factors such as current goals, task relevance and rewards (Folk et al., 1992; Yarbush, 1967; Hayhoe and Ballard, 2005) and recent selection history (see Awh et al., 2012, for a review) have been shown to influence gaze and attention in performance of a task and in presence of real-world scenes with clear semantic content, competing with and prevailing over bottom-up attention capture (Folk et al., 1992; Chen and Zelinsky, 2006). The computational model in Rao et al. (2002) captures such top-down effects by computing salience as a function of the similarity between the low-level perceptual features of the stimulus and a search target, creating a *top-down saliency map*. Top-down factors pose the problem of modeling local and global sources of information within the same framework (e.g., Navalpakkam and Itti, 2005; Torralba et al., 2006; Zelinsky et al., 2006), finding a suitable interaction between bottom-up models such as the salience-based model in Itti and Koch (2000) and top-down ones such as the target-based model in Rao et al. (2002).

Torralba et al. (2006) argue that a holistic representation of scene context needs to be taken into account when modeling gaze in search tasks on real-world scenes: their Contextual Guidance Model combines low-level saliency and global high-level and context features (e.g., scene priors and tasks) to create a *scene-modulated saliency map* selecting fixation sites. Similarly, Henderson et al. (2009) show that visually non-salient targets in expected locations are found more easily than salient regions that are not likely target locations. According to their Cognitive Relevance Framework, visual search is guided top-down by cognitive relevance, that is by the need of the cognitive system to make sense of the scene (based on task, semantic knowledge about the type of scene and episodic knowledge about the particular scene being viewed): objects will be prioritized depending on current information-gathering needs over their low-level visual salience.

Work in visual cognition has shown that the stimulus in itself can capture the perceiver's attention if it *pops out* from the background due to its low-level perceptual features (its visual salience), carrying information given its surround. Top-down factors such as the perceiver's goals, the features of a search target, relevance to the task, recent selection history, and cognitive relevance (prior semantic knowledge about the scene and expected objects) can override bottom-up factors in determining what locations capture attention. Linguistic salience can also be defined as a property of linguistic stimuli "standing out" from a ground. We will now show how this term has

been used in linguistics to refer to both low-level attention-capturing properties of the stimulus and to top-down activation of contextually-relevant elements.

3.2. Linguistic Salience as a Stimulus-Specific Property

A common use of the term *salience* in linguistics indicates a property of a sociolinguistic variable that makes it cognitively prominent (Trudgill, 1986; Kerswill and Williams, 2002). For example, Definite Article Reduction (DAR) in North England is the realization of the definite article as a glottal stop before consonants and vowels, which is cognitively salient (noticeable) to a speaker of a different variety of English (Rácz, 2013). What makes a variable in dialect *D* noticeable to a speaker of dialect *D'* is not its frequency per se, but a notable relative difference between its occurrence in *D* and its occurrence in *D'* that makes the variable "stand out." A speaker of *D'* would not commonly expect a glottal stop between vowels or before a stressed vowel: the DAR occurs in positions in *D* where it is much less likely to occur in *D'*, and therefore has a low transitional probability (large surprisal) for a speaker of *D'*. A variable that has cognitively salient realizations can, in turn, be a marker of social indexation, becoming socially salient.

These studies indicate that transitional probabilities may guide attention by selecting interesting parts of the acoustic signal, which crucially are those with high surprisal / high information content. Similarly, marked (and less frequent) prosodic or syntactic constructions (Lambrecht, 1994) can be used by the speaker to direct the listener's focus on a part of the signal, emphasizing it by way of the low predictability of the construction (e.g., *It was Moses who put two animals of each kind on the ark*, see also Givón, 1988). Acoustic salience and syntactic focus are low-level properties of the linguistic signal that capture the hearer's attention in a bottom-up fashion (similarly to *pop-out* effects in visual cognition) and that depend on the transitional probabilities of the relevant segments, that is on their *surprisal*. Identifying linguistic salience with surprisal is a tempting and, arguably, a theoretically elegant option. Salience in linguistics, on the other hand, has also been used to indicate aspects of processing that are not as easily accounted for by models of predictability and that we will now review.

3.3. Linguistic Salience as a Situation-Driven Property

The term *salience* has been used in linguistics not only to refer to the property of a stimulus that stands out from a perceptual ground, but also to qualify entities that are prominent in the discourse model or the situation and influence comprehension in a top-down fashion, as in the case of *discourse salience* and *situation-based salience* (also referred to as *semantic-pragmatic salience*, see also Giora, 2003). The idea behind these notions of salience is that, when understanding language, comprehenders maintain in their working memory a model of the evolving discourse context (Kamp, 1981; Asher, 1993; Kamp and Reyle, 1993; Grosz et al., 1995; Lascarides and Asher, 2007) or, in a simulation-view of language comprehension, they run a mental

simulation of the described situation (Zwaan and Radvansky, 1998). If perceptual attention is necessary because we cannot focus on every aspect of the stimulus simultaneously, here the focus is on a different cognitive necessity, that is the limited capacity of our working memory: “only a few elements of the situation are available at any one time, that is the most salient ones at a particular time during processing” (Claus, 2011). Salience is then accessibility in the discourse or situation model. High-accessibility entities are available for anaphoric binding and are likely to be mentioned in upcoming context (Grosz et al., 1995; Osgood and Bock, 1977; Levelt, 1989; Vogels et al., 2013). Discourse- and situation-based salience drive top-down predictions (derived from high-level information, be it the discourse model or the situation model) for what is going to be mentioned next, that is high-predictability entities.

Several factors may make an entity cognitively accessible / salient. An entity may be accessible because it is perceptually available in the shared visual context (Kelleher, 2011, see Section 3.4), because it is mentioned (and possibly highlighted) in discourse² (for example, if it is the subject, Vogels et al., 2013), or because of a mental simulation of the described situation. Consider this example discussed by Claus (2011):

1. *John was preparing for a marathon in August. After doing a few warm-up exercises, he put on / took off his sweatshirt and went jogging. He jogged halfway around the lake without too much difficulty.* (Glenberg et al., 1987).

In the first version (*put on*), the *sweatshirt* is still part of the situation involving John at the end of the story (it is part of the Here and Now of the protagonist, Claus, 2011), whereas in the second version (*took off*) it is not: the entity's accessibility depends on the situational representation. The Here and Now of the protagonist does not only include what is visible to her, but also what she can act upon, what is relevant to her goals and to her mental state (see also Carreiras et al., 1997; Radvansky and Curiel, 1998; Zwaan et al., 2000; Borghi et al., 2004), and determines which elements are accessible and likely to be mentioned next.

Situation-based salience can drive predictions that are different than those coming from lower-level representations. Consider the following examples:

1. *For breakfast the boys / the eggs would only eat / bury toast and jam.* (Kuperberg et al., 2003).
2. *A huge blizzard ripped through town last night. My kids ended up getting the day off from school. They spent the whole day outside building a big snowman / towel / jacket in the front yard.* (Metusalem et al., 2012).

As in visual cognition, when the context evokes a clear scenario (the *breakfast* scenario, the *playing in the snow* scenario), relevant elements, perfectly congruent with the scenario, are activated (*eggs* and *eating* in the first, *snowman* and *jacket* in the second). In one case, though, the scenario-fitting

element (*the eggs would only eat* and *building a big jacket*) does not fit the verb's selectional preferences: the higher-level predictions coming from the scenario are incompatible with lower-level predictions coming from the lexical semantic level. The congruity with the scenario reduces the N400 effect, which is evoked by a semantic violation due to the scenario-incongruent element (*They spent the whole day outside building a big towel*) and by a verb which is not supported by context (*For breakfast the boys would only bury*). High-level salient representations are activated and generate predictions for upcoming input even when they would be an anomalous continuation from the lower, lexical-semantic level of representation.

High-level predictions depend on generalized knowledge about real-world events and their typical participants, which is acquired both from first-hand participation or from second-hand experience (including language) and stored in our long-term memory (McRae and Matsuki, 2009). An interesting open question, in line with the discrepancy between frequency of events and frequency of memories which we brought up in 2.2, is how we map between our experience of these events and our representations. When we experience people *making coffee*, inferring the protagonist's goals and intentions may be as important as observing what things typically happen in the sequence. We might remember better to *use filtered water* rather than tap water if we know that the point is to avoid limestone deposits in our coffee machine: knowing *why* (inferring goals) may help us remember *what* is part of the scenario, making a difference between an uninteresting detail in the scenario and a relevant, even if infrequent, step in the process. Between experience and memory there is again a process of “attention, recognition, and coding,” mediated by the affect system (see Section 4) and shaped by hypotheses about what is relevant to us and to other people, that shapes our memory's probability distributions. Current models of surprisal, which work on the *linguistic signal* as it is, currently lack a mechanism to weight certain aspects of the signal more than other.

We have classified existing notions of salience in linguistics into two main categories, while also clarifying how they relate to predictability-driven language processing: stimulus-specific attributes, which attract the comprehender's attention in a bottom-up fashion, and situation- and discourse-driven accessibility of entities, which guides the comprehender's top-down predictions for upcoming stimuli. These two categories have something in common: they are properties of entities “standing out” from a ground (perceptual in one case, cognitive in the other) and are properties we rely on to deal with limitations of our cognitive resources (attention in one case and working memory in the other). Nevertheless, salience as a stimulus-specific property is characterized as high surprisal, whereas entities which are salient with regard to the discourse or to the situation are highly predictable (low surprisal). We will now clarify how one type of salience may influence the other and interact with visual salience, and we will then explain the interaction between bottom-up focus and top-down predictions.

²Arguably, highlighting an entity through syntactic focus affects its bottom-up salience. The acquired focus will then cause the entity to be salient in the discourse model, exerting a top-down influence on predictions, see also Section 3.4.

3.4. Interactions between Bottom-Up Visual and Linguistic Saliency and Situation-Driven Saliency

Given that language comprehension and production often take place within a non-linguistic, perceptual context, predictions in language processing will in many cases be shaped by a combination of linguistic and visual saliency. Indeed, there is ample evidence that speakers and listeners use stimulus-based properties of the visual environment in language planning and processing (e.g., Clark et al., 1983; Tanenhaus et al., 1995; Coco and Keller, 2009; Koolen et al., 2015). It is less clear how stimulus-specific visual cues interact with either bottom-up linguistic saliency or with top-down situation-driven saliency. Results from scene description experiments have suggested that visual cues can tap directly into the lexical-syntactic representation of the sentence, allowing them to interact with the lexical accessibility of a reference to an entity (e.g., Tomlin, 1997; Gleitman et al., 2007). More recent studies (e.g., Vogels et al., 2013; Coco and Keller, 2015), however, corroborate the view that visual cues only play a role in the high-level global apprehension of the scene, which in turn affects lower (lexical-syntactic) levels of linguistic processing (Griffin and Bock, 2000; Bock et al., 2004). Hence, stimulus-driven visual saliency influences the situation model, but only situation-driven saliency in turn affects linguistic formulation.

In this view, low-level visual features help “set the scene,” using attention to filter out what is important or relevant information. In language production, this influences how information is structured in an utterance (e.g., what is mentioned first). In language comprehension, visual saliency cues may be used to give weight to an entity (provided the listener has access to the same visual environment as the speaker), so as to adjust predictions about what will be mentioned next. Hence, what starts as a perceptual bottom-up, high-surprisal cue can become a top-down, high-predictability cue: a visually salient entity pops out as surprising, which gives it a salient status within the situation model; next, the mental representation of the salient entity will be highly accessible by virtue of its high news value. Consequently, this entity will be likely to be mentioned, and hence is predictable. Saliency is thus a way to describe what is in the current focus of attention, even though in one stage of processing this attentional focus may be due to a bottom-up surprising stimulus, whereas in a later stage of processing the same stimulus may be in focus because it is now highly predictable.

Top-down predictions arising from low-level visual cues may interact with predictions coming from other sources. For example, bottom-up linguistic saliency can also focus attention on a certain entity, as when it is marked as new information or as ‘in focus’ (in the information structural sense, as in “Once upon a time there was a *girl*”). As pointed out in Section 3.3, this may influence top-down accessibility at different levels of representation (situation-level, discourse-level, lexical-syntactic). In turn, each level of representation sprouts its own predictions and production choices, such as ‘which topic will be discussed next?’ (situation level) or “what linguistic form is appropriate here?” (lexical-syntactic level). These predictions may be either in

line or in conflict with predictions induced by the visual context (e.g., when the girl is either very visually prominent or not at all), and hence may lead to reduced or increased processing cost, respectively. In addition, linguistic saliency cues from different levels of representation may be either in line or in conflict with each other, which may show up as a modulation in correlates of processing cost (as with the *breakfast-eggs* example).

In general, when multiple saliency cues from different sources (visual, linguistic, bottom-up, top-down) can potentially be used to weight parts of the perceptual input, they may affect language planning and processing in different ways: they may influence either the same level or separate levels of processing, and their combined influence may show up as interactive or additive effects, or one cue may override the others. Hence, the effect of bottom-up saliency on processing difficulty and production choices can either be boosted or tempered by the integration with other stimulus-based cues or situation-driven predictions. Crucially, whether one cue takes precedence over another is highly dependent on current task goals. For example, visual saliency may play a different role in an object naming task than in a memorization task or a visual search task, because different parts of the scene will be relevant in each task (Coco et al., 2014; Montag and MacDonald, 2014). Comprehenders will also use their beliefs about the speaker’s intention to guide their focus of attention.

In sum, comprehenders’ predictions as well as speakers’ production choices are influenced by different stimulus-based and situation-based saliency cues at different levels of processing: saliency on a situation-model level may influence predictions about the likelihood of mention of an entity, while local linguistic predictions, such as which lexical form will be used, may be influenced by saliency on a more local, lexical-syntactic level (Kaiser and Trueswell, 2008; Vogels et al., 2013). At the same time, low-level, stimulus-based saliency (surprisal) may also exert an influence on high-level, situation-model saliency, resulting in a complex interplay between predictions at different levels of representation. Finally, the weighting of all those different saliency cues will be highly dependent on task goals and speaker intentions.

Summary

Work in visual cognition has shown that the stimulus low-level perceptual features (its visual saliency) as well as top-down factors (goals, tasks, cognitive relevance) determine what locations capture attention. Saliency-based approaches to language do not typically tackle the interaction between stimulus-specific properties of the linguistic signal and discourse- and situation-based saliency, often adopting a misleading terminology by calling both *saliency*, and ultimately are not explicit with regards to the relationship between saliency(s) and surprisal. We have shown that some aspects of linguistic saliency (e.g., acoustic saliency, markedness of prosodic or syntactic constructions), which capture the comprehender’s attention in a bottom-up fashion, can be easily conflated with surprisal, but discourse- and situation-based saliency cannot, as they are deeply intertwined with goals, tasks, and attention.

Predictability-based approaches go a long way in accounting for processing costs, but current surprisal-based models of language comprehension do not include a mechanism to focus on relevant levels of representation or on relevant parts of the stimulus based on the comprehender's task or on the recognition of the speaker's or the protagonist's goals. We will now review the Predictive Coding framework, illustrating how high- and low-level representations can influence expectations at the relevant level of processing, how top-down information can focus attention to particular stimuli and how stimulus properties can in turn capture attention and influence top-down predictions, and how attention, goals, and saliency can be reconciled with surprisal.

4. THE PREDICTIVE CODING FRAMEWORK

Early studies in visual cognition argued that “perception is no passive sampling from external events” (Mackworth and Morandi, 1967) and that there is “no perception without recognition” (Hake, 1957). With the Predictive Coding framework (Rao and Ballard, 1999; Friston, 2010; Clark, 2013) cognitive science completed a paradigm shift from the view of the brain as a “transformer of ambient sensations into cognition” to “a generator of predictions and inferences that interprets experience” (Mesulam, 2008, p. 368). Predictive coding is fully compatible with the results from predictability-based approaches to language reviewed in Section 2 and has been argued to be the most appropriate framework to shed light on the interaction between high- and low-level representations in prediction-driven language comprehension (van Berkum, 2010; Kuperberg, 2016; Kuperberg and Jaeger, 2016). Additionally, we argue that it provides a unique way to integrate surprise, surprisal and attention, and is thus an ideal candidate to model the interplay between saliency and predictability.

In the Predictive Coding framework, the brain is conceptualized as a hierarchical architecture in which high- and low-level representations can influence predictions for expected input, and top-down models predict the flow of sensory data by modeling the source of the sensory input, that is by actively generating a representation of the upcoming input before perceiving it. The information flow is bidirectional: perception involves explaining away the sensory input by cascading predictions from high-level units down to lower-level units, generating the desired activity in the units, and then matching the predictions against the input and transmitting only the prediction error back to the higher levels. The prediction error or surprisal is the mismatch between the expected representation and the perceived representation. For example, if we are watching a video, our brain prepares for the next frame by predicting a representation of the figure in motion in the next stage of its movement. If the next frame depicts the expected continuation of movement, then the prediction error will be low, if the motion is interrupted, or changes trajectory, or if the frame shows something completely unexpected, then the prediction error will be high. Perceptually similar items and

items that tend to occur in similar contexts will share a high degree of similarity in their representations. The prediction error is transmitted by dedicated “error units” and is used in turn to adjust future predictions to better match the input, resulting in a continuous cycle of prediction and error correction (Rao and Ballard, 1999).

The brain attempts to minimize prediction error, through perception, action and attention. *Perception* minimizes prediction error by trying to infer the nature of the signal source from the varying input signal and extracting repeating patterns and statistical regularities from its environment, guided by the statistical history of events in our environment, and *action* is used by the observer to move the sensors to resample the world by actively seeking expected stimuli (for example, by moving the body so to receive a better signal). But not all error-unit responses have the same weight: *attention* is a means to weight reliable / relevant error-unit responses more than non-reliable / irrelevant ones (Clark, 2013). We will now see how the brain encodes prediction as well as how it can use top-down information to inhibit bottom-up information, maximizing attention to task-relevant stimuli and suppressing task-irrelevant ones.

4.1. Neural Correlates of Top-Down and Bottom-Up Processes

Communication in the brain occurs through neural firing, but, in order to parallelize operations, the brain operates multiple simultaneous communication channels at different firing frequencies (*frequency-division multiplexing*). Bottom-up information from perceptual stimuli is generally thought to be processed using high-frequency brain waves, such as those found in the gamma band (30–100 Hz; e.g., Roux and Uhlhaas, 2014). Top-down information, on the other hand, is generally thought to be stored as low-frequency brain waves, as in the theta (4–7 Hz) or alpha (8–12 Hz) bands, and several studies have suggested that lower frequencies serve to gate higher frequencies as a top-down control mechanism (e.g., Klimesch et al., 2007; Sauseng et al., 2010; Jensen et al., 2012; Roux and Uhlhaas, 2014).

Theta-band frequencies are thought to provide top-down envelopes that modulate the activation of bottom-up sequential information (Lisman and Buzsáki, 2008; Sauseng et al., 2009; Holz et al., 2010; Roux and Uhlhaas, 2014). Essentially, the phase of the lower frequency encodes sequence positions, so when a high-frequency encoding of a stimulus is associated with a particular phase angle (sequence position) in the low-frequency signal, a corresponding association is made between the given stimulus and the selected sequence position. During each phase angle of the low-frequency brain wave, the amplitude of any associated bottom-up neural firing is boosted, producing a stronger signal for that percept. This mechanism, where the phase of a given frequency modulates the amplitude of a higher frequency, is called *phase-amplitude coupling* and uses frequency-division multiplexing to distinguish separate operations and time-division multiplexing to distinguish separate items (that is, each item corresponds to a separate point in the low-frequency phase).

In contrast to sequence-based prediction, perceptual saliency is controlled by phase-amplitude coupling between gamma-band and alpha-band frequencies (Jensen et al., 2002; Klimesch et al., 2007; Sauseng et al., 2009; Bonnefond and Jensen, 2015). Alpha-band waves generally inhibit other neural activation, so at the peak of an alpha wave, other signals can be completely suppressed. As the alpha wave transitions to a lower-power phase of its cycle, it exerts less inhibitory influence on other signals and can reveal those signals it would otherwise suppress (Klimesch et al., 2007; Jensen et al., 2012). Conversely, as the alpha wave transitions back to its peak, other signals will become increasingly (re-)suppressed, which can produce an effect known as *attentional blink*, whereby having an alpha-band signal at a certain phase can inhibit or completely suppress processing of a stimulus such that the subject will not perceive the stimulus at all (Raymond et al., 1992; Olivers, 2007). Subjects seem to exploit this mechanism by adjusting the phase and power of their alpha waves in response to bottom-up observations, maximizing exposure to task-relevant stimuli and maximally suppressing task-irrelevant distractors (e.g., Worden et al., 2000; Sauseng et al., 2005; Mathewson et al., 2009; Bonnefond and Jensen, 2012, though see Firestone and Scholl, 2015, for a dissenting review).

Phase-amplitude coupling thus uses the phase of top-down low-frequency control signals to increase the activation of select bottom-up high-frequency information signals, which literally increases the importance (saliency) of those signals. Therefore, the communication frameworks that underlie our neurological operations seem to rely on simultaneous but distinct top-down and bottom-up processing signals, which can be independently measured during processing. For example, a future study might test how the N400 is modulated by varying target predictability (measurable by theta-gamma phase-amplitude coupling) and by varying the amount of target perceptual saliency (measurable by alpha-gamma phase-amplitude coupling) afforded by the chosen task. Such a study would not have to rely on a priori, extrinsic measures of predictability (e.g., computed from *n*-gram statistics or incremental parsers) or saliency (e.g., the number of words since a previous referent mention) but could instead model the actual probability and saliency of each target and determine how those factors (as actually manifested during the experiment) influence processing.

Phase-amplitude coupling has already provided some support for the Predictive Coding framework (in addition to a wide array of other neurological evidence; see Lewis and Bastiaansen, 2015, for a review of evidence from other neural measures). Intracranial electroencephalography (iEEG) studies (e.g., Zion Golumbic et al., 2013; Fontolan et al., 2014) have shown that top-down neural firing entrains to task-relevant auditory input, amplifying relevant input while suppressing irrelevant input. These results also suggest that top-down attention in auditory association cortex is modulated as a function of bottom-up information from primary auditory cortex. Thus, top-down frequencies tune attention by focusing on aspects of bottom-up input that are made relevant both by the task and by accumulated sources of prediction error.

4.2. Attention and Goals

Attention balances the interaction between top-down predictions and bottom-up influences, weighting reliable / useful sources of prediction error more, and ultimately determining what levels and what parts of the stimulus are relevant at each moment. Attention is thus an ideal candidate to switch between levels of processing, which can account for a number of task- and goal-related effects in language comprehension.

Experimental work has shown that task influences the level of processing: Chwilla et al. (1995) contrasted a lexical decision task (*is the target a Dutch word?*) and a physical task (*did the target appear in uppercase?*) and observed a semantic priming effects (on the N400 and on reaction times) only when the task required accessing word meaning level (lexical decision task). Rayner and Raney (1996) showed that frequency effects found in a reading task disappeared if participants were given the task of searching for a target word in the text, while in Kaakinen and Hyönä (2010) and Schotter et al. (2014) the effect of frequency was instead increased in a proofreading task compared to a reading-for-comprehension task. Schotter et al. (2014) additionally showed that the size of the frequency effect increased in the proofreading if misspelled words were non-words, while the size of the predictability effect increased if the relationship between words was crucial to identify spelling errors (that is, if misspelled words happened to be real words and the spelling mistake was only revealed by context). Xiang and Kuperberg (2015) contrasted a reading-for-comprehension task and a coherence rating task, showing that the coherence rating task facilitated a deeper situation-level representation of context and subsequent prediction of upcoming words. Tasks and goals determine what level we pay attention to, which level is relevant in the architecture and ultimately how detailed and specified our predictions are.

4.3. Attention and Affect

Both the ability to predict what comes next and the ability to focus our attention on relevant stimuli are evolutionarily beneficial traits. The interoceptive and exteroceptive sensations perceived by our body (*affective* bodily changes, Barrett and Bar, 2009; Craig, 2009) determine the *valence* of perceived stimuli, that is their being perceived as pleasant and rewarding or painful and dangerous, which is possibly even more important for our survival. Valence is arguably also involved in language processing: van Berkum (2010) argues that language use, being an instance of social interaction, is entrenched in valence and affect, which arguably are part of the representations of not only emotionally-loaded lexical items, such as *abortion* or *euthanasia*, but of all lexical semantic content which is grounded in experience. The affect system is the neural circuitry that processes valence, and includes a broad set of cortical and subcortical brain areas such as the amygdala, the ventral striatum, the orbitofrontal cortex, the ventromedial prefrontal cortex, the cingulate cortex, the hypothalamus, and autonomic control centers in the brainstem (Barrett and Bar, 2009; LeDoux, 2000).

Valence is an integral dimension of perception and attention: the neurotransmitter dopamine, a key player in motivated and goal-directed behavior and in the resampling of stimuli that

have been associated with rewards (reinforcement learning, Wise, 2004), is also activated by surprising stimuli, such as sudden visual or auditory stimuli, that have never been associated with rewards (Horvitz, 2000). Kakade and Dayan (2002) have proposed that dopamine activations are *novelty bonuses* that increase the probability of re-sampling not only typically rewarding stimuli, but also surprising stimuli (see Barto et al., 2013, for a discussion of novelty vs. surprise), acting as a facilitator of exploratory action and perception. These properties make dopamine an ideal candidate for encoding precision of error units in the Predictive Coding framework (Fletcher and Frith, 2009; Clark, 2013). Interestingly, dopamine is also involved in the 'stamping-in' of memory (Wise, 2004), by loading environmental stimuli with motivational importance. Attention, affect and value drive learning, determining the strength of learned representations and ultimately making learning possible. The somatic marker hypothesis (Damasio, 1994) and, more recently, the affective prediction hypothesis (Vuilleumier, 2005; Barrett and Bar, 2009) and the interoceptive Predictive Coding model (Seth et al., 2011) suggest that affect and valence do not follow perception but instead are an integral part of it, for example driving object recognition. In a similar vein, Clark (2013) argues that nearly every aspect of perception is permeated by goal- and affect-laden expectations, and that the very division between emotional and non-emotional components may prove to be illusory. The affect system is arguably also the missing piece of the puzzle between physical experience and memory, reflecting a process which is not just reflective of frequency, but also of our attention processes and valence systems.

Summary

The studies reviewed here show that surprisal is not the only factor determining processing costs. The stimuli's relevance to the perceiver's goals, their valence and, crucially in the case of linguistic communication, their relevance to what we know of the speaker's abilities and preferences and their utility in confirming or correcting our hypotheses about the speaker's meaning determine what we pay attention to and what we are surprised by. At the two extremes of the predictability scale, stimuli can turn out to be too predictable (thus incompatible with what we assume to be relevant for the speaker's communicative goals), or too unpredictable (too costly and irrelevant, not worth the processing effort, or impossible to accommodate within our system of categories) and we may divert our attentions from both. On the other hand, relevant, unattended stimuli can be prioritized over task-irrelevant ones (for example, we can become aware of a deer by the side of the road, Jensen et al., 2012), or incongruent objects may capture our attention if their perceptual saliency is high enough (Coco et al., 2014). Tasks and goals determine what level of processing is relevant and thus what level we pay attention to. A linking hypothesis aimed at indexing predictability and saliency needs to account for these phenomena: high-level *surprise* may only be influenced by the relevant level of processing at each time, and *surprisal* at lower levels may not influence the behavioral response (unless it surpasses a certain threshold).

Predictive Coding provides an interesting framework for reconciling low-level computations of surprisal, high-level representations and hypotheses about the world and attentional focus mechanisms. We have reviewed recent work in neuroscience showing how our brain exploits multiple simultaneous channels at different firing frequencies to process perceptual stimuli bottom-up using high-frequency brain waves, while top-down information, at low-frequency brain waves, maximizes exposure to task-relevant stimuli by modulating the activation of relevant bottom-up information and suppressing task-irrelevant distractors. Attention is the mechanism we use to weight error-unit responses (in response to high-surprisal, attention-capturing input, or in response to relevant, interesting input, or as a function of the stimulus valence) over less interesting or informative ones. By weighting reliable sources of prediction error, attention and affect are the filter between perception and learned representations, and in the long-term shape our memories and beliefs. In the next section we will discuss in what way current surprisal models can be conceptually extended to yield more accurate accounts of language processing behavior.

5. IMPLICATIONS FOR MODELS OF PROCESSING DIFFICULTY: SURPRISE, ATTENTION, AFFECT

As discussed in this article, surprisal is a promising measure. Nevertheless, if our goal is not only to measure the amount of information contained in the linguistic signal, but also to describe how this amount of information relates to human processing difficulty, we need to also take into account effects of attention, namely (a) attention shifts from extremely predictable or too unpredictable stimuli, (b) the interplay of high- and low-level representations during language processing, mediated by attention and relevance, (c) the goal-dependent influence of higher-level representations, and (d) affect and valence and their influence on the learning of higher-level abstractions. We have argued that predictability and attention find a natural integration in the Predictive Coding framework, which accounts for how and why comprehenders generate predictions at multiple levels when processing language. In this framework, bottom-up properties of the signal are integrated with predicted percepts based on stored representations at multiple levels and grains of representation (van Berkum, 2010; Farmer et al., 2013; Kuperberg and Jaeger, 2016). During processing, a new percept will in turn be used to generate updated predictions about the next part of the input. The Predictive Coding framework is however not an implemented computational model that we can run on a new text (or multi-modal input) to obtain processing difficulty predictions. Therefore, we will now propose how a computational model of surprisal could be extended to account for effects of attention. In particular, we argue that each representational level (auditory / visual, lexical, structural, situational) might need its own attention modulation.

Surprisal models are trained to accurately account for upcoming words, that is, the objective function during training

of such models is to minimize prediction error. Consider for example an n -gram model, which predicts the surprisal of a word w_i based on the preceding sequence of n words, formalized as

$$\text{Surprisal}(w_i) = -\log P(w_i | w_{i-n}..w_{i-1})$$

In n -gram models there is no explicit modeling of syntax, semantic similarities, situational context representations or world knowledge. These models might therefore miss important generalizations or phenomena that are conditioned on words outside a window of n preceding words. However, with a lot of data and large contexts, many of the relevant statistics may be learned and represented by the model implicitly. N -gram models might therefore deliver good surprisal estimates for upcoming words, i.e., they might successfully predict upcoming words. Unfortunately, though, it is not clear how attention-based effects could be implemented in a model where the representation of linguistic knowledge is merely implicit. In such a model, the surprisal estimates would represent a combination of prediction errors and updates at all representation levels, i.e., they would be an approximation of the overall prediction error of a hierarchical architecture transmitting the prediction error up through all higher levels, and passing new updated anticipatory activations down. In order to adapt to a different task (e.g., reading for comprehension vs. spell checking), the model would have to be re-trained with a different objective function reflecting task-dependent costs of prediction errors.

A potential solution for modeling the hierarchical prediction process could therefore be in building models that also have a hierarchical architecture. Models with richer internal representations of linguistic structure and situational knowledge have been recently proposed. For instance, syntactic surprisal models internally represent syntactic structure (syntax tree $t \in T$) to estimate the predictability of upcoming words by calculating the difference in prefix probabilities (that is, the probability of observing sentence prefix $w_1..w_i$) before vs. after observing a word w_i . As Levy (2008) shows, the formula is equivalent to our the definition of surprisal $-\log P(w_i | w_1..w_{i-1})$.

$$\text{Surprisal}(w_i) = -\log \sum_{t \in T} P(t, w_1..w_i) + \log \sum_{t \in T} P(t, w_1..w_{i-1})$$

There have also been attempts to further extend computational models to capture topic context (e.g., Griffiths et al., 2007), semantic surprisal (e.g., Mitchell et al., 2010) or situation and event sequence knowledge (Frank et al., 2008; Venhuizen et al., 2016). A situation model representing situations S compatible with the prefix perceived so far and syntactic trees T that are consistent with the sentence prefix $w_1..w_{i-1}$ could be represented as³

$$\begin{aligned} \text{Surprisal}(w_i) = & -\log \sum_{s \in S} \sum_{t \in T} P(t, s, w_1..w_i) \\ & + \log \sum_{s \in S} \sum_{t \in T} P(t, s, w_1..w_{i-1}) \end{aligned}$$

³ S and T are chosen for the sake of the example, we do not intend to specifically argue for cognitive representations of syntax trees.

A hierarchical model (see also Farmer et al., 2013; Kuperberg, 2016; Kuperberg and Jaeger, 2016) then allows us to calculate the surprisal at each different level of representation. We can dissect the overall joint prefix probability that we use to calculate the information update from one word to the next in order to obtain prefix probabilities with respect to each level of representation:

$$\begin{aligned} -\log \sum_{s \in S} \sum_{t \in T} P(t, s, w_1..w_i) = & -\log \sum_{s \in S} \sum_{t \in T} P(s|t) \\ & \times P(t | w_1..w_i) \times P(w_1..w_i) \end{aligned}$$

The information update can thus be calculated separately for each specific level of representation, and is equivalent to Itti and Baldi's (2009) Bayesian *surprise* for that level. With such a hierarchical model, it would be possible to attach a separate linking theory to each level of representation. These could then be used to model the time course of processing, or specific ERPs.

In our review, we observed that attention is distributed among incoming stimuli and processing levels, that goals may affect processing and attention and that not all error signals, even if large, will necessarily affect higher-level representations. We will now briefly discuss how each of these aspects can be addressed by a hierarchical model with separable linking theories per representation level.

Attention is limited and hence has to be distributed among different stimuli. The reviewed evidence also supports the idea that not all representations and levels of processing need to be actively "at work" to the same extent in all tasks, i.e., for some tasks like spell-checking, others which are not relevant to the task (e.g. coherence, meaning) may not get much attention allocated to them, and contribute little to observable processing difficulty. Sanford and Sturt (2002) make the case for underspecified representations: we do not need to fully specify the linguistic signal at all possible levels, but we only need full specification for the levels of representation that are in the focus of attention, whereas those which are not in the focus of attention may be subject to more shallow processing or incomplete pattern specification. Sanford and Sturt (2002) also observe that sometimes underspecified representations lead to errors, such as semantic illusions, which are easily avoided by manipulating focus (e.g., *It was Moses who put two of each kind of animal on the ark*. Bredart and Modolo, 1988). In order to model phenomena like semantic illusions, the lexical semantic representation layer for the actor (*Moses/Noah*) would not be in the focus of attention during the critical region of this stimulus, and hence elicit only a small (or no) prediction error. The mismatch may therefore fail to propagate to other levels of representation, and not affect overall interpretation (that is, slip through unnoticed). The hierarchical model could specify a different linking function for each level of representation. It could then naturally account for task-dependent effects, such as the different strengths of predictability effects for different tasks.

Another apparent paradox that we discussed in Section 2.3 was the snow-screen paradox (Itti and Baldi, 2009): processing difficulty for an uninteresting fixed screen (e.g., a blue screen) and a randomly-changing snow screen are intuitively similar, even though the amount of surprisal of these two percepts

is extremely different. While prediction error when viewing a snow screen may be very large at the level of the visual cortex, this prediction error does not serve to update higher-level representations of the relevant semantics, as no interpretation of exact snow-screen patterns exists in the viewer's mind (the relevant categorization that can react to the incoming prediction error is not in place). The formulation of higher-level surprise also makes it explicit that a prediction error at a lower level only affects probability estimates at higher-level representations in as far as those prediction errors also change higher-level probability distributions: an exact pattern of snow might be very unpredictable, but the probability distribution over TV programmes $P(\text{TV_program}|\text{pixels})$ will not be affected by the likelihood of the exact pixel arrangement in the snow (at least not after already having perceived a few snow screens). Hence, these higher-level representations do not show any prediction error, and so the overall processing difficulty is low.

A similar situation could occur when a comprehender listens to somebody speaking in English (a language that the listener understands) and then switches to Finnish (a language she doesn't understand). In that case, processing difficulty would not go to infinity, but more likely she would stop predicting and processing the Finnish input in-depth: while there may be a very high prediction error at the word level, this prediction error does not serve to update any of the other representational layers, as it cannot be interpreted. During L2 language acquisition, new higher-level representations are learned. These can then "react" to certain input patterns from lower levels. This mechanism would then also naturally explain Goldilocks effects during learning, where learners only react to some types of prediction errors, most easily those that have representations in their own language as well, or those that are at the *just right* level of predictability, providing a theoretical explanation for observations in the language learning literature.

6. CONCLUSIONS

Prediction is a key aspect of cognition and in particular of language processing: comprehenders draw context-based expectations about upcoming input at different levels, relying and conditioning on multiple levels of representation at each point in processing, and experiencing a decrease in processing costs when the expectations are met and an increase when they are not. Current surprisal models go a long way in accounting for processing costs, but they still leave certain aspects unaccounted for, namely (1) phenomena at the extremes of the predictability scale (extremely high or low predictability), (2) the interaction between high- and low-levels of processing, (3) effects of task and goals, and (4) the influence of affect and valence. Work on linguistic saliency, by putting the emphasis on attention and relevance, has the potential of accounting for these aspects, but

has not exhaustively elucidated the interplay of saliency and surprisal.

We have resolved terminological inconsistencies related to saliency in linguistics by showing that, while perceptual acoustic saliency and prosodic or syntactic focus can be accounted for in terms of surprisal-driven bottom-up attentional capture, discourse- and situation-based saliency require an account of goal-driven attentional deployment that current models of surprisal lack. The Predictive Coding framework provides an integral account of prediction-driven perception, where perception, action, and attention share the common task of minimizing prediction error, respectively by trying to extract statistical regularities from the signal, by moving the sensors to resample the world to actively seek expected stimuli and by weighting reliable / goal-relevant and affect-laden error-unit responses more than non-reliable / irrelevant ones. The Predictive Coding framework is thus an ideal candidate to reconcile surprisal with attention and saliency and to account for how these guide comprehenders in expectation-driven language processing at different levels.

We argued that current models of surprisal need to be extended to account for the role played by attention and goals. This extension can potentially be achieved by providing the model with richer internal representations of linguistic structure, situational knowledge, event sequence knowledge, and beliefs and by weighting predictions at different levels with regard to their relevance, that is to the way they affect the interpretation at higher levels. These models would potentially be able to calculate surprisal at different levels, modeling the comprehension process in more detail and activating or inhibiting irrelevant processing levels or irrelevant parts of the stimulus in order to model processing difficulty as a function of task-mediated attentional focus.

AUTHOR CONTRIBUTIONS

AZ, VD, JV, and MV conceived the review; AZ wrote the paper with the exceptions of Section 3.4 (written by JV), Section 4.1 (written by MV), and Section 5 (written by VD). All authors contributed critical comments and revision of the review and agreed to the final content of the article.

ACKNOWLEDGMENTS

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 "Information Density and Linguistic Encoding" and the Cluster of Excellence "Multimodal Computing and Interaction" (EXC 284). This material is partially based upon work supported by the National Science Foundation 1476 Graduate Research Fellowship Program under grant no. DGE-1343012.

REFERENCES

- Altmann, G., and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73, 247–264.
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.
- Awh, E., Belopolsky, A. V., and Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends Cogn. Sci.* 16, 437–443. doi: 10.1016/j.tics.2012.06.010
- Balota, D. A., Pollatsek, A., and Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cogn. Psychol.* 17, 364–390.
- Bar, M. (2011). *Predictions in the Brain: Using Our Past to Generate a Future*. Oxford, UK: Oxford University Press.
- Barrett, L. F., and Bar, M. (2009). See it with feeling: affective predictions during object perception. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1325–1334. doi: 10.1098/rstb.2008.0312
- Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or surprise? *Front. Psychol.* 4:907. doi: 10.3389/fpsyg.2013.00907
- Bicknell, K., Elman, J. L., Hare, M., McRae, K., and Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *J. Mem. Lang.* 63, 489–505. doi: 10.1016/j.jml.2010.08.004
- Blumenthal-Dramé, A., Hanulíková, A., and Kortmann, B., eds. (2014). “Perceptual linguistic saliency: Modelling causes and consequences,” in *Proceedings of the FRIAS Workshop* (Freiburg).
- Bock, K., Irwin, D. E., and Davidson, D. J. (2004). “Putting first things first,” in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, eds J. M. Henderson and F. Ferreira (New York, NY: Psychology Press), 249–278.
- Bonnefond, M., and Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Curr. Biol.* 22, 1969–1974. doi: 10.1016/j.cub.2012.08.029
- Bonnefond, M., and Jensen, O. (2015). Gamma activity coupled to alpha phase as a mechanism for top-down controlled gating. *PLoS ONE* 10:e0128667. doi: 10.1371/journal.pone.0128667
- Borghi, A. M., Glenberg, A. M., and Kaschak, M. P. (2004). Putting words in perspective. *Mem. Cogn.* 32, 863–873. doi: 10.3758/BF03196865
- Bredart, S., and Modolo, K. (1988). Moses strikes again: Focalization effect on a semantic illusion. *Acta Psychol.* 67, 135–144.
- Bruce, N. D., and Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* 9, 1–24. doi: 10.1167/9.3.5
- Carreiras, M., Carriedo, N., Alonso, M. A., and Fernández, A. (1997). The role of verb tense and verb aspect in the foregrounding of information during reading. *Mem. Cogn.* 25, 438–446.
- Chen, X. and Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Res.* 46, 4118–4133. doi: 10.1016/j.visres.2006.08.008
- Chiarcos, C., Claus, B., and Grabski, M. (2011). *Saliency: Multidisciplinary Perspectives on Its Function in Discourse*, Vol. 227. Berlin: Walter de Gruyter.
- Chwilla, D. J., Brown, C. M., and Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology* 32, 274–285.
- Chwilla, D. J., and Kolk, H. H. (2005). Accessing world knowledge: evidence from N400 and reaction time priming. *Cogn. Brain Res.* 25, 589–606.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204.
- Clark, H. H., Schreuder, R., and Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *J. Verbal Learn. Verbal Behav.* 22, 245–258.
- Claus, B. (2011). “Establishing saliency during narrative text comprehension: a simulation view account” in *Saliency: Multidisciplinary Perspectives on Its Function in Discourse*, eds C. Chiarcos, B. Claus, and M. Grabski (Berlin: Walter de Gruyter), 251–277.
- Coco, M. I., and Keller, F. (2009). “The impact of visual information on reference assignment in sentence production,” in *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (Amsterdam).
- Coco, M. I., and Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Q. J. Exp. Psychol.* 68, 46–74. doi: 10.1080/17470218.2014.936475
- Coco, M. I., Malcolm, G. L., and Keller, F. (2014). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *Q. J. Exp. Psychol.* 67, 1096–1120.
- Craig, A. D. (2009). How do you feel – now? the anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555
- Damasio, A. R. (1994). *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York, NY: Grosset / Putnam.
- 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
- Demberg, V., and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109, 193–210. doi: 10.1016/j.cognition.2008.07.008
- Drenhaus, H., Demberg, V., Köhne, J., and Delogu, F. (2014). “Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (Québec City, QC).
- Ehrlich, S. F., and Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *J. Verbal Learn. Verbal Behav.* 20, 641–655. doi: 10.1016/S0022-5371(81)90220-6
- Elman, J. L., Hare, M., and McRae, K. (2005). “Cues, constraints, and competition in sentence processing,” in *Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*, eds M. Tomasello and D. Slobin (Mahwah, NJ: Lawrence Erlbaum Associates), 111–138.
- Farmer, T. A., Brown, M., and Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models in language processing. *Behav. Brain Sci.* 36, 211–212. doi: 10.1017/S0140525X12002312
- Federmeier, K. D., and Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495.
- Ferretti, T. R., Kutas, M., and McRae, K. (2007). Verb aspect and the activation of event knowledge. *J. Exp. Psychol.* 33, 182–196. doi: 10.1037/0278-7393.33.1.182
- Ferretti, T. R., McRae, K., and Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *J. Mem. Lang.* 44, 516–547. doi: 10.1006/jmla.2000.2728
- Firestone, C. and Scholl, B. J. (2015). Cognition does not affect perception: evaluating the evidence for ‘top-down’ effects. *Behav. Brain Sci.* doi: 10.1017/S0140525X15000965. [Epub ahead of print].
- Fletcher, P. C., and Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58. doi: 10.1038/nrn2536
- Folk, C. L., Remington, R. W., and Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J. Exp. Psychol.* 18, 1030–1044.
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5, 1–10. doi: 10.1038/ncomms5694
- Fossum, V., and Levy, R. (2012). “Sequential vs. hierarchical syntactic models of human incremental sentence processing,” in *Proceedings of the 3rd Annual Workshop on Cognitive Modeling and Computational Linguistics* (Montreal, QC).
- Frank, S. L. (2009). “Surprisal-based comparison between a symbolic and a connectionist model of sentence processing,” in *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, (Amsterdam).
- Frank, S. L., Koppen, M., Noordman, L. G., and Vonk, W. (2008). World knowledge in computational models of discourse comprehension. *Discourse Process.* 45, 429–463. doi: 10.1080/01638530802069926
- Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2013). “Word surprisal predicts N400 amplitude during reading,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Sofia).
- Frank, S. L., and Vigliocco, G. (2011). Sentence comprehension as mental simulation: an information-theoretic perspective. *Information* 2, 672–696. doi: 10.3390/info2040672
- Freyd, J. J. (1983). The mental representation of movement when static stimuli are viewed. *Percept. Psychophys.* 33, 575–581.
- Frison, S., Rayner, K., and Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *J. Exp. Psychol.* 31, 862–877. doi: 10.1037/0278-7393.31.5.862

- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Futrell, R. (2012). *Processing Effects of the Expectation of Informativity*. MA thesis, Stanford University, Stanford.
- Gahl, S., and Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language* 80, 748–775. doi: 10.1353/lan.2004.0185
- Giora, R. (2003). *On Our Mind: Salience, Context, and Figurative Language*. New York, NY: Oxford University Press.
- Givón, T. (1988). “The pragmatics of word order: predictability, importance and attention,” in *Studies in Syntactic Typology*, eds M. Hammond, E. A. Moravcsik, and J. Wirth (Amsterdam: John Benjamins), 243–284.
- Gleitman, L. R., January, D., Nappa, R., and Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *J. Mem. Lang.* 57, 544–569. doi: 10.1016/j.jml.2007.01.007
- Glenberg, A. M., Meyer, M., and Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *J. Mem. Lang.* 26, 69–83.
- Griffin, Z. M., and Bock, K. (2000). What the eyes say about speaking. *Psychol. Sci.* 11, 274–279. doi: 10.1111/1467-9280.00255
- Griffiths, T. L., Steyvers, M., and Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychol. Rev.* 114, 211–244. doi: 10.1037/0033-295X.114.2.211
- Griffiths, T. L., and Tenenbaum, J. B. (2007). From mere coincidences to meaningful discoveries. *Cognition* 103, 180–226. doi: 10.1016/j.cognition.2006.03.004
- Grosz, B. J., Weinstein, S., and Joshi, A. K. (1995). Centering: A framework for modeling the local coherence of discourse. *Comput. Linguist.* 21, 203–225.
- Hake, H. W. (1957). *Contribution of Psychology to the Study of Pattern Vision*. USAW WADC Tech. Rept. 1957.
- Hale, J. (2001). “A probabilistic Earley parser as a psycholinguistic model,” in *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics* (Pittsburgh, PA).
- Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., and Weber, A. (2012). When one person's mistake is another's standard usage: the effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103
- Hare, M., Elman, J. L., Tabaczynski, T., and McRae, K. (2009). The wind chilled the spectators, but the wine just chilled: sense, structure, and sentence comprehension. *Cogn. Sci.* 33, 610–628. doi: 10.1111/j.1551-6709.2009.01027.x
- Hare, M., McRae, K., and Elman, J. L. (2003). Sense and structure: meaning as a determinant of verb subcategorization preferences. *J. Mem. Lang.* 48, 281–303. doi: 10.1016/S0749-596X(02)00516-8
- Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194. doi: 10.1016/j.tics.2005.02.009
- Henderson, J. M., Malcolm, G. L., and Schandl, C. (2009). Searching in the dark: cognitive relevance drives attention in real-world scenes. *Psychon. Bull. Rev.* 16, 850–856. doi: 10.3758/PBR.16.5.850
- Holz, E. M., Glennon, M., Prendergast, K., and Sauseng, P. (2010). Theta-gamma phase synchronization during memory matching in visual working memory. *Neuroimage* 52, 326–335. doi: 10.1016/j.neuroimage.2010.04.003
- Horvitz, J. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* 96, 651–656. doi: 10.1016/S0306-4522(00)00019-1
- Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: a review of the findings. *Psychon. Bull. Rev.* 12, 822–851. doi: 10.3758/BF03196775
- Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res.* 49, 1295–1306. doi: 10.1016/j.visres.2008.09.007
- Itti, L., and Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40, 1489–1506. doi: 10.1016/S0042-6989(99)00163-7
- Izquierdo, J., and Collins, L. (2008). The facilitative role of L1 influence in tense-aspect marking: a comparison of hispanophone and anglophone learners of french. *Modern Lang. J.* 92, 350–368.
- Jaeger, T. F., and Levy, R. P. (2007). “Speakers optimize information density through syntactic reduction,” in *Advances in Neural Information Processing Systems Vol. 19*, eds B. Schölkopf, J. Platt, and T. Hoffman (Cambridge, MA: MIT Press), 849–856.
- Jensen, O., Bonnefond, M., and VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends Cogn. Sci.* 16, 200–206. doi: 10.1016/j.tics.2012.03.002
- Jensen, O., Gelfand, J., Kounios, J., and Lisman, J. E. (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. *Cereb. Cortex* 12, 877–882. doi: 10.1093/cercor/12.8.877
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Typol. Stud. Lang.* 45, 229–254. doi: 10.1075/tsl.45.13jur
- Kaakinen, J. K., and Hyönä, J. (2010). Task effects on eye movements during reading. *J. Exp. Psychol.* 36, 1561–1566. doi: 10.1037/a0020693
- Kaiser, E., and Trueswell, J. C. (2008). Interpreting pronouns and demonstratives in Finnish: evidence for a form-specific approach to reference resolution. *Lang. Cogn. Process.* 23, 709–748. doi: 10.1080/01690960701771220
- Kakade, S., and Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Netw.* 15, 549–559. doi: 10.1016/S0893-6080(02)00048-5
- Kamide, Y., Altmann, G. T., and Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* 49, 133–156. doi: 10.1016/S0749-596X(03)00023-8
- Kamp, H. (1981). “A theory of truth and semantic representation,” in *Formal Methods in the Study of Language*, eds P. H. Portner and B. H. Partee (Amsterdam: Foris), 277–213.
- Kamp, H., and Reyle, U. (1993). *From Discourse to Logic*. (Dordrecht: Kluwer).
- Kelleher, J. D. (2011). “Visual salience and the other one,” in *Salience: Multidisciplinary Perspectives on Its Function in Discourse*, eds C. Chiarcos, B. Claus, and M. Grabski (Berlin: Walter de Gruyter), 205–228.
- Kerswill, P. and Williams, A. (2002). “Salience” as an explanatory factor in language change: evidence from dialect levelling in urban England,” in *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, eds M. C. Jones and E. Esch (Berlin; New York, NY: Mouton de Gruyter), 81–110.
- Khalkhali, S., Wammes, J., and McRae, K. (2012). Integrating words that refer to typical sequences of events. *Can. J. Exp. Psychol.* 66, 106–114. doi: 10.1037/a0027369
- Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE* 7:e36399. doi: 10.1371/journal.pone.0036399
- Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003
- Koch, C., and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4, 219–227.
- Köhne, J., and Demberg, V. (2013). “The time-course of processing discourse connectives,” in *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (Berlin).
- Koolen, R., Krahmer, E., and Swerts, M. (2015). How distractor objects trigger referential overspecification: testing the effects of visual clutter and distractor distance. *Cogn. Sci.* doi: 10.1111/cogs.12297. [Epub ahead of print].
- Kravtchenko, E., and Demberg, V. (2015). “Semantically underinformative utterances trigger pragmatic inferences,” in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (Pasadena, CA).
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? what the N400 can tell us about the comprehension of events. *Lang. Cogn. Neurosci.* 31, 602–616. doi: 10.1080/23273798.2015.1130233
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- 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., 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 (Oxford, UK: Oxford University Press), 190–207.
- 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. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–163.
- Lambrecht, K. (1994). *Information Structure and Sentence Form: A Theory of Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, UK: Cambridge University Press.
- Lascarides, A., and Asher, N. (2007). “Segmented discourse representation theory: dynamic semantics with discourse structure,” in *Computing Meaning*, Vol. 3, (Berlin: Springer), 87–124.
- LeDoux, J. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184. doi: 10.1146/annurev.neuro.23.1.155
- Levelt, W. J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge, MA: The MIT Press.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006
- Lewis, A. G., and Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* 68, 155–168. doi: 10.1016/j.cortex.2015.02.01
- Lisman, J., and Buzsáki, G. (2008). A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophr. Bull.* 34, 974–980. doi: 10.1093/schbul/sbn060
- Loftus, G. R., and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol.* 4, 565–572.
- MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703.
- Mackworth, N. H., and Morandi, A. J. (1967). The gaze selects informative details within pictures. *Percept. Psychophys.* 2, 547–552.
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci.* 29, 725–732. doi: 10.1523/JNEUROSCI.3963-08.2009
- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., and McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *J. Exp. Psychol.* 37, 13–934. doi: 10.1037/a0022964
- McDonald, S. A., and Shillcock, R. C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychol. Sci.* 14, 648–652. doi: 10.1046/j.0956-7976.2003.psci_1480.x
- McRae, K., and Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Lang. Linguist. Compass* 3, 1417–1429. doi: 10.1111/j.1749-818X.2009.00174.x
- Mesulam, M. (2008). Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. *Ann. Neurol.* 64, 367–378. doi: 10.1002/ana.21534
- Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., and Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *J. Mem. Lang.* 66, 545–567. doi: 10.1016/j.jml.2012.01.001
- Mitchell, J., Lapata, M., Demberg, V., and Keller, F. (2010). “Syntactic and semantic factors in processing difficulty: an integrated measure,” in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (Uppsala).
- Montag, J. L., and MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Lang. Speech* 57, 163–180. doi: 10.1177/0023830913495656
- Morris, R. K. (1994). Lexical and message-level sentence context effects on fixation times in reading. *J. Exp. Psychol.* 20, 92–103.
- Navalpakkam, V., and Itti, L. (2005). Modeling the influence of task on attention. *Vision Res.* 45, 205–231. doi: 10.1016/j.visres.2004.07.042
- Nicenboim, B., Vasisht, S., Gattei, C., Sigman, M., and Kliegl, R. (2015). Working memory differences in long-distance dependency resolution. *Front. Psychol.* 6:312. doi: 10.3389/fpsyg.2015.00312
- Olivers, C. N. (2007). The time course of attention: it is better than we thought. *Curr. Direct. Psychol. Sci.* 16, 849–860. doi: 10.1111/j.1467-8721.2007.00466.x
- Osgood, C. E., and Bock, J. K. (1977). “Salience and sentencing: Some production principles,” in *Sentence Production: Developments in Research and Theory*, ed S. Rosenberg (Hillsdale, NJ: Erlbaum), 89–140.
- Palm, G. (2012). *Novelty, Information and Surprise*. Berlin: Springer.
- Pierrehumbert, J. B. (2006). The next toolkit. *J. Phon.* 34, 516–530. doi: 10.1016/j.wocn.2006.06.003
- Prat-Sala, M., and Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *J. Mem. Lang.* 42, 168–182. doi: 10.1006/jmla.1999.2668
- Rácz, P. (2013). *Salience in Sociolinguistics: A Quantitative Approach*. Berlin: Walter de Gruyter.
- Radvansky, G. A., and Curiel, J. M. (1998). Narrative comprehension and aging: The fate of completed goal information. *Psychol. Aging* 13, 69–79.
- Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202. doi: 10.1038/nrn1052
- Rao, R. P., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87.
- Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., and Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Res.* 42, 1447–1463. doi: 10.1016/S0042-6989(02)00040-8
- Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *J. Exp. Psychol.* 18, 849–860.
- Rayner, K., and Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychon. Bull. Rev.* 3, 245–248.
- Roark, B., Bachrach, A., Cardenas, C., and Pallier, C. (2009). “Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing,” in *Proceedings of the 14th Conference of Empirical Methods in Natural Language Processing* (Singapore).
- Rohde, H., and Horton, W. S. (2014). Anticipatory looks reveal expectations about discourse relations. *Cognition* 133, 667–691. doi: 10.1016/j.cognition.2014.08.012
- Roux, F., and Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn. Sci.* 18, 16–25. doi: 10.1016/j.tics.2013.10.010
- Sanford, A. J., and Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends Cogn. Sci.* 6, 382–386. doi: 10.1016/S1364-6613(02)01958-7
- Sauseng, P., Klimesch, W., Heise, K., Gruber, W., Holz, E., Karim, A., Glennon, M., Gerloff, C., Birbaumer, N., and Hummel, F. (2009). Brain oscillatory substrates of human visual short-term memory capacity. *Curr. Biol.* 19, 1846–1852. doi: 10.1016/j.cub.2009.08.062
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., and Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x
- Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., and Birbaumer, N. (2010). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x
- Sayeed, A., Fischer, S., and Demberg, V. (2015). “Vector-space calculation of semantic surprisal for predicting word pronunciation duration,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing* (Beijing).
- Schotter, E. R., Bicknell, K., Howard, I., Levy, R., and Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: evidence from eye movements in reading and proofreading. *Cognition* 131, 1–27. doi: 10.1016/j.cognition.2013.11.018
- Seth, A. K., Suzuki, K., and Critchley, H. D. (2011). An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2:395. doi: 10.3389/fpsyg.2011.00395
- Shirai, Y. (2009). “Temporality in first and second language acquisition,” in *The Expression of Time*, eds W. Klein and P. Li (Berlin: Mouton de Gruyter), 167–194.
- Simons, D. J., and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception* 28, 1059–1074.
- Smith, N. J., and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition* 128, 302–319. doi: 10.1016/j.cognition.2013.02.013
- Sperber, D., and Wilson, D. (1986). *Relevance: Communication and Cognition*. Cambridge, MA: Harvard University Press.

- Spivey-Knowlton, M. J., Trueswell, J. C., and Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Can. J. Exp. Psychol.* 47, 276–309.
- 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.
- Tatler, B. W., Hayhoe, M. M., Land, M. F., and Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting saliency. *Journal of Vision*, 11(5)(5):1–23.
- Tomlin, R. S. (1997). “Mapping conceptual representations into linguistic representations: the role of attention in grammar,” in *Language and Conceptualization*, eds J. Nuyts and E. Pederson (Cambridge, UK: Cambridge University Press), 162–189.
- Torralba, A., Oliva, A., Castelinho, M. S., and Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 113, 766–786. doi: 10.1037/0033-295X.113.4.766
- Trudgill, P. (1986). *Dialects in Contact*. Oxford, UK: Blackwell.
- Trueswell, J., Tanenhaus, M., and Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.* 33, 285–318.
- Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psychol.* 19, 528–553.
- van Berkum, J. J. (2009). “The neuropragmatics of ‘simple’ utterance comprehension: an ERP review,” in *Semantics and Pragmatics: From Experiment to Theory*, eds U. Sauerland and K. Yatsushiro (Basingstoke, UK: Palgrave Macmillan), 276–316.
- van Berkum, J. J. (2010). The brain is a prediction machine that cares about good and bad - any implications for neuropragmatics? *Ital. J. Linguist.* 22, 181–208.
- van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol.* 31, 443–467. doi: 10.1037/0278-7393.31.3.443
- van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591. doi: 10.1162/jocn.2008.20054
- van der Meer, E., Beyer, R., Heinze, B., and Badel, I. (2002). Temporal order relations in language comprehension. *J. Exp. Psychol.* 28, 770–779. doi: 10.1037/0278-7393.28.4.770
- van der Meer, E., Krüger, F., and Nuthmann, A. (2005). The influence of temporal order information in general event knowledge on language comprehension. *Z. Psychol.* 213, 142–151. doi: 10.1026/0044-3409.213.3.142
- Van Oostendorp, H., and De Mul, S. (1990). Moses beats Adam: a semantic relatedness effect on a semantic illusion. *Acta Psychol.* 74, 35–46.
- van Schijndel, M., and Schuler, W. (2015). “Hierarchic syntax improves reading time prediction,” in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Denver, CO).
- van Schijndel, M., Schuler, W., and Culicover, P. W. (2014). “Frequency effects in the processing of unbounded dependencies,” in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* Québec City, QC.
- Venhuizen, N., Brouwer, H., and Crocker, M. (2016). “When the food arrives before the menu: modeling event-driven surprisal in language comprehension,” in *Abstract Presented at Events in Language and Cognition, Pre-CUNY Workshop on Event Structure* (Gainesville, FL).
- Vogels, J., Krahmer, E., and Maes, A. (2013). Who is where referred to how, and why? the influence of visual saliency on referent accessibility in spoken language production. *Lang. Cogn. Process.* 28, 1323–1349. doi: 10.1080/01690965.2012.682072
- Vuilleumier, P. (2005). How brains beware: neural mechanisms of emotional attention. *Trends Cogn. Sci.* 9, 585–594. doi: 10.1016/j.tics.2005.10.011
- Wilson, D., and Sperber, D. (2004). “Relevance theory,” in *Handbook of Pragmatics*, eds G. Ward and L. Horn (Oxford, UK: Blackwell), 607–632.
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 5, 483–494. doi: 10.1038/nrn1406
- Wolfe, J. M., and Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 5, 495–501. doi: 10.1038/nrn1411
- Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific α -band electroencephalography increases over occipital cortex. *J. Neurosci.* 20:RC63.
- Xiang, M., and Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Lang. Cogn. Neurosci.* 30, 648–672. doi: 10.1080/23273798.2014.995679
- Yarbus, A. L. (1967). *Eye Movements and Vision*. New York, NY: Plenum.
- Zarcone, A., Padó, S., and Lenci, A. (2014). Logical metonymy resolution in a words-as-cues framework: Evidence from self-paced reading and probe recognition. *Cogn. Sci.* 38, 973–996. doi: 10.1111/cogs.12108
- Zelinsky, G., Zhang, W., Yu, B., Chen, X., and Samaras, D. (2006). “The role of top-down and bottom-up processes in guiding eye movements during visual search,” in *Advances in Neural Information Processing Systems*, Vol. 18, eds Y. Weiss, B. Schölkopf, and J. Platt (Cambridge, MA: MIT Press), 1569–1576.
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a ‘cocktail party’. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037
- Zwaan, R. A., Madden, C. J., and Whitten, S. N. (2000). The presence of an event in the narrated situation affects its availability to the comprehender. *Mem. Cogn.* 28, 1022–1028. doi: 10.3758/BF03209350
- Zwaan, R. A. and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychol. Bull.* 123, 162–185.

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.

Copyright © 2016 Zarcone, van Schijndel, Vogels and Demberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Salience of Complex Words and Their Parts: Which Comes First?

Hélène Giraudo^{1*} and Serena Dal Maso²

¹ CLLE, University of Toulouse, CNRS, Toulouse, France, ² Department of Cultures and Civilizations, University of Verona, Verona, Italy

This paper deals with the impact of the salience of complex words and their constituent parts on lexical access. While almost 40 years of psycholinguistic studies have focused on the relevance of morphological structure for word recognition, little attention has been devoted to the relationship between the word as a whole unit and its constituent morphemes. Depending on the theoretical approach adopted, complex words have been seen either in the light of their paradigmatic environment (i.e., from a paradigmatic view), or in terms of their internal structure (i.e., from a syntagmatic view). These two competing views have strongly determined the choice of experimental factors manipulated in studies on morphological processing (mainly different lexical frequencies, word/non-word structure, and morphological family size). Moreover, work on various kinds of more or less segmentable items (from genuinely morphologically complex words like *hunter* to words exhibiting only a surface morphological structure like *corner* and irregular forms like *thieves*) has given rise to two competing hypotheses on the cognitive role of morphology. The first hypothesis claims that morphology organizes whole words into morphological families and series, while the second sets morphology at a pre-lexical level, with morphemes standing as access units to the mental lexicon. The present paper examines more deeply the notion of morphological salience and its implications for theories and models of morphological processing.

Keywords: morphological salience, visual word recognition, morphological processing, masked priming, lexical access

OPEN ACCESS

Edited by:

Alice Julie Blumenthal-Dramé,
University of Freiburg, Germany

Reviewed by:

Joanna Morris,
Hampshire College, USA
J. P. Blevins,
University of Cambridge, UK

*Correspondence:

Hélène Giraudo
helene.giraudo@univ-tlse2.fr

Specialty section:

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

Received: 02 March 2016

Accepted: 31 October 2016

Published: 21 November 2016

Citation:

Giraudo H and Dal Maso S (2016)
The Salience of Complex Words
and Their Parts: Which Comes First?
Front. Psychol. 7:1778.
doi: 10.3389/fpsyg.2016.01778

WHAT IS SALIENT IN MORPHOLOGICAL PROCESSING?

In linguistics, the semiotic notion of salience has been applied to inflectional and derivational morphology from the 1980s onward, mainly in the framework of 'Natural Morphology' (NM; e.g., Dressler et al., 1987). In this approach, the idea of morphological salience refers to the relative importance or prominence of a morpheme (stem or affix) in a morphologically complex word, the underlying assumption being that the salience of morphological components drives the mechanisms underlying complex word processing as well as storage and lexical organization. More recently, in the domain of language acquisition, Goldschneider and DeKeyser (2001) defined morphological salience as referring to "how easy it is to hear or perceive a given structure" (p. 22).

In the Natural Morphology (henceforth: NM) approach, salience is one of the factors that contribute to the 'naturalness' of a linguistic item or structure, which in turn determines how easily it can be processed by the human brain (Dressler et al., 1987, p. 11). Thus, NM theory explicitly defines naturalness on psychological grounds and makes particular reference to cognitive limitations on perception and processing (e.g., on memory, information recall, and selective

attention). According to Natural Morphologists, psycholinguistic factors do not directly determine linguistic structures, but they limit the choice of available linguistic (in our case morphological) techniques, favoring the ones that are cognitively less demanding and disfavoring the more cognitively demanding ones. In a way, psycholinguistic factors ‘constrain’ the possibilities of languages.

Two kinds of factors are supposed to determine the saliency of the components of a morphologically complex word, thereby affecting the recognition of its morphological structure. The first group of factors relates to the strength of the mental representation of the whole complex word and its components, which is thought to be modulated by the following variables: (i) (token and type) frequency; (ii) numerosity (i.e., the number of distinct words with which a suffix occurs, cf. Burani and Thornton, 2003); (iii) productivity. Intuitively, the more frequently a form is heard and processed, the stronger a mental representation it has and the easier it is to recognize. The second group of factors relates to more formal characteristics of morphemes and involves a wide range of features, in particular: (i) their size and phonological features (e.g., stress); (ii) their position within the complex word (i.e., initial, final, or internal); (iii) their formal (in)variance (i.e., the less an item varies in a paradigm, the more recognizable it is); (iv) the morphotactic transparency of the complex word they are embedded in; (v) their formal distinctness [i.e., if a morphological component is salient, its form is distinct paradigmatically both with respect to forms of the same morphological family or paradigm and with regard to forms which are formally similar but semantically unrelated (i.e., the orthographic neighborhood, Andrews, 1989, 1992)].

More broadly, the saliency of a morphological item may also be influenced by semantic and functional properties, such as consistency (a formal component is recognized more easily if it always occurs with the same meaning or function) and morpho-semantic transparency (the constituents fully contribute to the meaning of the complex word, see Plag, 2003). The present paper will discuss the extent to which experimental psycholinguistic studies have confirmed the psychological plausibility of the notion of saliency and its effects in word processing and lexical organization.

THE WHOLE-WORD AND DECOMPOSITIONAL PERSPECTIVES

Differing stances on the nature and role of morphology within the mental lexicon have led to two opposite hypotheses about processing: either morphemic representations stand as access units to word representations, or word representations organize the mental lexicon into morphological families. According to the first view, which is often referred to as the “decompositional view,” the morphemic units correspond to concrete pieces of words (i.e., stems and affixes), coded at a sublexical level and processing complex words implies passing through a decomposition mechanism that strips off the affix in order to isolate the stem, so that the morphemic nature of the remaining letters can be checked by the system. Access to word representations (i.e., word forms coded in the

orthographic lexicon) thus operates via the pre-activation of their constituent morphemes. This mechanism is exemplified in the interactive activation model developed by Taft (1994), a model instantiating the decompositional view of morphology by integrating sublexical morphemic representations as access units.

According to the second view, called the “whole-word perspective,” morphology is represented at the interface of word and semantic representations and derives from lexemes as introduced by Aronoff (1994), i.e., lexeme units are coded at a *morphomic* level and have the function of organizing the lexicon in terms of morphological families. In processing terms, the recognition of any complex word initially triggers the activation of all word forms that can match with it, and a competition is then engaged between the pre-activated forms until the right lexical representation reaches its recognition threshold (determined by its surface frequency). During this competition phase, competitors send positive activation to their respective base lexemes, which send positive activation back to them. According to this account, exemplified in the supralelexical model of Giraudo and Grainger (2000), complex words are not “decomposed” following the procedure described by the sublexical/decompositional account, but are able to trigger the activation of their constituent morphemes.

Both sublexical and supralelexical approaches to morphological processing integrate a morphological level of processing, however, they differ with respect to the location of morphological units within the architecture of the mental lexicon, as well as the content of these units, both of which properties define their role of such units in word processing. According to the sublexical view, morphemic units stand as access units, situated between the letter/syllable level and the word level: consequently, these units can only correspond to concrete letter clusters that constitute words (i.e., bound stems, free stems, and affixes) and are insensitive to any semantic characteristics of words (i.e., transparent vs. opaque) or to their lexical environment (in terms of orthographic neighborhood or family size). On the other hand, the supralelexical view situates morphological units above the word-forms and before the semantic units. These intermediate units are supposed to be abstract enough to tolerate form variations induced by the processes of derivation and inflection. This implies that a morphemic unit does not need to exist in the real world in order to be coded in long-term memory, but that its existence/emergence depends on the interactions between the word-form and the semantic levels; it also implies that all morphemes of a given language are not necessarily represented within the mental lexicon: unknown words, neologisms, hapaxes, and nonce words are not necessarily connected with morphemic units.

However, determining which factors are involved in lexical access and which factors influence the organization of the mental lexicon are issues that have not been sufficiently explored so far, although they are highly relevant to lexical modeling. We suggest that it is crucial to keep these issues apart: the factors driving the early stages of processing are likely to be different from those coded in long-term memory. In our view, observing sensitivity to the internal structure of complex words can be interpreted as reflecting a central role of morphemes in lexical access, but the

factors influencing lexical access (e.g., lexical frequency) are likely to be different from those organizing the mental lexicon properly (e.g., morphological family size).

EVIDENCE TAKEN TO SUPPORT THE DECOMPOSITIONAL APPROACH

Numerous psycholinguistic studies have addressed the issue of morphological processing during word recognition. Using the lexical decision task (in which participants must make a decision about whether combinations of letters are words or not), these studies explore the factors influencing the processing of complex words as well as their internal structure. Among these factors, surface frequency (equivalent to token or lemma frequencies) and base frequency (the token or lemma frequency of a root), which measure the statistical occurrence of complex words, have been extensively studied in languages for which lexical databases are available (e.g., Taft, 1979, 2004; Burani et al., 1984; Burani and Caramazza, 1987; Colé et al., 1989, 1997; Baayen et al., 1997; Bertram et al., 1999, 2000a; Burani and Thornton, 2003; Ford et al., 2010; Xu and Taft, 2015). These studies show, among other findings, that when two words are matched in terms of surface frequency (SF), reaction times depend on their base frequency (BF), with high BF words being recognized faster than low BF words. The fact that recognition latencies for complex words depend on base frequencies has been taken as evidence that readers are sensitive to morphological structure and that a cognitive component of word processing is related to the perceptual saliency of both the whole word and its morphemic structure. These data gave rise to the decomposition hypothesis as reflecting the automatic processing of morphemes by the cognitive system.

Many studies have lent further support to the decomposition approach to complex word recognition, using priming, and, more recently, masked priming (Forster and Davis, 1984). In masked priming, a prime word is presented for a very short duration (under 60 ms) and is masked by a backward font (usually a string of hash marks), before a target word on which subjects have to perform a lexical decision task is presented. Because this duration does not allow the participants to identify the prime consciously, this paradigm has the advantage of examining very early automatic processes of lexical access as well as non-strategic responses based on the relationships shared by the prime-target pairs (see Forster, 1999, for a review). From the seminal repetition priming study conducted by Stanners et al. (1979) to the most recent studies investigating the brain correlates of masked priming (e.g., Morris et al., 2013), morphological priming effects have been extensively studied and have systematically revealed strong facilitation. Morphological effects (i.e., a morphologically complex prime like *hunter* facilitating the recognition of its morphologically related target *hunt*) differing significantly from formal (e.g., *hungry-hunt*) and meaning relationships (e.g., *pursuit-hunt*), have led the authors to conclude that independent morphological representations are coded somewhere within the mental lexicon in a similar way to orthographic, phonological,

and semantic representations. Therefore, until the beginning of the 21st century, experimental studies considered morphological effects to result from systematic form-meaning correlations.

However, between 2000 and 2005, many masked priming studies started to focus exclusively on formal aspects of complex words, that is on their so-called ‘morphological surface structure’ (e.g., Rastle et al., 2004, p. 1091) in order to examine whether processing is decompositional or holistic. The underlying hypothesis was that if significant priming effects can emerge only from the surface structure of words (i.e., from form only), whether morphologically complex or not, then morphology is not coded within the lexicon but rather in its access routes. It is important to highlight here that this approach to morphological complexity, which considers only the surface forms of words, is based on the assumption that morphology can be emptied of its meaning component. Consequently, according to this view separating morphology from semantics, morphological regularities within languages exclusively increase the ‘surface’ saliency of morphemes, the aim being to guide pre-lexical processes.

While the priming study carried out by Rastle et al. (2000) historically defines the starting point of this series of masked priming studies, the most striking ones were conducted, respectively, for French by Longtin et al. (2003) and for English by Rastle et al. (2004). Both manipulated word pairs involving primes with morphologically pseudo-complex surface forms (e.g., the English word *corner*, which cannot be decomposed into the morphemes *corn-* and *-er*). Using the masked priming paradigm, it was shown that pseudo-derived word primes (e.g., *corner*) as well as pseudo-derived non-word primes (i.e., non-words composed of two existing morphemes such as *cornering*) were able to produce significant priming effects on the recognition times of their pseudo-base (e.g., *corn*). Moreover, the studies found both the quality and the magnitude of these priming effects to be comparable to the priming effects produced by genuinely derived words (e.g., *banker-bank*). Finally, the systematic use of orthographic control primes (i.e., morphologically simple forms whose onset alone mimics a stem morpheme, such as *brothel*, whose ending *-el* never functions as a suffix in English) in these studies showed that these surface morphological effects could not be assimilated to mere formal overlap. Consequently, these effects were taken to result exclusively from the surface morphological structure of the primes.

Further masked priming studies have tested the effect of pseudo-derived non-words primes, and systematically found facilitation effects, lending strong support to the notion of an early mechanism of form decomposition that is applied to all morphologically structured stimuli (McCormick et al., 2009; Morris et al., 2013; Beyersmann et al., 2014; Crepaldi et al., 2016). In general, the logic behind such studies is that since non-words are not supposed to have lexical representation(s), any masked priming effect obtained must reflect activation of sublexical units, i.e., morphemes. Thus, in a recent review, Amenta and Crepaldi (2012) claimed that “morphological effects in non-words exclude the possibility that morphological information only comes into play after lexical identification” (p. 9), given that “it is clear

that non-words with a morphological structure are analyzed in terms of their morphemes, thus questioning seriously any theory that suggests morphological processing to kick off upon lexical identification" (p. 7). For example, Longtin and Meunier (2005) used pseudo-derived pseudo-words to test the robustness of early morphological decomposition. In their masked priming study, non-existent possible words created from two existing morphemes (for instance, the base *sport-* combined with the suffix *-ation* to produce *sport-ation*) were used as primes. The data revealed that pseudo-word primes like *sportation* facilitate the recognition of their base (e.g., *sport*) with no difference from the facilitation effects obtained using transparent primes (e.g., *sportif*, which is a licit and semantically transparent derivation from the base *sport*).

Studies showing masked morphological priming effects without semantic relationships have been broadly replicated in various languages (Spanish: Sánchez-Casas et al., 2003; German and French: Diependaele et al., 2005, 2009; French: Giraud and Voga, 2013; Arabic: Boudelaa and Marslen-Wilson, 2004a,b, 2005; English: Lavric et al., 2007; Marslen-Wilson et al., 2008; Feldman et al., 2009, 2015; McCormick et al., 2009; Lehtonen et al., 2011¹; Finnish: Järvikivi and Pyykkönen, 2011 and Russian: Kazanina et al., 2008; Kazanina, 2011).

All these studies led the authors to conclude that the morphological decomposition mechanism transcends stimuli and languages. A review by Rastle and Davis (2008) clearly set out that "morphological decomposition is a process that is applied to all morphologically structured stimuli, irrespective of their lexical, semantic or syntactic characteristics" (p. 949). Further evidence in support of this view was provided by a study by McCormick et al. (2008), who manipulated a particular category of derived stimuli that cannot be segmented perfectly into their morphemic components (e.g., *dropper-drop*, in which there is a duplicated consonant) in order to test the flexibility of the morpho-orthographic segmentation process described by compositional models. Once again, their results were interpreted as demonstrating the robustness of the decomposition process in the case of various orthographic alterations in semantically related (e.g., *adorable-adore*) as well as unrelated prime-target pairs (e.g., *fetish-fete*).

OBJECTIONS TO THE DECOMPOSITIONAL APPROACH

The results reported in the previous section have largely been taken to support a decompositional approach. However, in our view, there are also studies that are inconsistent with this interpretation.

Some masked priming studies have indeed demonstrated very early semantic influences in word recognition. Feldman et al. (2009) matched affixes across semantically transparent and opaque related (and unrelated) prime-target pairs and increased the proportion of identical prime-target filler pairs

(e.g., *artist-artist*) in order to enhance semantic facilitation (e.g., Bodner and Masson, 2003). They found that morphological facilitation was significantly greater for semantically transparent pairs (e.g., *coolant-cool*) than for opaque pairs (e.g., *rampant-ramp*). Giraud and Voga (2013) manipulated prefixed words (e.g., *prénom* 'name') and non-words (e.g., *dénom* = *dé-* + *-nom*) in French. They showed that when compared to unrelated primes, both prefixed words and prefixed non-words facilitate target recognition. However, when compared to an orthographic non-word condition (e.g., *danom*), pseudoprefixed primes do not differ from orthographic primes, suggesting a strong formal component in surface morphological priming with semantics. Finally, Feldman et al. (2015) tracked the time course of processing of the interaction between form and meaning using different prime exposure durations (increasing from 34 to 100 ms). They observed that the time course of facilitation varies for similar forms with and without semantic similarity, the transparency effect being evident even at an SOA of 34 ms (Experiment 3).

Other studies have explored the interaction of frequency effects with paradigmatic factors such as affix type and suffix productivity. In a series of lexical decision task experiments, Colé et al. (1989) and later Beauvillain (1996) with eye-movement recordings, showed that while suffixed word recognition in French is sensitive to the manipulation of both types of frequencies (SF and BF), prefixed word recognition is affected only by SF. The authors suggested that this asymmetry could simply reflect the left-to-right direction of the reading process, but studies using other paradigms such as masked priming refuted this physical explanation (e.g., Giraud and Grainger, 2003). Moreover, Bertram et al. (2000b) discovered that BF effects in Dutch emerge only for words with a very productive suffix. This interaction between BF and affix productivity was replicated for English by Ford et al. (2010), who found that this effect occurs independently of the morphological family size effect, suggesting the occurrence of both holistic and compositional effects during complex word recognition. Only three studies have so far investigated frequency effects using masked priming, and the results have been inconsistent. Giraud and Grainger (2000) manipulated the SF of derivatives used as primes for the same target (High SF *amitié* - *ami* 'friendship-friend'; Low SF *amiable* - *ami* 'friendly friend') and found an interaction between priming effects and the prime SF (Experiment 1), but no effect for the BF. Experiments 1 and 3 demonstrated that the SF of morphological primes affects the degree of morphological priming: high SF derived primes show significant facilitation relative to orthographic control primes (e.g., *amidon* - *ami* 'starch-friend'), whereas low SF primes do not. The results of Experiment 4 revealed, by contrast, that BF does not influence the size of morphological priming on free root targets. Suffixed word primes facilitate the processing of free root targets with low and high BF. These data support the relevance of the whole word form (as reflected by SF) over its parts, since the BF does not interact with priming. More recently, McCormick et al. (2009) re-investigated frequency effects during masked priming, though without mentioning the results of the earlier studies reported here. They compared the effects of High SF, Low SF

¹ These three masked priming paradigm studies associated ERP measures with RT recordings.

and pseudoword primes on target recognition, but contrary to Giraudo and Grainger (2000) they compared each priming effect on different targets (e.g., *brutal* – *brute* vs. *adorable* – *adore* vs. *agitatal* – *agitate*, respectively). They found facilitation effects on all three conditions relative to each of the three unrelated baselines (e.g., *verbal* – *brute*, *enviable* – *adore*, *corrodal* – *agitate*, respectively). In our view, the lack of orthographic controls that could separate formal from morphological effects constitutes a serious obstacle for the interpretation of their data, which thus only show that related primes facilitate target recognition. Furthermore, it is very surprising to see that despite an interpretation in favor of the decompositional hypothesis, these authors did not test BF effects, which should strongly determine decomposition and therefore the magnitude of priming effects.

Further evidence against the decompositional hypothesis comes from the studies conducted by Giraudo and Orihuela (2015) and Giraudo and Dal Maso (2016). These masked priming studies carried out for French and for Italian replicated the SF interference effect and revealed that while whole-word frequency speeds up lexical access, morphological priming effects are also modulated by the relative frequencies of the prime and the target. SF interference effects highlight the role of the whole word over its internal structure during the very early stages of word recognition, and indicate that whole-word characteristics are more important for morphological saliency than those of the word's subparts. However, this does not amount to claiming that morphological structure does not play a role. In our view, morphological saliency emerges from relationships between whole word forms and their parts. The whole word guides lexical access, while morphological relationships are expressed by the links that cluster together word forms belonging to the same family or series (which cluster complex words according to the affix they share in common, e.g., *cleaner*, *hunter*, *biker*).

Finally, a set of studies that, in our view, contradict the mandatory decomposition hypothesis, use non-word primes involving transposed letters (TL) that disrupt the morpho-orthographic structure. Masked priming experiments have compared the effects of complex non-word primes with TL at a morpheme boundary (e.g., *painetr-paint*) to effects of primes with TL outside the morpheme boundary (e.g., *paniter-paint*). Although priming effects were obtained independently of the position of the TL (at the morpheme boundary or not), this has not lead researchers to call the decompositional approach into question (Perea and Carreiras, 2006; Rueckl and Rimzhim, 2011; Beyersmann et al., 2012, 2013; Luke and Christianson, 2012; Diependaele et al., 2013b).

We take issue with this interpretation, since if morphological decomposition governs access to word forms coded in the mental lexicon, non-word primes which cannot be parsed into distinct surface morphemes should not be able to induce priming. Since their surface morphological structure is hidden by the TL (e.g., *painetr*), no morphemic units should be activated and therefore no priming is expected. And even if a sublexical mechanism was able to recode letter position (as suggested by Diependaele et al. (2013a)), the position of the TL should interfere with morphological priming: letter-transposed primes with intact morphemic boundaries should be more effective

for the recognition of their base (like *paniter* – *paint*) than those with disrupted morpheme boundaries (as in *painetr* – *paint*). Moreover, the mechanism of letter recoding must depend on a match between the prime and a whole-word representation coded at the word form level, which implies that the whole word guides access rather than its parts. In our view, rather than supporting decomposition, the data obtained with non-words constitute strong evidence in favor of holistic processing of the primes and, by extension, of all the stimuli, whatever their surface structure. We take the fact that words with jumbled letters can induce priming effects to provide sufficient grounds to reject the claim by Amenta and Crepaldi, according to which non-word effects cannot result from lexical activation. We interpret these data obtained with non-words in the opposite way: the pattern of systematic form-meaning correspondences that we call morphology (Bybee, 1988, 2001; Booij, 2010) has to be extended to all possible words.

Talking about morphological links implies taking into account another factor whose impact on complex word recognition has been demonstrated and replicated in various languages: morphological family size (i.e., the total number of words derived from the same morphological family; Bertram et al., 2000b; De Jong et al., 2000). It has been shown that complex words with many morphological relatives are processed faster than those with a small morphological family, suggesting that the locus of morphological effects is not exclusively the word to be processed, and that factors outside the word in question intervene in morphological processing. In the same line, Voga and Giraudo (2009) explored a novel variable, called the “pseudo-family size,” which is the opposite of the morphological family. The notion of pseudofamily size includes neighbors in the classic sense (i.e., members of the morphological family), but also all words sharing their stem with a given entry, even if what remains of the word once the stem is removed is not really an affix. Their working hypothesis was that pseudo-relatives should behave like competitors at the word level. This was tested in two masked inflectional priming experiments comparing two kinds of stimuli: verbs from large pseudo-families and verbs from small or non-existent pseudo-families. The first experiment studied the classic configuration, where the target is the easiest-to-activate member of a paradigm (e.g., *monté-monter* ‘climbed-climb,’ where the target *monter* has the highest SF in the family). By contrast, the second experiment took as targets less frequent inflected forms (e.g., *monté-montons* ‘climbed-we climb,’ where *montons* has a low SF within the family), thus reversing the typical design in which the target corresponds to the base, i.e., the member of the morphological family that already has the greatest residual activation because of its frequency. Under the conditions of the second experiment, only small pseudo-family-size verbs induce repetition and morphological priming, for both frequent and infrequent inflections, whereas large pseudo-family-size verbs fail to induce repetition or morphological priming. Moreover, inflectional priming for small pseudo-families verbs does not differ for the two types of primes, i.e., frequent or not frequent inflections. These data added new evidence to the

view that both the lexical frequency of word-forms and relative frequencies between primes and targets influence morphological processing.

THE SALIENCE OF WHOLE WORDS IN AN INTEGRATIVE PERSPECTIVE

All the data presented in Section “Objections to the Decompositional Approach” can be interpreted in a way that is straightforwardly compatible with the holistic view. In our view, advocates of the decompositional view of word recognition have systematically confused two types of results: On the one hand, data obtained on the basis of complex words and non-words whose surface morphemes can be rapidly and easily extracted have been interpreted as supporting automatic morphological decomposition. On the other hand, obstacles to a perfect morphological segmentation have been attributed to the robustness of the decomposition mechanism.

Returning to the notion of morphological saliency, this property as derived from the decompositional perspective is based only on the surface morphemic complexity of the stimuli and is opposed to another definition under which morphological saliency emerges from form-meaning correlations. While the former reduces morphological to formal effects, the latter stresses the role of paradigmatic relationships between words without denying the role of morphemes during word recognition. Aronoff (2007) claims with respect to this issue that “[t]here is plenty of evidence, linguistic and psycholinguistic, for morphemes and roots and for morphological relatedness. But none of this evidence, pace Stokall and Marantz (2006), supports a purely morpheme-based theory over one that recognizes lexemes but also recognizes roots and morphemes as morphologically significant elements, albeit not as reliable Saussurean signs” (p. 813). In line with this statement, we recognize the existence of morphemes, but only as secondary and derivative units of description.

As mentioned above, the empirical data from the psycholinguistic literature so far have mostly been interpreted in favor of a decompositional view, which reduces morphological effects to formal effects. But if morphological saliency only relates to the surface structure of words, this saliency, which seems to guide the early stages of word recognition, cannot be called ‘morphological’ since morphological relationships are, by definition, pairings of form, and meaning (Blevins, 2014). On the other hand, numerous studies have shown that ‘morphological’ priming is distinct from mere formal relationships: *freeze* does not prime *free* while both *hunter-hunt* and *corner-corn* show facilitation effects. The relevant priming effect must therefore take place at a level which is more than formal, but less than morphological. However, this structural saliency effect does not exclude a genuine morphological saliency effect emerging from paradigmatic relationships between the word representations coded within the mental lexicon. In other words, we assume the co-existence of both morphological structure and whole-word

saliency effects, but while the former depends on quantitative factors such as the statistical occurrence of letter clusters (including those that correspond to morphemes), the latter is determined by qualitative variables (e.g., the degree of semantic transparency) resulting from morphological relationships shared by words.

The present review has presented and discussed the factors which guide the processing and the lexical representation of morphologically complex words, and has given an overview of the highly controversial debate on possible interpretations of the results obtained so far. More specifically, we have shown that the issue of the relative prominence of the whole word and its morphological components has been overshadowed by the fact that psycholinguistic research has progressively focused on purely formal and superficial features of words, drawing researchers’ attention away from what morphology really is: systematic mappings between form and meaning. While we do not deny that formal features can play a role in word processing, an account of the general mechanisms of lexical access also needs to consider the perceptual and functional saliency of lexical and morphological items.

We hold that results obtained on the basis of masked priming are in line with holistic models of lexical architecture or models in which morphology emerges from the systematic overlap between forms and meanings (Baayen et al., 2011). In such models, saliency is not only a matter of internal structure, but also results from the organization of words in morphological families and series; as a consequence not only syntagmatic, but also paradigmatic relationships must be taken to contribute to morphological saliency.

Certainly, the notion of saliency refers primarily to formal aspects, because the perceptual body of the morpheme is necessarily the starting point of the processing mechanism. However, the notion of saliency makes sense for complex word processing only if the form it refers to is associated with a meaning or function. Saliency, in other words, is a property of the morpheme (i.e., a stable association of form and meaning), not simply of a phonetic or graphemic chain. We suggest that re-focusing attention on saliency, rather than on purely formal aspects, could lead to more interesting interpretations of the data observed so far in the psycholinguistic literature.

AUTHOR CONTRIBUTIONS

HG: Psycholinguistic contribution; SD: Linguistic contribution.

FUNDING

This work was funded by Agence Nationale de la Recherche. Be-SyMPHONic Project: Human Behavior and Machine Simulation in the Processing of (Mor)Phonotactics. International cooperation with: Austria/FWF (Fonds zur Förderung der wissenschaftlichen Forschung). Name and first name of the French coordinator: Basilio Calderone, CLLE Laboratory, CNRS, and University of Toulouse, France.

REFERENCES

- Amenta, S., and Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Front. Psychol.* 3:232. doi: 10.3389/fpsyg.2012.00232
- Andrews, S. (1989). Frequency and neighborhood size effects on lexical access: activation or search? *J. Exp. Psychol.* 15, 802–814.
- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: lexical similarity or orthographic redundancy? *J. Exp. Psychol.* 18, 234–254.
- Aronoff, M. (1994). *Morphology by Itself*. Cambridge: MIT Press.
- Aronoff, M. (2007). In the beginning was the word. *Language* 83, 803–830. doi: 10.1353/lan.2008.0042
- Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual route model. *J. Mem. Lang.* 37, 94–117. doi: 10.1006/jmla.1997.2509
- Baayen, R. H., Milin, P., Filipovic Durdevic, D., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–482. doi: 10.1037/a0023851
- Beauvillain, C. (1996). The integration of morphological and whole-word form information during eye fixations on prefixed and suffixed words. *J. Mem. Lang.* 35, 801–820. doi: 10.1006/jmla.1996.0041
- Bertram, R., Hyönä, J., and Laine, M. (2000a). The role of context in morphological processing: evidence from Finnish. *Lang. Cogn. Process.* 15, 367–388. doi: 10.1080/01690960050119634
- Bertram, R., Laine, M., and Karvinen, K. (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: evidence from a morphologically rich language. *J. Psycholinguist. Res.* 28, 213–226. doi: 10.1023/A:1023200313787
- Bertram, R., Schreuder, R., and Baayen, R. H. (2000b). The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *J. Exp. Psychol.* 26, 489–511.
- Beyersmann, E., Casalis, S., Ziegler, J. C., and Grainger, J. (2014). Language proficiency and morpho-orthographic segmentation. *Psychon. Bull. Rev.* 22, 1054–1061. doi: 10.3758/s13423-014-0752-9
- Beyersmann, E., Castles, A., and Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: evidence from masked priming. *Q. J. Exp. Psychol.* 65, 1306–1326. doi: 10.1080/17470218.2012.656661
- Beyersmann, E., McCormick, S., and Rastle, K. (2013). Letter transpositions within morphemes and across morpheme boundaries. *Q. J. Exp. Psychol.* 66, 2389–2410. doi: 10.1080/17470218.2013.782326
- Blevins, J. P. (2014). “The morphology of words,” in *The Oxford Handbook of Language Processing*, eds M. Goldrick, V. Ferreira, and M. Miozzo (Oxford: Oxford University Press), 152–164.
- Bodner, G. E., and Masson, M. E. J. (2003). Beyond spreading activation: an influence of relatedness proportion on masked semantic priming. *Psychon. Bull. Rev.* 10, 645–652. doi: 10.3758/BF03196527
- Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.
- Boudelaa, S., and Marslen-Wilson, W. D. (2004a). Abstract morphemes and lexical representation: the cv-skeleton in Arabic. *Cognition* 92, 271–303. doi: 10.1016/j.cognition.2003.08.003
- Boudelaa, S., and Marslen-Wilson, W. D. (2004b). Allomorphic variation in Arabic: implications for lexical processing and representation. *Brain Lang.* 90, 106–116. doi: 10.1016/S0093-934X(03)00424-3
- Boudelaa, S., and Marslen-Wilson, W. D. (2005). Discontinuous morphology in time: incremental masked priming in Arabic. *Lang. Cogn. Process.* 20, 207–260. doi: 10.1080/01690960444000106
- Burani, C., and Caramazza, A. (1987). Representation and processing of derived words. *Lang. Cogn. Process.* 2, 217–227. doi: 10.1080/01690968708406932
- Burani, C., Salmaso, D., and Caramazza, A. (1984). Morphological structure and lexical access. *Visible Lang.* 18, 342–352.
- Burani, C., and Thornton, A. M. (2003). “The interplay of root, suffix and whole-word frequency in processing derived words,” in *Morphological Structure in Language Processing*, eds R. H. Baayen and R. Schreuder (Berlin: Mouton de Gruyter), 157–207.
- Bybee, J. (1988). “Morphology as lexical organization,” in *Theoretical Morphology: Approaches to Modern Linguistics*, eds M. Hammond and M. Noonan (San Diego, CA: Academic Press), 119–142.
- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Colé, P., Beauvillain, C., and Segui, J. (1989). On the representation and processing of prefixed and suffixed derived words: a differential frequency effect. *J. Mem. Lang.* 28, 1–13. doi: 10.1016/0749-596X(89)90025-9
- Colé, P., Segui, J., and Taft, M. (1997). Words and morphemes as units for lexical access. *J. Mem. Lang.* 37, 312–330. doi: 10.1006/jmla.1997.2523
- Crepaldi, D., Hemsworth, L., Davis, C. J., and Rastle, K. (2016). Masked suffix priming and morpheme positional constraints. *Q. J. Exp. Psychol.* 69, 113–128. doi: 10.1080/17470218.2015.1027713
- De Jong, N. H., Schreuder, R., and Baayen, R. H. (2000). The morphological family size effect and morphology. *Lang. Cogn. Process.* 15, 329–365. doi: 10.1080/01690960050119625
- Diependaele, K., Lemhöfer, K., and Brysbaert, M. (2013a). The word frequency effect in first and second language word recognition: a lexical entrenchment account. *Q. J. Exp. Psychol.* 66, 843–863. doi: 10.1080/17470218.2012.720994
- Diependaele, K., Morris, J., Serota, R. M., Bertrand, D., and Grainger, J. (2013b). Breaking boundaries: letter transpositions and morphological processing. *Lang. Cogn. Process.* 7, 988–1003. doi: 10.1080/01690965.2012.719082
- Diependaele, K., Sandra, D., and Grainger, J. (2005). Masked cross-modal morphological priming: unraveling morpho-orthographic and morpho-semantic influences in early word recognition. *Lang. Cogn. Process.* 20, 75–114. doi: 10.1080/01690960444000197
- Diependaele, K., Sandra, D., and Grainger, J. (2009). Semantic transparency and masked morphological priming: the case of prefixed words. *Mem. Cogn.* 37, 895–908. doi: 10.3758/MC.37.6.895
- Dressler, W. U., Panagl, O., Mayerthaler, W., and Wurzel, W. U. (1987). *Leitmotifs in Natural Morphology*. Philadelphia, PA: John Benjamins Publishing Company.
- Feldman, L. B., Milin, P., Cho, K. W., Martin, F. M., and O’Connor, P. A. (2015). Must analysis of meaning follow analysis of form? A time course analysis. *Front. Hum. Neurosci.* 9:111. doi: 10.3389/fnhum.2015.00111
- Feldman, L. B., O’Connor, P. A., and Martin, F. M. (2009). Early morphological processing is morpho-semantic and not simply morpho-orthographic: An exception to form-then-meaning accounts of word recognition. *Psychon. Bull. Rev.* 16, 684–691. doi: 10.3758/PBR.16.4.684
- Ford, M. A., Davis, M. H., and Marslen-Wilson, W. D. (2010). Derivational morphology and base morpheme frequency. *J. Mem. Lang.* 63, 117–130. doi: 10.1016/j.cortex.2013.08.007
- Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain Lang.* 68, 5–15. doi: 10.1006/brln.1999.2078
- Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol.* 10, 680–698.
- Giraudo, H., and Dal Maso, S. (2016). Suffix perceptual saliency in morphological masked priming. *Lingue Linguaggio* 1, 85–106. doi: 10.1418/83655
- Giraudo, H., and Grainger, J. (2000). Effects of prime word frequency and cumulative root frequency in masked morphological priming. *Lang. Cogn. Process.* 15, 421–444. doi: 10.1080/01690960050119652
- Giraudo, H., and Grainger, J. (2003). “On the role of derivational affixes in recognizing complex words: evidence from masked priming,” in *Morphological Structure in Language Processing*, eds R. H. Baayen and R. Schreuder (Berlin: Mouton de Gruyter), 209–232.
- Giraudo, H., and Voga, M. (2013). “Prefix units within the mental lexicon,” in *Morphology in Toulouse. Selected Proceedings of Décembrettes 8*, eds N. Hathout, F. Montermini, and J. Tseng (München: Lincom Europa), 61–78.
- Giraudo, H., and Orihuela, K. (2015). “Visual word recognition of morphologically complex words: effects of prime word and root frequency,” in *Proceedings of the NetWordS Final Conference: Word Structure and Word Usage*, eds V. Pirrelli, C. Marzi, and M. Ferro (Pisa: CEUR-WS.org Team), 128–131.
- Goldschneider, J. M., and DeKeyser, R. M. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: a meta-analysis of multiple determinants. *Lang. Learn.* 51, 1–50. doi: 10.1111/1467-9922.00147
- Järvi, J., and Pykkönen, P. (2011). Sub- and supralexical information in early phases of lexical access. *Front. Psychol.* 2:282. doi: 10.3389/fpsyg.2011.00282
- Kazanina, N. (2011). Decomposition of prefixed words in Russian. *J. Exp. Psychol.* 37, 1371–1390. doi: 10.1037/a0024335
- Kazanina, N., Dukova-Zheleva, G., Geber, D., Kharlamov, V., and Tonciulescu, K. (2008). Decomposition into multiple morphemes during lexical access: a

- masked priming study of Russian nouns. *Lang. Cogn. Process.* 23, 800–823. doi: 10.1080/01690960701799635
- Lavric, A., Clapp, A., and Rastle, K. (2007). ERP evidence of morphological analysis from orthography: a masked priming study. *J. Cogn. Neurosci.* 19, 866–877. doi: 10.1162/jocn.2007.19.5.866
- Lehtonen, M., Monahan, P. J., and Poeppel, D. (2011). Evidence for early morphological decomposition: combining masked priming with magnetoencephalography. *J. Cogn. Neurosci.* 23, 3366–3379. doi: 10.1162/jocn_a_00035
- Longtin, C.-M., and Meunier, F. (2005). Morphological decomposition in early visual word processing. *J. Mem. Lang.* 53, 26–41. doi: 10.1016/j.jml.2005.02.008
- Longtin, C.-M., Segui, J., and Hallé, P. A. (2003). Morphological priming without morphological relationship. *Lang. Cogn. Process.* 183, 313–334. doi: 10.1080/01690960244000036
- Luke, S. G., and Christianson, K. (2012). Semantic predictability eliminates the transposed-letter effect: insights from a combined self-paced reading and masked-priming paradigm. *Mem. Cogn.* 40, 628–641. doi: 10.3758/s13421-011-0170-4
- Marslen-Wilson, W. D., Bozic, M., and Randall, B. (2008). Early decomposition in visual word recognition: dissociating morphology, form and meaning. *Lang. Cogn. Process.* 23, 394–421. doi: 10.1080/01690960701588004
- McCormick, S. F., Brysbaert, M., and Rastle, K. (2009). Is morphological decomposition limited to low-frequency words? *Q. J. Exp. Psychol.* 62, 1706–1715. doi: 10.1080/17470210902849991
- McCormick, S. F., Rastle, K., and Davis, M. H. (2008). Is there a 'fete' in 'fetish'? Effects of orthographic opacity on morpho-orthographic segmentation in visual word recognition. *J. Mem. Lang.* 58, 307–326. doi: 10.1016/j.jml.2007.05.006
- Morris, J., Grainger, J., and Holcomb, P. J. (2013). Tracking the consequences of morpho-orthographic decomposition using ERPs. *Brain Res.* 1529, 92–104. doi: 10.1016/j.brainres.2013.07.016
- Perea, M., and Carreiras, M. (2006). Do transposed-letter effects occur across lexeme boundaries? *Psychon. Bull. Rev.* 13, 418–422. doi: 10.3758/BF03193863
- Plag, I. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- Rastle, K., Davis, M., Marslen-Wilson, W. D., and Tyler, L. (2000). Morphological and semantic effects in visual word recognition: a timecourse study. *Lang. Cogn. Process.* 15, 507–537. doi: 10.1080/01690960050119689
- Rastle, K., Davis, M. D., and New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychon. Bull. Rev.* 11, 1090–1098. doi: 10.3758/BF03196742
- Rastle, K., and Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Lang. Cogn. Process.* 23, 942–971. doi: 10.1080/01690960802069730
- Rueckl, J. G., and Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Lang. Cogn. Process.* 26, 482–508. doi: 10.1080/01690965.2010.500020
- Sánchez-Casas, R., Igoa, J. M., and García-Albea, J. E. (2003). On the representation of inflections and derivations: data from Spanish. *J. Psycholinguist. Res.* 32, 621–668. doi: 10.1023/A:1026123315293
- Stanners, R. F., Neiser, J., and Painton, S. (1979). Memory representation for prefixed words. *J. Verbal Learning Verbal Behav.* 18, 733–743. doi: 10.1016/S0022-5371(79)90439-0
- Stokall, L., and Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *Ment. Lexicon* 1, 85–123. doi: 10.1075/ml.1.1.07sto
- Taft, M. (1979). Lexical access via an orthographic code: the Basic Orthographic Syllabic Structure (BOSS). *J. Verbal Learning Verbal Behav.* 18, 21–39. doi: 10.1016/S0022-5371(79)90544-9
- Taft, M. (1994). Interactive-activation as a framework for understanding morphological processing. *Lang. Cogn. Process.* 9, 271–294. doi: 10.1080/01690969408402120
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Q. J. Exp. Psychol.* 57, 745–765. doi: 10.1080/02724980343000477
- Voga, M., and Giraudo, H. (2009). "Pseudo-family size influences processing of French inflections: evidence in favor of a supralexical account," in *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*, eds F. Montermini, G. Boyé, and J. Tseng (Somerville, MA: Cascadia Proceedings Project), 148–155.
- Xu, J., and Taft, M. (2015). The effects of semantic transparency and base frequency on the recognition of English complex words. *J. Exp. Psychol.* 41, 904–910. doi: 10.1037/xlm0000052

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.

Copyright © 2016 Giraudo and Dal Maso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Toward a Unified Socio-Cognitive Framework for Salience in Language

Hans-Jörg Schmid* and Franziska Günther

Department of English and American Studies, Ludwig Maximilian University of Munich, Munich, Germany

Keywords: salience, expectation, contexts, language experience, social group

INTRODUCTION: OPPOSING VIEWS OF SALIENCE

To begin with, consider the following four statements, one by one:

- (1) The word seemed salient because it was the first word that came to my mind.
- (2) The word seemed salient because it was the first word that came to my mind in this context.
- (3) The word seemed salient because I had not expected to hear it in this context.
- (4) The word seemed salient because I had never heard it before.

It is not unlikely that all four statements seem plausible, although 1 and 2 are actually opposed to 4 and 3 respectively. Apparently, then, words can be considered salient because they are...

- (1) highly familiar and strongly entrenched,
- (2) highly expected in a given context,
- (3) highly unexpected in a given context, or
- (4) totally unfamiliar.

OPEN ACCESS

Edited by:

Alice Julie Blumenthal-Dramé,
University of Freiburg, Germany

Reviewed by:

Dagmar Divjak,
University of Sheffield, UK
Martin Hilpert,
University of Neuchâtel, Switzerland

*Correspondence:

Hans-Jörg Schmid
hans-joerg.schmid@lmu.de

Specialty section:

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

Received: 05 February 2016

Accepted: 11 July 2016

Published: 05 August 2016

Citation:

Schmid H-J and Günther F (2016)
Toward a Unified Socio-Cognitive
Framework for Salience in Language.
Front. Psychol. 7:1110.
doi: 10.3389/fpsyg.2016.01110

Surprisingly, linguists have actually relied on at least three of these four scenarios for defining the notion of salience (see also Bowman et al., 2013, for a psychological perspective). Scenario (1) lies at the heart of Giora's idea of salience as what is "foremost on one's mind [...] stored and coded in the mental lexicon" (Giora, 2003, p. 15). Scenario (2) accords with Geeraerts' view of onomasiological salience in terms of "the relative frequency with which a signifiant is associated with a given signifié" (Geeraerts, forthcoming), i.e., the frequency with which a word is used to denote a given piece of experience. Scenario (3) corresponds to understanding salience in terms of surprisal, as, e.g., proposed by Rácz: "A segment is cognitively salient if it has a large surprisal value when compared to an array of language input" (Rácz, 2013, p. 37; see also Friston, 2010; Clark, 2013; Fine et al., 2013; Divjak, 2016). Scenario (4) represents an extreme variant of type (3) which builds on memory-based *novelty* rather than context-based *surprise* (see Barto et al., 2013, for this distinction).

The four scenarios can be summarized systematically by a cross-tabulation of two types of sources of expectations, viz. long-term memory and current context, with two types of mechanisms of salience, viz. confirmation and violation of expectations:

- (1) Salience by context-free entrenchment: confirmation of expectations based on knowledge stored in long-term memory.
- (2) Salience by contextual entrenchment: confirmation of expectations derived from the probability of occurrence in the current context.
- (3) Salience by surprisal: violation of expectations derived from the probability of occurrence in the current context.
- (4) Salience by novelty: violation of expectations based on lack of stored knowledge.

In this paper we propose a unified framework for salience which reconciles these opposing conceptions by showing that they focus on different aspects of the interaction between knowledge, context, expectation, and external input.

EXPECTATION AND TYPES OF CONTEXTS

Recent theories of linguistic and general perceptual, cognitive, and/or neural systems and processing share the view that expectations primed by context are crucial for saliency effects to occur (see Levy, 2008; Friston, 2010; Clark, 2013; Fine et al., 2013; Jaeger and Snider, 2013; Divjak and Caldwell-Harris, 2015, pp. 59–60). The notions of *expectation* and *context* thus seem to hold the key to a better understanding of saliency.

We define *expectation* as the state of the cognitive system immediately prior to processing a given linguistic cue. This state represents the *immediate cognitive context* for the upcoming processing event. What becomes activated as immediate cognitive context results from the interaction between four types of input which we also regard as contexts:

- (1a) *Linguistic context*, i.e., what has been said before
- (1b) *Situational context*, i.e., the participants, time, place, setting, objects
- (1c) *Social context*, i.e., the type of social event, the social roles of and relations between participants
- (2) *General cognitive context*, i.e., general and linguistic knowledge and routines stored in long-term memory.

All four types of contexts cooperate in shaping the immediate cognitive context, and yet type (2) differs fundamentally from the other three types. Whereas types (1a–c) are based on the current perception of external events, general cognitive context is internal and based on long-term memory. However, as has been acknowledged in the psychological literature on saliency and attention (e.g., Wilder et al., 2011; Clark, 2013), the effects of perception-based external contexts on our immediate cognitive contexts are invariably modulated by our memory-based general cognitive contexts, because what we perceive, how we perceive it, and how we process it linguistically is strongly affected by what we already know. In addition to this interaction between current external contexts and long-term internal context (see also Fine et al., 2013), the three types of external contexts—linguistic, situational, and social—also influence each other. For example, the perception of the linguistic context will partly depend on that of the situational and social context in the use of deictic expressions such as *the book over there* or the use of forms of address like *Madam* or *Doctor*. A graphic representation of this view of expectation and context is provided in **Figure 1A**.

SALIENCE AS COMPARISON BETWEEN THE INCOMING LINGUISTIC CUE AND IMMEDIATE COGNITIVE CONTEXT

Saliency effects arise when an incoming linguistic cue is processed before the backdrop of the immediate cognitive context. Since saliency effects are considered to involve the confirmation or violation of expectations (see Introduction: Opposing Views of Saliency), the notion of saliency—both in perception and in language—logically depends on a comparison between expectations and the cue to be processed. This characteristic is shared by perceptual and linguistic saliency. A word that is

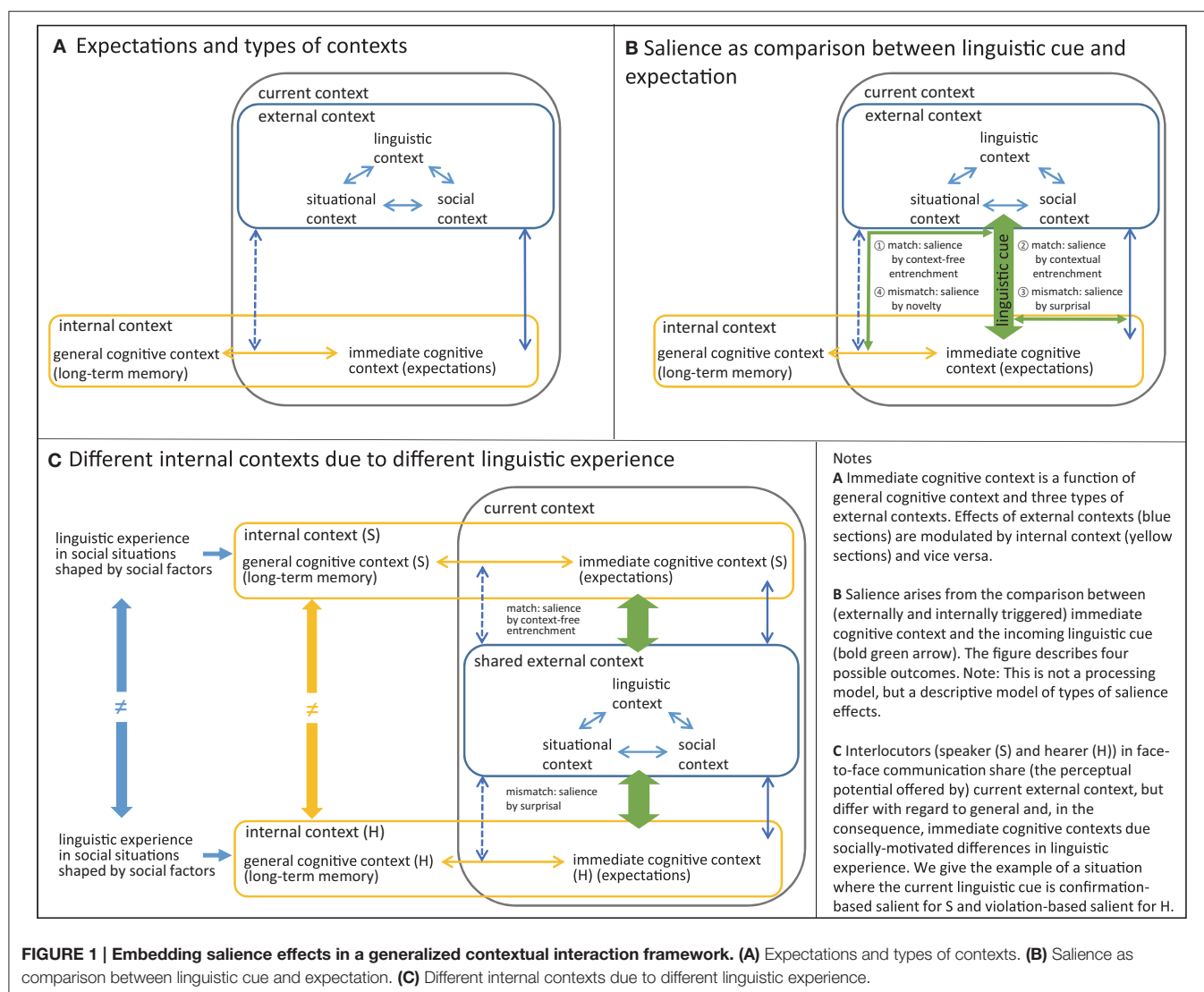
surprising in a given linguistic or situational context (see Scenario 3 in Introduction: Opposing Views of Saliency) is salient by virtue of the same principle as a green apple is in an array of red apples, i.e., through a comparison of a piece of information against its context. What is special about saliency in language is that linguistic context plays a key role, and that general long-term memory-based context includes the full range of entrenched linguistic knowledge and routines, i.e., the individual's current linguistic competence.

DIFFERENT VIEWS OF SALIENCE HIGHLIGHT DIFFERENT OUTCOMES OF THE COMPARISON

We would like to argue that the seemingly opposing types of saliency explained in the introduction correspond to four different outcomes of the comparison between the immediate cognitive context and its sources, and the incoming linguistic cue (see **Figure 1B**).

- (1) Saliency by context-free entrenchment: the incoming cue matches expectations that are mainly activated from general cognitive context, i.e., linguistic knowledge stored in long-term memory; this is the case for a very frequent word that is generally highly familiar and strongly entrenched, irrespective of the current context.
- (2) Saliency by contextual entrenchment: the incoming cue matches expectations whose activation is mainly triggered by current linguistic, situational, and/or social context; examples are words that are strongly suggested by what was said before (e.g., as part of a strong collocation), by situational circumstances (e.g., by reference to a salient object), or by social aspects of the speech event (e.g., in a ritualized speech event like a baptizing or wedding ceremony).
- (3) Saliency by surprisal: the incoming cue fails to match expectations that are mainly activated from current linguistic, situational, and/or social context; this could arise from violations of collocational restrictions or preferences, from unfamiliar ways of referring to objects, or from different conceptions of the social significance of words.
- (4) Saliency by novelty: the incoming cue completely fails to match up with expectations activated from long memory; the hearer encounters a word that he or she simply does not know.

The four different views of saliency thus highlight different interactions between internal and external contexts as sources of saliency on the one hand, and the mechanisms of confirmation and violation of expectations on the other. The main step forward made by the integrative and unified view we are proposing consists in the way in which it integrates internal and external as well as long-term and short-term contextual effects. This characteristic of the model opens up further options for explaining interactional and social saliency effects that we have neglected so far because we have focused on an individual idealized speaker.



VIOLATION-BASED SALIENCE IN INTERACTION CAN ARISE FROM EXPERIENCE-BASED SOCIAL DIFFERENCES BETWEEN SPEAKERS

Linguistic saliency effects arise in the interaction between two or more interlocutors. So the framework proposed thus far must be extended. **Figure 1C** represents the idealized case of two participants, a speaker (S) and hearer (H), engaged in face-to-face interaction. As is indicated in the Figure, in this case the participants largely share the same external linguistic, situational, and social context. The impact of these external contexts on their respective immediate cognitive contexts is not identical, however, partly because the participants may not have equal perceptual access to what was said before or to objects in the shared situation. More importantly, and as pointed out above, the effect of external context is modulated by internal long-term knowledge, which is by definition individual rather than shared (see Fine et al., 2013,

p. 1), and therefore differs from speaker to speaker (as is indicated by the arrows interrupted by the “is unequal” symbol).

The effect of these differences is that, despite shared external context, the current expectations of the two participants differ because the linguistic and encyclopedic knowledge they recruit for shaping their immediate cognitive contexts is not the same. **Figure 1C** illustrates a case where a linguistic cue (e.g., a word) that is highly familiar to the speaker is contextually surprising to the hearer. Such a word would be confirmation-based salient for the speaker, but violation-based salient for the hearer if the latter does not expect the word in this context or has never heard it before.

The likelihood of such situations correlates with the difference between the participants’ general cognitive contexts, i.e., their entrenched linguistic association patterns and routines. According to usage-based models of grammar (e.g., Barlow and Kemmer, 2000; The Five Graces Group, 2009) these patterns and routines are shaped by lifelong linguistic experience, which

is in turn shaped by social factors such as group-membership and participation in social networks and communities of practice (Schmid, 2015; see the left-hand side of in **Figure 1C**). At this point, the cognitive dimension of the framework is supplemented by the social dimension. While the cognitive dimension highlights the existence of individual differences, the social dimension licenses testable predictions concerning the sources of these differences and their effects on saliency. One such prediction is that interlocutors from distant social groups in terms of education, age, ethnicity, gender, and other classic social variables are more likely to experience violation-based saliency effects—"I have never heard this before," "I would not have expected this in this context"—than interlocutors who share their social background and linguistic experience. In this way, our framework naturally integrates saliency effects typically observed in sociolinguistic conceptions of saliency. We therefore regard it as an integrative and unifying socio-cognitive framework for understanding saliency. The paper by Jaeger and Weatherholtz (2016) in this special issue, which accords extremely well with the ideas presented here, provides more details and empirical evidence concerning the sociolinguistic aspects.

CONCLUSION

We have proposed a unified framework which reconciles the tension between opposing views of saliency by means of a differentiated conception of two central elements of saliency,

viz. *expectation* and *context*. Linguistic saliency emerges from a comparison between an incoming linguistic cue and expectations that are activated from the interaction between current perception-based linguistic, situational, and social context, and long-term memory-based cognitive context (i.e., linguistic and encyclopedic knowledge). Different existing conceptions of saliency highlight different aspects of this coherent framework. Experientially and socially motivated differences between the long-term memory-based cognitive contexts of individuals can be responsible for surprisal-based saliency effects. The framework proposed is thus *socio-cognitive* in the sense that it accommodates both cognitive and social causes of linguistic saliency effects.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We would like to thank Alice Blumenthal-Dramé, Adriana Hanulíková, and Bernd Kortmann for organizing the very stimulating workshop on *Perceptual linguistic saliency: Modeling causes and consequences* held in Freiburg, October 15th–17th 2014. We are grateful to the participants of this workshop for their illuminating talks and contributions to discussions, which were an important source of inspiration for this opinion article.

REFERENCES

- Barlow, M., and Kemmer, S. (eds.). (2000). *Usage-Based Models of Language*. Stanford, CA: CSLI Publications.
- Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or surprise? *Front. Psychol.* 4:907. doi: 10.3389/fpsyg.2013.00907
- Bowman, H., Filetti, M., Wyble, B., and Olivers, C. (2013). Attention is more than prediction precision. *Behav. Brain Sci.* 36, 26–28. doi: 10.1017/S0140525X12002324
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Divjak, D. (2016). The role of lexical frequency in the acceptability of syntactic variants: evidence from *that*-clauses in Polish. *Cogn. Sci.* doi: 10.1111/cogs.12335. [Epub ahead of print].
- Divjak, D., and Caldwell-Harris, C. L. (2015). "Frequency and entrenchment," in *Handbook of Cognitive Linguistics*, eds E. Dabrowska and D. Divjak (Berlin: de Gruyter), 53–75.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., and Quian, T. (2013). Rapid expectation adaption during syntactic comprehension. *PLoS ONE* 8:e77661. doi: 10.1371/journal.pone
- Friston, K. (2010). The free energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787
- Geeraerts, D. (forthcoming). "Entrenchment as onomasiological saliency," in *Entrenchment and the Psychology of Language: How We Reorganize and Adapt Linguistic Knowledge*, ed H.-J. Schmid (Boston, MA: American Psychological Association and Walter de Gruyter).
- Giora, R. (2003). *On Our Mind: Saliency, Context, and Figurative Language*. Oxford: Oxford University Press.
- Jaeger, T. F., and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition* 127, 57–83. doi: 10.1016/j.cognition.2012.10.013
- Jaeger, T. F., and Weatherholtz, K. (2016). What the heck is saliency? How predictive language processing contributes to sociolinguistic perception. *Front. Psychol.* 7:1115. doi: 10.3389/fpsyg.2016.01115
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006
- Rácz, P. (2013). *Saliency in Sociolinguistics*. Berlin: de Gruyter Mouton.
- Schmid, H.-J. (2015). "A blueprint of the entrenchment-and-conventionalization model," in *Yearbook of the German Cognitive Linguistics Association*, Vol. 3, eds P. Uhrig and T. Herbst (Berlin: de Gruyter Mouton), 1–27.
- The Five Graces Group [= Becker, C., Blythe, R. A., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al.] (2009). "Language as a complex adaptive system: position paper," in *Language as a Complex Adaptive System*, eds N. C. Ellis and D. Larsen-Freeman (Chichester: Wiley-Blackwell), 1–26.
- Wilder, M. H., Mozer, M. C., and Wickens, C. D. (2011). An integrative, experience-based theory of attentional control. *J. Vision* 11, 1–30. doi: 10.1167/11.2.8

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.

Copyright © 2016 Schmid and Günther. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



What the Heck Is Salience? How Predictive Language Processing Contributes to Sociolinguistic Perception

T. Florian Jaeger^{1,2,3*} and Kodi Weatherholtz¹

¹ Human Language Processing Lab, Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA, ² Department of Computer Science, University of Rochester, Rochester, NY, USA, ³ Department of Linguistics, University of Rochester, Rochester, NY, USA

Keywords: accent, dialect, idiolect, salience, surprisal, prediction, expectation, learning

INTRODUCTION

Some sociolinguistic variables are prone to hypercorrection, stigmatization and style shifting, while other variables are not. The status of the former type—sometimes called *stereotypes* and *markers* (Labov, 1972)—has been attributed to the increased meta-linguistic awareness language users seem to have of these variables. This awareness in turn is attributed to the *salience* of these variables, such that greater salience is assumed to cause greater meta-linguistic awareness (e.g., Trudgill, 1986). Salience has similarly been invoked when aiming to explain implicit social inferences about, or attitudes toward, speakers who exhibit certain variables in their speech (Babel, 2016; Drager and Kirtley, 2016; Squires, 2016). However, salience is a hard to define concept (for review, see Auer et al., 1998; Kerswill and Williams, 2002) and, partly as a consequence, “notoriously difficult to quantify” (Hickey, 2000). For a concept that plays such a central and ubiquitous role in sociolinguistic explanations, this is arguably a dangerous state of affairs.

This motivates the present commentary. We believe that advances in computational psycholinguistics offer definitions of sociolinguistic salience that are more concrete, both empirically and formally grounded, and quantifiable (and thus falsifiable). We propose that it is important to distinguish between the *initial salience* a listener experiences when first encountering a novel variant (e.g., because of exposure to a previously unfamiliar dialect, sociolect, or idiolect—henceforth *lects*; Schirmunski, 1930; Preston, 1996), and salience at later stages. Salience after the initial encounter is the cumulative product of an individual’s experience related to the lectal variant, including direct experience, as well as discourse about the variant (e.g., explicit stereotyping or enregisterment, Agha, 2003). Here we focus on the causes for initial salience, which we think *can* be defined in a principled and quantifiable way.

Specifically, we propose that salience in the first moments when a novel lect is encountered cannot be understood without reference to prior expectations based on listeners’ past language experience and the ensuing expectation violation that a listener experiences relative to those prior expectations—an idea explored in more depth by Rácz (2012, 2013). Here we contribute to these efforts. We draw on basic concepts from probability and information theory to define initial salience as a function of (top-down) prior expectations. This has several advantages. First, the proposed definition of salience is quantifiable (see also Rácz, 2013). Second, computational psycholinguistics has linked the very same quantities to language processing and learning. Recognizing this link offers the opportunity to ground sociolinguistic salience in human information processing—both empirically and theoretically—offering a parsimonious account of initial salience.

OPEN ACCESS

Edited by:

Adriana Hanulíková,
University of Freiburg, Germany

Reviewed by:

Lauren Squires,
Ohio State University, USA
Christian Langstrof,
University of Freiburg, Germany
Daniel Müller-Feldmeth,
University of Freiburg, Germany

*Correspondence:

T. Florian Jaeger
tjaeger@ur.rochester.edu

Specialty section:

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

Received: 07 April 2016

Accepted: 12 July 2016

Published: 03 August 2016

Citation:

Jaeger TF and Weatherholtz K (2016)
What the Heck Is Salience? How
Predictive Language Processing
Contributes to Sociolinguistic
Perception. *Front. Psychol.* 7:1115.
doi: 10.3389/fpsyg.2016.01115

After we have outlined our proposal, we briefly turn to an apparent puzzle that was raised during the workshop leading to this special issue: several presenters pointed out that saliency sometimes seems to be inversely related to the frequency of a variant and other times positively related. This puzzle readily dissolves once the view proposed here is taken into account.

FIRST ENCOUNTERS WITH A VARIANT: SURPRISAL AS A MEASURE OF INITIAL SALIENCY

Imagine a listener during the first moments of encountering a talker who speaks in an unfamiliar lect. The unfamiliar lect by definition differs from what the listener has previously experienced. Following the sociolinguistic literature, we can think of these differences as differences in the realization of linguistic variables, and the specific realization of the variables as lectal variants (Labov, 1966). What then makes a lectal variant salient in this hypothetical first encounter? Research in sociolinguistics has identified a number of perceptual features that can contribute to the perception of a variant as salient, such as a priori perceptual or articulatory distinctiveness (for review, see Auer et al., 1998). However, influences of prior experience are arguably as important or more important. Specifically, variants that are *unexpected* given the listener's *prior expectations about linguistic variables* (including, broadly speaking, the listener's language background) should be more salient in the moment they are experienced.

Events that we do not expect, or that are surprising to us, tend to stand out. There is now strong evidence that this anecdotal observation about strongly unexpected events extends to subtle and highly gradient differences in unexpectedness. During language processing, words and structures that are less expected are processed more slowly (e.g., MacDonald et al., 1994; Garnsey et al., 1997; McRae et al., 1998; MacDonald and Shillcock, 2003) and they are recognized accurately less often in noise (Cole and Perfetti, 1980; Grosjean, 1980). Critically, similar costs of unexpectedness are observed for unfamiliar lectal variants when comprehenders first encounter them (e.g., Kaschak and Glenberg, 2004; Squires, 2014a; Fraundorf and Jaeger, in press). Unexpectedness—or the degree to which something is violating our expectations based on previous experience—can be measured in a number of ways. One principled measure is referred to as *surprisal* (Hale, 2001; Levy, 2008). The surprisal associated with processing a certain input (e.g., a phonetic feature, phonological category, word, or syntactic structure) is identical to the amount of new information gained by processing the input, also known as the Shannon information (Shannon, 1948).

The surprisal of a unit is defined as the logarithm of the inverse of the contextual probability of the unit:

$$I(\text{unit}) = \log \frac{1}{p(\text{unit} | \text{context})} = -\log p(\text{unit} | \text{context}) \quad (1)$$

If the logarithm of the inverse contextual probability is taken to base 2, surprisal measures the number of bits of information gained by processing the input over and above what was expected

prior to processing the input. The surprisal of a word in (linguistic) context has been found to be proportional to its average reading time (Frank and Bod, 2011; Smith and Levy, 2013; Linzen and Jaeger, 2016). Surprisal has also been found to be correlated with neural signatures in ERP or MEG studies (Frank et al., 2015; for further references, see Kuperberg and Jaeger, 2016).

Recent studies have further linked surprisal to implicit learning operating during language processing (Fine and Jaeger, 2013; Jaeger and Snider, 2013; for a related view, see Dell and Chang, 2014). As is well-known from sociolinguistic research, talkers differ in their pronunciation, lexical, and syntactic preferences (among other things, Labov, 1972). As a consequence, efficient and robust language processing requires that linguistic expectations need to flexibly adapt to these differences (Fine et al., 2013; Kleinschmidt and Jaeger, 2015). Indeed, expectation adaptation has now been documented for speech perception (Clayards et al., 2008; for review, see Weatherholtz and Jaeger, in press), lexical (Creel et al., 2008), syntactic (Fine et al., 2013), and prosodic processing (Kurumada et al., under review), including adaptation to novel lectal variants (e.g., Kaschak and Glenberg, 2004; Bradlow and Bent, 2008; Kraljic et al., 2008; Fraundorf and Jaeger, in press). Adaptation to changes in the statistics of the environment should be sensitive to surprisal (or more generally to expectation violation): the degree to which inputs differ from prior expectations is informative about how and how much learners need to adapt their future expectations (Courville et al., 2006; Qian et al., 2012). Consistent with this prediction, there is evidence that the amount of expectation adaptation after processing unexpected linguistic input is proportional to that input's surprisal (Fine and Jaeger, 2013; Arai and Mazuka, 2014; for related evidence from production, see Bernolet and Hartsuiker, 2010; Jaeger and Snider, 2013).

Taken together, this research suggests that surprisal (or its generalization, Bayesian surprise; Itti and Baldi, 2009) is a plausible measure of “unexpectedness” and, as such, one factor that is likely to contribute to the initial saliency of newly encountered lectal variants. Specifically, it is the surprisal of the variant *given the prior expectations of the listener* that is expected to predict initial saliency. These prior expectations, we further submit, depend not only on *linguistic context* (e.g., the probability of a lectal variant given surrounding phonological or lexical information, including the presence or absence of other lectal variants) but also on *social context* (e.g., the probability of a lectal variant given socio-indexical information about the talker).

Consider, for example, a specific linguistic variable, such as /t/-deletion or flapping: if this variant occurs overall much more frequently in a newly encountered lect than a priori expected or in different phonological and lexical contexts than a priori expected, it will have high surprisal (this reasoning also extends to novel, not previously encountered, variables)¹. It is in this sense that the saliency of a lectal variant is *inversely* related

¹Under the naïve assumption that everything that has never been observed is considered to have a probability of 0, the surprisal of a novel variant would be infinite. This is avoided, if some probability mass is held out to account for the fact that we do, in fact, observe novel events even as adults.

to frequency—specifically to the *expected* relative contextual frequency of the variant².

Since the expectations that determine the surprisal of a lectal variant reflect the individual's previous language experience, it naturally follows that initial saliency can be “different for different social groups” (Kerswill and Williams, 2002) and individuals (see also Hickey, 2000; Campbell-Kibler, 2012). Specifically, initial saliency should depend on which lects the individual has previously been exposed to, the frequency of the novel lectal variant in those familiar lects, and perhaps the frequency of similar variants in familiar lects (see Squires, 2014b). Next we turn to the question of how the initial saliency of a variant is related to the probability that the variant will become associated with the lect, thereby acquiring social meaning.

BEYOND THE FIRST ENCOUNTER: FREQUENCY AND ASSOCIATION

What then happens over time, as a novel lectal variant is encountered again? Consider a novel talker producing a high surprisal variant only once, compared to producing that (equally high surprisal) variant repeatedly. Intuitively, listeners should be more likely to learn an association between the variant and the novel lect in the latter case: while the surprisal of a lectal variant determines how much it “stands out,” the frequency with which the lectal variant is observed increases the probability that the variant is perceived and learned—a prerequisite to becoming associated with the lect. It is in this sense that the resulting sociolinguistic saliency of a variant is *positively* related to its (*actually observed* relative) frequency *in the novel lect*. Note that this is not in conflict with our previous statement. Surprisal is predicted to cause the initial saliency experienced when observing a lectal variant that was unexpected based on prior experience. High frequency in the novel lect—or specifically the cumulative effect of the surprisal experienced whenever a variant is encountered again—is predicted to increase the likelihood that the listener learns that the variant is associated with the lect (this idea is closely related to the *mutual information* between the variant and lect).

This also predicts that lectal variants can become associated almost instantaneously with a new lect or social group *if the variant is particularly unexpected* (as seems to be the case, Squires, 2014a). Such *ad-hoc* associations should be even more likely when listeners have other reasons to believe (rightly or wrongly) that the producer belongs to a novel group—a prediction that, to the best of our knowledge has not been directly tested.

Viewed this way, we can think of the sociolinguistic saliency that a lectal variant acquires over time as being a function of its (perceived) informativeness about social group membership.

²There is one caveat to this prediction: prior expectations also affect *what* we perceive (cf. perceptual illusions or the perceptual magnet effect; Kuhl, 1991), and therefore can lead to a non-faithful representation of the perceptual input (cf. Feldman et al., 2009).

This raises an interesting question for future research. There is now evidence that listeners develop and store implicit models or expectations about different lects that they have been exposed to (Niedzielski, 1999; Strand, 1999; Bradlow and Bent, 2008; Walker and Hay, 2011; Hanulíková et al., 2012; Shaw et al., 2015; for review, see Foulkes and Hay, 2015; Kleinschmidt and Jaeger, 2015). It is, however, still an open question to what extent the features that these implicit expectations are conditioned on are the same that more explicit processes, such as stereotyping refer to.

CONCLUSION

We propose that research on sociolinguistic saliency needs to take into account what is known about language processing and learning (see also Rácz, 2013; for a related perspective that grew out of the same workshop, see Schmid and Günther, 2016). One consequence of this is that the surprisal and frequency of lectal variants are likely predictors of a variant's saliency. Specifically, surprisal is high when first encountering unfamiliar lectal variants. With further exposure, the association between the variant and the lect increases, while the surprisal evoked by the variant decreases.

One advantage of this approach to saliency is that it makes novel testable predictions, some of which we have derived above. A second benefit is that surprisal and frequency are quantitative measures that can—in principle (provided suitable corpora)—be estimated objectively from language database. Of course, other properties of lectal variants (e.g., differences in a priori perceptual saliency, such as loudness) or processes operating over them are likely to affect saliency (e.g., enregisterment, which will selectively strengthen the associations between a lectal variant and the lect; Agha, 2003; Schmid, 2007). However, these other contributors to saliency are generally difficult to measure reliably. We thus submit that the proposal outlined here should be taken into account *first*, providing a baseline for a variant's expected saliency.

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGMENTS

We thank Alice Blumenthal-Dramé, Adriana Hanulíková, and Bernd Kortmann for organizing the workshop that led to this special issue, *Perceptual linguistic saliency: Modeling causes and consequence* held in Freiburg, October 15th to 17th 2014. The ideas expressed here benefitted from stimulating discussions with participants at the workshop and from the reviewers' comments, who went beyond the expected. Work on this paper was partially supported by NSF CAREER award IIS-1150028 and NICHD grant R01 HD075797 to TFJ. The views expressed here are not necessarily those of the funding agencies.

REFERENCES

- Agha, A. (2003). The social life of cultural value. *Lang. Commun.* 23, 231–273. doi: 10.1016/S0271-5309(03)00012-0
- Arai, M., and Mazuka, R. (2014). The development of Japanese passive syntax as indexed by structural priming in comprehension. *Q. J. Exp. Psychol.* 67, 60–78. doi: 10.1080/17470218.2013.790454
- Auer, P., Birgit, B., and Beate, G. (1998). Subjective and objective parameters determining ‘saliency’ in long-term dialect accommodation. *J. Sociolinguist.* 2, 163–187. doi: 10.1111/1467-9481.00039
- Babel, A. (ed.). (2016). *Awareness and Control in Sociolinguistic Research*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9781139680448
- Bernolet, S., and Hartsuiker, R. (2010). Does verb bias modulate syntactic priming? *Cognition* 114, 455–461. doi: 10.1016/j.cognition.2009.11.005
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005
- Campbell-Kibler, K. (2012). Contestation and enregisterment in Ohio’s imagined dialects. *J. Eng. Linguist.* 40, 281–305. doi: 10.1177/0075424211427911
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809. doi: 10.1016/j.cognition.2008.04.004
- Cole, R. A., and Perfetti, C. A. (1980). Listening for mispronunciations in a children’s story: the use of context by children and adults. *J. Verb. Learn. Verb. Behav.* 19, 297–315. doi: 10.1016/S0022-5371(80)90239-X
- Courville, A., Nathaniel, D. D., and David, S. T. (2006). Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* 10, 294–200. doi: 10.1016/j.tics.2006.05.004
- Creel, S. C., Aslin, R. N., and Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition* 106, 633–664. doi: 10.1016/j.cognition.2007.03.013
- Dell, G. S., and Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20120394. doi: 10.1098/rstb.2012.0394
- Drager, K., and Kirtley, J. (2016). “Awareness, saliency, and stereotypes in exemplar-based models of speech production and perception,” in *Awareness and Control in Sociolinguistic Research*. ed A. Babel (Cambridge, UK: Cambridge University Press), 1–24. doi: 10.1017/CBO9781139680448.003
- Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (2009). The influence of categories on perception: explaining the perceptual magnet effect as optimal statistical inference. *Psychol. Rev.* 116, 752–782. doi: 10.1037/a0017196
- Fine, A., and Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cogn. Sci.* 37, 578–591. doi: 10.1111/cogs.12022
- Fine, A. B., Florian Jaeger, T., Thomas, A. F., and Ting, Q. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE* 8:e77661. doi: 10.1371/journal.pone.0077661
- Foulkes, P., and Hay, J. (2015). “The emergence of sociophonetic structure,” in *The Handbook of Language Emergence*, eds B. MacWhinney and W. O’Grady (Hoboken, NJ: John Wiley and Sons, Inc.), 292–313.
- Frank, S. L., Leun, J. O., Giulia, G., and Gabriella, V. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain Lang.* 140, 1–11. doi: 10.1016/j.bandl.2014.10.006
- Frank, S. L., and Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychol. Sci.* 22, 829–834. doi: 10.1177/0956797611409589
- Fraundorf, S., and Jaeger, T. F. (in press). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *J. Mem. Lang.* doi: 10.1016/j.jml.2016.05.006
- Garnsey, S. M., Neal, J. P., Elizabeth, M., and Melanie, L. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *J. Mem. Lang.* 37/1, 58–93. doi: 10.1006/jmla.1997.2512
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28, 267–283. doi: 10.3758/BF03204386
- Hale, J. (2001). “A probabilistic Earley parser as a psycholinguistic model,” in *Proceedings of NAACL* (Pittsburgh, PA). doi: 10.3115/1073336.1073357
- Hanulíková, A., van Alphen, P. M., van Gochnd, M. M., and Andrea, W. (2012). When one person’s mistake is another’s standard usage: The effect of foreign accent on syntactic processing. *J. Cogn. Neurosci.* 24, 878–887. doi: 10.1162/jocn_a_00103
- Hickey, R. (2000). “Saliency, stigma and standard,” in *The Development of Standard English 1300-1800: Theories, Descriptions, Conflicts*, ed L. Wright, (London: Cambridge University Press). 57–72.
- Itti, L., and Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Res.* 49, 1295–1306. doi: 10.1016/j.visres.2008.09.007
- Jaeger, T. F., and Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition* 127, 57–83. doi: 10.1016/j.cognition.2012.10.013
- Kaschak, M. P., and Glenberg, A. M. (2004). This construction needs learned. *J. Exp. Psychol. Gen.* 133, 450–467. doi: 10.1037/0096-3445.133.3.450
- Kerswill, P., and Williams, A. (2002). “Saliency’ as an explanatory factor in language change: evidence from dialect levelling in urban England,” in *Language Change: The Interplay of Internal, External and Extra-linguistic Factors*, eds M. C. Jones and E. Esch (Berlin: Mouton de Gruyter), 81–101. doi: 10.1515/9783110892598.81
- Kleinschmidt, D. F., and Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychol. Rev.* 122, 148–203. doi: 10.1037/a0038695
- Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008). Accommodating variation: dialects, idiolects, and speech processing. *Cognition* 107, 54–81. doi: 10.1016/j.cognition.2007.07.013
- Kuhl, P. (1991). Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do no. *Percept. Psychophys.* 50, 93–107. doi: 10.3758/BF03212211
- Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299
- Labov, W. (1966). *Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic Patterns*, Oxford: Blackwell.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006
- Linzen, T., and Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. *Cogn. Sci.* doi: 10.1111/cogs.12274. [Epub ahead of print].
- MacDonald, M. C., Pearlmutter, N., and Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703. doi: 10.1037/0033-295X.101.4.676
- McDonald, S. A., and Shillcock, C. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychol. Sci.* 14, 648–652. doi: 10.1046/j.0956-7976.2003.psci_1480.x
- McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38, 283–312. doi: 10.1006/jmla.1997.2543
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Lang. Awareness* 5, 40–74. doi: 10.1080/09658416.1996.9959890
- Qian, T., Jaeger, T. F., and Aslin, R. (2012). Learning to represent a multi-context environment: more than detecting changes. *Front. Psychol.* 3:228. doi: 10.3389/fpsyg.2012.00228
- Rácz, P. (2012). Operationalising saliency: Definite article reduction in the North of England. *Engl. Lang. Linguist.* 16, 57–79. doi: 10.1017/S1360674311000281
- Rácz, P. (2013). *Saliency in Sociolinguistics*. Berlin/New York: De Gruyter Mouton.
- Schirmunski, V. (1930). Sprachgeschichte und Siedlungsmundarten. *Germanisch-Romanische Monatsschrift* 18, 113–122, 171–188.
- Schmid, H. J. (2007). “Entrenchment, saliency, and basic levels,” in *The Oxford Handbook of Cognitive Linguistics*, eds D. Geeraerts and H. Cuyckens (New York, NY: OUP), 117–138.
- Schmid, H.-J., and Günther, F. (2016). Towards a unified socio-cognitive framework for saliency in language. *Front. Psychol.* 7:1110. doi: 10.3389/fpsyg.2016.01110
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 379–423, 623–656.

- Shaw, J. A., Catherine, T. B., Karen, E. M., Gerard, D., Bronwen, G. E., Paul, F., et al. (2015). "Effects of short-term exposure to unfamiliar regional accents: Australians' categorization of London and Yorkshire English consonants," in *Proceedings of the 15th Australasian International Conference on Speech Science and Technology* (Christchurch), 3–5.
- Smith, N. J., and Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition* 128, 302–319. doi: 10.1016/j.cognition.2013.02.013
- Squires, L. (2014a). Knowledge, processing, evaluation: Testing the sociolinguistic perception of English subject-verb agreement variation. *J. Engl. Linguist.* 42, 144–172. doi: 10.1177/0075424214526057
- Squires, L. (2014b). *Social Differences in the Processing of Grammatical Variation*. Penn Working Papers in Linguistics 20. Available online at: <http://repository.upenn.edu/pwpl/vol20/iss2/20/>
- Squires, L. (2016). "Processing grammatical differences: Perceiving versus noticing," in *Awareness and Control in Sociolinguistic Research*, ed A. Babel (Cambridge, UK: Cambridge University Press), 80–103.
- Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *J. Lang. Soc. Psychol.* 18, 86–100. doi: 10.1177/0261927X99018001006
- Trudgill, P. (1986). *Dialects in Contact*, Oxford: Blackwell.
- Walker, A., and Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Lab. Phonol.* 2, 219–237. doi: 10.1515/labphon.2011.007
- Weatherholtz, K., and Jaeger, T. F. (in press). Speech perception and generalization across talkers and accents. *Oxf. Res. Encycl. Linguist.*

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.

The reviewer DM and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Jaeger and Weatherholtz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Estimating the Relative Sociolinguistic Salience of Segmental Variables in a Dialect Boundary Zone

Carmen Llamas*, Dominic Watt and Andrew E. MacFarlane

Department of Language and Linguistic Science, University of York, York, UK

OPEN ACCESS

Edited by:

Bernd Kortmann,
University of Freiburg, Germany

Reviewed by:

Barbara Johnstone,
Carnegie Mellon University, USA
Patrick Honeybone,
University of Edinburgh, UK

*Correspondence:

Carmen Llamas
carmen.llamas@york.ac.uk

Specialty section:

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

Received: 04 May 2016

Accepted: 20 July 2016

Published: 15 August 2016

Citation:

Llamas C, Watt D and
MacFarlane AE (2016) Estimating
the Relative Sociolinguistic Salience
of Segmental Variables in a Dialect
Boundary Zone.
Front. Psychol. 7:1163.
doi: 10.3389/fpsyg.2016.01163

One way of evaluating the salience of a linguistic feature is by assessing the extent to which listeners associate the feature with a social category such as a particular socioeconomic class, gender, or nationality. Such ‘top–down’ associations will inevitably differ somewhat from listener to listener, as a linguistic feature – the pronunciation of a vowel or consonant, for instance – can evoke multiple social category associations, depending upon the dialect in which the feature is embedded and the context in which it is heard. In a given speech community it is reasonable to expect, as a consequence of the salience of the linguistic form in question, a certain level of intersubjective agreement on social category associations. Two metrics we can use to quantify the salience of a linguistic feature are (a) the speed with which the association is made, and (b) the degree to which members of a speech community appear to share the association. Through the use of a new technique, designed as an adaptation of the Implicit Association Test, this paper examines levels of agreement among 40 informants from the Scottish/English border region with respect to the associations they make between four key phonetic variables and the social categories of ‘Scotland’ and ‘England.’ Our findings reveal that the participants exhibit differential agreement patterns across the set of phonetic variables, and that listeners’ responses vary in line with whether participants are members of the Scottish or the English listener groups. These results demonstrate the importance of community-level agreement with respect to the associations that listeners make between social categories and linguistic forms, and as a means of ranking the forms’ relative salience.

Keywords: salience, perception, borders, isogloss, indexicality, nationality, accent, dialect

INTRODUCTION

The study of the salience of speech sounds and other linguistic units can be approached in a diversity of ways, each based on different sets of assumptions about the nature and relative magnitude of the effect that an external stimulus has on the subject who is exposed to it. For some purposes it may be appropriate to focus on what salience means in terms of differences in the response sensitivity of the human peripheral auditory system, or to investigate how patterns of neuronal activation reveal inequalities in the prominence of certain speech stimuli relative to others or to background noise. We will henceforth use the term ‘salience’ to refer to

that property of a spoken form which causes listeners to respond to the form in such a way as to indicate that it encodes information about the (presumed) social characteristics and/or geographical origins of the speaker, alongside the linguistic functions that the form simultaneously fulfills (e.g., to help to distinguish the word in which it appears from other plausible candidate words): sociolinguistic saliency, in other words. Forms with high saliency are, according to this definition, argued to index social information more unequivocally than do forms with lower saliency. Variation in the directness of the mapping of sounds to speakers' non-linguistic characteristics means that when we test the association between the form and the social category it evokes, listeners are likely to respond faster and more consistently to high saliency forms than they are to low saliency forms.

To deduce the relative strength of a phonetic form's sociolinguistic saliency we must establish that the form does indeed function as a vehicle for social meaning. Also, for the association to be meaningful in terms of its capacity to index social information, the association should be the property of the group rather than just the individual, as such meaning is shared meaning. The association, therefore, must be one that listeners will generally agree upon.

'Top-down' associations of this type will inevitably differ between listeners, so strictly one-to-one relationships between phonetic forms and social group associations are unlikely to exist. Linguistic features carry multiple social category associations depending on the variety in which the features are embedded, the listeners' experience of the variety, and the context in which the features are heard (see further Niedzielski, 1999; Clopper and Pisoni, 2004; Johnstone and Kiesling, 2008; McGowan, 2015). As a result, any phonetic form may potentially index multiple social factors, as different listeners may associate it with different social groups.

As a testbed for the above claims we focus upon the border zone between Scotland and England that our previous research in the area (Llamas, 2010; Watt et al., 2014b) has shown us to be a region in which the prevalence of linguistic markers of ingroup and outgroup status is of particular significance. Consensus levels among community members with respect to social category associations with phonetic forms are quantified in the current study via an innovative adaptation of the Implicit Association Test (IAT; Greenwald et al., 1998). The new Social Category Association Test (SCAT) we present here allows the strength of association to be measured through analysis of response times, with faster responses implying a higher degree of certainty on the listener's part about the association, and slower ones demonstrating a level of hesitancy from which we can infer that the association is weaker and less direct.

As well as looking at shared agreement on social meaning across the border zone as a whole, we investigate differences in the responses gathered from inhabitants of communities on either side of the border and at its ends, where the border meets the coast. Age- and gender-related differences are also examined. We begin by considering the role of saliency and social meaning in how languages vary synchronically and change over time, before outlining the benefits of utilizing border zones – and the

particular border zone under investigation in our study – as test sites for the operationalization of saliency as an observable quantity. After examination of the results, we assess the extent to which we can propose differing degrees of saliency among segmental forms, based on community-level agreement on the forms' social category associations.

Saliency and Social Meaning

The saliency of a variable linguistic feature, from a sociolinguistic point of view, relates to the level of awareness that speakers have of that variable, which in turn is connected to the social meanings that become attached to its variants. According to Rácz (2013), saliency is an essential predictor of whether an *indicator* (a linguistic unit which varies non-randomly according to speakers' social characteristics) will become a *marker* (a feature of which speakers are aware to the extent that they adjust their use of it in line with the amount of attention they are paying to their speech; see further Labov, 1972, pp. 178–180). Increasing awareness of a marker may lead to it becoming a *stereotype*, in Labov's taxonomy, or acquiring 'third order indexicality,' as per Silverstein's (2003) model. However, by the time features become the topic of overt social comment, they may have become recessive in actual speech production. An example of a form which has attained the status of a stereotype is the apical trilled /r/ in Scottish English, which is popularly associated very closely with that variety, but which in fact occurs in modern Scottish English only infrequently.

The explanatory potential of saliency as a motivating factor in language change has long been acknowledged, though the factors involved in a variable becoming salient are still much debated. Trudgill (1986, p. 11) describes a set of testable criteria, comprised of both external (language-extrinsic, social) and internal (language-intrinsic) factors, according to which saliency can be attributed to forms in interactional situations. External factors include whether or not a variable is currently undergoing change, and the degree of overt stigmatization of its variants. Stigmatization of this kind often relates to whether a high-prestige form is represented in orthography; an example from British English is the long-standing stigma attached to /h/-dropping in content words such as *hat* or *house*. Internal factors include the maintenance of phonological contrast and the phonetic distance between the variants of a variable, whereby variants that are highly distinct from one another are more salient. Certain features, Trudgill claims, possess 'extra-strong saliency' thanks to their 'overly strong' association with particular accents or dialects. Forms of this sort are so closely associated with certain varieties, perhaps to the point of iconicity, that they tend to inhibit accommodation in dialect contact situations (see Llamas et al., 2009; Watt et al., 2010 for discussion of accommodation effects and saliency in the border context under investigation in the present study).

Kerswill and Williams (2002, p. 83) criticize Trudgill's criteria, arguing for the inclusion of extra-linguistic cognitive, social-psychological and pragmatic factors in the pool of factors contributing to saliency. They do so as a way of attempting to eliminate the circularity inherent in definitions of saliency which claim that forms are salient by virtue of language users

being more highly aware of them than they are of other forms. The extra-linguistic factors listed above are, according to Kerswill and Williams (2002, p. 105), ‘ultimately the cause of salience.’ Any operationalization of the notion of salience must, Kerswill and Williams assert, involve consideration of three components: (1) some patterning of language change or language variation for which the explanation may lie in the salience of the feature in question; (2) language-internal factors, such as the maintenance of phonological contrast; and (3) extra-linguistic cognitive, pragmatic, interactional, social-psychological, and socio-demographic factors (Kerswill and Williams, 2002, p. 105). Awareness on the researcher’s part of the subjective evaluation of forms and their embedding in linguistic structure is also a crucial element, given that these phenomena are subject to change within and beyond the speech community.

The primacy of social factors argued for by Kerswill and Williams (2002) is challenged by Hollmann and Siewierska (2006), who contend that cognitive perceptual factors are paramount. Through an examination of the Lancashire dialect, they propose that linguistic forms are free from social values when the forms first come into existence, and it is only after the forms have emerged that social forces start working on them (Hollmann and Siewierska, 2006, p. 27). They identify properties such as token frequency and transparency of the form-meaning relation as examples of the perceptual-cognitive factors they have in mind. Hollmann and Siewierska (2006, p. 27) concede that social factors play an important role in the process by which social values come to be attached to forms, but conclude that ‘ultimately it is the cognitive-perceptual constraints that make a form more or less liable to becoming subject to social evaluation and patterning.’

A similar position is adopted by Rácz (2013), who distinguishes between *cognitive salience* and *sociolinguistic salience*. The former, he argues, stems from the perceived difference between the transitional probability patterns of the realization of the variable in one dialect as opposed to another, which leads to listeners’ ‘surprisal’ and noticing of the variable. A form accrues sociolinguistic salience, by contrast, if it is mobilized for the purposes of social indexation (Rácz, 2013, p. 10). One of the case studies drawn upon by Rácz in his examination of salience in sociolinguistics is rhoticity (r-fulness) in Scotland. This has obvious relevance to the present study, as we shall see. Rácz (2013, p. 21) argues that rhoticity is a phonetically fine-grained process, and that the extent of phonetic variation in coda /r/ is such that it prevents listeners from targeting the feature as a reliable carrier of social indexation. Rácz also claims that the salience of features in a given speech community can be determined by a number of independent measures, proposing that the ‘best tools are attitude studies, which clearly show whether listeners associate the presence versus absence of a variant with a particular geographical location or social stratum’ (Rácz, 2013, p. 8). We will suggest in the current study that, on the basis of overt comments drawn from attitudinal data collected from our informants, the presence or absence of coda /r/ in words such as *car* or *farm* is seen as a key indicator of whether a talker is from Scotland or from England. Surprisingly, however, we will see that in the perceptual testing strand of the

project the expected association of rhoticity with Scotland turns out not to be a robust one.

Though we do not deny its central relevance to the study of perceptual salience, we are less concerned here with the *process* by which salience becomes attached to a linguistic form than we are with how to establish the degree of salience that that feature possesses. We would argue that a form’s salience is linked both to its capacity to index social meaning and the functions that it actually fulfills in this respect. It is probably true to say that, in terms of their potential for perceptual salience, some speech sounds are *a priori* better candidates than others. Some, such as strident fricatives, trills, or clicks, may be intrinsically more conspicuous, perceptually speaking, than (say) back rounded vowels or nasals, such that irrespective of any social information these more prominent sounds might convey about the talker they simply stand out from the acoustic background more than other sounds do (‘bottom-up’ salience); in the parsing of the speech stream, more salient sounds, according to Goldstein (1977, p. vi), ‘constrain higher-level decision processes more than others,’ affording them special value as conduits for linguistically relevant information (see also, e.g., Narayan, 2008). The notions of markedness and frequency are implicated too, as Bardovi-Harlig (1987), Podesva (2006), Honeybone and Watson (2013), and Watson and Clark (2013) demonstrate. It seems prudent, in any event, to allow for a certain level of unpredictability, even arbitrariness, when it comes to the identity of speech features destined to become sociolinguistically salient, in view of the evidence showing that features apparently lacking much acoustic or auditory prominence can nonetheless acquire substantial sociolinguistic salience. An example might be ‘TH-fronting,’ whereby the English dental fricatives /θ/ and /ð/ are realized as [f] and [v] respectively. Miller and Nicely (1955, p. 347) find that under controlled experimental conditions the distinctions between the fricative pairs [θ]~[f] and [ð]~[v] seem especially difficult for listeners to hear reliably, and yet in contemporary UK English TH-fronting is a widely attested sociophonetic variable that has long attracted overt comment and at times considerable stigma from laypeople (e.g., Levon and Fox, 2014; Baranowski and Turton, 2015). Where there is little to distinguish speech sounds from one another acoustically, it becomes more challenging to identify reasons why listeners might treat the forms in question as more significant social information-bearing units than others that they hear.

As we have suggested above, the associations that listeners make between linguistic forms and speakers’ social characteristics, and the extent to which listeners agree on those associations, vary from community to community. These sets of associations are therefore dynamic rather than static. We should also take account of how closely they are tied to production patterns in the communities to which the listeners belong, given that these patterns are similarly variable from place to place, and in view of the mutual dependence of production and perception. Our aim, then, is to assess salience in respect of the social category associations that linguistic forms embody for community members, as well as to examine correspondences between patterns in listeners’ perceptual responses to a form, and

spoken productions of the same form within the listeners' speech communities.

The Operationalization of Saliency in a Border Zone

Contexts in which markers of ingroup and outgroup status are known to be particularly prominent present ideal test sites for the investigation of saliency. Such contexts can be found in border regions, where linguistic and non-linguistic markers of claimed and ascribed identities generally abound. These markers are described by Kiely et al. (2001, p. 33) as '[t]hose social characteristics presented to others to support a national identity claim and looked to in others, either to attribute national identity, or receive and assess any claims of attributions made.' One of the behaviors that is most accessible to observers as a marker of this kind, according to Kiely et al. (2001), is accent. It follows that accent or dialect differences between localities in close geographical proximity to one another may be particularly sharply demarcated if the localities are separated by a political border.

The saliency of linguistic features has an important function for the inhabitants of border regions, as it assists with the categorization of speakers as ingroup or outgroup members according to a superficially straightforward binary distinction: that of being from one side of the border versus the other. In certain cases, linguistic forms may mark speakers out as members of a transborder community in a zone which straddles the border and which is distinct in social and/or linguistic ways from regions further away from the border. But in either scenario, linguistic forms are key carriers of social meaning that pertains to national and regional identity groupings.

Even when movement across a border is not in any way impeded, a political border – by definition – marks a discontinuity of some kind. We can in many cases point to linguistic artifacts of the divide: there are numerous dialect isoglosses which coincide closely with political boundaries, for instance. When isoglosses bundle together like this, we can say we have evidence for a dialect boundary. Regions where marked accent or dialect differences exist, such as areas divided by political borders, have great potential in terms of their capacity to show us how those differences are exploited by members of the communities as a means of claiming or ascribing different national identities in casual spoken interactions. Clearly, there will be many features which contribute to the listener's classification of an interlocutor as a member of a group from one side of the border or the other, but some features are likely to weigh more heavily in this evaluation than others. As border zones lend themselves very naturally to this kind of dichotomous grouping of speakers in terms of one nationality versus another, it is justifiable to treat linguistic variables which elsewhere may have complex and multiple indexicalities as forms which embody a binary opposition association (*nationality X* versus *not-nationality X*). We do so under the assumption that, in a border zone, this opposition is one that is both highly relevant and frequently encountered by local inhabitants. We are then in a position to put to the test our hypotheses concerning the extent to

which people living close to the border share the perception that a form reliably marks one national identity but not the other, as well as to measure the strength – that is, the relative saliency – of that perception within their communities.

The Scottish/English Border Region

We chose as the context of the present study four communities lying close to the political border separating Scotland from England. Inhabitants of localities in Scotland and England, two of the constituent nations of a single state (UK), have the possibility of claiming identities (*Scottish* versus *English*) which serve to distinguish them from people from the other side of the border, as well as an identity which unites them as a single category (i.e., *British*). This particular border therefore offers a productive testing ground for theoretical models of the convergent and divergent linguistic processes that take place along and across national and regional borders, and how these processes are manifested in the domains of speech production, speech perception and the claiming and ascribing of identity groupings.

Stretching for approximately 100 miles (160 km), the border separating Scotland and England is short compared to many other political frontiers. Nonetheless, its importance in linguistic terms is considerable. It has, indeed, been claimed to coincide with one of the most significant dialect boundaries in the Anglophone world. So numerous are the discontinuities in the distributions of phonological features in the area that the border has been dubbed a 'strong linguistic barrier' (Ihalainen, 1994, p. 248), while Aitken (1992, p. 895) asserts that the political border aligns with the 'most numerous bundle of dialect isoglosses in the English-speaking world.' This isomorphy, according to Aitken, effectively turns Scotland into a 'dialect island.' Among the phonological features that Aitken lists as contributors to the distinctiveness of Scottish varieties are the realization of the STRUT¹ vowel as [ʌ], the distribution and pronunciation of /r/, and the presence of the velar fricative [x] in words such as *night*. The Scottish Vowel Length Rule (SVLR; see Materials and Methods), a coherent set of alternations affecting multiple vowels in the system, is also seen as a key diagnostic of Scots and Scottish English (Aitken, 1984).

Glauser's (1974) traditional dialectological survey of the region revealed that the political border also coincided with a substantial bundle of lexical isoglosses. Glauser surveyed 106 locations around the border by collecting data from one informant per locality. The most common type of isogloss Glauser recorded was one separating a dialect form on the Scottish side of the border from a non-localized or standard form used on the English side (Glauser (1974, p. 278). When analyzed together, the isoglosses in Glauser's survey clustered particularly densely in the central, upland stretch of the border, which then (as now) was much more sparsely populated than the areas at the border's eastern and western ends. Transition zones were found at either end of the border, and while at the western end the transition zone

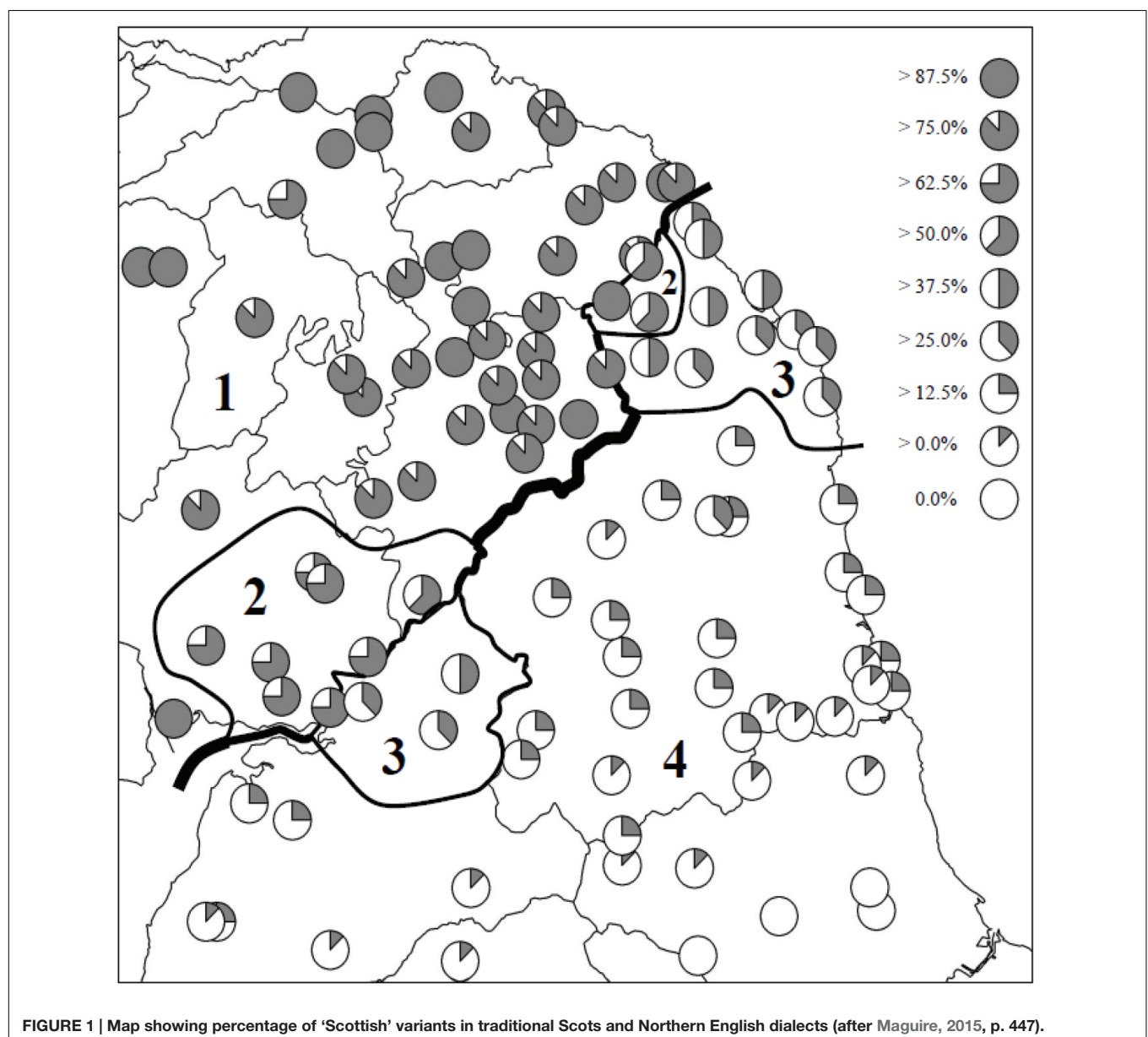
¹Throughout this article we make use of Wells' (1982) lexical set keywords, shown in small capitals (FLEECE, GOOSE, etc.), as a way of avoiding the ambiguity that often results from denoting vowel variables using symbols of the International Phonetic Alphabet.

straddled the border, in the east it occupied only the English side.

Of more relevance to the present study is Maguire's (2015) examination of phonological differences in the traditional dialects spoken on either side of the border. By plotting 22 of the dialects' phonological features, Maguire (2015) set out to investigate whether the same distributional patterns mapped by Glauser were also in evidence where phonological variation was examined. Of the phonological variables investigated, onset and coda /r/ were included, as was a vowel (PRICE) conditioned by the SVLR. For each locality, a 'Scottishness' index expressed as a percentage was calculated by pooling data collected from fieldwork sites sampled for volume 3 of *The Linguistic Atlas of Scotland* (LAS3; Mather and Speitel, 1986), the *Survey of English Dialects* (Orton and Dieth, 1962–1971), the *Orton Corpus*

(Rydland, 1998), and unpublished data gathered for the *Linguistic Survey of Scotland*. **Figure 1** presents Maguire's mapping of the degree of 'Scottishness' of the 22 phonological variables examined.

Maguire's findings reveal a pattern very similar to that which emerged from Glauser's lexical survey. A robust linguistic divide, more sharply delineated in its upland middle section than at its lowland endpoints, is resolved. Furthermore, the same transition zones – the western one spanning the border, the eastern confined to just the English side of the border – are visible in the phonological distributions. Evidence of the dividing effect of the border on the traditional dialects is clear: as Maguire puts it (Maguire, 2015, p. 448), 'we have two independent studies which confirm that the Scottish–English Border is the locus of a significant dialect discontinuity.'



Studies documenting phonological variation in the border zone since the traditional dialectological work was carried out have observed the erosion of traditional dialect forms in favor of patterning of a less localized nature under the influence of the standard Englishes of both England and Scotland (see further Johnston, 1980). Even so, linguistic distinctions between the border localities persist, as research by Maguire et al. (2010), and McMahon and Maguire (2011, 2013) demonstrates. Using an algorithm that generates a cross-dialectal distance metric, six varieties spoken in the border zone were compared. In line with the results described above, the analysis yielded evidence of a sharp distinction between the dialects from Scotland and those from England. Rhoticity is found to be a major contributor to the similarity measure, such that varieties cluster more tightly according to whether they are rhotic or non-rhotic than they do in respect of other similarities. In spite of the attrition of many of the features of traditional dialect which served to differentiate varieties from either side of the border, Maguire (2015, p. 452) concludes that ‘modern accents in the Border area are as complex as was the relationship between traditional dialects of the early 20th century.’

In addition to the border’s continuing status as a major spatial discontinuity in the distributions of traditional lexical and phonological features in the region, a perception among non-linguists that the border represents a deep and entrenched linguistic faultline is also readily apparent. Perceptual dialectological research by Montgomery (2014) reveals that the border has a psychological effect on the perception of dialect areas, as evidenced through a map drawing task. Montgomery’s data, gathered from informants living in towns on either side of the border, demonstrated a unidirectional proximity effect, with his English participants showing relatively little knowledge of variation in dialects of Scotland by comparison with their Scottish counterparts. Among the latter group, knowledge of variation in dialects of English spoken in England was similar to that possessed by respondents from the English side of the border.

The discontinuities in pronunciation features that align with the border are evidence of the halting or slowing of the progression of various sound changes that have spread toward the border, principally from the south. Patterns of phonological variation in the region imply that local people have formed strong associations between these features and relevant social groupings based on prominent in-/outgroups. As the forms in question index particular social categories of relevance, their use persists for as long as it is in speakers’ interests to mark social category memberships using linguistic resources. The perceptual dialectological research undertaken in the area suggests that the border represents a psychological divide linked to the placement of accent groups. Although language in the area undeniably undergoes change, the border’s political and ideological implications are such that the view prevails that the border continues to represent a potent linguistic boundary. Indeed, Maguire (2015, p. 454) states that ‘[w]ith the transition from traditional dialects to modern accents, the Border is continuing to act as an important linguistic boundary, not watertight but certainly an impediment

to change and indeed a focus of reinforcement of national identities.’

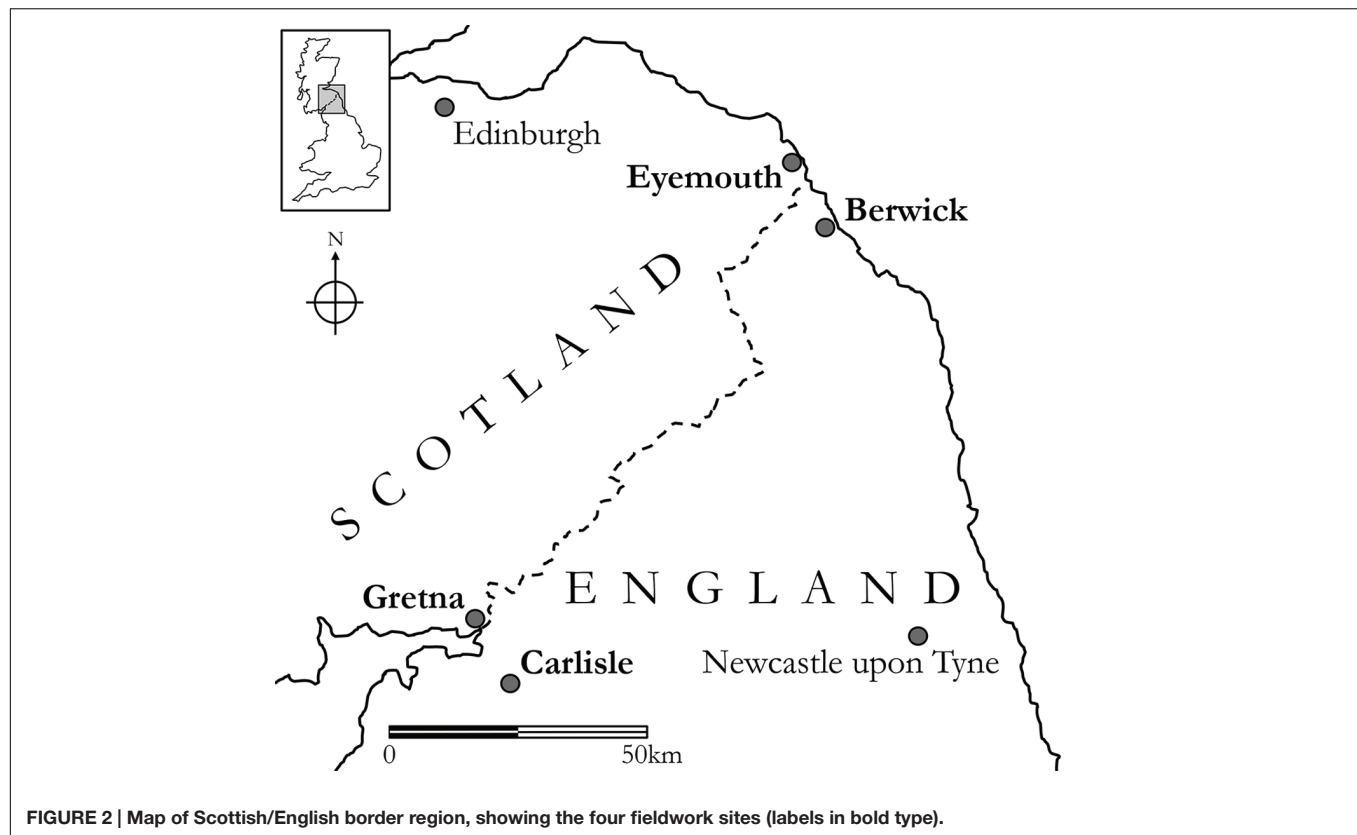
The AISEB Study

The *Accent and Identity on the Scottish/English Border* (AISEB) project was an empirical investigation of phonological variation and change in four border localities and of the social-psychological effects of the border in terms of how ingroup and outgroup categorizations were constructed and enacted by people living in the area. Four fieldwork sites were chosen – Gretna and Eyemouth in Scotland, and Carlisle and Berwick-upon-Tweed in England (see Figure 2). We chose ‘paired’ communities lying very close to the border and to their partner locality: the distance between Gretna and Carlisle, and between Eyemouth and Berwick, is less than 10 miles (16 km). The two Scottish localities are considerably smaller than the English ones [Gretna (2,700); Eyemouth (3,400); Carlisle (107,500); Berwick (12,000)].² The border does not inhibit movement – in physical terms it is invisible but for a few signs and flagpoles at the roadside – and in consequence there is plentiful contact between the paired cross-border localities. However, the population asymmetry in each pair of communities means that it is much more likely that residents of the Scottish towns will travel to the larger English localities than vice versa, a prediction which was confirmed very clearly in our informants’ interview responses. One might expect, then, that any linguistic changes taking place in the region would tend to go in the direction of the English model, with the Scottish speakers converging on the speech patterns of their English counterparts across the border.

The AISEB study took a tripartite approach to methodology, incorporating attitudinal and perceptual strands alongside the elicitation and analysis of production data. Our belief was that collecting data on the attitudinal positioning of the informants with respect to their identities and the socio-psychological effects of the border, and then combining these data with experimental evidence of the social meaning of key phonetic features, would yield a better understanding of the variation and change in the phonological patterns we uncovered. The current paper presents findings from just one of the tests used in AISEB’s perceptual strand, but relevant production and attitudinal data from the project are available in more detail elsewhere (see further Llamas, 2010; Llamas and Watt, 2014; Watt et al., 2014a,b).

As was found in previous surveys of the region (see The Scottish/English Border Region), production data collected for the AISEB project revealed marked differences between the Scottish and the English localities sampled. One of the main differences of particular relevance here was observed in rhoticity patterns. The speakers in the English localities were found to be effectively non-rhotic, while on the Scottish side of the border, rhotic forms were frequent. Eyemouth speakers, in particular, demonstrated near-categorical levels of r-fulness. In Gretna, at the border’s western end, rhoticity was much rarer and moreover was found to be decreasing considerably, with the younger

²Population figures are derived from Scotland’s Census Results OnLine (<http://www.scot.nhs.uk/>), Berwick-upon-Tweed Town Council (http://www.berwick-tc.gov.uk/town_council/), and the UK Office for National Statistics (<http://www.ons.gov.uk/>).



speakers using markedly lower rates of rhoticity than the older speakers (around 15% versus approximately 45%; see Watt et al., 2014b for further detail). These findings are in line with those from other varieties of Scottish English, particularly in Edinburgh and Glasgow (e.g., Lawson et al., 2014), where the process of derhoticity (coda /r/ loss) appears to be well underway.

Contrary to Rácz's assertion that coda /r/ is 'entirely ignored by the speaker' (Rácz, 2013, p. 147), we found that when our participants were asked to identify features associated with 'Scottish' as opposed to 'English' speech, they singled out /r/ more frequently than any other phonological feature in the border area, claiming it to be diagnostic of national and/or regional identity. Other pronunciation features were seldom mentioned, and were certainly not identified as consistently as /r/ is. Glauser's (2000, p. 75) suggestion that variation in /r/ is of primary importance among the set of features that inhabitants of the border area use to categorize speakers as Scottish or English appears to us a very reasonable stance. On balance then, and in light of the divergent production patterns mentioned above, we hypothesized that /r/ – particularly where it occurs in coda position – is the phonological form with the highest degree of saliency in the border zone. We turn next to the methods we used to test this prediction.

MATERIALS AND METHODS

The new and innovative SCAT formed part of a battery of tests designed to examine speakers' perceptions of social category

associations and of the geographical and social distributions of key phonetic forms in the border region. The tests were run on a subset of 40 of the original 160 speakers who had previously participated in the production and attitudinal strand of the project. For practical reasons, only a quarter of the full sample was invited to participate in the perception study, the time demands on individual participants having already been fairly heavy. The 40 subjects (10 in each of Eyemouth, Gretna, Berwick, and Carlisle) were split evenly by gender (male versus female) and into younger and older age groups (ages 16–24 and 57–82, respectively). The sample, therefore, can be divided into 20 older and 20 younger participants, as well as 20 Scottish and 20 English subjects. The fieldworker administered the perceptual tests in participants' homes. The study was carried out with approval from the Ethics Committee, University of York, UK. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

The SCAT ran as an adapted version of the IAT commonly used in psychological research (Greenwald et al., 1998). The IAT is typically used to access implicitly-held attitudes or associations by measuring the subject's automatic associations between different target categories (e.g., Black people versus White people) and positive or negative attributes, represented by adjectives with positive or negative meanings (e.g., beautiful, annoying, etc.). A series of sorting tasks is used to assess the automaticity of association between the target categories and positive or negative attributes. The difference in response times when the target category is sorted with positive as opposed to

negative attributes is taken as a measure of the difficulty of the task for the subject, and is argued to reveal differences in the subject's implicit attitudes between the target categories.

For the SCAT, the framework of the IAT was implemented in *PsyScope* (Cohen et al., 1993). Audio samples taken from word list readings from the larger production study sample were played to subjects through headphones, and their task was to indicate, as quickly as possible, whether they associated the sample with either *England* or *Not England* in one of the test blocks or *Scotland* or *Not Scotland* in the other. Unlike the IAT, there were no right or wrong answers in terms of the sorting task; any response was considered valid. The speed of the subject's decision and the degree of consensus across the group(s) to which the subject belonged were the metrics used to quantify saliency in this experiment.

Single-word audio files containing each of the target forms were extracted from recordings of authentic Scottish and English individuals who had participated in the production strand of AISEB. The phonetic forms chosen for the audio samples were:

- /r/ in coda position – *car* [kɑː] (rhotic) versus [kɑ:] (non-rhotic)
- tapped or approximant realizations of /r/ in onset position – *red* [rɛd] versus [ɹɛd]
- short or long realizations of FLEECE – *need* [nid] versus [ni:d]
- front or back realizations of GOOSE – *spook* [spʊk] versus [spʊk].

Although increasing derhoticization (also described as 'r/-vocalization') has been reported in Scottish varieties since the 1970s (see Reid, 1978; Romaine, 1978; Macafee, 1983; Johnston, 1997; Stuart-Smith, 2003, 2008; Stuart-Smith et al., 2014, among others), rhoticity is still considered one of the critical defining features of Scottish varieties of English (Wells, 1982). Northern England is effectively non-rhotic (Beal et al., 2012), and it was clear by the 1970s that derhoticization in Northumbria, England's northernmost county, was already practically complete (see Pålsson, 1972). As noted in The AISEB Study, findings from the production strand of AISEB confirm this discontinuity, in that the speakers we recorded in the English localities were almost uniformly non-rhotic, while those on the Scottish side of the border continued to exhibit high degrees of rhoticity (see further Llamas, 2010; Watt et al., 2014b). As is also noted in The AISEB Study, important east/west differences were revealed in the amount of rhoticity found among Scottish speakers, a factor we consider further in East/West Cross-Border Community Pairings.

In onset position, the alveolar tap [ɾ] occurs in varieties of English spoken in both Scotland and the far north of England (Llamas, 2001; Johnston, 2007; Stuart-Smith, 2008), but among the varieties spoken around the border is very much more frequent in the Scottish ones (Watt et al., 2014b). The approximant realization of /r/ can likewise be heard on either side of the border, but is more frequently and consistently used by speakers in England by virtue of their near-categorical avoidance of the tap and other available variants, and indeed is associated with England to the extent that in his work on phonological

variation in the border area Glauser (2000) refers to [ɹ] as the 'English /r/' in opposition to the 'Scottish' taps and trills.

The two vowel variables that we chose to include in the SCAT test, FLEECE and GOOSE, exemplify variation in quantity and quality respectively. The variants of FLEECE represent a difference of vowel length consistent with the durational conditioning that results from the operation of the SVLR (Agutter, 1988; Scobbie et al., 1999a,b; Pukli, 2004). The SVLR results in vowels that are phonetically long before voiced fricatives, before /r/ and before a boundary (including a morpheme boundary). Elsewhere, they are short. The SVLR operates alongside the 'voicing effect' (Chen, 1970; Lehiste, 1996) that is thought to condition vowel duration in all varieties of English, including Scottish ones. The voicing effect predicts that vowels preceding voiceless consonants will be phonetically shorter than vowels preceding voiced consonants. The SVLR, by contrast, takes account not just of the voicing of a following consonant, but also of its manner of articulation and the morphological structure of words. The vowel in the stimulus word *need* used in the SCAT is followed by a voiced stop consonant, predicting a phonetically short vowel in the SVLR-conditioned realization. Although evidence of complex context-conditioning of vowel length akin to the SVLR has been reported for locations south of the border (see, for example, Agutter, 1988; Glauser, 1988; Milroy, 1995; Krause, 1997; Watt and Ingham, 2000; Llamas et al., 2011), for our purposes we are testing perception of an association of the short FLEECE variant with Scotland rather than England. For clarity, we will refer to the variable henceforth as FLEECE, although it is in fact the sensitivity of listeners to SVLR-conditioned vowel duration alternations we are attempting to test here. It would have been possible, for instance, to have instead used GOOSE for these purposes, GOOSE being the other monophthong that exhibits SVLR conditioning the most markedly and consistently in Scottish English.

The variants of the GOOSE vowel in the present case were chosen to illustrate a difference of quality rather than of length, however. In Scottish varieties of English, GOOSE is realized as a close, central vowel transcribed as [ʊ] (Stuart-Smith, 2008), though it can also in fact be fully fronted. The North of England (particularly the North East), on the other hand, is one of an apparently dwindling number of places in the English-speaking world where close, back and fully rounded realizations of the vowel – i.e., [u] – can still be heard (Beal et al., 2012). The GOOSE item chosen for inclusion in the SCAT is *spook*, a word in which the vowel is predicted to be short in both English and Scottish varieties, as vowels preceding /k/ are exempt from SVLR-conditioned lengthening. Measurements of the vowel durations of the stimuli bear this prediction out, and the overall word durations also match closely.

We made every effort to ensure that the other characteristics of the stimulus words were as neutral and as closely comparable to one another as possible. That is, we checked carefully that there were no clear differences between the Voice Onset Time duration or degree of aspiration of /k/ in *car* (see e.g., Docherty et al., 2011), and that the vowel qualities in the two exemplars of this word matched closely. We detected no differences in the rhymes of the exemplars of *red* that might reinforce or confound listeners'

judgements of the stimuli based on the quality of the initial rhotic; neither did the consonants in our *need* and *spook* stimuli exhibit any dissimilarities that would concern us. The non-target parts of the test words are not absolutely identical, of course, but this is an unavoidable aspect of using natural stimuli rather than synthetic or spliced ones.

Based on previous literature, then, along with findings from the production strand of the research, the expected associations are those shown in **Table 1**.

Productions of forms predicted to be associated with Scotland were selected from recordings of informants from the Scottish localities Gretna and Eyemouth, while those predicted to be associated with England were taken from English informants from Carlisle and Berwick. All speakers were male, and were matched as closely as possible to one another for age and voice quality. Isolated tokens were drawn from word list readings, so as to ensure that all audio samples were clear and unambiguous. Listeners heard two forms of each target word (one rhotic and one non-rhotic token of *car*, one token of *red* beginning with the alveolar tap and a second beginning with an approximant, and so on), and were required to indicate using a computer keyboard the associations the forms evoked by pressing a key corresponding to the listener's choice. The options the participants were presented with were the binary oppositions *Scotland/Not Scotland* or *England/Not England*.

As with the IAT design, the SCAT consisted of several blocks, and began with a practice block in which participants familiarized themselves with the layout of the computer screen and keyboard. The screen showed 'PRESS 'd' FOR SCOTLAND' in the top left-hand corner, and 'PRESS 'k' FOR NOT SCOTLAND' in the top right-hand corner. Either 'SCOTLAND' or 'NOT SCOTLAND' would then appear in the middle of the screen, and the participant had to press the relevant key as quickly as possible. The next block followed the same format, but this time the audio samples were introduced, and were accompanied by a visual representation of the stimulus word (a block of the color red and stylised pictures of a car, a ghost and a begging bowl for the words *red*, *car*, *spook* and *need* respectively). Participants, who listened to the samples through high-quality closed-cup headphones, were instructed to press either the key indicating 'SCOTLAND' or the one indicating 'NOT SCOTLAND' as quickly as possible after having heard a sample. This was also a practice block. The block that followed it was ostensibly the same as the practice block, but was the block from which the results were taken. For the fourth block the setup was again the same, but in this case 'SCOTLAND' and 'NOT SCOTLAND' were replaced by 'ENGLAND' and 'NOT ENGLAND.' As before, there was a practice block followed by the test block (block 6) from which results were taken. Half of

the participants began with the *Scotland/Not Scotland* opposition and half began the SCAT with *England/Not England*, so as to compensate for any fatigue effects. Each sound file representing each variant was heard three times in random order in each block, making 24 stimuli in total per block (i.e., 3 repetitions \times 4 words \times 2 forms of each word). The keypress prompted the next screen and audio stimulus. In total, the six blocks of the test took between 5 and 10 mins for each participant to complete. As noted above, all participants had taken part in the earlier part of the study during which the production and attitudinal data were collected.

Because the target forms appeared in different positions in the stimulus words, it was necessary to give listeners sufficient time to hear the form but also to respond to it as quickly as possible after exposure. We therefore adjusted the zero point from which the response time was measured depending on where in the stimulus word the target form appeared. Where the target form was word-initial, we allowed one third of the duration of the stimulus word to elapse before the zero point was reached. For word-medial forms, the zero point was placed two thirds of the way through the word, and for word-final forms, the zero point was placed at the end of the word (see **Figure 3**). In analyzing our results, as is common practice, we applied a lower cutoff at 200 ms to eliminate any values that were likely to be spurious. An upper cutoff at 3517 ms (=2.5 SD above the mean) was also identified. This resulted in a loss of only 2.6% of the data, a value falling well below the threshold recommended by Ratcliff (1993, p. 517), who advises that it is reasonable to apply a cutoff that eliminates not more than 15% of the total data.

The non-linguistic variables that were used to model the results were the listener's age, gender, nationality, and the geographical location of his/her speech community of origin (East or West). The results were subjected to linear mixed effects modeling and logistic mixed effects regression in R, as appropriate.

RESULTS

The results of the SCAT are considered firstly in terms of the performance of each phonetic variant. We then turn to examine the influence of the listener's social characteristics on how the social meaning carried by the form was perceived.

Variation by Phonetic Form Categorization

The degree to which the individual phonetic variants were more or less likely to be categorized according to the expected patterns was examined, and an overall model was initially run to test whether the phonetic variants had an effect on the predicted categorization. Using log likelihood comparisons, we compared a fully fit model with nationality, East/West, age, gender and phonetic variant as fixed effects to one without phonetic variant. The inclusion of individual phonetic variants significantly improves the power of the model ($p < 0.001$). A logistic mixed effects regression with phonetic variant as a fixed effect and individual participant and stimulus (word) as random

TABLE 1 | Expected associations based on previous literature and AISEB production data.

Scotland	England
Realized /r/ in coda	No realized /r/ in coda
Tapped /r/ in onset	Approximant /r/ in onset
Short FLEECE	Long FLEECE
Front GOOSE	Back GOOSE

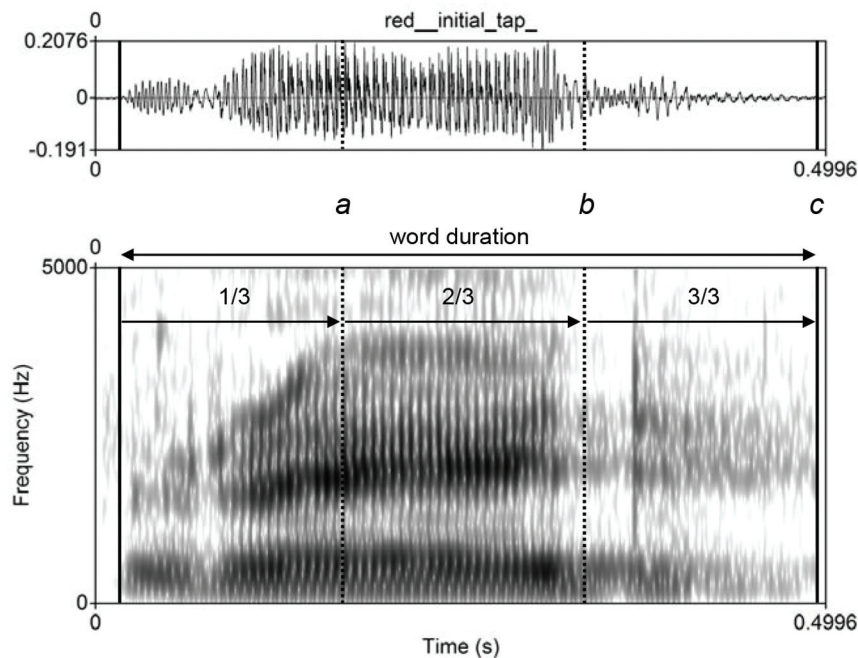


FIGURE 3 | Waveform and spectrogram of *red* produced with an initial alveolar tap [ɾ] by a speaker of Scottish English. The superimposed solid vertical lines represent the start and end points of the word, while the dashed vertical lines dividing the word into three equal portions represent the alternative placements of the zero time point relative to which listeners' response times were measured. Where the target form was word-initial (as in *red*), the response time was logged relative to point *a*, i.e., after one third of the word had elapsed; for a word-medial form (as in *need* and *spook*), the zero point was at *b*; and for word-final target sounds (as in *car*), the listener heard the entire word before he/she had the opportunity to respond (time point *c*).

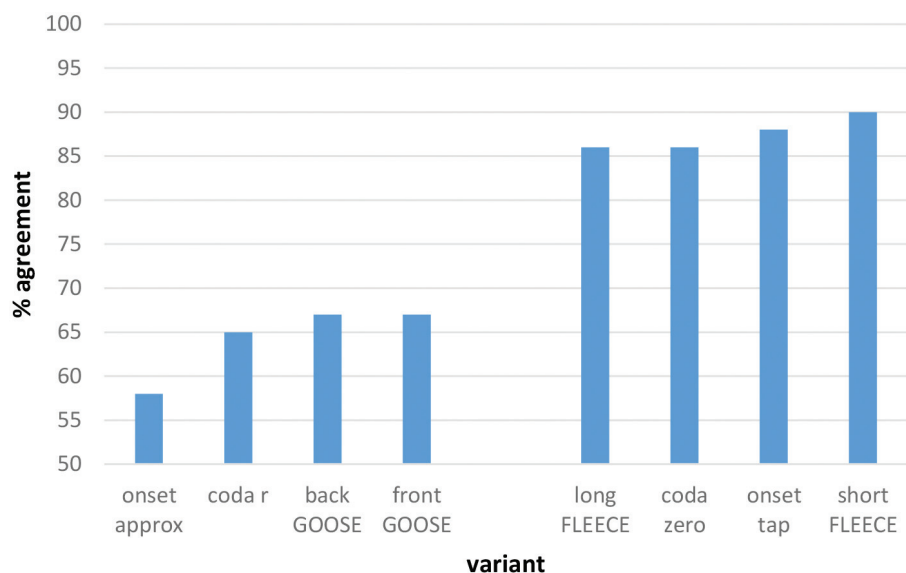


FIGURE 4 | Predicted levels of community agreement (%) on 'correct' association made between social category and phonetic form based on overall SCAT results. Note that 50% represents chance level.

effects was run. Onset approximant in the stimulus *red* (i.e., the pattern of responses for [ɹɛd]) was set as the baseline. **Figure 4** reveals the plot of the model's predicted associations based on the raw data.

Figure 4 reveals a cluster of forms that elicit high levels of community agreement with respect to associated social meaning, and a second group for which levels of agreement are substantially lower. These two clusters differ substantially

for predicted levels of community agreement ($p < 0.001$). The onset approximant [ɹ] is the phonetic form with the lowest probability of being categorized according to the expected pattern (i.e., association with England). Indeed, categorization of [ɹ] is at around chance level, showing no association with one category more than the other. Describing the onset approximant as the 'English /r/', as per Glauser (2000), would therefore be misleading, and reflective of a view that is apparently no longer held by people from the border area, assuming that it ever was. Surprisingly, the r-ful coda realization (in [kɹɪ]) also falls within the low-agreement cluster. As noted in The AISEB Study, rhoticity was anticipated to be the feature with the highest degree of saliency among the forms considered, yet it appears not to be marking an agreed social meaning of *Scotland* in the present case. Conversely, the non-rhotic form [kɹ:] falls within the high-agreement cluster in **Figure 4**, suggesting that this phonological environment is salient after all, even if the expected association of an approximant realization of /r/ in coda position and *Scotland* is not in fact agreed upon. It is possible that for these participants the use of the coda approximant [ɹ] is also associated with other varieties of English, such as American English, a variety to which participants are regularly exposed through the media. This lack of exclusivity might serve to dilute the association between coda [ɹ] and *Scotland*.

Figure 4 also makes it clear that variants of the FLEECE vowel are highly salient, to judge from the level of community consensus about its social category associations. The phonetic variant that has the highest probability of being categorized according to the expected pattern is the short variant of the FLEECE vowel. The long variant is likewise agreed upon in the anticipated manner. The fourth form in the high agreement cluster is the onset tap (in [rɛd]). This appears to be highly salient in terms of agreement on its social meaning. Compared with the onset approximant, every other variant yielded statistically significant levels of community agreement on social category association according to the predicted pattern.

Response Time

The second measure used to estimate the saliency of our target forms was response time. We expected that more salient forms would elicit faster responses than less salient ones. An overall linear mixed effects model was run with all phonetic variants included. As we did for testing categorization, log likelihood comparisons were run on a fully fit model with phonetic variant, age, gender, nationality, and East/West, and the same model without phonetic variant. The retention of phonetic variant significantly improves the power of the model ($p < 0.001$). Individual variant was entered as a fixed effect, and participant and individual stimulus (word) as random effects. Again, the onset approximant was set as a baseline.

Figure 5 is a plot of the predicted response times. It can be seen that the onset tap [ɹ] and onset approximant [ɹ] are reacted to faster than all other variants. The tapped form elicits an especially fast response time. **Figure 5** shows a marked difference between the response times for the variants of onset /r/ and the other variables, which cluster together in the ~1000–1200 ms range (the difference in RTs between these clusters was substantial; $p < 0.001$). The difference between the results for onset /r/ and other variables suggests that participants possess a higher degree of certainty about the associations they make with onset /r/ than those they make with coda /r/ and the vowel variables. The slowest response time is found for coda /r/, suggesting a degree of hesitancy about the associations made with this form (cf. the discussion in Categorization, above).

Hierarchy of Saliency

Ranking of each variant's performance in the SCATs, as measured by community consensus and speed of association, reveals a hierarchy of saliency. Taking both measures into account, we can say that the form with the highest level of saliency among those examined is the tapped form of /r/ in onset position (as in *red* [rɛd]), given the very high level of community agreement on its association with *Scotland* and *Not England* combined with

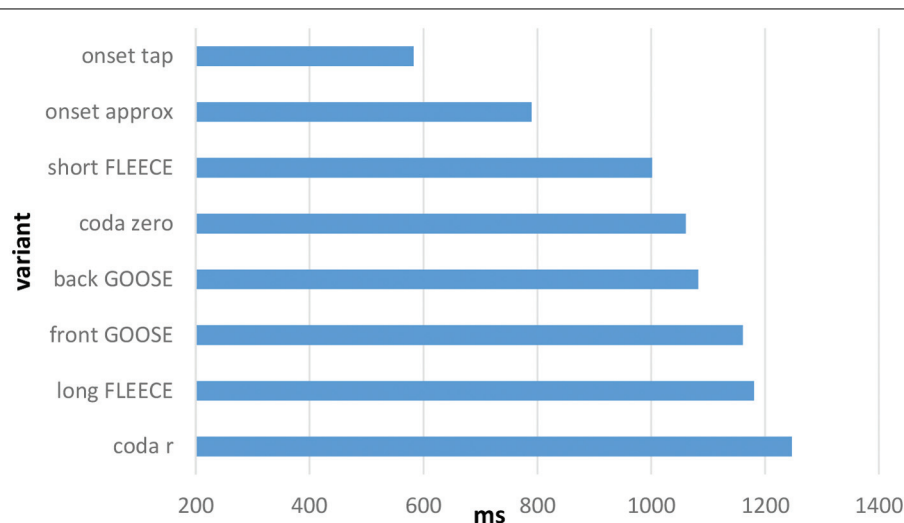


FIGURE 5 | Predicted response time (ms) to phonetic forms based on overall SCAT results.

the speed with which the association was made. Variants of the FLEECE vowel are also imbued with a very high degree of saliency as markers of Scottish versus English identities. The short variant of FLEECE and the *Scotland* association, contrary to our expectations, is the combination which predicts the highest levels of community agreement on social meaning. Non-rhoticity is also highly salient in terms of its agreed social meaning as a marker of *England* and/or *Not Scotland*.

Unexpectedly, the model does not predict high levels of agreement on the association of the r-ful realization of coda /r/ with any social category. Of the set of features examined, this is the feature that had been predicted to be the most salient. However, not only is there a lack of community agreement on its association, it also elicits the longest response time of all variants, suggesting even more strongly a degree of uncertainty around what social category it connotes. The other surprising result was the lack of association of the onset approximant with *England*, in spite of its treatment in the literature as the 'English /r/' (Glauser, 2000).

Variation by Listener Characteristic

So far, we have considered the results of this experiment as though the participants were interchangeable members of a single monolithic community. We turn now to see whether age, gender, nationality (Scottish and English) or cross-border community pairing (East versus West) predict any differences in the reported degree of saliency of the phonetic forms under investigation.

Nationality

In order to test overall rates of association according to the expected patterns, a logistic mixed effects model was run with nationality as a fixed effect and individual participant and stimulus (word) as random effects. Nationality was found not to

be a significant predictor across all the variables when these were treated *en masse* ($p = 0.107$).

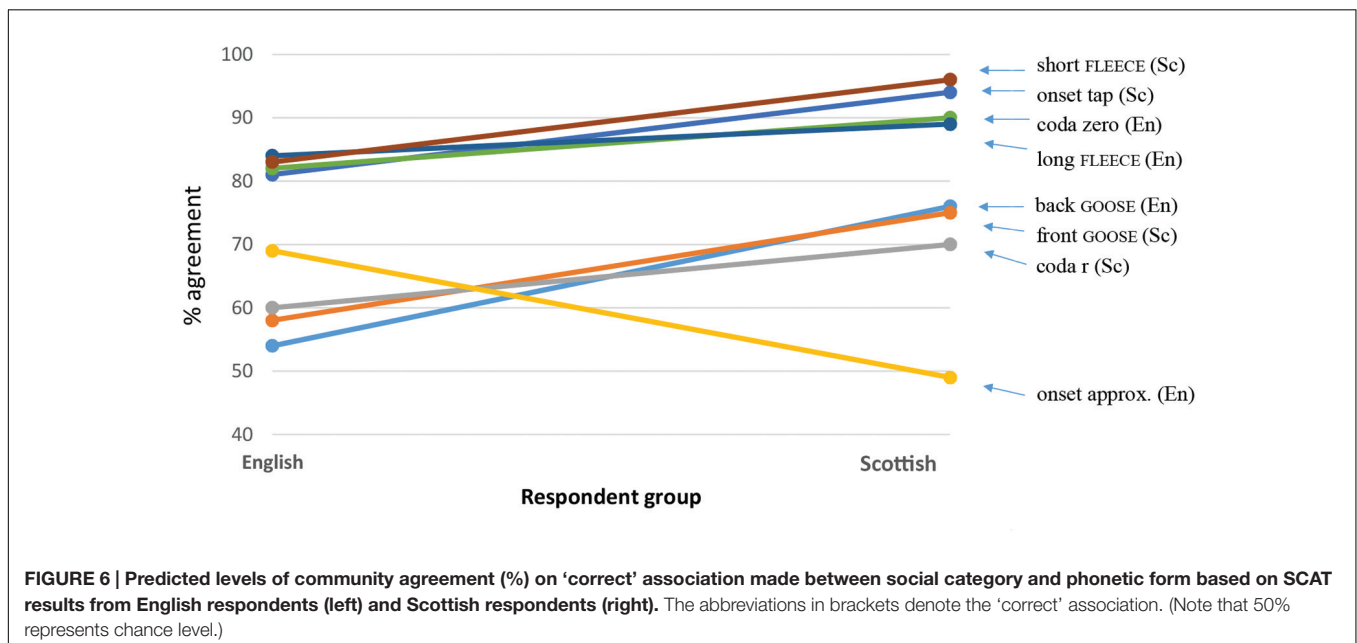
Whether consensus of association across individual variants differed as a function of participant nationality was then tested by fitting a logistic mixed effects regression model with individual variant and nationality as fixed effects, and participant and stimulus as random effects. **Figure 6** is the plot of the model's predictions.

Figure 6 shows that for all but one of the variants under examination, the Scottish listeners are in closer agreement than are the English respondents about the 'correct' (expected) social category association made with the phonetic form. The only variant that fails to follow this pattern is the onset approximant (as in *red* [ɹɛd]), for which Scottish listeners are predicted to perform at around chance levels. English listeners are, by contrast, predicted to exhibit a moderate level of agreement where this variant is concerned.

As with the model based on the overall community results (see Variation by Phonetic Form), we see here a clustering of high-performing variants, and, although Scottish listeners perform more uniformly than English listeners in terms of community consensus, agreement about these forms (viz., variants of FLEECE, onset tap and r-less coda) is still very high, at over 80%, among English listeners.

East/West Cross-Border Community Pairings

Although the individual localities in the two pairs of communities (i.e., Gretna/Carlisle and Eyemouth/Berwick) are separated by the political border such that the nationality of participants from each of the four towns is a relevant factor, we can justifiably also view them as pairings which share the defining characteristic of being located at either the western end or the eastern end of the border. It seems natural to think that because they are both in Scotland the towns of Gretna and



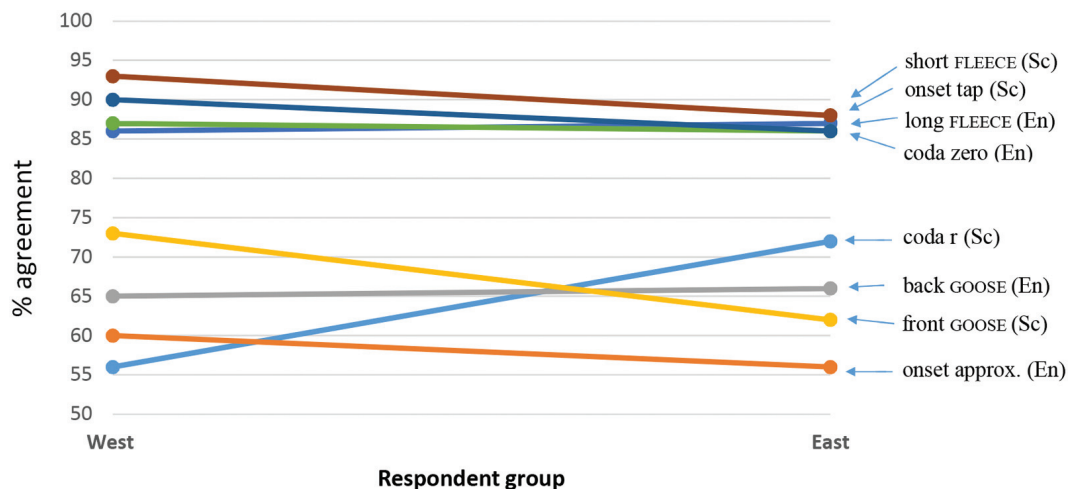


FIGURE 7 | Predicted levels of community agreement (%) on 'correct' association made between social category and phonetic form based on SCAT results from West respondents (left) and East respondents (right). The abbreviations in brackets denote the 'correct' association. (Note that 50% represents chance level.)

Eyemouth are somehow more similar to one another than they are to their respective nearby English partner communities on the other side of the border. But Gretna and Eyemouth, just like Carlisle and Berwick, are separated from one another by a relatively long distance, at least by British standards. Travel between the two same-nation localities along the length of the border is indirect and time-consuming even using private transport, so direct face-to-face contact between members of these communities is not likely to occur very often. By contrast, the conditions are very favorable for high levels of contact between inhabitants of the paired communities at either end of the border, in view of the fact that they live less than 10 miles (16 km) apart and experience no hindrances to their cross-border movement, as we noted in The AISEB Study. For these reasons, we turn now to a consideration of the two paired cross-border communities (Gretna/Carlisle and Eyemouth/Berwick) at each end of the border. **Figure 7** shows predicted differences between participants from the border's eastern and western ends.

Despite the short distance between the two communities in each pair, and the separation of the same-nation localities lying at the border's extreme ends, more difference is discernible between the respondents when they are classed by nationality than when they are grouped into cross-border communities. In terms of how they perform in the present experiment, the Gretna respondents have more in common with their fellow Scots in Eyemouth than they do with their English near-neighbors in Carlisle, for example. We do nevertheless see fairly close cross-border correspondences, particularly with respect to the high-performing cluster of phonetic variants. Where slight differences are in evidence, the tendency is for respondents from the western end of the border to categorize the target phonetic forms 'correctly' according to social meaning more often than is the case for those from the eastern end. There is one notable exception to this trend, however. Participants

from the east are more likely to make the 'correct' association of overtly realized coda /r/ with *Scotland* than are their western counterparts. A logistic mixed effects regression model was fit with East/West and individual variant (fixed effect) tested as an interaction, and individual participant and stimulus as random effects. The presence of coda /r/ and a 'correct' association with *Scotland* was the only variant of the set to be affected by location ($p = 0.034$); there was no overall effect of East/West ($p = 0.650$).

This finding ties in closely with the production differences noted in The AISEB Study. With regard to the production of r-ful realizations, frequency of usage among Scottish speakers at the western end of the border (Gretna) is much lower than that recorded for the eastern Scottish (Eyemouth) speakers. In the AISEB sample, regular rhoticity is really only found in Eyemouth (see e.g., Watt et al., 2014b for further details of the production data).

Gender and Age

In order to test overall rates of association according to the expected patterns, a logistic mixed effects model was run with age as a fixed effect and individual participant and stimulus (word) as random effects. Age was found not to be a significant predictor across any of the variables ($p = 0.857$). Additionally, the effect of participant gender was tested for and was found not to be significant either as a main effect ($p = 0.35$) or as an interaction.

In order to test for whether participant age significantly affected response times, a linear mixed effects model with participant age as a fixed effect and individual participant and stimulus as random effects was run. The difference was not significant ($p = 0.692$). The predicted response time for younger participants was found to be 1003 ms, while for older participants it was 976 ms.

Although we found no significant effects for age overall, further inspection of the raw data revealed marked age differences

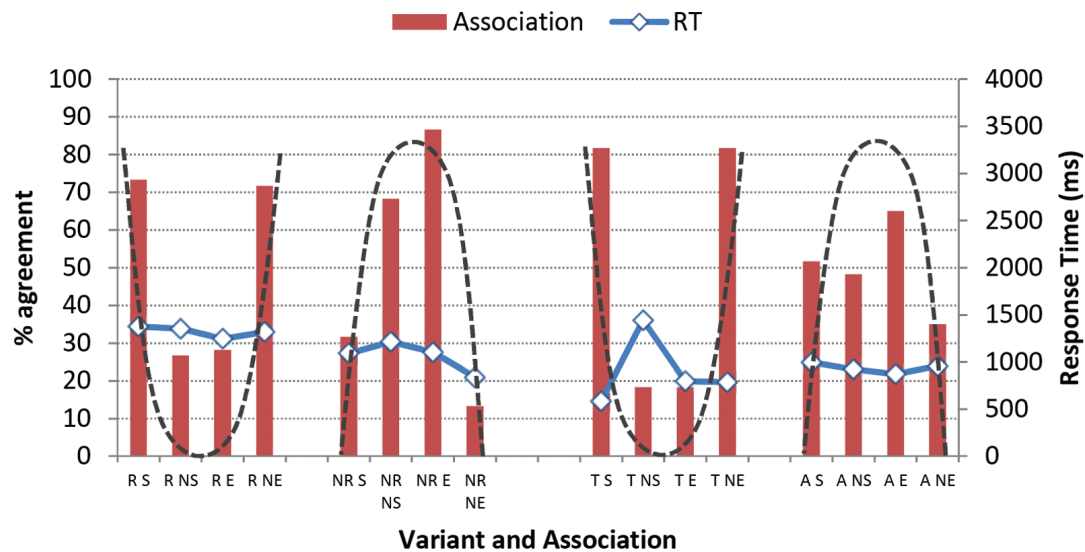


FIGURE 8 | Associations (bars) and Response Times (RT; solid lines) of older judges for variants of /r/. (Variants are indicated as R = rhotic, NR = non-rhotic, T = tap, A = approximant; Social categories are indicated as S = Scotland, NS = Not Scotland, E = England, NE = Not England; Dashed line indicates shape of 'correct' pattern.) (Note that 50% represents chance level.)

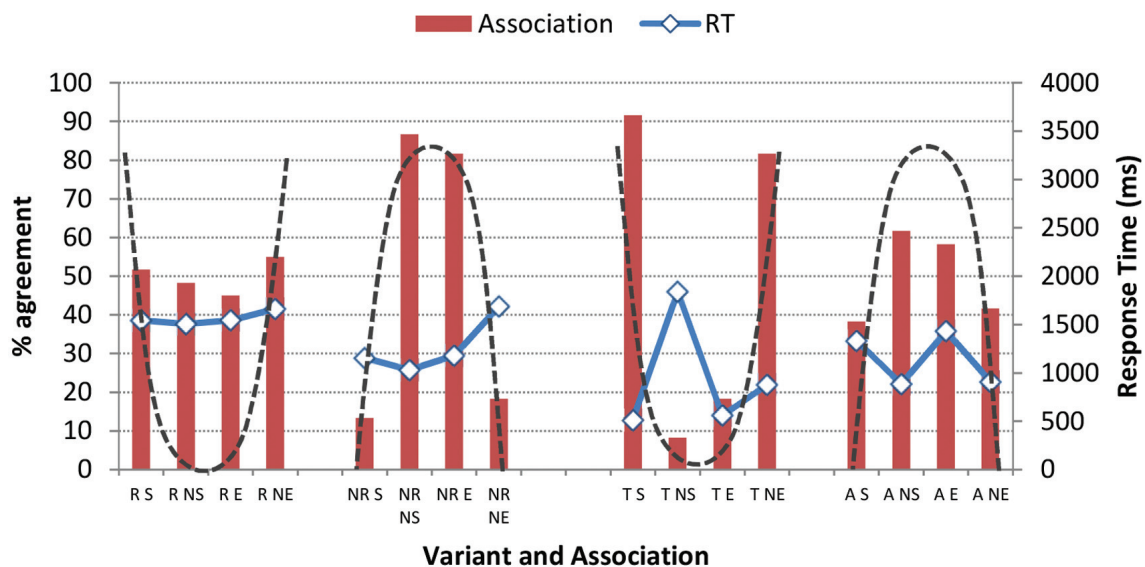


FIGURE 9 | Associations (bars) and Response Times (RT; solid lines) of younger judges for variants of /r/. (Variants are indicated as R = rhotic, NR = non-rhotic, T = tap, A = approximant; Social categories are indicated as S = Scotland, NS = Not Scotland, E = England, NE = Not England; Dashed line indicates shape of 'correct' pattern.) (Note that 50% represents chance level.)

in the degree of group consensus about association. **Figures 8 and 9** present the raw data for the associations listeners made with the two /r/ variables. The dashed line superimposed on each figure approximates the shape of the pattern predicted if a high proportion of 'correct' associations was made by listeners.

While the listeners' agreement on the association made between tapped /r/ in onset position and *Scotland* remains stable across listener age, it is also apparent that the association with *Scotland* we expected to see when the listeners heard the

rhotic form [kɑ.ɪ] drops to chance levels, indicating that the association between r-ful realizations and the social category *Scotland* has become recessive. To test this hypothesis, a general linear model was run on the rhotic variant only, with participant age as a fixed effect. Younger participants were predicted to be significantly less likely to make the association between the rhotic form and *Scotland* ($p = 0.001$). They were, moreover, also predicted to be less likely to make the association between the non-rhotic form and *England* ($p = 0.012$). We know from

the AISEB production data and the findings of other studies (see Materials and Methods) that derhoticization is underway in varieties of Scottish English, including the influential urban varieties of central Scotland, in that younger speakers produce fewer r-ful realizations than do their older counterparts. Here we see a loosening of the association of the r-ful pronunciation with Scotland, and a consequent diminution of the salience of the form.

Examination of the effects of listener characteristics has revealed that, in general, the patterns we observe hold across all listener groupings. However, the Scottish listeners in our sample are more likely to exhibit the anticipated associations between the high-consensus forms and the social categories *Scotland/not Scotland/England/not England* than are the English listeners. The other notable finding in the results broken down by listener characteristics is that perception mirrors production patterns, in the sense that the association between coda /r/ and *Scotland* – which we had hypothesized to be the strongest of any of the associations we set out to test – is weakening, just as overtly realized /r/ in syllable codas is becoming less frequent in Scottish English.

DISCUSSION

The approach taken in the present study rests, firstly, on the use of community consensus concerning the social categories that listeners associate with phonetic forms as a measure of the salience of those forms. Secondly, the speed with which subjects respond when making the association between a form and a social category is treated as an indicator of the association's strength, and therefore of the degree of salience of the form in question. The results from the SCATs reveal that salience is a gradient property, such that salience-bearing forms can be ranked with respect to their relative salience. Certain forms, notably the short FLEECE variant and the realization of /r/ as the tap [ɾ] in onset position, are almost categorically associated with the social category *Scotland*. In the case of [ɾ], the association is made extremely quickly by listeners. Other forms, such as the onset approximant [ɹ], appear to possess negligible levels of salience.

The results presented above also demonstrate that salience is not a static property of phonetic forms or an inherent attribute of units of this kind. As we have seen, the degree of salience of a form, as estimated using measures of shared social meaning, can differ between speech communities separated by very small geographical distances, and also appears to change over time. Among other things, our findings strongly imply that a loosening of the association between r-ful pronunciations and the social category *Scotland* is underway in the region. Furthermore, the association of the r-less pronunciation with the category *England* robustly persists, demonstrating that the lack of a form can carry at least as much salience as its presence in equivalent contexts.

The SCAT results also demonstrate clear connections between linguistic production and perception. As noted above, our findings show a relaxation of the association between rhoticity

and the social category *Scotland*, accompanied by a degree of hesitancy in making this association, as revealed through participants' longer response times. These results coincide with changes in production patterns found in the larger AISEB study, whereby rhoticity appears to be decreasing rapidly in one of the Scottish localities (Gretna). In the broader context, we see that the process of derhoticization is well underway in the varieties spoken in Scotland's dominant urban centers, Edinburgh and Glasgow (Stuart-Smith et al., 2014). This change appears to be most strongly linked to younger, working-class speakers. Mirroring these production patterns, we see in the results of the present experiment that, in terms of perception, the younger participants respond only at chance levels to the rhotic stimulus (*car* [kɑ.ɹ]), demonstrating no agreement on its association with Scotland. We see further evidence of the interconnectedness of production and perception when we compare the results from the western end of the border to those from the eastern end. Western respondents are less inclined to make the 'correct' association than are their eastern counterparts. In the AISEB production data, levels of rhoticity are much lower for the western group than for the eastern group, and are decreasing over apparent time (see The AISEB Study), providing compelling evidence that the process of derhoticization is in progress in the west. Another example of the parallels between perceptual associations and production patterns is apparent in the lack of a strongly-held association between the approximant in onset position (in *red* [ɹɛd]) and the social category *England*. The high and increasing use of the approximant realization of /r/ in the Scottish localities is documented in the production strand of the study (e.g., Watt et al., 2014b). The reduction in the mutual distinctiveness of Scottish and English varieties brought about by this change in the distribution of [ɹ] is a probable contributor to the loss of the association of the approximant with England.

The findings of the present study, then, reveal a number of close links between production patterns and the perception of social meaning attached to a form. The salience of a phonetic form can increase or decrease depending on the usage patterns of the form. Thus, we would not argue that forms acquire salience, and remain salient thereafter, solely by virtue of their intrinsic phonetic properties. Rather, the strength of their socio-indexical value as seen through the lens of shared social meaning dictates how salient the forms will be. As production patterns change, so may the agreed social meaning of the form. Whether this is a direct causal relationship, or a bidirectional one whereby the one phenomenon acts as a trigger for the other, are matters for further investigation.

Finally, our findings lead us to sound a note of caution with regard to the prior assumptions that researchers bring to investigations of the sort represented in the current paper. Deciding in advance on which features are likely to have the strongest sociolinguistic salience in a given speech community may in general be inadvisable. Claims regarding the importance of /r/ to taxonomies of the subvarieties of English are abundant in the literature (e.g., Maguire et al., 2010, p. 97), and if we couple these claims with the frequency with which /r/ is mentioned as a stereotype of Scottish English by informants in the larger

AISEB study, we could easily be led to form the expectation that the association between the r-ful pronunciation and the social category *Scotland* would be the most strongly-held association of those we tested. This prediction is, however, not borne out: in the statistical models, rhoticity was shown to yield *low* community agreement on its social meaning. As we noted earlier, the use of the approximant in onset position has been referred to as the ‘English /r/’ in this regional context (Glauser, 2000, p. 75), but a strong association of this nature is not observed in the results of the present study. We do, however, observe a robust connection between the non-rhotic form and *England*, so to this extent we do have evidence for the saliency of (non-)rhoticity as a marker. Additionally, the use of tapped /r/ in onset position is extremely salient, according to the measures applied here. The phonetic feature which yielded the highest level of community consensus was, however, found to be the SVLR-conditioned alternation in the length of the FLEECE vowel, which is not a feature mentioned in any of the overt comments made by the AISEB informants.

The complex findings presented here clearly demonstrate the utility of the technique we used to collect them. The test we present here opens up new ways of investigating sociolinguistic saliency. By using levels of community consensus about the association of phonetic forms and social categories as measures of the saliency of the forms, we can posit a hierarchy of saliency among key phonetic forms, and at the same time observe how features arrayed on this hierarchy may be re-ranked by members of the speech community in parallel with changes in production patterns.

CONCLUSION

We have argued in this paper that, from a sociolinguistic perspective, the choice of features which become salient is in large part an arbitrary one. Saliency depends on listeners initially noticing a feature and then collectively assigning social meaning to it. Under this definition, investigations of saliency are examinations of perceptual aspects of the linguistic forms of which members of a given community or group have conscious or subconscious awareness (i.e., as stereotypes or markers, and indicators, respectively). Cognizance of a linguistic form may come about because the form is unusual in some way, and perhaps (but not necessarily) infrequent. It may also be occasioned because the form is an important marker of relevant ingroup or outgroup status within a speech community. It will only become an important marker of social category membership, however, if there is sufficient agreement among members of the speech community with respect to its function

as a signal of group-membership meanings of this kind. Information about the association between phonetic forms and social categories among speech community members is usually not accessible via overt discussion. A way of operationalizing the saliency of phonetic forms such that it can be empirically investigated, therefore, is by examining the extent to which the social meaning carried by the form, in terms of its social group associations, is shared by members of a speech community. This paper set out to test a method of estimating the relative saliency of segmental variables, and has shown that not only is it possible to do so, it is also feasible using these techniques to examine the mutual dependencies between production, perception, and changes in saliency over time.

Focusing on multiple localities in a border zone, a region in which social category divisions may be sharper and more prominent than in other places, enables us to see how phonetic forms are used to categorize speakers by social group, and permits us to identify those features which have sociolinguistic saliency as group markers. Many linguistic forms are mobilized in the marking of social categories by speakers and listeners. Some forms, however, do more work in this regard than others. Comparison of levels of consensus about social category associations within and between communities, and of the speed with which these associations are made in the minds of listeners, gives us a means of estimating how salient a marker is relative to other markers. Estimating and tracking these changing levels of saliency can then yield further insights into how and why language changes.

AUTHOR CONTRIBUTIONS

CL and DW designed the experiments, analyzed the data and wrote the manuscript. AM did the statistical modeling of the data.

FUNDING

The AISEB project was funded by the Economic and Social Research Council (ESRC), UK (RES-062-23-0525).

ACKNOWLEDGMENTS

We are very grateful to Robin Lindop Fisher, Damien Hall, Jen Nycz, Gerry Docherty, Daniel Redinger, and Daniel Ezra Johnson for their contributions to the research upon which this article is based.

REFERENCES

- Agutter, A. (1988). “The dangers of dialect parochialism: the Scottish vowel length rule,” in *Historical Dialectology: Regional and Social*, ed. J. Fisiak (The Hague: de Gruyter), 1–21.
- Aitken, A. J. (1984). “Scottish accents and dialects,” in *Language in the British Isles*, ed. P. Trudgill (Cambridge: Cambridge University Press), 94–114.
- Aitken, A. J. (1992). “Scots,” in *The Oxford Companion to the English Language*, ed. T. McArthur (Oxford: Oxford University Press), 893–899.
- Baranowski, M., and Turton, D. (2015). “Manchester English,” in *Researching Northern Englishes*, ed. R. Hickey (Amsterdam: Benjamins), 293–316.
- Bardovi-Harlig, K. (1987). Markedness and saliency in second language acquisition. *Lang. Learn.* 37, 385–407. doi: 10.1111/j.1467-1770.1987.tb00577.x
- Beal, J., Burbano-Elizondo, L., and Llamas, C. (2012). *Urban North-Eastern English: Tyneside to Teesside*. Edinburgh: Edinburgh University Press.
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129–159. doi: 10.1159/000259312

- Clopper, C. G., and Pisoni, D. B. (2004). Homebodies and army brats: some effects of early linguistic experience and residential history on dialect categorization. *Lang. Var. Change* 16, 31–48. doi: 10.1017/S0954394504161036
- Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behav. Res. Methods Instrum. Comput.* 25, 257–271. doi: 10.3758/BF03204507
- Docherty, G. J., Watt, D., Llamas, C., Hall, D., and Nycz, J. (2011). “Variation in voice onset time along the Scottish-English border,” in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 591–594.
- Glauser, B. (1974). *The Scottish-English Linguistic Border: Lexical Aspects*. Bern: Francke Verlag.
- Glauser, B. (1988). “Aitken’s context in Northumberland, Cumberland and Durham: a computer-assisted analysis of material from the Survey of English Dialects (SED),” in *Methods in Dialectology: Proceedings of the Sixth International Conference held at the University College of North Wales, August 1987*, ed. A. R. Thomas (Clevedon: Multilingual Matters), 611–624.
- Glauser, B. (2000). The Scottish/English border in hindsight. *Int. J. Sociol. Lang.* 145, 65–78. doi: 10.1515/ijsl.2000.145.65
- Goldstein, L. (1977). Three Studies in Speech Perception: Features, Relative Salience, and Bias. *UCLA Working Papers in Phonetics*, 39. Available at: <http://escholarship.org/uc/item/08c9j6cm>
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Hollmann, W., and Siewierska, A. (2006). Corpora and (the need for) other methods in a study of Lancashire dialect. *Z. Anglist. Am.* 54, 203–216.
- Honeybone, P., and Watson, K. (2013). Salience and the sociolinguistics of Scouse spelling: exploring the phonology of the contemporary humorous dialect literature of Liverpool. *Engl. World Wide* 34, 305–340. doi: 10.1075/eww.34.3.03hon
- Ihalainen, O. (1994). “The dialects of England since 1776,” in *The Cambridge History of the English Language: English Language in Britain and Overseas*, Vol. 5, ed. R. M. Burchfield (Cambridge: Cambridge University Press), 197–274.
- Johnston, P. (1980). *A Synchronic and Historical Study of Border Area Bimoric Vowel Systems*. PhD thesis, University of Edinburgh, Edinburgh.
- Johnston, P. (1997). “Regional variation,” in *The Edinburgh History of the Scots Language*, ed. C. Jones (Edinburgh: Edinburgh University Press), 433–513.
- Johnston, P. (2007). “Scottish English and Scots,” in *Language in the British Isles*, ed. D. Britain (Cambridge: Cambridge University Press), 105–121.
- Johnstone, B., and Kiesling, S. F. (2008). Indexicality and experience: exploring the meanings of /aw/-monophthongization in Pittsburgh. *J. Sociolinguist.* 12, 5–33. doi: 10.1111/j.1467-9841.2008.00351.x
- Kerswill, P., and Williams, A. (2002). “‘Salience’ as an explanatory factor in language change: evidence from dialect levelling in urban England,” in *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, eds M. C. Jones and E. Esch (Berlin: Mouton de Gruyter), 81–110.
- Kiely, R., Bechhofer, F., Stewart, R., and McCrone, D. (2001). The markers and rules of Scottish national identity. *Sociol. Rev.* 49, 33–55. doi: 10.1111/1467-954X.00243
- Krause, C. B. (1997). Vowel duration in Lowick/Northumberland: a sociophonetic approach to dialectology. *Paper Presented at the 1st UK Language Variation Workshop*, Reading.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2014). “A socio-articulatory study of Scottish rhoticity,” in *Sociolinguistics in Scotland*, ed. R. Lawson (London: Palgrave Macmillan), 53–78.
- Lehiste, I. (1996). “Suprasegmental features of speech,” in *Principles of Experimental Phonetics*, ed. N. J. Lass (St. Louis, MO: Mosby), 226–244.
- Levon, E., and Fox, S. (2014). Social salience and the sociolinguistic monitor: a case study of ING and TH-fronting. *J. Engl. Linguist.* 42, 185–217. doi: 10.1177/0075424214531487
- Llamas, C. (2001). “The sociolinguistic profiling of (r) in Middlesbrough English,” in *r-at-ics: Sociolinguistic, Phonetic and Phonological Characteristics of /r/*. *Etudes & Travaux* 4, eds H. Van de Velde and R. van Hout (Brussels: ILVP), 123–140.
- Llamas, C. (2010). “Convergence and divergence across a national border,” in *Language and Identities*, eds C. Llamas and D. Watt (Edinburgh: Edinburgh University Press), 227–236.
- Llamas, C., and Watt, D. (2014). Scottish, English, British? Innovations in attitude measurement. *Lang. Linguist. Compass* 8, 610–617. doi: 10.1111/lnc3.12109
- Llamas, C., Watt, D., French, P., and Roberts, L. (2011). “Effects of the Scottish Vowel Length Rule on vowel quantity in Tyneside English,” in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 1282–1285.
- Llamas, C., Watt, D., and Johnson, D. E. (2009). Linguistic accommodation and the salience of national identity markers in a border town. *J. Lang. Soc. Psychol.* 28, 381–407. doi: 10.1177/0261927X09341962
- Macafee, C. (1983). *Varieties of English around the World: Glasgow*. Amsterdam: John Benjamins.
- Maguire, W. (2015). “The north above the North: Scotland and Northern English,” in *Researching Northern English*, ed. R. Hickey (Amsterdam: John Benjamins), 437–458.
- Maguire, W., McMahon, A., Heggarty, P., and Dediu, D. (2010). The past, present and future of English dialects: quantifying convergence, divergence and dynamic equilibrium. *Lang. Var. Change* 22, 1–36. doi: 10.1017/S0954394510000013
- Mather, J., and Speitel, H. (1986). *The Linguistic Atlas of Scotland, Scots Section: Phonology*, Vol. 3. Beckenham: Croom Helm.
- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Lang. Speech* 58, 502–521. doi: 10.1177/0023830914565191
- McMahon, A., and Maguire, W. (2011). “Quantitative historical dialectology,” in *Analysing Older English*, eds D. Denison and R. Bermúdez-Otero (Cambridge: Cambridge University Press), 140–158.
- McMahon, A., and Maguire, W. (2013). “Quantifying phonetic similarity between accents of English: the case of rhoticity,” in *Research Methods in Language Variation and Change*, eds M. Krug and J. Schlüter (Cambridge: Cambridge University Press), 421–423.
- Miller, G. A., and Nicely, P. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338–352. doi: 10.1121/1.1907983
- Milroy, J. (1995). Investigating the Scottish vowel length rule in a Northumbrian dialect. *Newcastle Durh. Work. Pap. Linguist.* 4, 187–196.
- Montgomery, C. (2014). “Perceptual ideology across the Scottish/English border,” in *Language, Borders and Identity*, eds D. Watt and C. Llamas (Edinburgh: Edinburgh University Press), 118–136.
- Narayan, C. R. (2008). The acoustic-perceptual salience of nasal place contrasts. *J. Phon.* 36, 191–217. doi: 10.1016/j.wocn.2007.10.001
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Orton, H., and Dieth, E. (eds) (1962–1971). *Survey of English Dialects: The Basic Materials*. Leeds: Edward Arnold.
- Påhlsson, C. (1972). *The Northumbrian Burr: A Sociolinguistic Study*. Lund: Gleerup.
- Podesva, R. (2006). *Phonetic Detail in Sociolinguistic Variation*. Ph.D. dissertation, Stanford University, Stanford, CA.
- Pukli, M. (2004). *Scottish English and the Scottish Vowel Length Rule – An Empirical Study of Ayrshire Speakers*. Available at: <http://w3.pac.univ-tlse2.fr/pdf/Pukli0709.pdf>
- Rącz, P. (2013). *Salience in Sociolinguistics: A Quantitative Approach*. Berlin: de Gruyter.
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychol. Bull.* 114, 510–532. doi: 10.1037/0033-2909.114.3.510
- Reid, E. (1978). “Social and stylistic variation in the speech of children: some evidence from Edinburgh,” in *Sociolinguistic Patterns in British English*, ed. P. Trudgill (London: Edward Arnold), 158–171.
- Romaine, S. (1978). “Post-vocalic /r/ in Scottish English: sound change in progress?,” in *Sociolinguistic Patterns in British English*, ed. P. Trudgill (London: Edward Arnold), 144–158.
- Rydland, K. (1998). *The Orton Corpus: A Dictionary of Northumbrian Pronunciation, 1928–1939*. Oslo: Novus Press.
- Scobbie, J., Hewlett, N., and Turk, A. (1999a). “Standard English in Edinburgh and Glasgow: the Scottish vowel length rule revealed,” in *Urban Voices: Accent*

- Studies in the British Isles*, eds P. Foulkes and G. J. Docherty (London: Arnold), 230–245.
- Scobbie, J., Turk, A., and Hewlett, N. (1999b). “Morphemes, phonetics and lexical items: the case of the Scottish Vowel Length Rule,” in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, CA, 1618–1620.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Lang. Commun.* 23, 193–229. doi: 10.1016/S0271-5309(03)00013-2
- Stuart-Smith, J. (2003). “The phonology of modern urban Scots,” in *The Edinburgh Companion to Scots*, eds J. Corbett, J. D. McClure, and J. Stuart-Smith (Edinburgh: Edinburgh University Press), 110–137.
- Stuart-Smith, J. (2008). “Scottish English: phonology,” in *Varieties of English: The British Isles*, eds B. Kortmann and C. Upton (Berlin: Mouton de Gruyter), 48–70.
- Stuart-Smith, J., Lawson, E., and Scobbie, J. M. (2014). “Derhoticisation in Scottish English: a sociophonetic journey,” in *Advances in Sociophonetics*, eds C. Celata and S. Calamai (Amsterdam: Benjamins), 59–96.
- Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell.
- Watson, K., and Clark, L. (2013). How salient is the NURSE~SQUARE merger? *Engl. Lang. Linguist.* 17, 297–323. doi: 10.1017/S136067431300004X
- Watt, D., and Ingham, C. (2000). Durational evidence of the Scottish Vowel Length Rule in Berwick English. *Leeds Work. Pap. Linguist. Phon.* 8, 205–228.
- Watt, D., Llamas, C., Docherty, G. J., Hall, D., and Nycz, J. (2014a). “Language and identity on the Scottish/English border,” in *Language, Borders and Identity*, eds D. Watt and C. Llamas (Edinburgh: Edinburgh University Press), 8–26.
- Watt, D., Llamas, C., and Johnson, D. E. (2010). Levels of linguistic accommodation across a national border. *J. Engl. Linguist.* 38, 270–289.
- Watt, D., Llamas, C., and Johnson, D. E. (2014b). “Sociolinguistic variation on the Scottish-English border,” in *Sociolinguistics in Scotland*, ed. R. Lawson (Basingstoke: Palgrave Macmillan), 79–102.
- Wells, J. C. (1982). *Accents of English*, Vol. 3. Cambridge: Cambridge University Press.

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.

Copyright © 2016 Llamas, Watt and MacFarlane. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Linking Place and Mind: Localness As a Factor in Socio-Cognitive Salience

Marie M. Jensen *

Department of Culture and Global Studies, Aalborg University, Aalborg, Denmark

This paper investigates the salience of vernacular Tyneside forms on the basis of theories of enregisterment and exemplar processing. On one level, exemplar theory provides a psycholinguistic account of how the link between social value and linguistic features is possible. Conversely, integrating the notion of social value into exemplar theory extends the value of this originally cognitive theory to social domains. It is suggested that the association of social value and particular local, linguistic forms may contribute to the salience of these forms among local speakers. The empirical work reported here takes the form of a questionnaire study, which aims to uncover Tyneside inhabitants' awareness of forms as well as their affiliation with the local community. Results showed differences in frequency perceptions between participants themselves and others which indicate that speakers can identify local forms as such, but that the variety is stigmatized. The strength of local affiliation correlated with participants' own language use and it is suggested that this can be accounted for by employing a social personae explanation, where speakers use certain salient forms to index local belonging despite overt stigma.

OPEN ACCESS

Edited by:

Bernd Kortmann,
University of Freiburg, Germany

Reviewed by:

Joan Christine Beal,
University of Sheffield, UK
Willem B. Hollmann,
Lancaster University, UK

*Correspondence:

Marie M. Jensen
mariemj@cgs.aau.dk

Specialty section:

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

Received: 28 April 2016

Accepted: 18 July 2016

Published: 29 July 2016

Citation:

Jensen MM (2016) Linking Place and
Mind: Localness As a Factor in
Socio-Cognitive Salience.
Front. Psychol. 7:1143.
doi: 10.3389/fpsyg.2016.01143

Keywords: Tyneside English, language variation, social value, socio-cognitive salience, exemplar theory, enregisterment

INTRODUCTION

Within sociolinguistics, studies of the meaning of place (often in local or regional terms) to speakers' language use and identity are many. Place is seen as a natural external variable in sociolinguistic studies, mainly because research in this field has always been engaged in the study of language variation across different localities (early dialectology being a prime example). In addition, sociolinguistic studies have investigated the meaning of places as a factor which shape speakers' linguistic identities, their sense of self and, importantly, their self-representation. Borrowings from linguistic anthropology have further enriched the area of study, most recently with the terms indexicality (Silverstein, 2003) and enregisterment (Agha, 2003) becoming commonplace in sociolinguistic studies (Johnstone et al., 2006; Johnstone and Kiesling, 2008; Beal, 2009; Johnstone, 2009, 2010).

While these two terms, which account for processes at play on the social level, work well in underpinning sociolinguistic patterns of variation and change, especially when these are concerned with matters of identity, as such they do not present ideas which have not already been posited by earlier sociolinguists (such as Labov, 1972). In addition, what they ultimately aim to capture can be summed up by the term *salience*; a term which has many uses and connotations in many fields and which is not, in itself, easily accounted for. Finally, what is perhaps lacking is an account of how these processes can operate from a cognitive perspective. How can we support these ideas of locality having an impact on speakers' language use through arguments about their identity and not as mere reflections of variation due to differences in locality, e.g., Manchester vs. Liverpool?

This article presents a sociolinguistic study of the role of local attachment by Tyneside English speakers in their awareness and perceptions of local forms' frequency of use and local status.

The data was collected via questionnaires which asked participants to rate example sentences with regards to their frequency of use. In addition, participants were also tested on their ability to identify local forms and they were assessed with regards to their local affiliation. Five variables were included in the study: (do+NEG), (our), (told), (throw), and (go). In the interpretation of results, I will suggest that the perception of the forms as unique to Tyneside (and thus encapsulating localness) makes them occupy an especially salient position in speakers' minds (see Honeybone and Watson, 2013 for a similar argument for phonological forms in Liverpool English based on an analysis of contemporary dialect literature). We can find support for this suggestion in exemplar theory, if we view language as a complex adaptive system (CAS), where social and cognitive factors both play equal roles in the shaping of language use, both on the individual and on the community level (Beckner et al., 2009; Bybee, 2010).

First, I set up the theoretical underpinnings for the study of local vernacular forms in Tyneside English presented here and briefly introduce the topic of salience from a sociolinguistic perspective and link it to indexicality and enregisterment. I then place the sociolinguistic approach to salience within an exemplar theoretical framework (and a wider conceptualization of language as a CAS) in order to show how the sociolinguistic approach can be supported from a psycholinguistic point of view. In the third section, I introduce the questionnaire study, which forms the empirical basis for this paper, and briefly account for the five vernacular variables under study. The data is then analyzed quantitatively and, in section four, I discuss the results in relation to salience and suggest the concept of *social personae* as a way to account for the patterning found.

SALIENCE IN SOCIOLINGUISTICS

While the topic of salience is hardly new, finding common ground between the many publications on this topic can be difficult as many approach the topic from vastly different perspectives. Within sociolinguistics, the early work of Labov (1972, 1994) and Trudgill (1986) seems to form the basis on which definitions and later studies of salience have been based. Both Labov and Trudgill take as their focus the speech community as a whole and aimed to describe how forms were salient (or not) both within a community (in-group) as well as to out-group members and how this, then, could be linked with language change. According to Labov and Trudgill, features of which speakers are aware are *salient* variants and these can be classed as either *markers* or *stereotypes*. Variables which are non-salient in the speech community or to the individual speaker are called *indicators*. The difference between indicators and the other classifications is that indicators only display variation on the social level (i.e., among the different social classes) but not stylistic variation. Their status, however, can change over time. Markers, on the other hand, are salient but

only to in-group members and display variation on both the social and stylistic levels (Labov calls this "consistent stylistic and social stratification," 1994, p. 78). Markers are subject to change due to their salience (assuming that when a feature is salient it can be controlled which gives the speaker a choice when constructing utterances). Lastly, stereotypes are salient to both in-group and out-group members and often have an extra high level of awareness attached to them. However, due to their status as stereotype, they often function as a basis for negative comments and are often misrepresentations of vernacular speech. Stereotyped features, though, might enjoy widespread prestige among in-group speakers. This dual status of stereotyped features means that they not only are subject to correction and hypercorrection (Labov, 1994, p. 78) but also that they may not necessarily be likely to change, due to their ultra-salient status as this "may inhibit accommodation." (Trudgill, 1986, p. 125).

According to Kerswill and Williams (2002), salience is "a notion which seems to lie at the cusp of language internal, external and extra-linguistic motivation [...] which we can provisionally define rather simply as a property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent." (ibid.: 81). In their (2002) paper, Kerswill and Williams review several empirical studies of salience (including Trudgill, 1986) and conduct their own study investigating vowels, consonants and non-standard grammatical features in Milton Keynes, Reading and Hull. Based on their results and a discussion of the social embedding of forms, Kerswill and Williams conclude that it is not possible to set up any conditions which are either necessary or sufficient in order for a linguistic phenomenon to be salient and that the only prerequisite for salience seems to be that "its presence and absence must be noticeable in a psychoacoustic sense" (2002, p. 105). So "while language-internal factors play a part, it is in the end sociodemographic and other extra-linguistic factors that account for the salience of a particular feature" (ibid.: 81).

Branching out from pure sociolinguistic research, Hollmann and Siewierska (2006) take a socio-cognitive approach to salience. They agree with Kerswill and Williams' emphasis on the importance of social factors but "see cognitive-perceptual factors as primary" (ibid.:209) because "linguistic items are will normally be more or less free from social values when they come into existence. It is only after they have emerged that social forces can start working on them" (ibid.). Thus, they place emphasis on cognitive-perceptual factors in determining salience as they see them as not only prior to any social factors but also as governing whether a form becomes subject to social evaluation.

In one of the more recent publications on salience within sociolinguistics (Rácz, 2013), we find a differentiation between cognitive (primary) and social (secondary) salience. Rácz' study is based in the area of sociophonetics and he sees salience as ultimately connected with surprisal. While related, cognitive salience is seen as separate from social salience and he defines the relationship between the two as follows: "Cognitive salience is an attribute of variation that allows language users to pick up on it, whereas social salience means that variation is already used

to carry social indexation.” (ibid.: 37). This conceptualization of salience seems to support that presented by Hollmann and Siewierska (2006) above and brings in a useful distinction: that between the individual and the community level. It is clear that any consideration of the cognitive level must be concerned with individuals only, but also that individuals form communities, which allows us to extend our focus from the individual to the community. We return to this below in the conceptualization of language as a CAS.

The Enregisterment of Social Meaning

Rácz is not the only one to consider the role of social meaning in the study of salience. Honeybone and Watson (2013) in their study of Liverpool English phonology based on Contemporary, Humorous, Localized Dialect Literature suggest that a likely factor of the social salience of linguistic forms is the form's status as a local variant, indexing local identity. Similar results were also found for morphosyntactic and lexical forms in Tyneside English in Jensen (2013) who defines salience as the association of social content and linguistic forms in the cognitive domain. Thus, we see here that the social aspect is seen as crucial in the degree of salience of a number of non-standard forms.

Linked to the role of social meaning of local forms in speakers' identity constructions and often invoked in sociolinguistic studies as explanations of language variation and change are Silverstein's social indexicality (2003) and Agha's process of enregisterment (2003). Silverstein (2003, p. 217) directly maps his idea of different levels of social indexicality onto Labov's indicators, markers and stereotypes. Labov's *indicators*, Silverstein argues, are forms used by all members of a particular social group and they thus index only the speakers' macro-social identity (ibid.). *Markers*, on the other hand, are more intricate as they index not only macro-social identity but also style. He concludes on the topic of markers that “[w]hat Labov and followers have graphed in the so-called sociolinguistic marker is the dialectical process of indexical order for members of the standard-register informed language community as an articulated macro-social/micro-social fact” (ibid.: 220). Finally, Silverstein comments that *stereotypes* are markers whose interpretation is now wholly in the $n + 1$ st order indexical field, i.e., the social connotations of the linguistic form are presupposed before the original (n -th order) interpretation (ibid.: 220). Connected to the notion of indexical order and the social indexicality of forms is enregisterment which describes “processes through which a linguistic repertoire becomes differentiable within a language as a socially recognized register of forms” (Agha, 2003, p. 231). Indeed, it can be argued that the $(n + 1)$ st order indexical value of a linguistic form expresses the enregistered meaning of the form.

Johnstone (2009, p. 164), who investigates the indexicality of Pittsburghese, presents an overview of Silverstein's levels of indexicality and links them, very helpfully, with Agha's (2003) processes of enregisterment. We can summarize these in the following way:

- n th order indexicality/first order: this describes a linguistic form whose frequency of use patterns according to the

socio-demographic background of the speakers (gender, class, region, age).

- $n+1$ st order indexicality/second order: this describes a linguistic form which has acquired a social meaning which reflects dominant ideologies in the speech community (e.g., language correctness). At this stage, the form and social meaning are noticed by speakers.
- $(n+1)+1$ st order indexicality/third order: this describes a linguistic form which has acquired an additional indexical meaning (in addition to its first order index) which results in it being interpreted in light of a different ideology (than the second order index). It is on this level that we find the additional layer of social value and where the form has been enregistered in the community. A link has been established between the use of the form and the social value (e.g., localness of the speaker).

As we can see, Silverstein's indexicality gives an account for how the social meaning of linguistic forms emerges on the level of the community and Agha's enregisterment describes the processes which cement the third order indexical values of these forms in the community.

But why do linguistic forms suddenly become enregistered in a community? Beal (2000; 2009) and Johnstone (2009, but most explicitly 2010) have argued that it is in times of change that the re-interpretation or resemiotization (that is, re-indexing of social meaning and enregisterment) of linguistic forms in terms of third order indexical meaning takes place. Johnstone's (2010) main argument is that in times of disruption (she focuses on globalization), very local forms come to index different social meanings. The features become the topic of conversation and they are used to differentiate members of different speech communities. However, most importantly, Johnstone argues that the idea of local speech as unique (and thus enregistered) solidifies the link between speech and place which renders other indexicalities (such as class, gender or age) less accessible.

If we acknowledge the cognitive aspect of salience (as is done e.g., by Jensen, 2013 and Rácz, 2013), then indexicality and enregisterment are useful aspects to consider. However, in processes of both indexicality and enregisterment, the attachment of social value to linguistic forms must take place on the level of the individual first and then spread to the community level from this point. Below, I bring in a psycholinguistic perspective as a way of unifying the social and cognitive aspects of language use.

Social Meaning in an Exemplar Framework

By viewing language as a CAS (Beckner et al., 2009), we can account both for the link between the social and the cognitive aspects of language via exemplar theory (on the level of the individual) as well as the link between the individual and the community-level patterns of enregistered social meaning.

According to Beckner et al (ibid.: 2), the key features of language as a CAS are:

- (a) The system consists of multiple agents (the speakers in the speech community) interacting with one another.

- (b) The system is adaptive; that is, speakers' behavior is based on their past interactions, and current and past interactions together feed forward into future behavior.
- (c) A speaker's behavior is the consequence of competing factors ranging from perceptual mechanics to social motivations.
- (d) The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive processes.

What this means, then, is that speakers make choices about their own language (idiolect) but that these individual choices across a community result in emergent patterns of language use on a community level (ibid.: 14–15). Within this conceptualization of language, speakers' individual grammars are constructed as exemplar frameworks (ibid.:7).

Exemplar theory was first introduced in psychology in the 1970s as a model of perception and categorization and it has since then been adopted by linguistics and extended to the study of speech sounds and word recognition (Bybee, 2001, 2010; Pierrehumbert, 2001, 2003, 2006) among other areas. In short, exemplar models posit that “people represent categories by storing individual exemplars in memory, and classify objects on the basis of their similarity to these stored exemplars” (Nosofsky and Johansen, 2000, p. 375). Thus, exemplar theory presupposes richly detailed memory of exemplars, it is nonanalytic and works instead to match exemplars in a network fashion and it relies on probabilities and frequencies to do so (Mendoza-Denton, 2007; Barsalou, 2012; Fowler and Magnuson, 2012).

Pierrehumbert (2001) proposes that memories of tokens are stored in cognitive clouds where similar exemplars are stored close together and dissimilar ones far apart. The individual tokens or exemplars can be stored in several cognitive clouds depending on their categorization. In this way, the remembered tokens represent the range of variation encountered. A token can, for instance, be a word stored with information about particular acoustic features perceived (with phoneme-level exemplars stored separately, Drager, 2015, p. 154), the linguistic context in which it occurred and the social situation of when it was encountered (including formality levels and social information about the person who uttered it). If exemplars are frequently activated (either in production or perception), they remain at the forefront of the network “cloud” and are more easily activated again (they “carry the highest weight values,” Drager, 2015, p. 155). Both perception and production can be biased by the attachment of non-linguistic information to stored linguistic exemplars. In other words, social characteristics of interlocutors and the attitudes a speaker holds toward an interlocutor affect how we perceive their speech and how we address them (Niedzielski, 1999; Hay et al., 2006; Drager, 2015, p. 155–156).

According to Campbell-Kibler (2011), exemplar theory has appealed to linguistic theory generally, but the link between extralinguistic information and linguistic forms has been adopted and explored by sociolinguists and sociophoneticians in particular. She further states that “(e)xemplar theory's emphasis on the details of individual linguistic tokens makes it straightforward to link social information to extremely specific linguistic units and it is a compelling framework for further exploration of the linguistic character of sociolinguistic

connections.” (ibid.: 437). And while an exhaustive survey of all studies exploring the attachment of social meaning to linguistic variables is impossible to undertake here (even if focusing only on studies which couch their interpretation of results in exemplar theoretical terms), I will here summarize a few which have been selected to show exemplar-based accounts pertaining to both production and perception as well as different linguistic levels.

Hay et al. (2006) investigated the effect of perceived speaker identity on the perception of NEAR/SQUARE diphthongs which are currently merging in New Zealand English. Listeners were shown a photo of a speaker (older/younger, middle class/working class) and listened to a pre-recorded wordlist of unmerged NEAR/SQUARE items. While the results of the study were quite complex, overall, listeners seemed to be influenced by the social characteristics displayed by the photos. When listeners thought they were listening to an older speaker (who would be likely to produce unmerged diphthongs), they performed more accurately on the word identification task than when they thought they were listening to a younger speaker (who would be more likely to use merged forms), even though the auditory input was the same. According to the authors, this indicates that listeners treat the words as being ambiguous (when they think they are produced by a younger speaker) as they expect the vowels to be merged to a greater extent. Their results for the manipulation of the speakers' social class were less clear, but listeners seemed to expect middle class speakers to be less merged than working class speakers (2006, p. 479). Hay, Warren and Drager suggest that these results support an exemplar-based model of speech perception where exemplars are linked to social characteristics.

More recent work by Drager (2015) investigates both perception and production of *like* among adolescents in a New Zealand all girls' school. She takes a qualitative, ethnographic approach to the investigation of identity construction among the different social groups in the school (all centered on the use or non-use of the school Common Room) but also employs quantitative acoustic analyses and experimental designs. Her variable, *like*, can have both grammatical (verb, adverb, noun, etc.) and discursive (discourse marker, quotative, approximative adverb, etc.) functions (ibid.: 76–77), and she investigates both grammatical and acoustic differences in the production, use and perception of this single lemma. I will just focus on her results for the production aspects here, where Drager found that the girls' use of phonetic variants was related to whether they used the school Common Room (and thus were part of the “normal” social groups) or not (and thus identified as “weird” and as different from the “normal” groups). She states that “this finding provides evidence that linguistic variables are correlated with a speaker's stance and that speakers actively adopt and reject linguistic variants as part of the construction of their identity.” (ibid.: 148).

Campbell-Kibler (2011) investigated the perception of variants of the variable (ING), *-in* and *-ing*, through a matched guise experiment which contained three guises: *-in*, *-ing*, and a neutral guise which contained no (ING) tokens. Her initial hypothesis was that listeners' expectations would be influenced by speakers' regional accent and that this would impact the

perceptions of (ING). However, instead she found that the two variants were associated with different social features: *-ing* speakers were seen as more intelligent/educated and more articulate (than *-in* and neutral speakers) whereas *-in* speakers were perceived as being more informal and less likely to be gay (than *-ing* and neutral speakers). Thus, Campbell-Kibler concludes that “in some cases, variants of the same variable function independently as loci of indexically linked social meaning” (ibid.: 423).

Finally, also within sociolinguistic studies, both Rác (2013) and Jensen (2013), who specifically investigate the topic of salience, suggest exemplar theory as a way of explaining the link between the social and the linguistic in the cognitive, and Foulkes and Docherty (2006) argue that an exemplar-based model of phonological knowledge offers the most productive means of modeling socio-phonetic variation as it offers a unified account of how socio-phonetic and linguistic material might be learned and stored. They conclude that “the interweaving of sociophonetic and linguistic information in speech is so complete that no natural human utterance can offer linguistic information without simultaneously indexing one or more social factor” (ibid.: 419). Indeed, Foulkes (2010) goes as far as stating that “[e]xemplar theory appears to be the most promising candidate to construct a cognitively-realistic, integrated theory of phonological knowledge, speech production, and speech perception in which indexical knowledge is not marginalized but central.” (ibid.: 32). We see that indexical knowledge, then, again appears and is deemed to be central to the organization of an exemplar network.

QUESTIONNAIRE STUDY

This section reports on the variables under study (in Section Linguistic Variables), the design of the research instrument and the data yielded from the collection of questionnaires (Section Questionnaire Design and Output). The aim of the questionnaires was to investigate whether the local forms of the variables (do+NEG), (our), (told), (throw), and (go) are salient to Tyneside speakers and to investigate if participants’ *affiliation* with Newcastle and the wider Tyneside area had any impact on their awareness and frequency ratings of speech containing Tyneside vernacular features.

Linguistic Variables

This section will briefly introduce the linguistic variables (the vernacular forms) studied here. While this section aims to introduce the variables to the reader, the main focus will be on how they can be formally described as well as how frequent they are. Further descriptions, including etymology, can be found elsewhere (e.g., Beal, 1993, 2004, 2010; Beal et al., 2012; Jensen, 2013, 2015).

As a way to gauge the frequency of use of the different forms, a mini-corpus of Tyneside speech was compiled consisting of 24 dyadic interviews collected in Newcastle and Gateshead by local interviewers. The interviews selected were collected in the period 2007–2009 and are part of the Diachronic Electronic

Corpus of Tyneside English¹. More information about this corpus can also be found in Jensen (2013, 2015). The 48 speakers were distributed across social class, age and gender in the following way: 27 working class speakers and 29 middle class speakers, 29 young speakers (ages 17–34) and 27 older speakers (35+), 23 male speakers and 25 female speakers. The tokens were extracted using AntConc and included a variety of spellings² for each variable, in order to find all tokens in the corpus.

The frequencies of forms are given here first and foremost to help readers unfamiliar with the variety. Secondly, the corpus frequencies given below are also compared to the perceived frequencies given in the questionnaire study in Section Analysis and Results of Frequency Judgments below. As such, this paper does not attempt to investigate links between actual frequencies and perceived frequencies or hypothesize on the role of relative or absolute frequencies of vernacular forms to their level of salience. Indeed, the topic of interest in this paper is the link between forms’ perceived frequencies and salience.

(do + NEG)

Sentential negation with *do* in Tyneside English is realized as *divn’t* (see examples below) and this form dominates the full present tense paradigm apart from the third person singular, which is *doesn’t* (possibly realized as *dizn’t*, see Rowe, 2007). The mini-corpus contained a total of 1663 tokens of sentential negation with *do*; 96 of these were in a vernacular form (5.8%).

- (1) Ah I just *divn’t* want to get kidnapped. [07-08/N/ML/159]
- (2) The bars open late now *divn-t* they [07-08/N/RM/512]

(our)

The first person plural possessive pronoun in Tyneside English is *wor* and while this form is unique to the Tyneside area (Jensen, 2013), indeed the first person standard pronoun paradigm has been nearly completely re-organized in Tyneside English (this includes the use of *us* in both the plural subject and singular object, for instance). The mini-corpus contained 236 tokens of the first person plural possessive pronoun, 70 (29.7%) of which were *wor*.

- (3) Me and Kerry have known each other like, all *wor* life [07-08/T/BB/929]
- (4) Oh yeah, we’re great friends with *wor* next door neighbors [07-08/N/VL/3892]

(told)

In Tyneside English, the past tense of the verb *to tell* is *telt*, which occurs both in the simple past as well as in constructions with the past participle. The compiled mini-corpus contained only 84 tokens of this variable out of which 5 (6%) were local forms.

- (5) I *telt* O’Brien about them [07-08/N/ML/159]
- (6) *didn-t* want to be *telt* what to do [07-08/N/PS/243]

¹<http://research.ncl.ac.uk/decte/>.

²This was necessary for two reasons: first to collect all morphological forms of the words (e.g., *hoy*, *hoyed*, and *hoying*) but, two, also because there is some variability in the transcription conventions used within the corpus (so *divn’t* may be found as *divn’t*, *divvent*, *divn-t*).

(throw)

In Tyneside English we find a different lexical verb for the verb *to throw*, namely *to hoy*. This verb follows the regular paradigm and also occurs in the present participle (as *hoying*) and the past participle (*hoyed*). The corpus featured a total of 40 tokens with 11 (27.5%) being vernacular forms.

(7) they hoy it in the microwave and all [07-08/N/PM/85]

(8) the police used to hoy you over the wall so you'd get in free when you were little. [08-09/N/TS/556]

(go)

Finally, the verb *to go* is realized as *gan* in Tyneside English (present tense and present participle only) and is considered a separate verb (rather than a reflection of phonological differences between Standard English and Tyneside English; for more on this see Jensen, 2015). There is some variability in the vernacular paradigm as it seems to occur both with *-s* in all persons (as is common for some Northern verbs in the present tense, see Beal, 2010) and without (possibly following either the regular Standard paradigm or as subject to the Northern Subject Rule, Beal, 2010; Jensen, 2015). The mini-corpus featured a total of 2289 tokens of this variable; 202 (8.8%) of these were vernacular forms.

(9) Every-time you gan somewhere in that castle, shotgun shell! [07-08/G/DM/456]

(10) They constantly had me mam ganning up to the school to talk about us and stuff like that [07-08/N/PS/243]

Questionnaire Design and Output

The questionnaire consisted of three separate tasks. Task one was a frequency judgment task which asks participants to indicate how frequent they believe certain forms are. Task 2 asked participants about their own language use and tested whether they can identify Tyneside features, and task 3 aimed to establish the participants' affiliation with the local area. The original questionnaire tested 12 different vernacular variables as well as four filler variables, but the part reported here will focus on only the five included in this paper (the full account can be found in Jensen, 2013).

The format of the questionnaire was inspired by Burbano-Elizondo (2008), who carried out a study of Sunderland English (another North Eastern British variant). In her study, she implemented an "affiliation"-score which she matched against informants' assessments of sentences featuring non-standard forms. She found a correlation between the informants' level of positivity toward Sunderland and their assessments of non-standard forms.

The section below gives further information about the general considerations of the questionnaire design including the counterbalancing scheme, the construction of example sentences and the use of filler sentences and controls overall. Section Analysis and Results of Frequency Judgments describes each task in more detail and includes information about the number of example sentences and fillers used and the type of output generated.

Overall Questionnaire Design

The questionnaire featured a brief introduction to its objectives and what participants were required to do. Each of the three tasks also featured a brief description of the task at hand and an example of how the participants should indicate their answers. Due to the high number of variables in the original questionnaire (12 vernacular variables + 4 filler variables), three overall versions of the questionnaire were created (A, B, C) each of which tested only four vernacular variables in task 1. For each version, two subversions were created which featured different example sentences containing the different variables (resulting in A1, A2, B1, B2, C1, C2). Finally, for each of these subversions 2 editions were created which featured the example sentences in random order (thus giving A1a, A1b, A2a, A2b, etc.). The tasks presented participants with both sentences containing Tyneside English forms, sentences containing standard forms and filler sentences containing either common non-standard forms (i.e., not local to the Tyneside area) or ungrammatical forms. The counterbalancing scheme can be found in **Figure 1** mentioned below. Note that this is based on example sentences in task 1.

The example sentences in tasks 1 and 2 were given in direct speech which formed part of small scenarios in order to make them more pragmatically acceptable (Schütze, 1996; Buchstaller and Corrigan, 2011; Buchstaller et al., 2013). This strategy also helps in making the written forms of the dialect variables less odd to the participants as they occur in the form of direct speech, and informants may then be more likely to judge them without prescriptivist influence. In addition, the example sentences used simple vocabulary (Cowart, 1997) in order to avoid sentences being rated negatively due to participants' unfamiliarity with the vocabulary used. The context in which the direct speech example sentences occurred was based on interactions between four fictional characters (John, Peter, Emily, and Betty) and described everyday set in everyday situations.

As mentioned above, the questionnaires also contained four filler variables, which functioned as control sentences in tasks 1+2 (in addition to the Standard English sentences). Fillers prevent participants from remembering and deliberating prior ratings and perhaps realizing what the underlying variable being tested is (Buchstaller and Corrigan, 2011). The fillers used took the form of two common non-standard forms (use of *ain't* and *they was*) and two ungrammatical forms (missing past tense inflection on verbs in combination with the adverb *yesterday* and erroneous use of the past tense form of an irregular main verb in negative sentences with *didn't*). Cowart (1997) also suggests that the fillers used represent different levels of unacceptability. In this study, the control sentences can be grouped on three levels of unacceptability. The standard forms of the vernacular sentences (which can be classed as a type of control too) would be expected to be rated as most frequent, as they are fully well-formed sentences. Participants would be expected to rate the common non-standard filler sentences as less frequent, as they are likely to be seen as less well-formed than the standard sentences but possible to some speakers. Finally, the ungrammatical filler sentences would be expected to be rated as most infrequent as they are likely to be completely non-acceptable to participants.

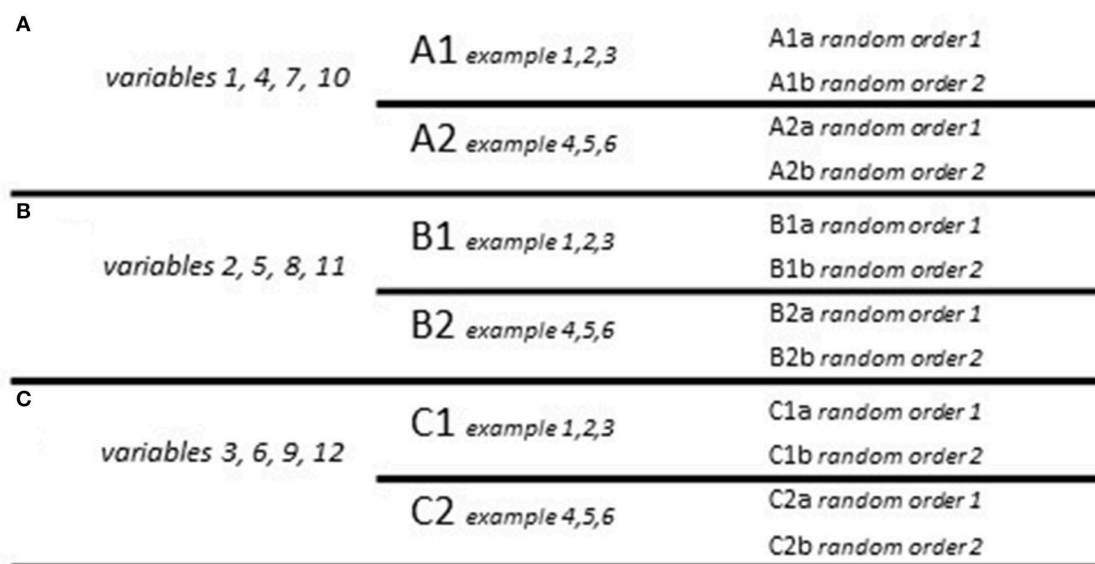


FIGURE 1 | Counterbalancing scheme.

The example sentences used were all taken from either the DECTE corpus (for Tyneside English forms) or the BNC (for the fillers) and modified to fit the example context and edited for simplicity to avoid ratings based on structural complexity (Schütze, 1996). For the non-grammatical fillers, this meant actually making them ungrammatical and, for the Standard English forms, this meant converting the original Tyneside English form to the standard form.

Task Structure and Output

This section will provide further information about the structure of the individual tasks, what their aims are and what kind of output they yield.

Task 1

The aim of task 1 was to uncover how frequent participants believe certain forms to be. As mentioned above, there are three versions of the questionnaire (versions A, B, C) and task 1 tests four different variables on each of these versions (each variable is featured three times in order to increase reliability of ratings, Cowart, 1997). In total, task 1 featured 36 sentences (12 sentences in Tyneside English, 12 in Standard English and 12 fillers). Participants were asked to rate each sentence on a scale from 1 to 7. A rating of 1 was described as “This sentence is never used here” and a rating of 7 as “I hear this all the time. People use this a lot.” There were no verbal descriptions given to the ratings in between. A 4-point scale with verbal descriptors was used in Buchstaller et al. (2013), and while this is perhaps more appealing to participants (as it may be easier to identify with verbal descriptors as opposed to numbers) and it avoids a median value, the use of an interval scale allows for the use of parametric tests in the analysis phase. In addition to running the risk of being perceived as an ordinal scale (Cowart, 1997, p. 70–72), the use of verbal descriptors would also yield data unsuitable for parametric

testing and thus non-parametric (i.e., less powerful) statistical methods would have to be used. The output of this task takes the form of numerical ratings from 1 to 7, which can then be averaged for each variable.

Task 2

The second task consisted of two parts: firstly, it aimed to establish how participants rate the frequency of their own use of particular forms and, secondly, if they can correctly identify local variants. The questionnaires tested all 12 variables in this task and included only the Tyneside English variants and the filler variables. This task featured 12 Tyneside English sentences (one for each variable) and 12 filler sentences (each of the four fillers occurred three times). Like task 1, task 2 also asked participants to use a 7-point scale to rate the example sentences. In this task, the verbal descriptors were 1: “I would never say this” and 7: “I say this all the time.” Due to prescriptivist pressure, participants were probably more likely to find this direct approach more invasive (compared to task 1), as they were asked to rate their own language. However, collecting both direct and indirect frequency judgments allows us to investigate how different variables are viewed in a community (Buchstaller and Corrigan, 2011). In the second part, participants were asked to indicate if the example sentences contained any local forms and to circle the word(s). This taps into their language awareness and requires that participants can be explicit about which features can be classified as belonging to the local area.

The output generated by this task is two-fold: the first output is similar to that of task 1, only this is a reflection of participants’ own use (to the extent that they are able to gauge it). This allows for comparisons between perceived “other” use and perceived “own” use with results telling us something about how forms are perceived in the community. The second output, the “awareness

score,” describes participants’ performance on the identification task and summarizes participants’ answers to the two parts (first a yes/no question and, second, the identification itself). The “awareness score” is thus simply a numeric expression of the total number of correct identifications, i.e., a correct indication of YES in the first question and a correctly circled form in the second part of this task yielded a score of one. This score was calculated for each variable (the average number of correct identifications of this variable across participants) as well as for the participants as a group (the average number of correct identifications across all variables). The awareness scores tell us if participants are explicitly aware of local forms and connect them with the area.

Task 3

The third task measured participants’ attitudes toward their local area, including the extent to which they feel an affiliation with the area. In this task, participants were asked to indicate to what extent they agreed with 10 statements which fell into five categories: opinion of the local area, orientation, network, self-definition, and attitude to dialect. These 10 statements also had to be rated on a scale from 1 to 7, where 1 was described as “I disagree strongly” and 7 as “I completely agree.” No verbal descriptions were given for the intermediate values. This section also featured background questions about participants’ gender, age, education, socio-economic class, area where they grew up and if they had ever lived outside the Tyneside area.

The 10 statements and their categories were:

- (1) I am proud to be from Newcastle (opinion of Newcastle/Gateshead)
- (2) I’m more interested in local news than national or international news (orientation)
- (3) I consider myself to be a Geordie (self-definition)
- (4) I like the way people speak in Newcastle (attitude to dialect)
- (5) Most of my friends are from Newcastle (network)
- (6) I have a Geordie accent (self-definition)
- (7) I feel I have more in common with people from Newcastle (network)
- (8) I like the way I speak (attitude to dialect)
- (9) I’m more likely to strike up a conversation with a stranger if I know they are from Newcastle (orientation)
- (10) I like living in Newcastle (opinion of Newcastle/Gateshead)

The output generated by this task is a “local affiliation score” which was calculated as an average of participants’ ratings of the 10 statements. This score can be compared to participants’ performance on the other tasks in order to investigate whether a locally-rooted social identity is linked with perceived frequency of; perceived own use of; and identification of vernacular forms. It is this affiliation score which allows us to explore possible links between social identity and language perceptions.

As mentioned earlier, the composite affiliation score generated by the responses to this task is based on Burbano-Elizondo’s work on Sunderland English. In her 2008 study, she employed a combination of different qualitative and quantitative methods in the construction of her *Index of Sunderland affiliation (ISA)* (2008, p. 126). While the present questionnaire study does not have a qualitative component, by incorporating questions about

participants’ orientation and opinion of the local vernacular in the local affiliation score, it aims to cover, in a quantitative manner, a similar range of topics.

Overview of Collected Data

Participants for the questionnaire study were recruited using the snowball method and, in total, 143 questionnaires were collected (summer of 2012). No particular social or age groups were targeted; the only criterion for participation was that participants identified themselves as Tyneside locals. The data was split into age groups after collection following the median of the participants’ reported age (median age = 47), which also gave the best distribution across the other social categories (class and gender). Class is operationalized in terms of the informants’ own definition of themselves (6 participants did not indicate class). The social stratification of the participants can be seen in **Table 1** below.

While this study will not further discuss the different behaviors of members of different social categories in detail, the above table provides the reader with an overview of the participants in the study. Overall, we can see that males were the hardest participants to reach, older males especially and middle class older males in particular. As a general observation, it should be added that middle class participants were harder to find when relying on people’s own definition of themselves; however, many participants who identified as working class indicated high levels of education such as university degrees (see Jensen, 2013, in press for a discussion of this).

Analysis and Results

This section describes the collected questionnaire data and presents the different analyses and results based on the output described above.

Analysis and Results of Frequency Judgments

Comparing the ratings of the vernacular example sentences in tasks 1 and 2 gives us an indication of the status of the variables (see **Figure 2**). The reader should bear in mind that the ratings for task 1 are based on 46–49 responses as not all variables were included in each questionnaire version in task 1. The means for task 2 are based on 138–143 responses.

Dependent *t*-tests found significant differences between participant ratings for all variables and an overview of results is given in **Table 2** below.

As we can see from the table, participants rate the use of vernacular forms by others as more frequent compared to their own use and significantly so. This indicates that participants are aware of the stigma surrounding non-standard forms.

TABLE 1 | Distribution of questionnaire participants based on social information.

WC				MC			
Younger (15–47)		Older (48+)		Younger (15–47)		Older (48+)	
Male	Female	Male	Female	Male	Female	Male	Female
12	34	7	32	11	14	4	23

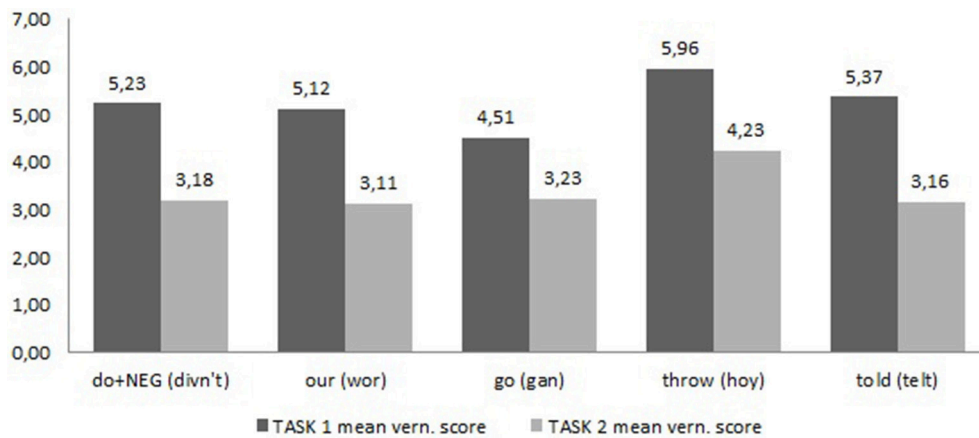


FIGURE 2 | Perceived frequency ratings compared.

TABLE 2 | T-test analysis of mean vernacular scores for tasks 1 and 2.

Variables	Ns	TASK 1 mean	TASK 1 st.dev	TASK 2 mean	TASK 2 st.dev	Difference (1–2)	t	p	95% C.I.	Cohen's d (effect size)
(do + NEG)	45	5.20	1.51	3.38	2.34	1.82	5.521	0.000	1.2–2.5	0.92
(our)	46	5.12	1.29	3.80	2.26	1.32	4.203	0.000	0.7–1.9	0.72
(go)	45	4.54	1.76	3.27	2.32	1.27	3.947	0.000	0.6–1.9	0.62
(throw)	47	6.04	0.92	4.51	2.18	1.53	5.259	0.000	0.9–2.1	0.91
(told)	47	5.32	1.30	2.43	1.98	2.89	10.673	0.000	2.3–3.4	1.73

Interestingly, the perceived frequencies of forms do not match up particularly well with the actual frequencies from the mini-corpus. Across all variables, questionnaire participants generally overstate the use of the local forms. **Table 3** below summarizes the frequencies from the corpus and also gives the corresponding means of tasks 1 and 2 from the questionnaire. In addition, the means from the questionnaires (which fall between 1 and 7) have been calculated into percentages (i.e., scores out of 100) to ease the comparison.

Correlational tests (Pearson product-moment) showed large positive correlations between the corpus frequencies and both task means, however, the results are not significant with an alpha level of 0.05. Task 1: $r = 0.475$, $n = 5$, $p = 0.419$ with a shared variance of 22.6%. Task 2: $r = 0.801$, $n = 5$, $p = 0.103$, 64.2% shared variance.

Analysis and Results of Identification Task

The output of this task was two “awareness scores”; one for the participants and one for the individual variables. Overall, participants were good at correctly identifying the Tyneside forms with a mean score of 9.08 ($N = 143$, standard deviation = 2.55, minimum = 0, maximum = 12). With regards to the individual variables, we can see from **Table 4** below that all five variables were identified over 90% of the time.

The awareness scores of the variables capture the degree to which participants were aware of them and connected them with the local area. In that way, they tell us something about the salience of the variables as participants have to be aware of the

TABLE 3 | Corpus frequencies.

Variables	N	Vernacular forms N/%	TASK 1 mean/%	TASK 2 mean/%
do + NEG (<i>divn't</i>)	1663	96/5.8	5.20/74.29	3.38/48.29
(<i>wor</i> for <i>our</i>)	236	70/29.7	5.12/73.14	3.80/54.29
go (<i>gan</i>)	2289	202/8.8	4.54/64.86	3.27/46.71
throw (<i>hoy</i>)	40	11/27.5	6.04/86.29	4.51/64.43
told (<i>telt</i>)	84	5/6	5.32/76	2.43/34.71

forms and link them to the area in order to be able to identify them.

Analysis and Results of Affiliation Task

As outlined above, the tasks consisted of 10 statements (in five categories) and participants had to indicate the extent to which they agree by using a 7-point scale. **Table 5** below shows participants' ratings of the different categories. We can see that they have a generally positive opinion of their local area, that they generally identify as *Geordies*, and that they have a favorable opinion of the local variety. Finally, while they have local networks, their orientation is not focused on the local area.

Before exploring the correlations between participants' affiliation score and their performance on the other tasks, a principal components analysis (Oblimin/oblique rotation) was carried out in order to test if the affiliation score can actually be perceived as a composite index at all. A PCA works by

TABLE 4 | Identification of vernacular forms.

Variables	N	Mean	Correct identification	Correct identification in %
do + NEG (<i>divn't</i>)	143	0.93	133	93
(<i>wor</i> for <i>our</i>)	143	0.91	130	91
go (<i>gan</i>)	143	0.91	130	91
throw (<i>hoy</i>)	143	0.94	134	94
told (<i>tel't</i>)	143	0.94	134	94

TABLE 5 | Affiliation ratings.

Task 3	Ns	Minimum	Maximum	Mean	Standard deviation
Opinion	143	1.0	7.0	6.11	1.23
Self-definition	143	1.0	7.0	5.46	1.38
Attitude	143	1.0	7.0	5.17	1.46
Network	143	1.0	7.0	4.67	1.51
Orientation	143	1.0	7.0	3.44	1.54
Scores across all five categories	143	1.0	7.0	4.97	1.00

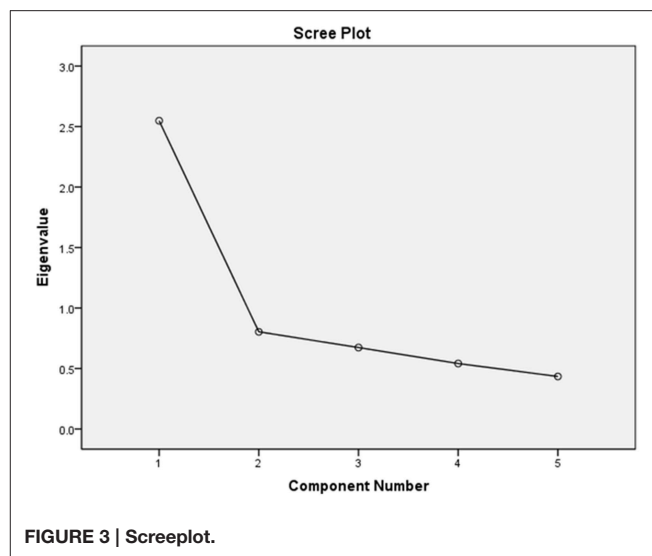
TABLE 6 | Components found in principal component analysis of the five categories.

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.548	50.962	50.962	2.548	50.962	50.962
2	0.803	16.064	67.026			
3	0.673	13.468	80.493			
4	0.541	10.821	91.315			
5	0.434	8.685	100.000			

reducing data and revealing underlying structures in large sets of variables. Here, it was used to investigate the extent to which the categories in the “affiliation index” cluster together, i.e., the extent of their association (Pallant, 2007, p. 179) and thus the extent to which they can be seen as parts of a composite score.

The data passed the initial suitability assessment (Kaiser-Meyer-Olkin value = 0.774, Bartlett's Test of Sphericity = $p < 0.000$). The coefficients of the correlation matrix were mainly above 0.3 and a high positive correlation ($r = 0.520$) between the categories “attitude” and “opinion” was found, clearly linking these two categories. The PCA of the five categories showed the presence of only one component with an eigenvalue exceeding 1.0 (2.548) explaining 50.962% of the variance as we see from Table 6 below.

This was further supported by the screeplot which showed a clear break after the first component, shown here in Figure 3.

**TABLE 7 | Correlations: frequencies and local affiliation.**

	N	Task 1	Task 3	r	p	N	Task 2	Task 3	r	p
(do + NEG)	46	5.22	4.94	0.46	0.001	142	3.18	4.96	0.41	0.000
(our)	48	5.11	5.14	0.44	0.002	141	3.11	4.96	0.31	0.000
(go)	46	4.51	4.94	0.24	0.115	142	3.23	4.97	0.34	0.000
(throw)	48	5.96	5.15	0.48	0.001	141	4.23	4.97	0.36	0.000
(told)	49	5.37	4.82	0.13	0.382	140	3.16	4.96	0.33	0.000

The component matrix showed that all variables loaded strongly on this single factor (over 0.4). The factor weights indicate that “attitude” loads most strongly (and is thus the most important in the composite score) with a score of 0.764, followed by “opinion” (0.751), “network” (0.749), “self-definition” (0.697), and finally “orientation” (0.595.). Because only one component was found, rotation could not be performed. On the basis of this analysis, we can accept the affiliation score as a composite index.

The affiliation score was correlated (using Pearson's Product-Moment Correlation) with the ratings in task 1 (perceived frequency of other people's use) and task 2 part 1 (perceived frequency of own use). Table 7 below gives the correlations between participants' affiliation score and their ratings in the two tasks, respectively. Variability in the mean values of task 3 (affiliation index) and the N-values is due to missing answers in either task 1 or task 2 as variables with missing responses were excluded from the analysis.

For all variables, we see that the correlation between the ratings and the affiliation index is positive, i.e., the higher the affiliation score, the higher the rating of the vernacular forms. The most important result here is the r -value as that describes the level of correlation between the two scores. Usually, a value above 0.3 is interpreted as a medium value (which will be the threshold used here). While it is important that the p -value is low (below 0.05 to indicate a significant and reliable result), the value itself does not indicate the importance of the r -value (Dancey and Reidy, 2011, p. 188, Pallant, 2007, p. 132–33). In the table, cells which

feature an r -value above 0.3 and a p -value below 0.05 have been shaded. We can see that there are significant correlations between the ratings for all variables in task 2 (participants' own use) and participants' affiliation scores and for three out of five variables in task one (frequency in other's use) and the affiliation index scores. In short, the more attached participants feel to the local area, the higher they rate both other people's use of vernacular forms but in particular their own. This indicates that local affiliation may influence perceptions of both other people's language use but also of own language use. This will be discussed further in Section Discussion and Conclusion below.

Finally, another Pearson test was run to see if there was any correlation between participants' affiliation score and their ability to correctly identify the vernacular variables. This was calculated on the basis of the responses to the individual variables (i.e., it was a point-biserial correlation with a bivariate variable, either correct or incorrect identification of the variable, and a continuous variable, the participants' affiliation score). As the identification task is a dichotomous variable, the mean values indicated are simply the mean of the coding, where 1 represented a correct identification and 0 an incorrect identification (either an erroneous identification or simply a missing answer). Again, cells with significant results ($p < 0.05$ and $r > 0.3$) have been shaded.

Table 8 above shows that, for three of the five variables, there is a significant correlation between participants' ability to correctly identify vernacular forms and the expression of local affiliation (as measured in the affiliation index). While none of the tests returned correlations above 0.3, we can see that (throw) came the closest with 0.220 (and also showed a highly significant correlation with $p = 0.008$) followed by (our), $r = 0.203$, $p = 0.015$). We can interpret these results as meaning that, at least for some vernacular features, there may be a tendency for level of local affiliation to positively impact explicit awareness of local vernacular forms.

DISCUSSION AND CONCLUSION

To summarize the above section, we saw that there was a difference in how speakers rate their own speech vs. that of others. The questionnaire participants rated all five variables as more frequently occurring on the speech of others than in their own to a significant degree. Furthermore, we saw that participants were very competent in identifying the five vernacular variables (all identified correctly over in 90% of occurrences) and connected them with the Tyneside area. The affiliation index allowed comparisons between participants' performance in the different tasks with a composite measure of their attachment to the local area. While not conclusive across all five variables, these comparisons showed that there may be a connection between speakers' affiliation with their local area and their awareness of the use of local features, in particular in their perceptions of the extent to which they themselves use local forms.

We can see, then, that the variables investigated here seem to be enregistered as unique to Tyneside (cf. Johnstone, 2010; Honeybone and Watson, 2013). Their status as indices of Tyneside local identity can become strengthened over time with

TABLE 8 | Correlations: identification and local affiliation.

	<i>N</i>	Identification	Affiliation	<i>r</i>	<i>p</i>
(do + NEG)	143	0.93	4.97	0.092	0.277
(our)	143	0.91	4.97	0.203	0.015
(go)	143	0.91	4.97	0.178	0.034
(throw)	143	0.94	4.97	0.220	0.008
(told)	143	0.94	4.97	0.075	0.377

use and increased exposure. In this way, we can see speakers as active participants in the construal of the social meaning of linguistic forms. From an exemplar theoretical perspective, we can argue that this enregistered status affects their storage in the exemplar network cloud. If unique local forms are stored as separate entries (rather than exemplars of standard forms), they are perhaps in a better position to be imbued with social value. This would also mean that they escape the pressure of prescriptive rules which face non-standard forms otherwise. This would presuppose, however, that the speakers perceive the vernacular forms as being unique to Tyneside, something which the results reported here indicate is the case. We can perhaps then also suggest there is a close link between salience and social value and that they are important factors in a model of language meaning, with unique forms (or forms perceived to be unique, rather) being the best carriers of social meaning as they are more positively viewed in the community and not stigmatized to the same extent as non-unique forms (Jensen, 2013). This link between the social value of the form and the linguistic form itself is what we can capture by the term salience if salience is defined as the association of social content and linguistic forms in the cognitive domain.

As mentioned in Section The Enregisterment of Social Meaning, it has been suggested that processes of enregisterment are set in motion by disruption in some form. In the case of the Tyneside area, this catalyst could be the transformation which the area has seen over the last several years. A hundred years ago, the Tyneside area was an area defined by heavy industry (such as shipbuilding) and the town of Newcastle was the retail center for the whole of the north of England. When the heavy industry began to wane in the mid-1900s, Newcastle strengthened its position as a retail center. More recently, focus has shifted to the consumption of culture with both a modern art gallery and an all-glass concert hall as well as several bars and pubs lining the banks of the river Tyne. Finally, Newcastle is also a popular student city and has the fifth largest student population in England and Wales (Beal et al., 2012; Jensen, 2013). It can thus be argued that this transformation of the Tyneside conurbation which the Tyneside speakers have witnessed (but which has not influenced the stereotypical associations held by out-group members, see Watt, 2002) provides the optimal conditions for enregisterment processes of certain local forms to happen.

An additional aspect of the social value argument is that attachment of meaning to particular local forms (in this case localness) allows forms to participate in the stylization of social personae (Podseva, 2011a,b; Drager, 2015, p. 157). Drager (2015, p. 157–163) gives a step-by-step account of how the construction

of a social persona through the adoption and non-adoption of different features (linguistic and otherwise) may be “understood within an exemplar-based hybrid model.” (ibid.: 157). In short, both the presence and absence of different features are part of creating a social persona, that is, different features can index different personae to different extents and sometimes it is the combination of variants over a range of variables which delimit one persona from another. Not all features which could become parts of a social persona do, however, and speakers are still influenced by social convergence and divergence (Giles and Powesland, 1975) and they are free to shift their personae over the course of an interaction.

In the study presented here, consideration of speakers’ creation of social personae (which in this case are centered around signaling localness) may explain the full correlation between all variables in task 2 part 1 and the affiliation score; speakers with a high affiliation score also want to present themselves as “true Geordies” (which can be done by claiming to use features perceived to be local to a large extent). This presupposes, of course, that participants can identify the local features in the examples sentences (task 2 part 2) and thus that they are aware of them. As we saw from the results of the identification task, all variables in this study were correctly identified as local over 90% of the time. Not only do participants’ ratings then indicate that they are aware of which features are local, but also that an awareness of what being a Geordie might entail and how to enact it. Additionally, the adoption of a Geordie persona also indicates a positive attitude both toward Geordie as an identity (and with that the local area) but also about showing it. This suggestion is backed up by findings reported in Beal (1999) and Jensen (2013). Indeed, Beal (1999:45) states that “[p]erhaps the preservation of stereotypical pronunciations in key words like “Toon,” along with the leveling toward supraregional rather than national norms reported by Watt (2002), represent a strategy for maintaining the positive aspects of the “Geordie” stereotype: friendliness and a strong sense of regional identity, whilst dissociating oneself from the negative, “grim up north” aspects of that stereotype.”

REFERENCES

- Agha, A. (2003). The social life of cultural value. *Lang. Commun.* 23, 213–273. doi: 10.1016/S0271-5309(03)00012-0
- Barsalou, L. W. (2012). “The human conceptual system,” in *The Cambridge Handbook of Psycholinguistics*, eds M. Spivey, M. Joanisse, and K. McRae (New York, NY: Cambridge University Press), 239–258.
- Beal, J. (1993). “The grammar of Tyneside and Northumbrian English,” in *Real English: The Grammar of English Dialects in the British Isles*, eds J. Milroy and L. Milroy (Harlow: Longman), 187–213.
- Beal, J. (1999). Geordie Nation: Language and identity in the North-east of England. *Lore Lang.* 171, 33–48. Available online at: <http://collections.mun.ca/PDFs/lorelang/LoreandLanguageVol17No01021999.pdf>
- Beal, J. (2000). From Geordie Ridley to Viz: popular literature in Tyneside English. *Lang. Lit.* 9, 343–359. doi: 10.1177/096394700000900403
- Beal, J. (2004). “English dialects in the North of England: morphology and syntax,” in *A Handbook of Varieties of English*, Vol. 2, eds B. Kortmann, E. W. Schneider, K. Burridge, R. Meshtrie, and C. Upton (Berlin: Mouton de Gruyter), 114–141.
- Finally, it should be self-evident that language exists on two levels; the individual level and the community level. We saw in Section Social Meaning in an Exemplar Framework how CAS theory suggests that speakers make choices about their own language but that these individual choices result in emergent patterns of language across a community. Similarly, we can also see language, or, rather, meaning, as operating on two levels; the first is the denotational level (which captures the communicative meaning of the speech signal) and the second is the sociolinguistic meaning, which is tied to speakers’ linguistic identities. If we see speakers’ individual grammar as constructed as exemplar frameworks, then the merger of these two levels of meaning is unproblematic. This is also supported by the literature reviewed in Section Social Meaning in an Exemplar Framework. As for the local Tyneside variables investigated here, we can thus see them as carrying heavy indexes of “locality” within the individuals’ exemplar clouds and that this will affect the way speakers and listeners use and perceive the forms. On the community level, this will then result in different patterns of use across groups and across time. I will leave it up to future studies to investigate how these patterns might emerge and develop.
- ## AUTHOR CONTRIBUTIONS
- The author confirms being the sole contributor of this work and approved it for publication.
- ## ACKNOWLEDGMENTS
- The research reported here was carried out as part of my doctoral thesis which was fully funded by Northumbria University. I extend a special thank you to my supervisor, Ewa Dabrowska, for her support in matters theoretical as well as empirical. I would also like to thank my former colleague, Kim Ebensgaard Jensen, for his tireless encouragement in the writing of this paper, as well as the two reviewers for their comments. Any shortcomings remain my own.
- Beal, J. (2009). Enregisterment, commodification, and historical context: “Geordie” versus “Sheffieldish”. *Am. Speech* 84, 138–156. doi: 10.1215/00031283-2009-012
- Beal, J. (2010). *An Introduction to Regional Englishes*. Edinburgh: Edinburgh University Press.
- Beal, J. L., Burbano-Elizondo, C., and Llamas, C. (2012). *Urban North-Eastern English: Tyneside to Teesside*. Edinburgh: Edinburgh University Press.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, H. M., Croft, W., Ellis, C. N., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59(Suppl. 1), 1–26. doi: 10.1111/j.1467-9922.2009.00533.x
- Buchstaller, I., and Corrigan, K. P. (2011). “Judge not lest ye be judged: exploring methods for the collection of socio-syntactic data,” in *Language Variation - European Perspectives III*, eds F. Gregersen, J. K. Parrott, and P. Quist (Amsterdam: John Benjamins), 149–160.
- Buchstaller, I., Corrigan, K. P., Holmberg, A., Honeybone, P., and Maguire, W. (2013). T-to-R and the Northern Subject Rule: questionnaire-based spatial, social and structural linguistics. *English Lang. Linguist.* 17, 85–128. doi: 10.1017/S1360674312000330
- Burbano-Elizondo, L. (2008). *Language Variation and Identity in Sunderland*. Ph.D. thesis, Sheffield University, Sheffield.

- Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Campbell-Kibler, K. (2011). The sociolinguistic variant as a carrier of social meaning. *Lang. Var. Change* 22, 423–441. doi: 10.1017/S0954394510000177
- Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publications.
- Dancey, C. P., and Reidy, J. (2011). *Statistics Without Maths For Psychology, 5th Edn*. Harlow: Pearson Education.
- Drager, K. (2015). *Linguistic Variation, Identity Construction and Cognition*. Berlin: Language Science Press.
- Foulkes, P. (2010). Exploring social-indexical knowledge: a long past but a short history. *J. Lab. Phon.* 1, 5–39. doi: 10.1515/labphon.2010.003
- Foulkes, P., and Docherty, J. G. (2006). The social life of phonetics and phonology. *J. Phon.* 34, 409–438. doi: 10.1016/j.wocn.2005.08.002
- Fowler, C. A., and Magnuson, J. S. (2012). “Speech perception,” in *The Cambridge Handbook of Psycholinguistics*, eds M. Spivey, M. Joanisse, and K. McRae (New York, NY: Cambridge University Press), 3–25.
- Giles, H., and Powesland, P. F. (1975). *Speech Style and Social Evaluation*. London: Academic Press.
- Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* 34, 458–484. doi: 10.1016/j.wocn.2005.10.001
- Hollmann, W. B., and Siewierska, A. (2006). Corpora and (the need for) other methods in a study of Lancashire dialect. *Zeitschrift für Anglistik und Amerikanistik* 54, 203–216. doi: 10.1515/zaa-2006-0210
- Honeybone, P., and Watson, K. (2013). Salience and the sociolinguistics of Scouse spelling: exploring the contemporary, humorous, localised dialect literature of Liverpool. *English World Wide* 34, 305–340. doi: 10.1075/eww.34.3.03hon
- Jensen, M. M. (2013). *Salience in Language Change: A Socio-Cognitive Study of Tyneside English*. Ph.D. thesis, University of Northumbria, Newcastle.
- Jensen, M. M. (2015). Two directions of change in one corpus: phonology vs morphosyntax in Tyneside English. *Globe* 1, 43–71. doi: 10.5278/ojs.globe.v1i0.698
- Jensen, M. M. (in press). “Education, class and vernacular awareness on Tyneside,” in *Perspectives on Northern British Englishes*, eds S. Hancil and J. Beal (Berlin: De Gruyter).
- Johnstone, B. (2009). Pittsburghese shirts: commodification and the enregisterment of an urban dialect. *Am. Speech* 84, 157–175. doi: 10.1215/00031283-2009-013
- Johnstone, B. (2010). “Indexing the local,” in *Handbook of Language and Globalization*, ed N. Coupland (Chichester: Wiley-Blackwell), 386–404.
- Johnstone, B., Andrus, J., and Danielson, A. E. (2006). Mobility, indexicality, and the enregisterment of “Pittsburghese.” *J. English Linguist.* 34, 77–104. doi: 10.1177/0075424206290692
- Johnstone, B., and Kiesling, F. S. (2008). Indexicality and experience: exploring the meanings of /aw/-monophthongization. *J. Sociolinguist.* 12, 5–33. doi: 10.1111/j.1467-9841.2008.00351.x
- Kerswill, P., and Williams, A. (2002). “Salience” as an explanatory factor in language change: evidence from dialect levelling in urban England,” in *Language Change: The Interplay of Internal, External and Extra-Linguistic Factors*, eds M. C. Jones and E. Esch (Berlin: Mouton de Gruyter), 81–110.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1994). *Principles of Linguistic Change. Volume one: Internal Factors*. Oxford: Blackwell Publishers.
- Mendoza-Denton, N. (2007). “Socio-linguistic extensions of exemplar theory: comments on Flege, Khatlab, and Darcy, Peperkamp and Dupoux,” in *Laboratory Phonetics* 9, eds J. Cole and J. I. Hualde (Berlin: Mouton de Gruyter), 443–453.
- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psychol.* 18, 62–85. doi: 10.1177/0261927X99018001005
- Nosofsky, R. M., and Johansen, M. K. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychon. Bull. Rev.* 7, 375–402. Available online at: http://www.cogs.indiana.edu/nosofsky/pubs/2000_rmn-mkj_pbr_Exemplar.pdf
- Pallant, J. (2007). *SPSS Survival Manual, 3th Edn*. Maidenhead: Open University Press, McGraw-Hill Education.
- Pierrehumbert, J. B. (2001). “Exemplar dynamics: word frequency, lenition and contrast,” in *Frequency Effects and the Emergence of Linguistic Structure*, eds J. L. Bybee and P. J. Hopper (Amsterdam: John Benjamins), 137–157.
- Pierrehumbert, J. B. (2003). “Probabilistic phonology: discrimination and robustness,” in *Probabilistic Linguistics*, eds R. Bod, J. Hay, and S. Jannedy (Cambridge, MA: The MIT Press), 177–228.
- Pierrehumbert, J. B. (2006). The next toolkit. *J. Phon.* 34, 516–530. doi: 10.1016/j.wocn.2006.06.003
- Podseva, R. J. (2011a). Salience and the social meaning of declarative contours: three case studies of gay professionals. *J. English Linguist.* 39, 233–264. doi: 10.1177/0075424211405161
- Podseva, R. J. (2011b). The California vowel shift and gay identity. *Am. Speech* 86, 32–51. doi: 10.1215/00031283-1277501
- Rącz, P. (2013). *Topics in English Linguistics [TiEL]: Salience in Sociolinguistics: A Quantitative Approach*. Berlin: Walter de Gruyter.
- Rowe, C. (2007). He didn’t gan tiv a college ti di that, man! A study of do (and to) in Tyneside English. *Lang. Sci.* 29, 360–371. doi: 10.1016/j.langsci.2006.12.013
- Schütze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Lang. Commun.* 23, 193–229. doi: 10.1016/S0271-5309(03)00013-2
- Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell Publishers.
- Watt, D. (2002). “I don’t speak with a Geordie accent, I speak, like, the Northern accent”: contact-induced levelling in the Tyneside vowel system. *J. Sociolinguist.* 6, 44–63. doi: 10.1111/1467-9481.00176

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.

Copyright © 2016 Jensen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The *Penefit* of Salience: Salient Accented, but Not Unaccented Words Reveal Accent Adaptation Effects

Ann-Kathrin Grohe* and Andrea Weber

Psycholinguistics and Applied Language Studies, English Department, Faculty of Humanities, University of Tübingen, Germany

OPEN ACCESS

Edited by:

Bernd Kortmann,
University of Freiburg, Germany

Reviewed by:

Melissa Michaud Baese-Berk,
University of Oregon, USA
Helmut Spiekermann,
Westfälische Wilhelms-Universität
Münster, Germany

*Correspondence:

Ann-Kathrin Grohe
ann-kathrin.grohe@uni-tuebingen.de

Specialty section:

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

Received: 11 February 2016

Accepted: 24 May 2016

Published: 07 June 2016

Citation:

Grohe A-K and Weber A (2016) The
Penefit of Salience: Salient Accented,
but Not Unaccented Words Reveal
Accent Adaptation Effects.
Front. Psychol. 7:864.
doi: 10.3389/fpsyg.2016.00864

In two eye-tracking experiments, the effects of salience in accent training and speech accentedness on spoken-word recognition were investigated. Salience was expected to increase a stimulus' prominence and therefore promote learning. A training-test paradigm was used on native German participants utilizing an artificial German accent. Salience was elicited by two different criteria: production and listening training as a subjective criterion and accented (Experiment 1) and canonical test words (Experiment 2) as an objective criterion. During training in Experiment 1, participants either read single German words out loud and deliberately devoiced initial voiced stop consonants (e.g., *Balken*—"beam" pronounced as **Palken*), or they listened to pre-recorded words with the same accent. In a subsequent eye-tracking experiment, looks to auditorily presented target words with the accent were analyzed. Participants from both training conditions fixated accented target words more often than a control group without training. Training was identical in Experiment 2, but during test, canonical German words that overlapped in onset with the accented words from training were presented as target words (e.g., *Palme*—"palm tree" overlapped in onset with the training word **Palken*) rather than accented words. This time, no training effect was observed; recognition of canonical word forms was not affected by having learned the accent. Therefore, accent learning was only visible when the accented test tokens in Experiment 1, which were not included in the test of Experiment 2, possessed sufficient salience based on the objective criterion "accent." These effects were not modified by the subjective criterion of salience from the training modality.

Keywords: native accents, adaptation, eye-tracking, salience

INTRODUCTION

Languages typically consist of a number of regional dialects—that is, native accents. In the southwestern German state of Baden-Württemberg, for example, one does not have to travel very far to encounter various native accents, as Spiekermann documented in 2008 (Spiekermann, 2008). This variation can pose a problem for non-locals. When non-locals hear a native accent for the first time, they often do not understand what is being said as easily as do locals, who are experienced with the regional varieties. Recent studies have indeed shown that listeners process accents in their native

language more easily when they are familiar with the accents than when they are unfamiliar with them (e.g., Adank et al., 2009). Adaptation by non-locals to native accents is, however, possible. Adaptation has been found for longer periods of exposure to a novel accent (Evans and Iverson, 2007), but it can even be observed after just 4 min of listening to a new accent (Trude and Brown-Schmidt, 2012). This is also true for second language (L2) learners. Producing a new accent for only 7 min can facilitate subsequent accent understanding for L2 learners, even more so than listening to the accent does (Grohe and Weber, in press). The act of speech production arguably makes an accent more salient than listening to that accent does. Can the advantage of production experience also be observed in a listener's native language (L1)? Next to signal modality (production and listening), salience can also emerge from concrete properties of the speech signal itself. Acoustic distinctiveness of a speech signal can enhance its salience (e.g., Cho and Feldman, 2013). The present study used a training-test paradigm in German in which salience was induced by either a subjective or an objective criterion and looked at the role of salience in native accent adaptation. The subjective criterion was implemented through two different accent trainings (production and listening) and the objective criterion through the featuring of either accented (Experiment 1) or canonical (Experiment 2) target words during test in an eye-tracking study.

Adaptation to Native Accents

Familiarity with a native accent facilitates accent processing. For example, listeners with extensive experience with the New York City (NYC) English accent show greater priming effects for words with the NYC-English-typical final r-dropping than listeners with limited experience (Sumner and Samuel, 2009). Similarly, Adank et al. (2009) found that only listeners who were familiar with both Southern Standard British English (SSBE) and Glaswegian English (GE) showed equal performance on both accent types in a sentence verification task. The familiarity advantage probably results from adaptation processes, as demonstrated by Evans and Iverson (2007). In their study, students who were originally from Northern England adapted to SSBE over the course of their university studies in Southern England, as shown through production and comprehension tasks. Processing advantages for participants' own accents over an unfamiliar accent were also found for French listeners (Floccia et al., 2006). In a lexical decision task, reaction times to items in long sentences were faster when sentences were presented in the participants' own accent (Northeastern French) than when they were presented in the unfamiliar Southern French accent. Furthermore, participants did not adapt to the unfamiliar accent during the course of the experiment (see also Floccia et al., 2009). Additionally, Adank and McQueen (2007) found no short-term adaption in a study with regionally-accented Dutch. In their study, Dutch participants who were not familiar with the Flemish accent had to make animacy decisions on single words spoken by two different speakers, one with a Flemish accent and one with the same accent as the participants. Then, participants were exposed to another speaker with the Flemish accent before having to repeat the

animacy decision task. Decision times in the second animacy task were not faster than in the original task.

Short-term adaptation was, however, found by Trude and Brown-Schmidt (2012). Participants were first trained on a native English accent and then tested in an eye-tracking paradigm. During training, participants listened to scripted dialogs with an accented male speaker and an unaccented female speaker. The male speaker raised the /æ/ before /g/ to [e], i.e., *bag* was pronounced /beɪg/. During test, target words were either spoken by the male or the female speaker. When *back*, a word unaffected by the accent, was the target and *bag* the competitor, *bag* was ruled out more quickly as a candidate word for trials with the male speaker than it was with the female speaker. When *bake*, a word with /eɪ/ in its canonical form, was the target and *bag* acted as competitor, *bake* was fixated less often when it was spoken by the male. This effect, however, was less strongly pronounced, i.e., competitor inclusion was more difficult than competitor exclusion.

Specific properties of the tested accents could account for the missing effects of adaptation in the studies discussed above, but, more importantly, speaker-specificity can explain it too. In contrast to Adank and McQueen, who had different speakers with the same accent during exposure and test and did not find accent adaptation, Trude and Brown-Schmidt used the same accented speaker in both of their two experimental phases. Short-term adaptation to native-accented speech may therefore be rather speaker-specific. This problem has also been addressed in studies on foreign accent adaptation, with mixed results. Using a training test paradigm, Bradlow and Bent (2008) found that generalization of accent learning (Chinese-accented English) to new voices is only possible if the listener is exposed to multiple speakers during training (for similar findings see also Sidaras et al., 2009). Kraljic and Samuel (2007), on the other hand, found with L1 listeners that the nature of the tested material has an effect on whether perceptual learning can generalize to new speakers. They found generalization effects for plosives but not for fricatives.

Adaptation with Production

Speaker specificity raises the question of whether there is a way to make training more effective, i.e., allowing for generalization across speakers, and potentially rendering competitor inclusion more robust. This might be possible through production training. In a recent study by Grohe and Weber (in press), the production of a foreign accent in participants' L2 promoted adaptation to that accent in a subsequent lexical decision task. Participants first either listened to an English short story that featured the replacement of all dental fricatives ("th") with /t/ (e.g., *theft* pronounced as * *teft*), or they read the same story aloud with the same substitutions. A control group had no accent training. Afterwards, all participants completed a lexical decision task on words with the th-substitutions. The production group accepted the accented words significantly more quickly than the control group did. The listening group produced only a weak training effect. When the same experiment was run with L1 participants, no training effect was observed. Referring to speaker effects, L1 participants in the production group produced the critical accent marker, but they were listening to an L2 speaker in the test phase.

According to Pickering and Garrod (2013), listeners are more likely to refer to their own previous production experience if it is highly similar to the speaker they are listening to (e.g., in terms of sex, L1 background, dialect). Less similarity leads listeners to draw on their experience with others' speech. Since only the L2 participants in Grohe and Weber had the same L1 background as the recorded test speakers, speaker-listener similarity was smaller for L1 participants than for L2 participants.

Facilitatory effects of producing an accent were also found in an accent imitation study with L1 speakers of Dutch (Adank et al., 2010). Baese-Berk and Samuel (2016), however, found that imitating a newly learned L2 sound can even inhibit learning. In their study, participants had to imitate sounds from a sound continuum ranging from /ʃa/–/fʌ/, which is arguably difficult for speakers to imitate correctly. A potential acoustic discrepancy between the sound prompt that was presented and the participants' productions may therefore have inhibited learning effects. A recent discrimination study with Danish vowels (Kartushina et al., 2015) supports this interpretation. In that study, production accuracy was increased by concrete feedback on productions, which in turn resulted in better sound discrimination performance after production than after listening training.

Salience in Adaptation

We now turn to the concept of salience, which can potentially explain both the results of accent adaptation and the advantage of produced compared to listened-to tokens. Salience has been generally defined as “the property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent” (Kerswill and Williams, 2002, p. 81). An important question, however, is what exactly makes a linguistic item salient.

First, sociolinguistic research on salience suggests that an accent can increase a word's salience. As suggested by Trudgill (1986), the phonetic difference between two variant forms affects their salience; the greater the difference, the more a dialect speaker is aware of it. Phonetic distance can also be considered within the framework of distinctiveness which assumes that isolated, i.e., distinct, words are more salient than others during encoding—provoking additional processing and, therefore, better memory (McDaniel and Geraci, 2006, for a review). Geraci and Manzano (2010), for example, had participants study a list of semantically related words that also included a few semantically unrelated, i.e., distinct, words. In ensuing tests, more unrelated than related words were recalled. Accordingly, Siegel (2010) claims that salience requires a listener to notice a contrast between two linguistic tokens. In terms of phonetic variability, words that carry an accent are distinct from their unaccented counterparts and bear greater salience. Therefore, they can be learned more easily than unaccented words¹. This was tested in a different memory study (Cho and Feldman, 2013). L1 English participants listened to either Dutch-accented or native-accented English words during a training

phase. In a subsequent word recognition task, there was an advantage for Dutch-accented words.

Second, factors beyond linguistic or structural properties may also affect salience. For the case of dialect accommodation, Kerswill and Williams (2002) suggest intensity of dialect contact as one of several factors. Considering the findings on production effects on accent adaptation, we can extend the list of extra-linguistic factors toward cognitive mechanisms by introducing accent learning modality (production vs. listening) as an additional factor. Several studies have found an advantage of production over listening for dialect accommodation; this has been named the *production effect*. It predicts that overt production facilitates word recollection when compared with studying a word silently (MacLeod et al., 2010) and also when compared with listening to others producing a word aloud (MacLeod, 2011). It has been suggested that produced words are more easily recalled because they are more distinctive and therefore more salient. Distinctiveness results from listeners focusing more on their own than on others' productions, which are, in the sense of the embodiment hypothesis (for an overview: Glenberg, 2010), more embodied than others' productions.

Salience, as described above, has been further specified in sociolinguistic research. Referring to Schirmunski (1928) and Trudgill (1986), Auer et al. (1998) differentiate objective and subjective criteria of salience. For example, *articulatory distance* is described as an objective criterion and *perceptual distance* as its subjective counterpart. The two relate in that articulatory distance describes the magnitude which a linguistic token deviates acoustically from the canonical realization, whereas perceptual distance describes which way a listener perceives this distance. Based on this information, we can conclude that subjective criteria increase the salience of a stimulus, for example, due to regular practice, and the resulting cognitive pre-activation. Objective criteria refer to properties of a stimulus that itself attracts attention because of its distinct, physical characteristics. Under this view, the production effect relies on the presence of an objective criterion. A self-produced word can be physically more distinct compared to a word read silently or a word that is produced by others because these words were only tested in mixed lists, i.e., one participant had to listen to/silently read and produce words in the same session.

In summary, prior research has shown that native accents are more easily processed if they are familiar to a listener than if they are new. Short-term adaptation to native accents is possible, and production alone can positively affect foreign accent learning, at least in L2 learners. Both robust accent adaptation and the role of production in accent adaptation may be related to salience. The role and concrete nature of salience in learning accented vs. canonical words through different forms of training, however, is not yet clear.

Present Study

The present study takes a closer look at this issue by investigating subjective and objective criteria of salience separately, using modality and accent as criteria. In an exposure-test paradigm, German participants first underwent native accent training before adaptation was tested by a printed word eye-tracking

¹For example, in their account on social weighing in speech perception, Sumner et al. (2013), predict better memory for accented words—but only if the accent is socially prestigious.

task. A subjective criterion was established by having two different types of training (production and listening), while the objective criterion featured accented vs. canonical test words. Accented test words had their initial voiced bilabial or velar stop devoiced. In Experiment 1, accented words (**Palken* for *Balken*—“beam”) were presented during training and test. In Experiment 2, the same accented words were presented during training (**Palken* for *Balken*), but target items during test were canonical words that overlapped in onset with the trained accented word (*Palme*—“palm tree”). Old word pairs from the training phase as well as new word pairs that had not been included in the training list were tested. This manipulation was included to test generalization of learning, i.e., whether the accent is only learned for trained words or also for new accented words.

A subjective criterion of saliency was tested by comparing effects of individuals’ accent productions with that of listening during training. In contrast to MacLeod (2011), the current study did not manipulate training modality in mixed lists within participants but rather between participants. This permits the comparison of the magnitude of saliency based on a subjective criterion of the production modality with that of the listening modality. Individual participants are exclusively trained with one modality. If producing an item in fact constitutes a subjective criterion for saliency compared to listening to an item, production training with that item would result in greater saliency than listening training.

An objective criterion of saliency, on the other hand, was manipulated by the presence of both accented and canonical test tokens. In Experiment 1, the presentation of accented words assigned saliency to the test tokens due to their great degree of inherent distinctiveness. Effects of accent as an objective criterion have been previously shown (Cho and Feldman, 2013), but with a memory experiment in which generalization effects were not examined. In the learning phase, the accented words were embedded in a list of filler target words that featured no particular accent marker. This made the accented words distinct from the fillers. Contrarily, in the present Experiment 2, the canonical test words were expected to be less salient. Experiment 2 tested whether the saliency inherent in the learned accent can modify the processing of words that include the accent target sound in their canonical form. A difference in learning effects can be reflected in the activation differences of canonical target words starting with the manipulated accent’s target sound. Learning that *Balken* is pronounced as **Palken* potentially increases lexical competition for the canonical *Palme*, which, in turn, slows down recognition of *Palme*. This is based on Trude and Brown-Schmidt (2012), who found that accent learning can imply the inclusion of new competitors. In the present study, *Balken* could be included as a new competitor for *Palme* after training, resulting in fewer target looks to *Palme*.

The pattern of target and competitor activation is especially important during the segmental overlap of target and competitor words. Referring to the principles of an abstract mental lexicon, we assume that accent learning is based on learning pre-lexical rules. When hearing **Palken* in Experiment 1, successful word recognition requires the application of a specific accent rule (/b/ → /p/). If the accent rule is learned robustly during training, it

is applied by default as soon as the auditory input potentially matches the accent, i.e., from initial /p/ presentation onward. When, in an eye-tracking experiment, the display includes *PALME* and *BALKEN* and **Palken* is the auditory target, both *PALME* and *BALKEN* should be fixated from word onset until disambiguation (including /pal/). Only after disambiguation should *BALKEN* be fixated more often than *PALME*. If the accent rule is not learned robustly enough, the candidates that require the rule (*BALKEN*) have a weaker activation than those that do not require the rule (*PALME*). Consequently, during the overlapping word portion, *PALME* will still be more strongly activated than *BALKEN*, and *BALKEN* will only be preferred after disambiguation.

Successful recognition of a canonical word (*Palme*), as in Experiment 2, does not require accent rule application. However, successful accent learning should result in increased competitor activation of words with a /b/ in initial position. This increase in competitor activation might adversely affect canonical target activation. The rule should be applied by default as soon as the auditory input potentially matches the accent, i.e., also when *Palme* is presented. Having *PALME* and *BALKEN* on the visual display, both words should be equally fixated during /pal/. Only after disambiguation should *PALME* be preferred. The same predictions as above emerge if the accent rule is not learned strongly enough—the candidates that require the rule (*BALKEN*) are activated less strongly than canonical words (*PALME*). Consequently, *PALME* will be more strongly activated than *BALKEN* even from the beginning of word processing.

The accent in the present study is an artificial accent that centers on one specific phonological accent marker and therefore must be differentiated from a dialect. Participants and pre-recorded speakers are not L2 speakers, and all used standard German pronunciation in the experimental context. “Standard” here means that the pre-recorded speakers did not have a noticeable dialect that could allocate them to a specific region in Germany, and the participants’ speech did not include specific local (e.g., Swabian) accent properties during the experiment. The tested accent affected German stop consonants and has, to our knowledge, not been documented as an existing native accent of German. It refers to the lenis/fortis-contrast in German bilabial and velar plosives (/b, p/ and /g, k/). In Standard German, fortis plosives are always aspirated in word initial position, while lenis plosives are never aspirated (Jessen, 1998; Kleiner and Knöbl, 2015). Our accent neutralized this contrast, i.e., lenis velar and bilabial stops were aspirated (/g/ pronounced [k^h] and /b/ pronounced [p^h]: *Gitter* pronounced **Kitter*, and *Balken* pronounced **Palken*). The accented sound was always aspirated, similar to the canonically fortis stops. For simplification, we refer to aspirated, fortis plosives (*Palme*) as “voiceless” and to the lenis plosives with the additional aspiration in the accented version as “devoiced” (**Palken*).

We opted for an accent with a target sound that is included in the German sound inventory (Kohler, 1999). This makes it easy to produce for native German participants and promises relatively stable acoustic properties of the target sounds across participants. The accent under investigation has to be differentiated from middle-Bavarian dialects where

bilabial, alveolar, or velar plosives are not realized with an aspiration contrast before /r, l, n, m/; they are always voiceless and unaspirated and therefore neutralize with their lenis counterpart, e.g., *Preiselbeeren*—“cranberries” pronounced **Breiselbeeren* (Moosmüller and Ringen, 2004). Likewise, in Austrian German, the fortis plosives /p/ and /t/ are not aspirated (Siebs et al., 1969), e.g., *Pinzel*—“brush” pronounced **Binsel*. In contrast, the accent presented in this study neutralized all bilabial plosives to [p^h] and all velar plosives to [k^h]. Since the accent tested in our study describes a voicing shift in the opposing direction of existing native German accents, we can assume that none of our participants had had experience with the accent. This ensured the observation of only laboratory-specific training effects.

We predict that accent training will result in accent learning effects. The training modality can determine the amount of saliency based on one subjective criterion. This would be in line with prior findings where producing rather than listening to a word resulted in better memory (MacLeod, 2011). Saliency that relies on an objective criterion of the target token is expected to affect looking patterns such that the learned accent affects processing of highly salient, accented devoiced tokens more than that of canonical voiceless tokens.

EXPERIMENT 1

Experiment 1 tested if saliency can result from training as subjective criterion. Critical test words had a native accent and were assumed to be highly salient based on the objective criterion “accent.” During training, native German participants either read aloud or listened to single German words that had their initial /b, g/ devoiced to /p, k/, e.g., *Balken* pronounced as **Palken*, while the control group had no training.

In the test phase of the experiment, participants accomplished the printed word variant (McQueen and Viebahn, 2007; Weber et al., 2007) of the eye-tracking task (Alloppenna et al., 1998). Participants saw four printed words in their canonical spelling (including a target, a competitor, and two distractors) on a computer screen and were auditorily instructed to click on a target word while their eye movements were recorded. They listened to devoiced words (**Palken*) and had to click on a visual display that included the target word (*BALKEN*) and a competitor (*PALME*). The competitor allows the investigation of whether activation of the devoiced token can be as strong as activation of voiceless word forms without an accent. The proportion of target fixations was measured and compared between the production, listening, and control groups.

Participants

Seventy-four native German speaking female students (19–30 years, mean age = 23.8, *SD* = 2.7; 5 left-handed) from the University of Tübingen participated for a small monetary remuneration. Only women were tested in order to account for the fact that the recordings were exclusively made by female speakers. German was their only mother tongue², they did

²Fifty participants indicated dialect familiarity (42 specifically with a Southern German dialect, mostly Swabian). As most dialect speakers had exposure to a

not suffer from any hearing disorders and had normal or corrected-to-normal vision. Two participants were excluded due to unsuccessful calibration, resulting in the collection of data from 72 participants (26 production group, 22 listening group, and 24 control group).

Materials

Words during the Test Phase

We presented 92 word quadruplets during test, each containing four German nouns. Twenty-eight quadruplets were based on critical word pairs; 64 quadruplets were based on filler word pairs. The 28 critical word pairs were each composed of a target word with an initial voiced stop and a competitor word starting with the corresponding voiceless stop. Only target words were presented auditorily during the experiment. Fourteen had a bilabial onset (e.g., target *BALKEN* “beam”—competitor *PALME* “palm tree”), and 14 had a velar onset (e.g., target *GITTER* “grid”—competitor *KITTEL* “tunic”). We opted for plosives, because it has been shown that perceptual learning of plosives can generalize across speakers (Kraljic and Samuel, 2007), arguably because they contain hardly no talker-specific information in comparison to fricatives, for example. This was important because participants in the training groups were trained with a different voice than was heard during test. Voiced stops occurred only in the initial position of target words. The initial stop consonant was always followed by a vowel³. Apart from the initial consonant, target and competitor overlapped in at least two segments. When the target words were presented auditorily, the initial voiceless plosives were devoiced (*Balken* was pronounced as **Palken*), resulting in overlapping word onsets of target and competitor for at least three segments. Auditory words with the native accent (**Palken*) were never existing words of German (see **Table A1** for target-competitor pairs).

Mean log-frequencies of target words were 0.61 per million for velar stop words, 0.85 for bilabial stop words, and of competitors 0.67 for velar stop words, and 0.88 for bilabial stop words according to the CELEX word form dictionary (Baayen et al., 1995). In order to form quadruplets, each of the 28 critical target-competitor pairs was paired with two semantically unrelated distractor words that matched in frequency with the target-competitor pair. Distractor words never had a stop in initial position but could contain stop consonants in other word positions.

The 64 filler word pairs also had a target and a competitor. There were 8 targets with initial /k/, 8 with initial /p/, 16 with initial /t/, and 32 targets with no initial stop in onset position (the “no-stop targets”). For the total of 32 targets with initial /k/, /p/, and /t/, half of the competitor word onsets overlapped with the target word onset by at least three segments, and half were phonologically unrelated. Two phonologically and

Southern German dialect, the variable “Southern Dialect” was tested in initial data analyses, resulting in no significant effect. Participants were not selected based on dialect competence, and it was not counter-balanced across conditions; therefore this factor was not included in the methods section.

³In some varieties of German, speakers tend to devoice initial voiceless stops when they are followed by a liquid (e.g., *grillen* can be pronounced as *krillen*). By always having a vowel following the initial consonant, potential previous experience with the accent was avoided.

semantically unrelated distractors were added to each target-competitor pair. The 32 no-stop targets were paired with competitors that also did not have stops in initial position. However, half of them overlapped in onset with the target for at least two segments (e.g., target *Seife* “soap”—competitor *Seite* “side/page”). There were four types of distractors for the 32 no-stop target-competitor pairs, each containing eight distractor pairs. The bilabial (b/p), velar (g/k), and alveolar (d/t) distractor pairs followed the same prerequisites as the corresponding critical target-competitor pairs. As they were not presented auditorily, stop+consonant onsets (e.g., *Brosche* “brooch”—*Prospekt* “brochure/leaflet”) were allowed. The fourth group had two semantically and phonologically unrelated initial sounds that were never stops.

Altogether, the test included 92 critical trials and four practice trials. Half of the critical targets and their corresponding competitors had been included in the preceding training phase, and half were new to participants. Likewise, half of the targets not starting with a stop (other-group) were new to participants and half were familiar from the training. Every participant had her own experimental list, each starting with the same four practice trials. Filler and critical trials were equally distributed across the lists, and a critical trial was always followed by at least one filler trial. There were never more than two old and not more than five new trials in a row. The various filler conditions were equally distributed across the lists.

Words during the Training Phase

Seventy-two single words from the above described target-competitor pairs were used for training. They included half of the devoiced targets (7 targets with bilabial onset, e.g., **Palken*, and 7 targets with velar onset) and their respective competitor (*Palme* for target **Palken*). The devoiced and voiceless items were included twice in the training list, resulting in 28 devoiced and 28 voiceless trials. Additionally, 16 filler targets from the no-stop targets were included, resulting in 72 training trials in total. Training trials with the same initial sound did not occur more than twice in a row, and each devoiced item was followed by at least one canonical item.

Recordings of Test and Training Tokens

All tokens used for training and test were recorded by two female native speakers of Standard German without a noticeable regional accent (speaker A: 23 years; speaker B: 28 years). The speakers did not differ significantly in F0-range or speaking rate, and the authors judged their pronunciation to be comparable. Two different speakers were recorded to have different voices for both the training and test phases in the listening group. This permitted constant conditions across the training groups because the production condition always involved a different speaker during the training (the participant) and the test (the pre-recorded talker). Nevertheless, speaker-listener similarity was granted by having participants and pre-recorded speakers with the same sex and L1 background in the test phase. Acoustic differences between the training and test tokens are held as small as possible. Moreover, it can be tested whether speaker specific

effects, as observed by Trude and Brown-Schmidt (2012), can generalize to new speakers of the same sex (both female).

Recordings were carried out in a sound proof cabin with an Olympus LS-11 sound recorder (44.1 kHz; 16-bit). Every target word was recorded in the context of the carrier sentence *Klicken Sie jetzt auf* —“Now click on.” The devoiced target words (**Palken*) and the voiceless words (*Palme*) were all recorded naturally, that is, the speakers were explicitly instructed to pronounce the /b/ in *Balken* the way they would pronounce the /p/ in *Palme*. The best exemplar of the carrier sentence was chosen for each voice, and the duration of the carrier sentence was matched between both voices. Then, the carrier sentence was added to each target word recording.

Procedure

An SR-Research *Eyelink 1000* set-up was used for data collection with a sampling rate of 1 kHz, and the experiment was programmed with *Experiment Builder* (SR Research Ltd., Canada). Before the experiment started, the dominant eye of each participant was determined. Then, participants were seated in front of a computer screen and placed their chin on a chin rest. They were brought to a position in which they could stay comfortably for the duration of the experiment (~30 min). The eye-tracker was calibrated, then written instructions were shown on the screen. Participants had as much time as they needed to read them and initiated the experiment with a mouse click.

Training

The same training list was presented to each participant from the two training groups, while the control group received no training. The training tokens were presented either visually and auditorily (listening group) or visually only (production group). The listening group first saw a fixation cross for 1000 ms, then the orthographic transcript of the training word (black Arial font, size 24) appeared in the center of the screen. It corresponded to German spelling rules (*BALKEN*). The initial letter that corresponded to the devoiced sound was colored red. Five hundred milliseconds later, the training word was played (**Palken*). Participants listened to the single words (devoiced, voiceless, and fillers) through noise-canceling headphones (Sennheiser HD 215 II) and at the same time fixated the transcript on the screen. There was a 2000 ms inter-trial interval. Participants in the listening group were explicitly told to listen attentively to the words and to be aware of the speaker's accent while fixating the orthographic version of the words. Witteman et al. (2013) have shown that a single word context is sufficient for listeners to learn an accent. In their cross-modal priming task participants without previous accent exposure had increased priming effects in the second half of the experimental list compared to the first half.

The production group did not wear headphones during the training. They saw the same orthographic transcript of the words on the screen and had to read every single word out loud while their productions were recorded. Participants were asked to pronounce the initial red letter “B” as /p/ and the initial red letter “G” as /k/. Before every single trial, there was a fixation cross for 1000 ms, and then the word was shown for 3500 ms

(accounting for the timing in the listening condition: 500 ms before the sound + 1000 ms mean word duration + 2000 ms pause). The next trial would then start. Between training and test, the written instructions for the eye-tracking task were shown on the screen. The production group had about 5 s to put on their headphones, and the listening group waited for 5 s until the initiation of the test phase. Overall, the training took about 7 min for each participant.

Test Phase

The test phase started with four practice trials. A fixation cross preceded each trial for 1000 ms, then four printed words from a word quadruplet were shown on the screen for 500 ms. The words were printed in black Times New Roman font, size 34 on a screen with a white background. Screen resolution was 1024 × 768 pixels, and the words were distributed across four different positions (256 × 576, 768 × 576, 256 × 192, and 768 × 192 pixels), see **Figure 1**. Display positions of target and competitor were pseudo-randomized, and the target never appeared in the same display position more than three times in a row. The mouse cursor (represented by a small circle) was located in the center of the screen at the beginning of each trial. Then the carrier sentence (about 1300 ms) followed by the target word was played auditorily. Participants clicked on the target word with the mouse. Visually, participants saw the target word in its correct spelling (*BALKEN*); auditorily, it had the same accent as presented during the training phase (**Palken*). A small fixation circle appeared on the screen after every six trials to initiate an automatic drift correction in the calibration of the eye-tracker. The experiment concluded with a language background questionnaire based on the *LEAP-Q* (Marian et al., 2007).

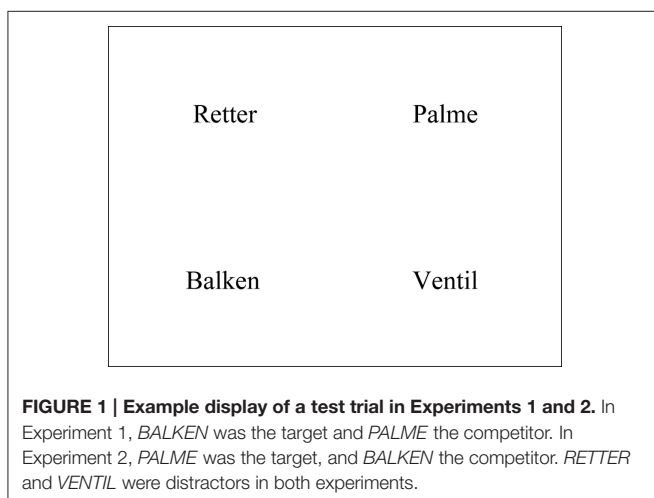
Analysis and Results

During training, the production group performed the instructed accent quite well. The experimenter decided based on perceptual judgments whether the critical training tokens were devoiced as communicated in the instructions. Every instance where the devoicing was not clearly perceivable was documented and subsequently validated by means of acoustic measurements of the

recordings. On average, only 0.7 out of 28 critical trials were not devoiced as instructed. The proportion of correct clicks on the target during the test phase was 94.3% (equally distributed across the training groups). However, five participants did not see the mouse cursor due to technical problems. We extracted fixation reports with the software *Data Viewer* (SR Research) and then further processed the data with the software *R* (R development core team, 2015). The data from each participant's dominant eye was used to determine the coordinates and timing of fixations. Only fixations that fell within a cell of one of the four interest areas—target, competitor, and two distractors—were analyzed (exclusion of 3.4% of the data). The interest areas each had a cell size of 472 × 344 pixels with a distance of 40 pixels between vertical cells and 60 pixels between horizontal cells. Saccades (20.8% of the data) were not added to fixation times. We then analyzed the fixations for the four interest areas in 20-ms steps in a time window from 0 to 1000 ms after target word onset. The dependent variable “target” indicated whether in the respective 20-ms step a participant fixated the target; “competitor” indicated a competitor fixation, and “distractor” a distractor fixation. This resulted in three variables with binary values. Target and competitor fixation proportions were calculated with the empirical logit function. The plotted fixation proportions were inspected visually to determine the critical time window to which linear mixed effects regression models (Baayen et al., 2008; Bates et al., 2015) were then applied. For each analysis we built an individual, best fitting model that included a particular choice of fixed and random factors. Random effect structure included random intercepts for participants and items as well as those random slopes that significantly improved the model fit as tested by likelihood ratio tests. Significance of factors was indicated by $t > |2|$. Corresponding p -values, as reported in the text below, were determined with likelihood ratio tests. As fixed effects, we considered training (production vs. listening vs. no training), familiarity (old, i.e., included in the training, vs. new), sound condition (bilabial vs. velar word initial sound), and speaker (speaker A vs. speaker B). Proportion of target fixations was the dependent variable.

Descriptive Analysis

Not surprisingly, the distractors were ruled out as potential target words very early by the participants (from about 200 ms, see **Figure 2**), i.e., the fixation proportion of distractors decreased very quickly. As launching a programmed eye movement usually takes about 200 ms (e.g., Altmann and Kamide, 2004), word processing is reflected in fixation proportions from this point in time on. Competitors were preferred over targets by all three groups from about 280 ms on until about 700 ms. Target fixations show that the two training groups started to fixate the target more often than the control group from about 250 ms on. The advantage of both training groups became more pronounced and started being robust from about 350 ms on. Visually, there was no difference between the production and listening groups. Statistical analyses were run for the time window 250–750 ms because it included the whole process of target-competitor disambiguation, and here it became evident that the two training groups had a stable advantage of target fixations compared to the



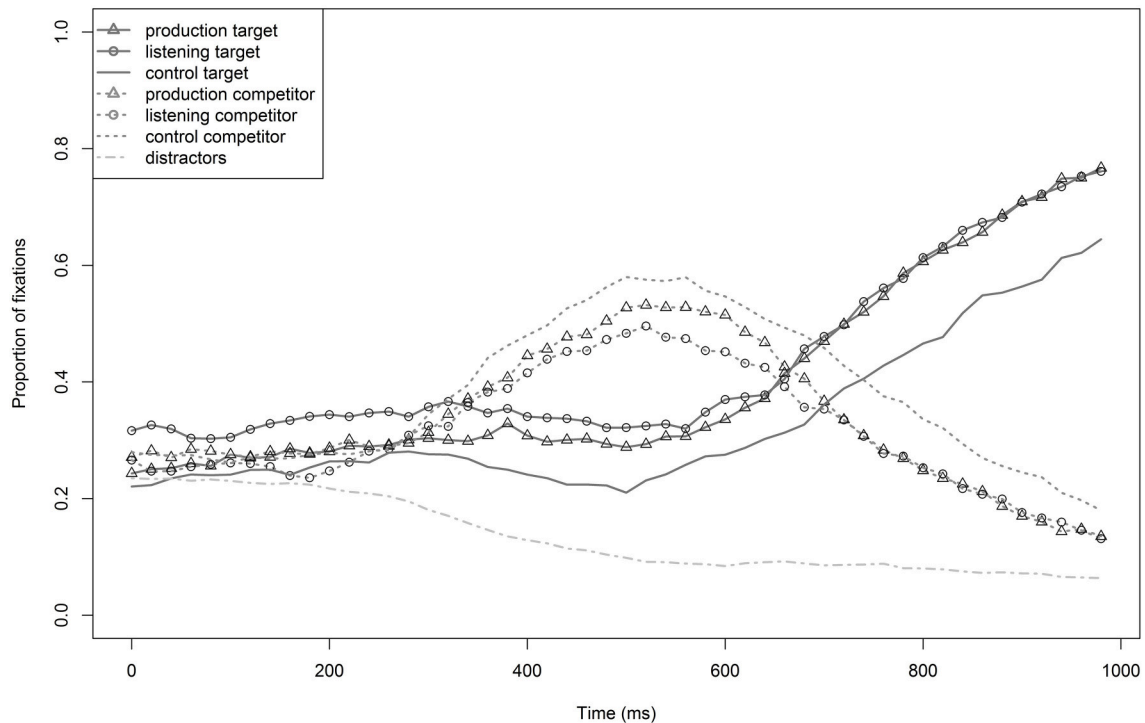


FIGURE 2 | Proportions of target (BALKEN) and competitor (PALME) fixations of the production, the listening, and the control group in Experiment 1.
The bottom line describes the mean number of distractor fixations of all three groups.

control group. As can be seen in **Figure 2**, the actual advantage lasted much longer—at least until 1000 ms.

Statistical Analysis

First, a model with data in the time window 0–200 ms was run. This tested looking biases before processing of the actual target word began. Training group was the fixed effect, and participant and item were random intercepts. There was a significant effect by training ($\chi^2 = 7.2$, $p < 0.03$); the results of the mixed model show that the listening group had more target fixations than both the control group ($\beta_{\text{training}} = 0.39$, $SE = 0.15$, $t = 2.6$) and the production group ($\beta_{\text{training}} = 0.31$, $SE = 0.15$, $t = 2.1$), hinting at a target bias for this group.

The second model analyzed data between 250–750 ms. It included training group and sound condition as fixed effects as well as participant and item as random intercepts. Training was significant ($\chi^2 = 10.7$, $p < 0.005$); both the listening group ($\beta_{\text{training}} = 0.48$, $SE = 0.15$, $t = 3.2$) and the production group ($\beta_{\text{training}} = 0.33$, $SE = 0.14$, $t = 2.3$) fixated the target more often than the control group. There was no difference between the two training groups ($t = 1.0$). Furthermore, there was a main effect of sound condition ($\chi^2 = 7.5$, $p < 0.007$), resulting in more target fixations for bilabial than velar items ($\beta_{\text{condition}} = 0.35$, $SE = 0.13$, $t = 2.8$). Due to the bias for the listening group found from 0–200 ms, the critical time window was further examined. On average, from 0–200 ms the proportion of target fixations was 8% higher for the listening group than for the control group. To account for this early bias, we subtracted

8% from listening group data between 250–750 ms and re-ran the same model with the modified data. Despite the reduction of the listening group's target fixation data, training was still significant ($\chi^2 = 6.2$, $p < 0.05$): the listening group still fixated the target more often than the control group ($\beta_{\text{training}} = 0.30$, $SE = 0.15$, $t = 2.0$), and there was no difference between the production and listening groups ($t = 0.2$). This suggests robust differences between the control group and the two training groups.

Discussion

We found that accent adaptation was possible after both listening and production training. The proportion of target looks in both training groups was higher than in the control group. The listening group, however, fixated targets more often than the other groups, even before actual target word processing began, which might be argued to have affected the listening group advantage in the subsequent critical time window. This, however, can be excluded because the pattern of results persisted even when the fixation data of the listening group in the larger, later time window were penalized for its advantage in the initial, smaller time window. There were no effects of speaker, i.e., learning occurred equally well with speaker A and B. The main effect for sound condition may be related to specific sound properties but does not further affect the general pattern of results. Moreover, the same pattern was observed for old tokens from the training phase and new tokens, indicating learning of a rule that generalizes to new words.

Our results suggest accent learning for the production and listening groups, with no difference between the two training groups. Thus, we found robust effects of accent training when testing single accented words, hinting at a great effect by target words' accent as objective criterion of saliency. Production and listening training seemingly do not differ from one another for L1 in terms of saliency.

Experiment 1 provides evidence for successful accent adaptation after listening to or producing an accent. However, the canonical competitors (*Palme*) were activated for a very long time (until about 700 ms) before the devoiced target word was fixated more often. This time window covers the entire initial portion of the word before disambiguation (average disambiguation point: $280 + 200$ ms for launching the eye movement = 480 ms; earliest disambiguation point: $150 + 200$ ms = 350 ms; longest disambiguation point: $420 + 200$ ms = 620 ms) and even longer. This suggests that, despite successful accent adaptation, canonical word forms still remained more easily accessible than accented word forms. There was potentially not enough accent exposure for the accented forms to be able to fully compete with canonical word forms. We suggest that if a learned accent were to be able to have effects on the access of canonical, voiceless words with the same onset as the accent's target form, a greater amount of training is required.

Experiment 2 examines whether accent learning can be strong enough as to affect the processing of voiceless, canonical words with double the amount of accent training. Successful accent learning could imply competition effects from words that were previously not included as competitors. Thus, accent training has potentially effects not only for understanding accented word forms, but accented forms can function as competitors and affect the recognition of canonical word forms. As opposed to Experiment 1 where highly salient accented target words were tested, in Experiment 2, we focused on test words that are expected to have a much smaller degree of saliency based on the objective criterion "accent," i.e., standard German canonical words. Training effects of devoiced words (**Palken*) were tested on words that canonically start with the accent's target sound (*Palme*). In order to increase the likelihood that accented forms could influence target recognition in their function as competitors, the training was doubled. If the accent is robustly learned, we would expect fewer target fixations by the training groups than without accent training.

EXPERIMENT 2

Again, three participant groups were tested. The training involved the same tokens as in Experiment 1, but the amount of training with the devoiced tokens was doubled. During test, participants did not hear the devoiced words (**Palken*) this time, but voiceless, canonical words (*Palme*), while seeing the same printed words on the display as the participants from Experiment 1.

Participants

Seventy-eight female students from the University of Tübingen participated for monetary reimbursement. Six had to be excluded

due to calibration problems, resulting in 72 participants (18–31 years, mean = 23.2, $SD = 3.2$; 14 left-handed) who successfully completed the experiment. None of them suffered from any hearing disorders, all had normal or corrected-to-normal vision, and German was their only mother tongue⁴. The participants were randomly assigned to one of the three experimental groups (24 production, 24 listening, and 24 control group).

Methods and Material

The training list was based on that of Experiment 1. However, devoiced (**Palken*) items were presented twice in a row (rather than just once), resulting in 100 training trials in total (twice the amount of training with the devoiced tokens compared to Experiment 1). Due to the greater amount of training, the training phase took 1 min longer.

During test, the same word quadruplets were presented on the screen—92 critical trials and 4 practice trials with the same properties as in Experiment 1. However, the roles of target and competitor words were switched. Targets were now voiceless tokens (*Palme*) in their canonical form, and competitors were words that have a voiced onset in their canonical form (*Balken*). Auditorily, voiceless words were presented (*Palme*) that matched in their onset with the target word on the screen (*PALME*). Voiced tokens (*BALKEN*) that had been devoiced during the training (**Palken*), were visually presented competitors. All target words had already been recorded in the recording session for Experiment 1 by the same female speakers.

Analysis and Results

The same procedure for analysis as in Experiment 1 was applied. During training, the production group performed quite well in accomplishing the substitutions (mean: 0.8 errors out of 56 devoiced word trials). The accuracy of clicks during the test phase was 99.8% (equally distributed across training groups). Saccades (17% of the data) and fixations that did not fall into one of the four interest areas (3%) were removed prior to analysis.

Descriptive Analysis

Figure 3 illustrates the proportions of target, competitor, and distractor fixations of the production, listening, and control groups. The distractors were ruled out from the beginning of word processing (200 ms) on, meaning that the proportion of fixations decreased. Target (*PALME*) preference started very early (at about 250 ms), and competitors (*BALKEN*) were quickly ruled out as potential target words. The competitors stayed at a relatively stable level of activation from 200–400 ms, and then fixations decreased noticeably. This represents approximately the interval where target and competitor still overlap (mean overlap: 273 ms). Target fixations by training group did not differ from one another from the beginning until the overall mean end of word processing (measurements of the voiceless target words resulted in a mean word duration of 767 ms). Disambiguation between targets and competitor occurred relatively early, and there was no clear advantage of one of the training groups in

⁴Fifty-three participants indicated dialect familiarity, 50 of whom had exposure to a Southern German dialect (mostly Swabian). The variable "Southern Dialect" was tested in initial data analyses, resulting in no significant effect, as in Experiment 1.

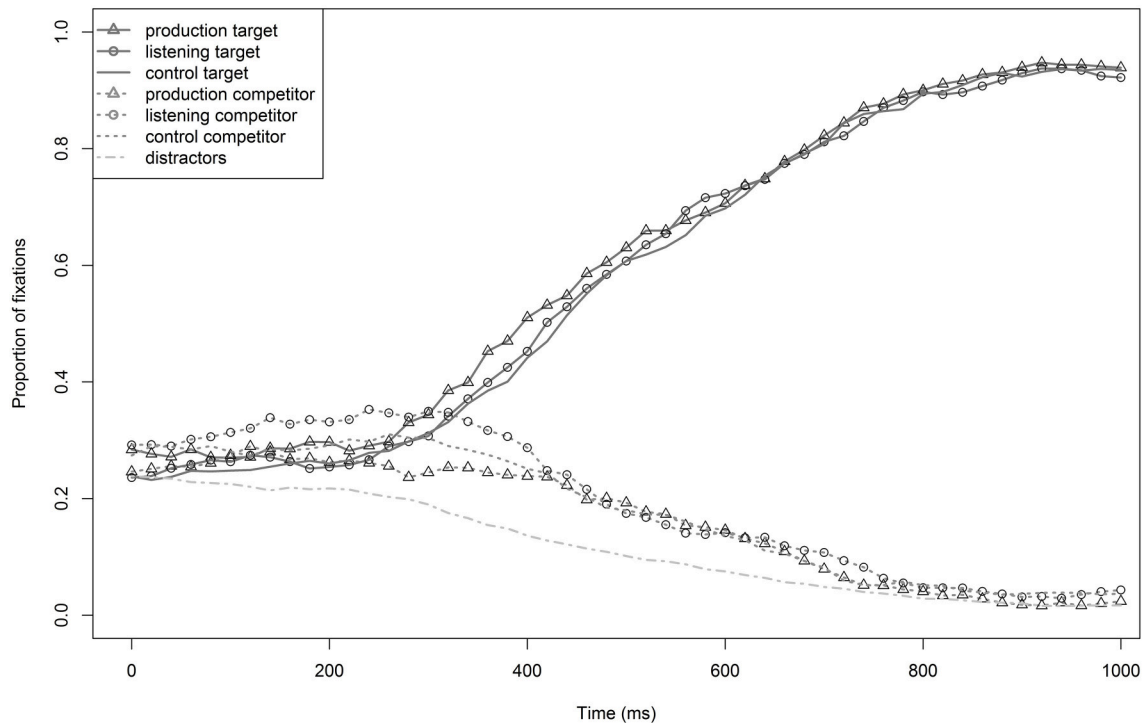


FIGURE 3 | Proportions of target (*PALME*) and competitor (*BALKEN*) fixations of the production, listening, and control groups in Experiment 2. The bottom line describes the mean number of distractor fixations of all three groups.

target fixations. Statistical analyses were run for the time window from 250–750 ms, which included the entire disambiguation process between targets and competitors and parallels analyses in Experiment 1.

Statistical Analysis

The baseline model for target fixations (0–200 ms) revealed no significant effect by training ($t < 1$). Mixed effects models revealed no significant effect of any of the considered fixed effects (all $t < |1.3|$) in the critical time window (250–750 ms). Auditorily presented voiceless words (*Palme*) that start with the same onset as the trained, devoiced words (**Palken*) triggered strong target activation from the beginning of word processing on. There was no effect of learning, neither by the production nor the listening group. In contrast to Experiment 1, the test words did not have the critical accent, but the voiceless paired words with the same sound onset as the devoiced, accented words were tested. As the devoiced training words included a sound substitution, the question is, especially for the production group: How much did the acoustic realizations of the devoiced tokens encountered during training differ from those of the voiceless tokens encountered during test? In other words, did the participants' own productions of the accent differ enough from the productions of the test speaker to prevent generalization across speakers? The missing training effect for both groups reinforces the question of effects of single tokens' acoustic properties. Therefore, acoustic properties of both the training materials and the test materials were analyzed in a next step.

Acoustic Analyses

Pre-recorded target and training stimuli as well as the tokens produced by the production group during training were analyzed acoustically. This tested if the difference between training and test stimuli was too great for adaptation effects to be observed. Particularly in the production group, the acoustic properties of the accented plosives were likely to vary individually. The stops that mark the manipulated accent were focused on in the analysis. Voice onset time (VOT) and burst intensity (relative to total word intensity) were measured for each token that was part of an old critical word pair, i.e., word pairs that were included in both the training and the test phase. Only old word pairs were included in analyses, because there was no reference to the training phase for new words. Each critical voiceless word (*Palme*) and its devoiced paired word (**Palken*) was considered for analysis. Both instances were taken as separate reference points in order to calculate the differences of the respective acoustic property value between the training and the test token (*Palme*). In the following, we refer to the *Palme-Palme* comparison as the voiceless word pair and the **Palken-Palme* comparison as the devoiced word pair. During training, one word was presented several times (devoiced: four times, voiceless: twice). This did not pose any problem for the listening group items because the same recording was presented several times. For the production group, however, single tokens differed individually. This issue was solved by taking average values. Two VOT- and two burst intensity difference values were assigned to each critical word for each participant—one with the values from the voiceless word as a reference point (*Palme*) and

one with the values from the devoiced word as a reference point (**Palken*). Voices differed between training and test in both the listening and the production condition, so minor differences were inevitable.

First, we compared the absolute training-test differences of the acoustic properties of the initial stops [i.e., $\text{dif}(\text{stop value}) = |\text{stop value}_{\text{Test(Palme)}} - \text{stop value}_{\text{Training(*Palken or Palme)}}|$]. Measurements for all old word pairs were made for VOT (min = 0.14 ms, max = 71.8 ms, mean = 19.9 ms, $SD = 16.6$) and intensity ratio of the burst (min = 0, max = 0.35, mean = 0.08, $SD = 0.06$). These values were compared between the listening and production groups as well as between the devoiced and voiceless training words that included a stop.

For each VOT difference and burst intensity difference mixed effects models were run. Each model included the acoustic variable of interest as the dependent variable. Training (listening vs. production) and word pair (devoiced vs. voiceless) were considered fixed effects, and participant and item were random effects. The model for VOT differences also included by-participant random slopes for training and word pair, as well as by-item random slopes for training. None of the factors resulted in significant effects for VOT difference (all $t < |0.7|$). The model for burst intensity included by-subject random slopes for word pair and by-item random slopes for training. There was a significant interaction between training and word pair ($\chi^2 = 5.6$, $p < 0.02$) illustrated in **Figure 4**. Examining the results of the mixed model (see **Table 1**), this interaction is based on the smaller burst intensity difference for devoiced word pairs in the production group than the listening group ($t = -2.15$), and there was no difference for voiceless word pairs between training groups ($t = 1.23$). Within training groups, there was no training-test difference between devoiced and voiceless word pairs ($t < 1.8$).

Discussion

Neither training group fixated the target less often than the control group without training did. They did not differ from one another in their amount of target fixations. The recognition of

voiceless *Palme* was not affected by previously having learned that *Balken* is pronounced as **Palken*. This occurred despite the fact that accent training was intensified by presenting devoiced tokens twice as often as in Experiment 1. This is good news for native accent listeners, because it shows that learning a new accent does not immediately distort comprehension of canonical forms. Concrete acoustic analyses tested whether this effect was due to greater inherent saliency based on an objective criterion of devoiced (as tested in Experiment 1) compared to voiceless tokens or rather because of greater acoustic differences between the devoiced training and the voiceless test tokens. There was no VOT difference between training groups, thus the production group was quite good at accomplishing the substitutions. The few production errors that occurred did not affect the overall pattern. This was also supported by the observation that burst intensity differences were even smaller for the production group than the listening group.

GENERAL DISCUSSION

The present study investigated whether different forms of native accent training and different token realizations (unaccented vs. accented) differ in saliency for L1 participants. This was measured by the amount of adaptation to the native accent. As a subjective criterion of saliency, the training phase was varied by having production and listening accent training (vs. no training), and an objective criterion was tested by the nature of the test tokens (accented/devoiced words in Experiment 1 vs. canonical/voiceless words in Experiment 2). In Experiment 1, native German participants produced or listened to single German words that featured the devoicing of initial voiced stops (*Balken* pronounced as **Palken*). In the subsequent eye-tracking task, participants from both training groups fixated the

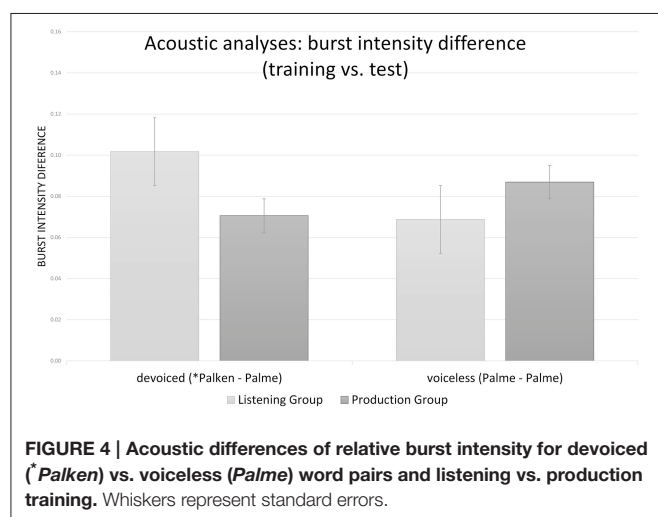


TABLE 1 | Results for burst intensity differences between training and test words as calculated by the model lmer (burst difference~word pair*training + (1 + word pair|participant) + (1 + training|item)).

Predictor	β	SE	t
Intercept (devoiced, listening)	0.10	0.01	7.51
word pair = voiceless	-0.03	0.02	-1.75
training = production	-0.03	0.01	-2.15
word pair*training	0.05	0.02	2.49
Intercept (voiceless, listening)	0.07	0.01	5.00
word pair = devoiced	0.03	0.02	1.75
training = production	0.02	0.01	1.23
word pair*training	-0.05	0.02	-2.49
Intercept (devoiced, production)	0.07	0.01	9.31
word pair = voiceless	0.01	0.01	1.61
training = listening	0.03	0.01	2.15
word pair*training	-0.05	0.02	-2.49

The factors were relevelled in order to calculate the model with different intercepts. This allows displaying t-values for all relevant level comparisons. β = Estimate, SE = Standard Error. Bold levels are significantly different from the intercept ($t > |2|$).

devoiced target more often than participants without training did, with no difference between the two training groups. This was true whether the accented target word had been included in the preceding training or if it was presented for the first time. Experiment 2 started with the same accent training and in the test standard German canonical words with the same onset as the devoiced tokens (*Palme*) were targets. The proportion of target looks was not affected by training. Acoustic analyses showed that devoiced training words (**Palken*) and voiceless test words (*Palme*) did not differ strongly in their onset.

Saliency and Adaptation

In Experiment 1, there were significantly more looks to devoiced targets after production and listening accent training than without training. In Experiment 2, which featured voiceless target words, target looks did not reveal accent adaptation. This can be explained by the role of saliency in accent adaptation. Two criteria for saliency were manipulated and tested in our study. First, an objective criterion was tested through the nature of the test tokens (accented/devoiced test words in Experiment 1, canonical/voiceless test words in Experiment 2). The devoiced test words were predicted to be more salient than the voiceless words, thereby resulting in greater adaptation effects. Second, a subjective criterion was tested by having an accent training session based on different modalities (production and listening).

In terms of the objective criterion, adaptation only showed effects for devoiced, and not for voiceless, target words that had the same word onset (**Palken* vs. *Palme*). This suggests that devoiced tokens are more salient than voiceless tokens. Acoustic analyses of Experiment 2 support our interpretation. There was no evident acoustic difference between devoiced and voiceless word onsets that could have inhibited learning. Training was still effective, though not visible, because the test tokens were not as salient as in Experiment 1. Test tokens in Experiment 1 and 2 therefore only differ perceptually from their disambiguation point onward (after /pal/ for **Palken* and *Palme*). This implies that training effects emerged only in later stages of processing, after word disambiguation. This is supported by the analysis in Experiment 1, where the training group advantage admittedly was already detectable from about 250 ms on (see **Figure 2**); however, the plot of fixation proportions suggests that the two training groups' advantage increased over time and became stable from about 350 ms on. The shortest duration of the ambiguous word section (i.e., overlapping with the competitor) measured approximately 150 ms in Experiment 1 (*ger* in *Germane* "Teuton," speaker A). The moment where the information after the disambiguation point is processed is then reflected from about 350 ms after the stimulus onset onward (150 + 200 ms eye movement launching). Cho and Feldman (2013) found a memory advantage for accented compared to canonical words. They argue that accented speech is more variable in terms of acoustic and phonetic detail, and, based on an episodic account of the mental lexicon, they suggest that difference between accented speech input and stored exemplars is greater than the difference between unaccented input and stored exemplars. Accordingly, this greater difference enriches the form-meaning relationship. This reasoning essentially follows the same principles as the

distinctiveness account of saliency. More distinct tokens are more salient, which results in memory advantages. It can be argued that saliency of accented tokens in the present study was artificially increased by the fact that there was only one specific accent marker and no more natural, global accent. However, a cross-modal priming study by Eisner et al. (2013) found that L1 English listeners adapt to final devoicing in English (*seed*, pronounced [si:t^h]) when it was produced either by a native British English speaker or by a native Dutch speaker with L2 English (with global Dutch accent features). Moreover, the findings from the Cho and Feldman study are in line with ours. They incorporated a global accent (Dutch-accented English) and still found a memory advantage of accented over canonical tokens.

A subjective criterion of saliency, on the other hand, was implemented through the training session. The production group was compared to the listening group as well as the control group without training. Accent adaptation worked equally well with both listening and production training in Experiment 1 (target **Palken*), and effects were not visible with voiceless (*Palme*) targets in Experiment 2. There was no difference between the two training groups in either experiment. This suggests that both production and listening accent training imply a similar amount of saliency in the fostering of accent adaptation, and adaptation effects become visible only when the test token receives sufficient saliency through an objective criterion.

Interestingly, we found that in L1, saliency elicited by the subjective criterion of producing an accent was as large as that of listening to the accent. In a previous study (Grohe and Weber, in press), the effects of production vs. listening training on accent adaptation were tested for both L1 and L2 participants. L2 participants adapted to the accent most easily with production training. L1 participants did not adapt, neither with listening nor production training. Importantly, all speech in the present study was produced by L1 speakers, but in Grohe and Weber, test items were always produced by an L2 speaker of English. Thus, for L1 participants in the production training group there was a switch in nativeness of the speaker between training (L1) and test (L2). L2 speakers likely involve a greater amount of variability (Wade et al., 2007) in their speech than L1 speakers, including more accent markers which probably require additional processing. Moreover, the similarity between listener and speaker is emphasized by the *integrated theory of language comprehension and production* (Pickering and Garrod, 2013), according to which a listener's previous production experience can affect comprehension. This experience is predicted to have greater effects with increasing speaker-listener similarity. The present results, however, do not necessarily support this suggestion. In spite of greater speaker-listener similarity (same sex, same L1 background, mostly similar dialects), the production group did not have greater training effects than the listening group. Nevertheless, having an L1, not L2, speaker produce the accent helped L1 participants to adapt to an accent after both listening and production training. Contrary to L2 participants in Grohe and Weber, however, accent adaptation was not stronger after production training. Producing an accent is only a more important subjective criterion of saliency than listening, because of specific L2 properties (e.g., greater perceptual flexibility). There

is no general advantage exhibited by producing compared to listening.

Taken together, there was arguably no advantage of production over listening training for L1 listeners, because production might only make a linguistic token more salient if it can act as objective, not subjective, criterion of saliency. This would additionally include that the concrete situation determines saliency. Furthermore, the studies that have found robust production effects (MacLeod, 2011; Cho and Feldman, 2013) were all memory studies that tested active and conscious word recall, thus later stages of processing. Contrarily, the present eye-tracking study tested online word processing. It is therefore also possible that the production advantage may not arise in the earliest stages of processing. Other studies conducted a repetition experiment rather than a listening-only task as we did (e.g., Cho and Feldman; Kartushina et al., 2015). Repetition includes listening and producing the critical token, possibly implying a greater amount of saliency than only production. Finally, concrete feedback may have affected the results of the study by Kartushina and colleagues.

Referring again to the definition of saliency established in the beginning of this article, MacLeod (2011) suggests that for mixed lists (including items both listened to and produced), produced items are more distinct and therefore more salient. This kind of saliency likely relies on an objective criterion—the stimulus itself attracts attention because of its distinct physical characteristics. In the present study, on the other hand, it was asked if the nature of training (production vs. listening) could act as a subjective criterion of saliency. Our results do not support a production advantage *per se*, but they also do not exclude the possibility of a production advantage. The production advantage may function within the scope of saliency that relies on an objective, but not subjective, criterion, even with L1 participants and in an online task. In summary, we have found saliency effects that rely on an objective criterion and no effects that rely on a subjective criterion. Previous studies that support the production effect have always tested saliency arising from an objective criterion. We hypothesize that the nature of saliency is the crucial factor in the adaptation process and that, in short-term adaptation, objective criteria are more powerful than subjective criteria.

This contrasts at first sight with findings on dialect accommodation by Auer et al. (1998), who emphasize the importance of subjective criteria of saliency. Note, however, that the researchers refer to change in production over the long term rather than to comprehension in the short term, as was tested in the present study. Therefore, different criteria of saliency might result in saliency that is most efficient at different stages of adaptation and in different modalities. On the other hand, these results are good news regarding short-term comprehension adaptation in language change contexts. These contexts mostly involve new and old pronunciation variants, resulting in contrasts between the two. This provides well-suited conditions for an objective criterion of saliency in terms of contrasts in phonetic realizations. Adaptation will be easier in contact situations than it would be in potential accent-only situations. At the same time, as accent comprehension improves, comprehension abilities of the canonical pronunciation are not

impaired. If we apply our results and those from Grohe and Weber (in press) to concrete L2 learning situations, we can conclude that for learning new variations, L2 learners, thanks to their greater cognitive flexibility, can still achieve reasonable results without switching between production and listening. It would, however, probably be even more beneficial to integrate variation between the two modalities.

Competitor Inclusion As a Further Step in Adaptation

Adaptation was observed not only for old words that had been part of the training phase; it also generalized to new words with the same accent and furthermore from the voice of the training speaker to the unfamiliar voice of the test speaker. This finding supports abstractionist accounts of the mental lexicon (McClelland and Elman, 1986; Norris, 1994) rather than episodic accounts (e.g., Goldinger, 1998). Whereas episodic accounts suggest the storage of every concrete exemplar of a speech unit encountered by a listener (including speaker-inherent details, e.g., voice properties), in abstractionist models, abstract representations of a word's canonical representation build the lexicon. Variations of the canonical form, such as accents, can be accounted for by pre-lexical mapping rules. These rules are built on the basis of a few exemplars that are no longer stored. When, for example, an accented token is encountered after accent training, the learned rule is applied to the respective abstract entry in the lexicon. This can explain why learning a specific variation can generalize across many different words (McQueen et al., 2006). However, we do not want to rule out the existence of exemplars in the lexicon. Hybrid models (e.g., McLennan et al., 2003) attempt to integrate exemplars and pre-lexical rules into a single account. In contrast to Bradlow and Bent (2008) and Sidaras et al. (2009), who observed speaker generalization only if training was conducted with multiple voices, one voice was sufficient for generalization in Experiment 1. The globally accented speakers in those studies likely featured many different accent markers, resulting in a stronger accent than the accent we presented. With only few accent markers, it is easy to build pre-lexical accent rules allowing for generalization to new talkers. With many different accent markers, however, multiple exemplars from multiple talkers might help successful rule-building as argued by Bradlow and Bent.

Moreover, Trude and Brown-Schmidt (2012) tested competitor exclusion and inclusion and found talker-specific adaptation effects. Competitor exclusion and inclusion describes modifications in the cohort of words initially activated when a word starts to be processed. Potential candidates can be excluded, or new candidates can be added (competitor inclusion). Effects of accent training on competitor activation are indirect training effects—the effects of the accent on other words (presented as targets) are then tested. These tokens seem less salient than accented tokens. It seems that if less salient targets are tested, the role of aspects such as talker specificity increases. In other cases, these aspects may be training intensity or prior accent familiarity, as shown in Trude et al. (2013). The design of their study was similar to the eye-tracking study discussed above

(Trude and Brown-Schmidt, 2012). Talker-specific effects of accent learning on competitor exclusion were again tested, but this time with a Québec-French accent that participants had never been exposed to before the experiment. The talker replaced every /i/ with an /I/ in English words, i.e., *weak* became *wick*. An accent training session did not help participants rule out unlearned competitors more easily if pronounced by the accented talker than the unaccented talker. As suggested by Trude et al. (2013), competitor exclusion failed seemingly due to the accent being completely new to participants. Considering the small amount of target word saliency, more previous accent exposure (as shown in Trude and Brown-Schmidt, 2012) or greater training intensity could have helped. This interpretation is also supported by Experiment 2 of the present study. In contrast to the accented, devoiced targets from Experiment 1, the canonical, voiceless targets in Experiment 2 implied smaller overall objectively induced saliency. Additionally, the accent was completely new to the participants.

We found that after a few minutes of training, an accent can be learned so that it is more easily processed than without training. Only highly salient target tokens made learning effects visible. Therefore, accent training does not always exhibit robust accent learning. As shown by Trude and Brown-Schmidt (2012), this does not mean that more robust accent learning is not impossible. They found effects of both competitor exclusion and inclusion when non-salient target tokens were tested. The effect was talker-specific, and the participants already had prior (pre-experimental) experience with the accent. Accent adaptation seems to occur in various steps, ranging from unadapted to partially adapted (effects can be observed for accented, salient words) all the way to fully adapted (effects can be observed for unaccented, non-salient words). Full accent adaptation would mean that the way that accented word forms function as competitors is similar to the functioning of canonical word forms. However, the amount of looks to the targets in Experiment 2 was the same with and without training, indicating that full adaptation had not occurred. It likely requires more intense training, pre-experimental accent familiarity, identical talkers during training and test, or even multiple talkers during training (Bradlow and Bent, 2008). The adaptation effects that we found seem to reflect partial accent adaptation, which is still important because it allows a listener to better understand the accented form itself. The reason why we did not find full adaptation can also lie in the native language background of our listeners. Bent and Bradlow (2003) found that non-native listeners performed equally well in a sentence recognition task while listening to a speaker with the same L1 as when listening to a native speaker. This advantage has even generalized to unrelated accents that were new to the listener. Native listeners, on the other hand, as shown in a training-test study by Baese-Berk et al. (2013), are only able to generalize accent learning to a new accent if they are trained on many different accents. This finding is in line with the results on generalization of accent learning to new voices by Bradlow and Bent (2008).

Basic assumptions from abstractionist accounts on lexical processing support our conclusion that the accent rule was not learned strongly enough to be applied to all tokens from

word onset onward. In Experiment 1, the voiceless competitors (*PALME*) of the target word **Palken* were considered as potential candidates for a long period, and only after disambiguation was the target *BALKEN* fixated more often than the competitor. With the auditory target *Palme* in Experiment 2, the pre-lexical rule was not learned strongly enough to establish additional competition by *BALKEN* during the /pal/-segment. One could assume that the results of Experiment 2 are due to increased competitor (*BALKEN*) activation. Participants learned that *Balken* becomes **Palken*, so they might have concluded that *Palme* becomes **Balme*. This is rather unlikely, however, because the training also included canonical words starting with the voiceless sound (*Palme*). Therefore, when hearing *Palme*, they did not interpret the word input as **Balme* and thus did not fixate the competitor more often than the target.

Native and Foreign Accent Adaptation

In our discussion, we included studies that tested adaptation to native accents as well as studies on adaptation to foreign accents. Research on foreign accent adaptation clearly shows that in their L1, listeners quickly adapt to foreign accents produced by L2 speakers and maintain long-lasting processing advantages (e.g., Clarke and Garrett, 2004; Maye et al., 2008; Witteman et al., 2013, 2015). Similar findings arise from native accent studies (Trude and Brown-Schmidt, 2012). It is therefore possible that a dichotomy between native and foreign accents is unjustified. Similar mechanisms could apply to both native and foreign accent processing. Clarke and Garrett (2004) suggested an accent processing classification that depend on the accent's acoustic distance from native speech. Foreign and native accents follow the same principles, but the strength of an accent could determine the nature of accent adaptation. Arguably, native accents can be closer to standard native speech than foreign accents. Processing of regional and foreign accents could then rely on similar mechanisms, but stronger accents induce greater processing effort than mild accents do. As a consequence, adaptation to regional accents would be easier than adaptation to foreign accents. This account would explain why, on the one hand, similar results were found if the same accent was produced by an L2 or L1 talker (Trude et al., 2013), and, on the other hand, greater processing difficulties were found for foreign than for native accents (Flocchia et al., 2006, 2009). Likewise, we found adaptation for L1 participants when an L1 speaker produced the contrived accent in the present study, but in a previous study (Grohe and Weber, in press), adaptation was not found when an L2 speaker produced the accent. We suggest that accent strength is very likely linked with the amount of different accent markers that a speaker produces, which varies among individuals. Some L2 talkers do not exhibit many accent features, whereas others do. Therefore, concrete acoustic features could be an important variable which the magnitude of accent learning depends upon.

CONCLUSION

In conclusion, our study suggests that native accent adaptation can be fast and easy, including generalization to new voices and new lexical tokens as well as learning through individual

production. However, the accent requires saliency that relies on an objective criterion during test in order to display its adaptation effects. The strength of accent learning is therefore limited; an accent is not learned well enough to affect the processing of other, non-salient canonical tokens. It is not integrated as strongly into the lexicon as canonical tokens. Learning was not affected by our training manipulation, which relied on a subjective criterion of saliency. There are, however, studies that have found an advantage of production over listening when training functioned as objective criterion of saliency. We therefore conclude that in short-term accent adaptation listeners might be more sensitive to objective than to subjective criteria of saliency.

ETHIC STATEMENT

This study was carried out in accordance with the recommendations for ethical guidelines of the English

Department (Psycholinguistics and Applied Language Studies), University of Tübingen, Germany. All participants gave informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contribution to the work and approved it for publication by agreeing to be accountable for all aspects of the work. Concretely, AG and AW developed the conception of the study based on prior findings and created the experimental design, discussed stimuli, interpreted results and revised prior versions of the manuscript together. Moreover, AG compiled and recorded experimental stimuli, ran participants, processed the data, conducted statistical analyses, and wrote the first draft of the manuscript.

REFERENCES

- Adank, P., Evans, B. G., Stuart-Smith, J., and Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 520–529. doi: 10.1037/a0013552
- Adank, P., Hagoort, P., and Bekkering, H. (2010). Imitation improves language comprehension. *Psychol. Sci.* 21, 1903–1909. doi: 10.1177/0956797610389192
- Adank, P., and McQueen, J. M. (2007). “The effect of an unfamiliar regional accent on spoken word comprehension,” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, eds J. Trouvain and W. J. Barry (Dudweiler: Pirrot), 1925–1928.
- Alloppenna, 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
- Altmann, G. T. M., and Kamide, Y. (2004). “Now you see it, now you don’t: mediating the mapping between language and the visual world,” in *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, eds J. M. Henderson and F. Ferreira (New York, NY: Psychology Press), 347–386.
- Auer, P., Barden, B., and Grosskopf, B. (1998). Subjective and objective parameters determining ‘saliency’ in long-term dialect accommodation. *J. Sociolinguist.* 2, 163–187. doi: 10.1111/1467-9481.00039
- 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
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database (Release 2) [CD-ROM]*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- Baese-Berk, M. M., Bradlow, A. R., and Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *J. Acoust. Soc. Am.* 133, EL174. doi: 10.1121/1.4789864
- Baese-Berk, M. M., and Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *J. Mem. Lang.* 89, 23–36. doi: 10.1016/j.jml.2015.10.008
- Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bent, T., and Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *J. Acoust. Soc. Am.* 114, 1600–1610. doi: 10.1121/1.1603234
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106, 707–729. doi: 10.1016/j.cognition.2007.04.005
- Cho, K. W., and Feldman, L. B. (2013). Production and accent affect memory. *Ment. Lex.* 8, 295–319. doi: 10.1075/ml.8.3.02cho
- Clarke, C. M., and Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J. Acoust. Soc. Am.* 116, 3647–3658. doi: 10.1121/1.1815131
- Eisner, F., Melinger, A., and Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Front. Psychol.* 4:148. doi: 10.3389/fpsyg.2013.00148
- Evans, B. G., and Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *J. Acoust. Soc. Am.* 121, 3814–3826. doi: 10.1121/1.2722209
- Floccia, C., Butler, J., Goslin, J., and Ellis, L. (2009). Regional and foreign accent processing in english: can listeners adapt? *J. Psycholinguist. Res.* 38, 379–412. doi: 10.1007/s10936-008-9097-8
- Floccia, C., Goslin, J., Girard, F., and Konopczynski, G. (2006). Does a regional accent perturb speech processing? *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1276–1293. doi: 10.1037/0096-1523.32.5.1276
- Geraci, L., and Manzano, I. (2010). Distinctive items are salient during encoding: delayed judgements of learning predict the isolation effect. *Q. J. Exp. Psychol. (Hove)*. 63, 50–64. doi: 10.1080/17470210902790161
- Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdiscipl. Rev. Cogn. Sci.* 1, 586–596. doi: 10.1002/wcs.55
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychol. Rev.* 105, 251–279. doi: 10.1037/0033-295X.105.2.251
- Grohe, A. K., and Weber, A. (in press). Learning to comprehend foreign-accented speech by means of production and listening training. *Lang. Learn.*
- Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German. Studies in Functional and Structural Linguistics*, Vol. 44. Amsterdam: J. Benjamins.
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., and Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *J. Acoust. Soc. Am.* 138, 817–832. doi: 10.1121/1.4926561
- Kerswill, P., and Williams, A. (2002). “‘Saliency’ as an explanatory factor in language change: evidence from dialect levelling in urban England,” in *Language Change: the Interplay of Internal, External and Extra-Linguistic Factors*, eds M. C. Jones and E. Esch (Berlin; New York, NY: Walter de Gruyter), 81–110.
- Kleiner, S., and Knöbl, R. (2015). *Duden, das Aussprachewörterbuch* (6., überarbeitete und aktualisierte Auflage). Der Duden in zwölf Bänden, Vol. 6. Berlin: Dudenverlag.
- Kohler, K. (1999). “Illustrations of the IPA: German,” in *Handbook of the International Phonetic Association*, ed International Phonetic Association (Cambridge: Cambridge University Press), 86–89.
- Kraljic, T., and Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *J. Mem. Lang.* 56, 1–15. doi: 10.1016/j.jml.2006.07.010
- MacLeod, C. M. (2011). I said, you said: the production effect gets personal. *Psychon. Bull. Rev.* 18, 1197–1202. doi: 10.3758/s13423-011-0168-8

- MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., and Ozubko, J. D. (2010). The production effect: delineation of a phenomenon. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 671. doi: 10.1037/a0018785
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50, 940. doi: 10.1044/1092-4388(2007/067)
- Maye, J., Aslin, R. N., and Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cogn. Sci.* 32, 543–562. doi: 10.1080/03640210802035357
- McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0
- McDaniel, M. A., and Geraci, L. (2006). “Encoding and retrieval processes in distinctiveness effects: Toward an integrative framework,” in *Distinctiveness and Memory*, eds R. R. Hunt and J. B. Worthen (New York, NY: Oxford University Press), 65–88.
- McLennan, C. T., Luce, P. A., and Charles-Luce, J. (2003). Representation of lexical form. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 539–553. doi: 10.1037/0278-7393.29.4.539
- McQueen, J. M., Cutler, A., and Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cogn. Sci.* 30, 1113–1126. doi: 10.1207/s15516709cog0000_79
- McQueen, J. M., and Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Q. J. Exp. Psychol.* 60, 661–671. doi: 10.1080/17470210601183890
- Moosmüller, S., and Ringen, C. (2004). Voice and aspiration in Austrian German plosives. *Folia Linguistica* 38, 1–2. doi: 10.1515/flin.2004.38.1-2.43
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition* 52, 189–234. doi: 10.1016/0010-0277(94)90043-4
- Pickering, M. J., and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav. Brain Sci.* 36, 329–347. doi: 10.1017/S0140525X12001495
- R development core team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Schirmunski, V. (1928). Die schwäbischen Mundarten in Transkaukasien und Südukraine. *Teuthonista*, 5, 38–60 (part I), 157–171 (part II).
- Sidaras, S. K., Alexander, J. E. D., and Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *J. Acoust. Soc. Am.* 125, 3306. doi: 10.1121/1.3101452
- Siebs, T., Boor, H. de, Moser, H., and Winkler, C. (1969). *Siebs Deutsche Aussprache: Reine und Gemäßigte Hochlautung mit Aussprachewörterbuch*. Berlin: de Gruyter.
- Siegel, J. (2010). *Second Dialect Acquisition*. Cambridge; New York, NY: Cambridge University Press.
- Spiekermann, H. (2008). *Sprache in Baden-Württemberg: Merkmale des Regionalen Standard*. Tübingen: Max Niemeyer Verlag.
- Sumner, M., Kim, S. K., King, E., and McGowan, K. B. (2013). The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Front. Psychol.* 4:1015. doi: 10.3389/fpsyg.2013.01015
- Sumner, M., and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *J. Mem. Lang.* 60, 487–501. doi: 10.1016/j.jml.2009.01.001
- Trude, A. M., and Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Lang. Cogn. Process.* 27, 979–1001. doi: 10.1080/01690965.2011.597153
- Trude, A. M., Tremblay, A., and Brown-Schmidt, S. (2013). Limitations on adaptation to foreign accents. *J. Mem. Lang.* 69, 349–367. doi: 10.1016/j.jml.2013.05.002
- Trudgill, P. (1986). *Dialects in Contact*. Oxford: Blackwell.
- Wade, T., Jongman, A., and Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica* 64, 122–144. doi: 10.1159/000107913
- Weber, A., Melinger, A., and Lara Tapia, L. (2007). “The mapping of phonetic information to lexical presentations in Spanish: evidence from eye movements,” in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, eds J. Trouvain and W. J. Barry (Dudweiler: Pirrot), 1941–1944.
- Witteman, M. J., Bardhan, N. P., Weber, A., and McQueen, J. M. (2015). Automaticity and stability of adaptation to a foreign-accented speaker. *Lang. Speech* 58, 168–189. doi: 10.1177/0023830914528102
- Witteman, M. J., Weber, A., and McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Atten. Percept. Psychophys.* 75, 537–556. doi: 10.3758/s13414-012-0404-y

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.

Copyright © 2016 Grohe and Weber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

TABLE A1 | Critical target-competitor pairs for Experiments 1 and 2.

	Voiced			Voiceless		
	German token	IPA transcript	English translation	German token	IPA transcript	English translation
bilabial	Butter	'bʊtə	butter	Putzer	'pʊt̚sə	cleaner
	Bistum	'bɪstʊ:m	diocese	Piste	'pɪstə	ski slope
	Beize	'baitsə	marinade/stain	Peitsche	'paɪt̚ʃə	whip
	Beifall	'baɪfal	acclaim	Peiler	'paɪlə	detector
	Baron	ba'ro:n	baron	Paris	pa'ri:s	Paris
	Balken	'balkŋ	beam	Palme	'palmə	palm tree
	Becher	'bɛçə	beaker/mug	Pächter	'pɛçtə	tenant
	Benzin	bɛn'tsi:n	gas	Pension	pɛn'zɪo:n	guest house/pension
	Bilanz	bi'lants	balance	Pilot	pi'lo:t	pilot
	Ballett	ba'lɛt	ballet	Palast	pa'last	palace
	Banner	'banə	banner	Panne	'panə	breakdown
	Bazille	ba'tsɪlə	bacillus	Pazifik	pa'tsi:fɪk	Pacific
	Bettler	'bɛtlə	beggar	Petzer	'pɛtsə	telltale
velar	Bauwerk	bauvɛrk	building	Pause	'pauzə	break
	Gorilla	go'rɪlə	gorilla	Korea	ko're:a	Korea
	Gulasch	'gʊləʃ	goulash	Kuli	'kʊli	ballpoint pen
	Galerie	galə'ri:	gallery	Kalorie	kalə'ri:	calorie
	Gasthaus	'gasthaus	guest house	Kasten	'kastŋ	box
	Gürtel	'gʏrtl̩	belt	Kürzung	'kʏrtsʊŋ	abridgement
	Gitter	'gɪtə	grid/fencing	Kittel	'kɪtl̩	tunic
	Ganove	ga'no:və	crook	Kanone	ka'no:nə	cannon/rod
	Gammler	'gamlə	loafer	Kammer	'kame	small room/chamber
	Germane	gɛr'ma:nə	Teuton	Keramik	kɛ'ra:mɪk	ceramics
	Geltung	'gɛltʊŋ	validity/prestige	Kälte	'kɛltə	cold
	Garant	ga'rant	guarantor	Karat	ka'ra:t	carat
	Garage	ga'ra:ʒə	garage	Karaffe	ka'rafə	carafe
	Gassenjunge	'gasɐn'jʊŋə	street urchin	Kassenzettel	'kasɐn'tsɛtl̩	(sales) receipt
	Gartenzaun	'gartŋsaun	garden fence	Kartenspiel	'kartŋʃpi:l	game of cards

Voiced items were used as targets in Experiment 1 (initial plosive was devoiced, i.e., /b/→/p/ and /g/→/k/), and voiceless items were competitors. In Experiment 2, voiceless items were targets, and voiced items were competitors.



Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus

Myrna C. Cintrón-Valentín^{1*} and Nick C. Ellis^{1,2,3}

¹ Department of Psychology, University of Michigan, Ann Arbor, MI, USA, ² Department of Linguistics, University of Michigan, Ann Arbor, MI, USA, ³ English Language Institute, University of Michigan, Ann Arbor, MI, USA

We consider the role of physical form, prior experience, and form focused instruction (FFI) in adult language learning. (1) When presented with competing cues to interpretation, learners are more likely to attend to physically more salient cues in the input. (2) Learned attention is an associative learning phenomenon where prior-learned cues block those that are experienced later. (3) The low salience of morphosyntactic cues can be overcome by FFI, which leads learners to attend cues which might otherwise be ignored. Experiment 1 used eye-tracking to investigate how language background influences learners' attention to morphological cues, as well as the attentional processes whereby different types of FFI overcome low cue salience, learned attention and blocking. Chinese native speakers (no L1 verb-tense morphology) viewed Latin utterances combining lexical and morphological cues to temporality under control conditions (CCs) and three types of explicit FFI: verb grammar instruction (VG), verb salience with textual enhancement (VS), and verb pretraining (VP), and their use of these cues was assessed in a subsequent comprehension test. CC participants were significantly more sensitive to the adverbs than verb morphology. Instructed participants showed greater sensitivity to the verbs. These results reveal attentional processes whereby learners' prior linguistic experience can shape their attention toward cues in the input, and whereby FFI helps learners overcome the long-term blocking of verb-tense morphology. Experiment 2 examined the role of modality of input presentation – aural or visual – in L1 English learners' attentional focus on morphological cues and the effectiveness of different FFI manipulations. CC participants showed greater sensitivity toward the adverb cue. FFI was effective in increasing attention to verb-tense morphology, however, the processing of morphological cues was considerably more difficult under aural presentation. From visual exposure, the FFI conditions were broadly equivalent at tuning attention to the morphology, although VP resulted in balanced attention to both cues. The effectiveness of morphological salience-raising varied across modality: VS was effective under visual exposure, but not under aural exposure. From aural exposure, only VG was effective. These results demonstrate how salience in physical form, learner attention, and instructional focus all variously affect the success of L2 acquisition.

OPEN ACCESS

Edited by:

Adriana Hanulíková,
University of Freiburg, Germany

Reviewed by:

Patrick Rebuschat,
Lancaster University, UK
Karin Madlener,
University of Basel, Switzerland

*Correspondence:

Myrna C. Cintrón-Valentín
mcintron@umich.edu

Specialty section:

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

Received: 29 February 2016

Accepted: 11 August 2016

Published: 29 August 2016

Citation:

Cintrón-Valentín MC and Ellis NC
(2016) Salience in Second Language
Acquisition: Physical Form, Learner
Attention, and Instructional Focus.
Front. Psychol. 7:1284.
doi: 10.3389/fpsyg.2016.01284

Keywords: second language acquisition, morphology, tense, learned attention, focus on form, grammar instruction, form-focused instruction, perceptual linguistic salience

INTRODUCTION

Psychological Aspects of Saliency

Psychological research uses the term saliency to refer to the property of a stimulus to stand out from the rest. Salient items or features are more likely to be perceived, to be attended to, and are more likely to enter into subsequent cognitive processing and learning. Saliency can be independently determined by physics and the environment, and by our knowledge of the world. It is useful to think of three aspects of saliency, one relating to psychophysics, the other two to what we have learned. (1) The physical world, our embodiment, and our sensory systems come together to cause certain sensations to be more intense (louder, brighter, heavier, etc.) than others. (2) As we experience the world, we learn from it, and our resultant knowledge values some associations higher than others. We know that some stimulus cues have affordances: they are associated with outcomes or possibilities that are important to us, while others are negligible (Gibson, 1977; James, 1890a, chap. 11). (3) We also have expectations about what is going to happen next in known contexts, we are surprised when our expectations are violated, and we pay more attention as a result. Each of the three phenomena is explained in more detail below.

Psychophysical Saliency

Loud noises, bright lights, and moving stimuli capture our attention. Saliency arises in sensory data from contrasts between items and their context. Stimuli with unique features compared to their neighbors (e.g., Os in a field of Ts, a red poppy in a field of yellow) “pop out” from the scene but in a shared feature context will not (Os among Qs; Treisman and Gelade, 1980). These are aspects of bottom-up processing (Shiffrin and Schneider, 1977).

Salient Associations

Attention can also be driven by top-down, memory dependent, expectation-driven processing. Emotional, cognitive, and motivational factors affect the saliency of stimuli. These associations make a stimulus cue “dear.” A loved one stands out from the crowd, as does a stimulus with weighty associations (\$500000.0 versus \$0.000005, however, similar the amount of pixels, characters, or ink in their sensation). The units of perception are influenced by prior association: “The chief cerebral conditions of perception are the paths of association irradiating from the sense-impression, which may have been already formed” (James, 1890b, p. 82). Psychological saliency is experience-dependent: *hotdog*, *sushi*, and 寿司 mean different things to people of different cultural and linguistic experience. This is why, *contra* sensation, the units of perception cannot be measured in physical terms. They are subjective. Hence George Miller’s definition of the units of short-term memory as “chunks”: “We are dealing here with a process of organizing or grouping the input into familiar units or chunks, and a great deal of learning has gone into the formation of these familiar units” (Miller, 1956, p. 91).

Context and Surprisal

The evolutionary role of cognition is to predict what is going to happen next, given that anticipation affords survival value. We find structure in time (Elman, 1990). The brain is a prediction machine (Clark, 2013). One consequence is that it is surprisal – when prediction goes wrong – that maximally drives learning from a single trial. Otherwise, the regularities of the usual course of our experiences sum little by little, trial after trial, to drive our expectations. Cognition is probabilistic, its expectations a conspiracy tuned from statistical learning over our experiences (Ellis, 2002).

Saliency and Learning

Rescorla and Wagner (1972) presented a formal model of conditioning which expresses the capacity of any cue [conditioned stimulus (CS), for example a bell in Pavlovian conditioning] to become associated with an outcome [unconditioned stimulus (US), for example food in Pavlovian conditioning] on any given experience of their pairing. This formula summarized over 80 years of research in associative learning, and it elegantly encapsulates the three factors of psychophysical saliency, psychological saliency, and surprisal. The role of US surprise and of CS and US saliency in the process of conditioning can be summarized as follows:

$$dV = ab(L - V).$$

The associative strength of the US to the CS is referred to by the letter V and the change in this strength which occurs on each trial of conditioning is called dV . On the right hand side, a is the saliency of the US, b is the saliency of the CS, and L is the amount of processing given to a completely unpredicted US. Thus both the saliency of the cue (a) and the psychological importance of the outcome (b) are essential factors in any associative learning. As for $(L - V)$, the more a CS is associated with a US, the less additional association the US can induce. As Beckett (1954) put it: “habit is a great deadener.” Alternatively, with novel associations where V is close to zero, there is much surprisal, and consequently much learning: first impressions, first love, first time...

This is arguably the most influential formula in the history of learning theory. Physical saliency, psychological saliency, and surprisal interactively affect what we learn from our experiences of the world.

Saliency in Second Language Acquisition (SLA)

Naturalistic second language (L2) learners tend to focus more in their language processing upon open-class words (nouns, verbs, adjectives, and adverbs) than on grammatical cues. Their language attainment often stabilizes at a “Basic Variety” of interlanguage that predominantly comprises open-class words; closed-class items—in particular, grammatical morphemes and prepositions—are rare if present at all (Noyau et al., 1995).

Although naturalistic second language learners are surrounded by language *input*, the available target language, not all of it becomes *intake*, that subset of input that actually gets in and which the learner utilizes in some way (Corder, 1967).

A classic case study is that of the naturalistic language learner, Wes, who was described as being very fluent, with high levels of strategic competence, but low levels of grammatical accuracy: “using 90% correct in obligatory contexts as the criterion for acquisition, none of the grammatical morphemes counted has changed from unacquired to acquired status over a 5 years period” (Schmidt, 1984, p. 5).

Although the Basic Variety is sufficient for everyday communicative purposes, grammatical morphemes and closed-class words tend not to be put to full use (e.g., Van Patten, 1996, 2006; Clahsen and Felser, 2006). Many untutored L2 learners initially make temporal references mostly by use of temporal adverbs, prepositional phrases, serialization, and calendric reference, with the grammatical expression of tense and aspect emerging only slowly thereafter, if at all (Meisel, 1987; Bardovi-Harlig, 1992, 2000; Noyau et al., 1995).

Psychophysical Saliency

One factor determining the learning of cues is psychophysical saliency: prepositional phrases, temporal adverbs, and lexical linguistic cues are salient and stressed in the speech stream. Verb inflections are not. In his landmark study of first language acquisition, Brown (1973, p. 343) breaks down the measurement of perceptual saliency, or “clarity of acoustical marking,” into “such variables as amount of phonetic substance, stress level, usual serial position in a sentence, and so on” Brown (1973, p. 463).

Many grammatical form-function relationships in English, like grammatical particles and inflections such as the third person singular *-s*, are of low saliency in the language stream. This is a result of the well-documented effect of frequency and automatization in the evolution of language. The basic principles of automatization that apply to all kinds of motor skills (like playing a sport or a musical instrument) are that through repetition, sequences of units that were previously independent come to be processed as a single unit or chunk (Ellis, 1996). The more frequently speakers use a form, the more they abbreviate it: this is a law-like relationship across languages. Zipf (1949) summarized this in the *principle of least effort* – speakers want to minimize articulatory effort so they tend to choose the most frequent words, and the more they use them, automatization of production causes their shortening. Frequently used words become shorter with use. Grammatical functors are the most frequent elements of a language, thus they lose their emphasis and tend to become abbreviated and phonologically fused with surrounding material (Bybee, 2000; Jurafsky et al., 2001; Zuraw, 2003).

Thus grammatical function words and bound inflections tend to be short and low in stress, with the result that these cues are difficult to perceive. In a corpus study by Cutler and Carter (1987), 86% of strong syllables occurred in open class words and only 14% in closed-class words; for weak syllables, 72% occurred in closed-class words and 28% in open-class words. When grammatical function words (*by, for, no, you*, etc.) are clipped out of connected speech and presented in isolation at levels where their open-class equivalents (*buy, four, know, ewe*, etc.) are perceived 90–100% correctly, adult native speakers can recognize

them only 40–50% of the time (Herron and Bates, 1997). Clitics, accent-less words or particles that depend accentually on an adjacent accented word and form a prosodic unit together with it, are the extreme examples of this: the /s/ of ‘he’s’, /l/ of ‘I’ll’, and /v/ of ‘I’ve’ can never be pronounced in isolation.

In sum, grammatical functors are difficult to perceive from bottom-up auditory evidence alone. Fluent language processors can perceive these elements in continuous speech because their language knowledge provides top-down support. But this is exactly the knowledge that learners lack. Thus the low psychophysical saliency of grammatical functors contributes to L2 learners’ difficulty in learning them (Goldschneider and DeKeyser, 2001; Ellis, 2006b).

Saliency as Modulated by Modality

Spoken and written language are very different media, with spoken language being fleeting while writing provides more permanent visual substance on the page, allowing the reader to attend linguistic form at their discretion. Attention to language form may therefore pose different challenges in written and spoken modalities. VanPatten (1990) showed that L2 learners of Spanish have difficulty simultaneously attending to meaning and form of aural input. He had them process spoken Spanish passages for meaning while simultaneously monitoring the input for either lexical content words like *inflación* or for grammatico-morphological forms like the definite article *la* or the verb morpheme *-n*. Monitoring grammatico-morphological forms negatively affected comprehension, whereas attention to lexical items did not. Wong (2001) replicated this study while also adding conditions in the written modality. She showed that comprehension was worse from aural language than from written language. Furthermore, while the results from the aural conditions replicated the patterns found by VanPatten, the number of idea units recalled by readers who had to pay attention to the definite article in the written input was not significantly less than those who read the passage for content only or for those who had to attend to the lexical item *inflación*. Thus modality can differentially affect the saliency of forms and their input processing: written language can make grammatical forms more salient and more easily processed.

Learned Attention

In addition to psychophysical factors, there are attentional factors which affect the saliency of grammatical functors. The first relates to their redundancy. Grammatical morphemes often appear in redundant contexts where their interpretation is not essential for correct interpretation of the sentence (Terrell, 1991; Van Patten, 1996; Schmidt, 2001). Tense markers often appear in contexts where other cues have already established the temporal reference (e.g., “*yesterday he walked*”), plural markers are accompanied by quantifiers or numerals (“*27 cats*”), etc. Hence their neglect does not result in communicative breakdown, they carry little psychological importance of the outcome (term *b* in the Rescorla-Wagner equation), and the Basic Variety satisfies (Simon, 1957) for everyday communicative purposes.

Still again, there are attentional biases that result from L2 learners’ history of learning – from their knowledge of a prior

language. Ellis (2006a,b) attributes L2 difficulties in acquiring inflectional morphology to an effect of *learned attention* known as *blocking* (Kamin, 1969; Mackintosh, 1975; Kruschke and Blair, 2000; Kruschke, 2006). Blocking is an associative learning phenomenon, occurring in animals and humans alike, that shifts learners' attention to input as a result of prior experience (Rescorla and Wagner, 1972; Shanks, 1995; Wills, 2005). Knowing that a particular stimulus is associated with a particular outcome makes it harder to learn that another cue, subsequently paired with that same outcome, is also a good predictor of it. The prior association "blocks" further associations.

All languages have lexical and phrasal means of expressing temporality. So anyone with knowledge of any first language is aware that there are reliable and frequently used lexical cues to temporal reference (words like German *gestern*, French *hier*, Spanish *ayer*, English *yesterday*). Such are cues to look out for in an L2 because of their frequency, their reliability of interpretation, and their saliency. Learned attention theory holds that, once known, such cues block the acquisition of less salient and less reliable verb tense morphology from analysis of redundant utterances such as *Yesterday I walked*. The Input Processing (IP) theory of SLA (Van Patten, 1996) includes a *Lexical Preference Principle*: "Learners will process lexical items for meaning before grammatical forms when both encode the same semantic information" (Van Patten, 2006, p. 118), and a *Preference for Non-redundancy Principle*: "Learners are more likely to process non-redundant meaningful grammatical markers before they process redundant meaningful markers" (Van Patten, 2006, p. 119).

Summing up, grammatical functors abound in the input, but, as a result of their low saliency, their redundancy, the low contingency of their form-function mappings, and adult acquirers' learned attentional biases and L1-tuned automatized processing of language, they are simply not implicitly learned by many naturalistic learners whose attentional focus is on communication.

Prior Experiments on Learned Attention and Blocking in SLA

Ellis and Sagarra (2010, 2011) and Ellis et al. (2014) report a series of experimental investigations of learned attention in SLA involving the learning of a small number of Latin expressions and their English translations. We sketch them in some detail here because they introduce key concepts and because we build on their design in the present study.

In Ellis and Sagarra (2011) there were three groups: Adverb Pretraining, Verb Pretraining, and Control. In Phase 1, Adverb Pretraining participants learned two adverbs and their temporal reference – *hodie* today and *heri* yesterday; Verb Pretraining participants learned verbs (shown in either first, second, or third person) and their temporal reference – e.g., *cogito* present or *cogitavi* past; the Control group had no such pretraining. During Phase 2, Sentence Exposure, all participants were shown sentences which appropriately combined an adverb and a verb (e.g., *heri cogitavi*, *hodie cogitas*, *cras cogitabis*) and learned whether these sentences referred to the past, the present, or

the future. In Phase 3, the Reception test, all combinations of adverb and verb tense marking were presented individually and participants were asked to judge whether each sentence referred to the past, present, or future. The logic of the design was that in Phase 2 every utterance contained two temporal references – an adverb and a verb inflection. If participants paid equal attention to these two cues, then in Phase 3 their judgments should be equally affected by them. If, however, they paid more attention to adverb (/verb) cues, then their judgments would be swayed toward them in Phase 3.

The Control Group illustrate the normal state of affairs when learners are exposed to utterance with both cues and learn from their combination. Multiple regression analysis, where the dependent variable was the mean temporal interpretation for each of the Phase 3 strings and the independent variables were the information conveyed by the adverbial and verbal inflection cues showed in standardized β coefficients, Control Group Time = 0.93 Adverb + 0.17 Verb. The adverb cues far outweighed the verbal inflections in terms of learnability. We believe this is a result of two factors (i) the greater saliency of the adverbial cues, and (ii) learned attention to adverbial cues which blocks the acquisition of verbal morphology.

The two other groups reacted to the cues in quite different ways – the Adverb pretraining group followed the adverb cue, the Verb pretraining group tended to follow the verb cue: Adverb Group Time = 0.99 Adverb – 0.01 Verb; Verb Group Time = 0.76 Adverb + 0.60 Verb. Pretraining on the verb in non-redundant contexts did allow acquisition of this cue when its processing was task-essential, but still, the adverb predominated.

Ellis and Sagarra (2010, Experiment 2) and Ellis and Sagarra (2011, Experiments 2 and 3) also illustrated long-term language transfer effects whereby the nature of learners' first language (+/– verb tense morphology) biased the acquisition of morphological versus lexical cues to temporal reference in the same subset of Latin. First language speakers of Chinese (no tense morphology) were less able than first language speakers of Spanish or Russian (rich morphology) to acquire inflectional cues from the same language experience where adverbial and verbal cues were equally available, with learned attention to tense morphology being in standardized β coefficients: Chinese (–0.02) < English (0.17) < Russian (0.22) < Spanish (0.41) (Ellis and Sagarra, 2011, p. 612). These findings suggest that there is a long-term attention to language, a processing bias affecting subsequent cue learning that comes from a lifetime of prior L1 usage.

Enhancing Attention to Non-salient Forms: The Role of Form-Focused Instruction

Several theories of SLA (e.g., Schmidt, 2001; Ellis, 2005) emphasize the centrality of attention. Schmidt's (2001) *Noticing Hypothesis* holds that conscious attention to linguistic forms in the input is an important precondition to learning: "people learn about the things they attend to and do not learn much about the things they do not attend to" (Schmidt, 2001, p. 30).

Form focused instruction (FFI) attempts to encourage noticing, drawing learners' attention to linguistic forms that

might otherwise be ignored (Spada, 1997; Spada and Tomita, 2010; Ellis, 2012). Variants of FFI vary in the degree and manner in which they recruit learner consciousness and in the role of the learner's metalinguistic awareness of the target forms (Ellis, 1994; Rebuschat, 2015). Explicit instruction traditionally centers upon "some sort of rule being thought about during the learning process" (DeKeyser, 1995). This type of instruction can be deductive, when learners are presented with grammar rule explanation, or inductive, when they are asked to attend to a particular set of forms with the purpose of inferring the rules on their own. This may include explicit metalinguistic feedback, which provides "comments, information, or questions, related to the well-formedness of the learner's utterance" (Lyster and Ranta, 1997, p. 47). Conversely, through more implicit instruction, learners are expected to infer regularities of form-meaning patterns without awareness. Having laid out the bare contrast like this, we emphasize that there is no simple binary divide between explicit and implicit instruction, that implicit and explicit knowledge interact, and that this is still an area of considerable research inquiry (e.g., Ellis, 1994, 2005; Rebuschat, 2015).

Long (1991) and Doughty and Long (2003) describe how a focus on meaning can be improved upon by periodic attention to language as object: during otherwise meaning-focused lessons, learners' attention is briefly shifted to linguistic code features, in context, to induce noticing. This is known as focus-on-form. Doughty and Williams (1998) give the following examples of focus-on-form techniques, ranging from less to more explicit: input flood, where texts are saturated with L2 models; input elaboration; input enhancement, where learner attention is drawn to the target through visual highlighting or auditory stress; corrective feedback on error, such as recasting; and input processing, where learners are given practice in using L2 rather than L1 cues.

Norris and Ortega's (2000) meta-analysis comparing the outcomes from studies that employed differing levels of explicitness of L2 input demonstrated that FFI instruction results in substantial target-oriented L2 gains, that explicit types of instruction are more effective than implicit types, and that the effectiveness of L2 instruction is durable. More recent meta-analyses of effects of type of instruction by Spada and Tomita (2010) and Goo et al. (2015) likewise report large advantages of explicit instruction in L2 acquisition. However, the studies gathered in these meta-analyses used a wide variety of types of instruction, learner, targeted feature, and method of assessment. There is need to compare FFI methods upon the processing of the same target feature in similar populations of learners.

This is one of the aims of our current study, which employs a series of explicit and implicit FFI techniques to contrast and illuminate the processes by which these different methods help learners refocus their attention to non-salient forms in the input. In the following sections we will discuss and operationalize the different types of FFI included in our design: (1) Verb grammar (VG), (2) Textual enhancement (TE), and (3) Verb pretraining (VP) in isolation in task-essential rather than redundant contexts.

Verb Grammar (VG)

One method that has been widely investigated both in SLA research and practice is that of explicit grammar instruction (EGI) which Terrell (1991, p. 53) defines as "the use of instructional strategies to draw the students' attention to, or focus on, form and/or structure," with instruction targeted at increasing the saliency of inflections and other commonly ignored features by, first, pointing them out and explaining their structure and, second, providing meaningful input that contains many instances of the same grammatical meaning-form relationship. Ellis (2006) reviews studies of EGI demonstrating that learning through explicit means alone, that is, without the provision of tasks requiring the learner to practice the target features before being tested on their knowledge of these forms, seems to be ineffective (e.g., Ellis, 1993; VanPatten and Oikennon, 1996). We therefore operationalized VG as short metalinguistic description of simple regular tense morphology in Latin which was followed by a sentence exposure phase where learners were presented with phrases combining adverbs and verb cues to temporality and were asked to determine the appropriate tense before proceeding to comprehension.

Textual Enhancement

Another common FFI technique is the use of Textual Enhancement such as color-coding, boldfacing and underlining, to increase learners' awareness of non-salient forms in the input (Sharwood-Smith, 1993; Doughty and Williams, 1998). Han et al. (2008) and Lee and Huang (2008) review studies of TE and conclude that there are conflicting findings regarding its effectiveness. They suggest that these discrepancies may be explained by differences between studies in such factors as learners' target and native languages, the type, complexity and communicative value of target forms, learner proficiency, treatment intensity, and the measures used to assess noticing and processing of these forms.

In the present study, we used boldfacing and color to make verb-tense inflections more salient. This condition is therefore called verbal saliency (VS). Contrary to the VG condition, we did not explicitly direct learners to attend to the enhanced verb-inflections. Nevertheless, given that we did provide VS participants with explicit feedback on their correctness during the exposure phase, we consider VS an explicit FFI technique designed to promote induction of the target form.

Verb Pretraining

The effect of blocking is particularly potent whenever the cue to be processed is met in a redundant context where other cues have the same interpretation and have been learned previously or are more salient. One way to counteract this type of blocking is to ensure that early in L2 experience, the cue is experienced on its own in situations in which it must be processed for successful interpretation (VanPatten and Oikennon, 1996). Ellis and Sagarra's (2010, 2011) VP conditions tested the effects of this and demonstrated that once the cue has been consolidated into the processing system, it continues to contribute to processing in subsequent situations of potential cue competition. For continuity, replication, and comparison, we include VP here to

compare its efficiency and operation with VG and VS conditions. VP does not explicitly provide learners with a metalinguistic description of verb-tense morphology, but rather gives them opportunity to infer how verb tense morphology works by processing Latin verb forms for temporality and providing feedback on their correctness.

Eye-Tracking as a Measure of Attention

Second Language Acquisition research is increasingly recognizing eye tracking as a research tool (Winke et al., 2013) because it “allows for the study of moment-by-moment processing decisions during natural, uninterrupted comprehension, and critically, without the need to rely on participants’ strategic or metalinguistic responses” (Roberts and Siyanova-Chanturia, 2013, p. 214). Ellis et al. (2014) used eye movement recordings to measure participants’ overt attention to adverb and verb cues and found that pretraining on different cue dimensions (adverb pretraining versus verb pretraining) led to differences in learners’ overt attention to these cues during processing, and that these in turn led to differences in their covert attention to these cues during the comprehension and production tasks.

Aims

The current studies extend previous research on saliency and learned attention in SLA by (i) exploring and comparing the degree to which VG, VS, and VP methods of FFI might serve to counteract learned attention effects whereby learners’ prior experience with adverbial cues in their L1 block their processing of verb inflections in the L2, and (ii) comparing their effects in aural and visual modalities of language.

In Experiment 1 we use eye-tracking to measure Chinese L1 speakers’ visual attention to form in these various FFI conditions of visual language exposure. The control condition (CC) and VP conditions allow us to replicate Ellis and Sagarra (2011), as well as to extend the findings in Ellis et al. (2014) using a more complex verbal system. The inclusion of VG and VS, additionally allow us to further compare the effects of these manipulations to VP.

Experiment 1 focuses upon several research questions:

Research Question 1 (RQ1): do the effects of physical saliency and learned attentional biases toward adverbial cues, under normal conditions of exposure (CC), prejudice the acquisition of verbal tense morphology, as indexed in participants’ relative reliance on these cues in subsequent language comprehension?

Research Question 2 (RQ2): does early experience of morphological cues to temporal reference, through each of the FFI treatments VG, VS, VP, counteract the effects of physical saliency and learned attentional biases, as indexed by participants’ relative reliance on these cues in subsequent language comprehension?

Research Question 3 (RQ3): does early experience of morphological cues to temporal reference lead to biases in subsequent overt perceptual attention (as indexed by number of fixations) during Sentence Exposure, where there are both adverbial and morphological cues to the same interpretation?

Research Question 4 (RQ4): does any bias in overt attention to input cues in turn lead to subsequent attentional biases to the adverbial or morphological cues in subsequent language comprehension?

In Experiment 2 we compare the processing of auditory and visual input to assess effects of modality on saliency, and again we contrast the effectiveness of VG, VS and VP methods of FFI in counteracting learned attention effects. The research questions of Experiment 2 are:

Research Question 5 (RQ5): as in Experiment 1, does early experience of morphological cues to temporal reference, through each of the FFI treatments VG, VS, and VP, counteract the effects of physical saliency and learned attentional biases, as indexed by participants’ relative reliance on these cues in subsequent language comprehension.

Research Question 6 (RQ6): are each of the FFI treatments VG, VS, and VP equally effective in reattuning learners’ attention to the non-salient morphological cues through visual and auditory modalities of exposure?

EXPERIMENT 1

Introduction

Cintrón-Valentín and Ellis (2015) used eye-tracking to investigate the attentional processes whereby different types of FFI instruction overcome learned attention and blocking effects in learners’ online processing of L2 input. English native speakers viewed Latin utterances combining lexical and morphological cues to temporality under control conditions (CC) and three types of explicit FFI: verb grammar instruction (VG), verb saliency with textual enhancement (VS), and verb pretraining (VP). All groups participated in three phases: exposure, comprehension test, and production test. VG participants viewed a short lesson on Latin tense morphology prior to exposure. VS participants saw the verb inflections highlighted in bold and red during exposure. VP participants had an additional introductory phase where they were presented with solitary verb forms and trained on their English translations. Instructed participants showed greater sensitivity to morphological cues in comprehension and production testing. Eye-tracking measures revealed how FFI affects learners’ attention during online processing and thus modulates long-term blocking of verb morphology.

This experiment aims to replicate these findings in another population of learners, L1 Chinese speakers, whose L1 does not exhibit verb-tense morphology. In Chinese languages, “gender, plurality and tense are either indicated by lexical choice or not indicated at all” (Li and Thompson, 1987, p. 825). As a result, L1 speakers of Chinese languages are particularly prone to long-term attentional blocking of verb tense morphology (Ellis and Sagarra, 2011).

Participants

Chinese native speakers who had not learned Latin or Italian previously were recruited from a major university in the USA

($n = 58$) or its local community ($n = 9$). They were volunteers and either participated as part of an undergraduate Psychology course requirement ($n = 3$) or they were compensated with 10 dollars for their time ($n = 64$). All were bilingual with high-level English language proficiency sufficient to admit them to study in English. However, all had learned English as a L2 after the age of 5 years. They were randomly assigned to one of four conditions: CC, $n = 19$ (12 females and 7 males), age range 19–35 years ($M = 24.58$); VG, $n = 18$ (13 females and 5 males), age range 20–26 years ($M = 22.50$); VS, $n = 14$ (11 females and 4 males), age range 19–30 years ($M = 23.13$) and; VP, $n = 15$ (10 females and 4 males), age range 20–34 years ($M = 24.80$). Of these participants, seven (CC = 4; VG = 2; VS = 1) were excluded from the eye-tracking analyses due to poor data quality. All participants received oral instructions in their native language prior to the start of the experiment, with the exception of three participants in the Chinese CC group and four participants in the Chinese VG group. Although it was originally intended that all participants would receive these additional instructions in their native language, to ensure that they were indeed bilingual, the research assistants were not all fluent in Chinese.

Procedure

The experiment was programmed in E-Prime (Schneider et al., 2002). It took less than 1 h to complete. There were three phases:

Pretraining, Sentence Exposure, and Comprehension testing. The procedure of these phases is shown in **Table 1**.

Pretraining

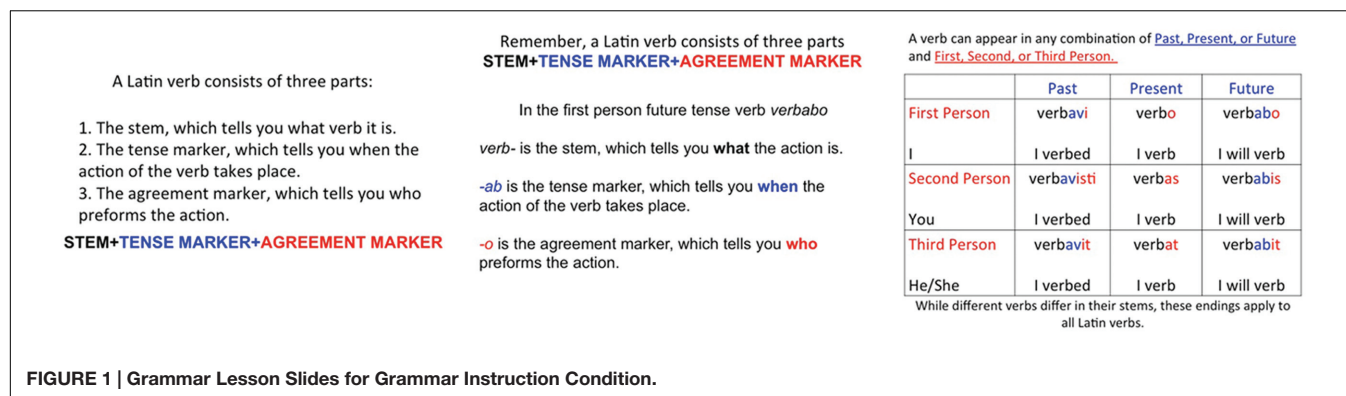
Verb pretraining participants engaged in a phase that involved training on verb inflections. On each trial they saw one of the past (*cogitavi*, *cogitavisti*, *cogitavit*) or present (*cogito*, *cogitas*, *cogitat*) inflected verbs and learned that each corresponded to either X think(s) or X thought by clicking the appropriate alternative with the mouse. A correct choice returned the feedback “Correct” or “Incorrect – the meaning of [Latin word] is [English word].” The 36 trials thus involved each of the three persons singular of present and past tense being presented six times in random order. Keeping the same number of trials of pretraining for all participants allows evaluation of what is gained from that amount of experience. This permits comparison across contents and conditions of pretraining, for example auditory versus visual modality as in Experiment 2 which follows here, and those of Ellis and Sagarra (2010, 2011) which vary with regard to the different levels of grammatical number and person. We report performance levels at the end of training in 2.4.1.

Pretraining for the VG participants involved a brief lesson on Latin verb-tense morphology using the three slides shown in **Figure 1**. Although they could view each of the slides for an undetermined amount of time, they were not allowed to take notes and could not regress to previous slides.

TABLE 1 | The design of Phases 1–3 of Experiment 1.

Pretraining (Phase 1) (+ feedback)	Sentence Exposure (Phase 2) (+ feedback) 36 (18 × 2) randomized trials	Comprehension Test (Phase 3) (– feedback) 66 randomized trials	
		Stimulus	Semidiem
Control group	Present		
No pretraining	<i>hodie cogito</i>	<i>hodie</i>	3
Verb Pretraining group	<i>hodie cogitas</i>	<i>heri</i>	1
(36 randomized trials)	<i>hodie cogitat</i>	<i>cras</i>	5
<i>cogito</i> “I think”	<i>cogito hodie</i>	<i>cogito/as/at</i>	3
<i>cogitas</i> “you think”	<i>cogitas hodie</i>	<i>cogitavi/visti/vit</i>	1
<i>cogitat</i> “X thinks”	<i>cogitat hodie</i>	<i>cogitabo/bis/bit</i>	5
<i>cogitavi</i> “I thought”	Past	<i>hodie cogito/as/at</i>	3
<i>cogitavisti</i> “you thought”	<i>heri cogitavi</i>	<i>hodie cogitavi/visti/vit</i>	2
<i>cogitavit</i> “X thought”	<i>heri cogitavisti</i>	<i>hodie cogitabo/bis/bit</i>	4
Verb Saliency group	<i>heri cogitavit</i>	<i>heri cogito/as/at</i>	2
No pretraining	<i>cogitavi heri</i>	<i>heri cogitavi/visti/vit</i>	1
Verb Grammar group	<i>cogitavisti heri</i>	<i>heri cogitabo/bis/bit</i>	3
Brief Grammar Lesson	<i>cogitavit heri</i>	<i>cras cogito/as/at</i>	4
See Figure 1	Future	<i>cras cogitavi/visti/vit</i>	3
	<i>cras cogitabo</i>	<i>cras cogitabo/bis/bit</i>	5
	<i>cras cogitabis</i>	<i>cogito/as/at hodie</i>	3
	<i>cras cogitabit</i>	<i>cogitavi/visti/vit hodie</i>	2
	<i>cogitabo cras</i>	<i>cogitabo/bis/bit hodie</i>	4
	<i>cogitabis cras</i>	<i>cogito/as/at heri</i>	2
	<i>cogitabit cras</i>	<i>cogitavi/visti/vit heri</i>	1
		<i>cogitabo/bis/bit heri</i>	3
		<i>cogito/as/at cras</i>	4
		<i>cogitavi/visti/vit cras</i>	3
		<i>cogitabo/bis/bit cras</i>	5

The rating scale for the Comprehension Test ranged from 1 (past) to 5 (future). The correct answer for each trial is shown in the semidiem column.



Sentence Exposure

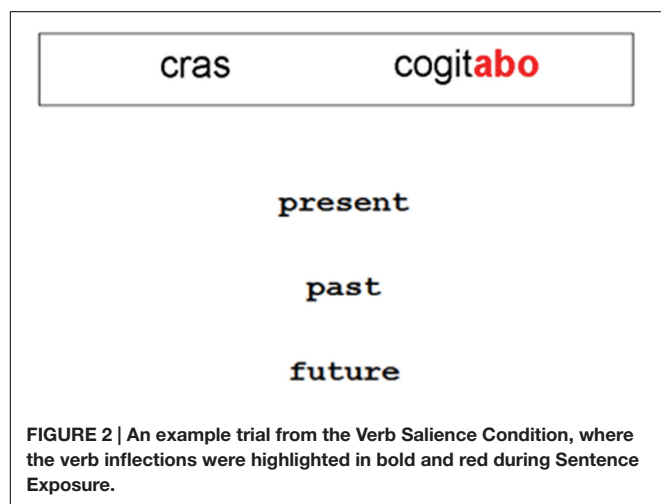
During Sentence Exposure, participants were exposed to 18 sentences (see **Table 1**) that appropriately combined the adverb with a verb (half in adverb-verb word order and half in verb-adverb order) and had to choose whether these sentences referred to the present, the past, or the future. Both word orders were used to counterbalance which cue was experienced first across sentences. Each of the 18 sentences was presented twice during this phase of the experiment. Feedback was given for both correct and incorrect choices. For correct answers, the word “correct” would appear on the screen, whereas for incorrect answers, participants would see the word “wrong” accompanied by the correct answer (e.g., “Wrong – [*heri cogitavisti*] is [past]”). The Sentence Exposure procedure was identical for the CC, VG, and VP groups. For VS participants only, the stimuli were textually enhanced so that the verbal inflections were highlighted in bold and red to increase the saliency of these items (see **Figure 2**). Participants were not made aware of this beforehand and were given the same instructions for this task as were the other groups.

Comprehension Test

In this phase, participants were presented with all single-word items (verbs and adverbs) and all possible combinations of

adverbs and verb tenses for a total of 12 single-word items and 54 two-word items (comprised of 27 unique combinations), respectively (a grand total of 66 trials; see **Table 1**). The two-word items were presented in two different word orders, counterbalancing the cue participants would experience first. The presentation of all possible combinations meant that participants experienced sentences that were familiar to them from the previous task and also combinations in which the verb and adverb were incongruent in their time reference. Before the start of the task participants were told that there would be both congruent and incongruent sentences. They were asked to judge their temporal reference on a five-point scale by using their mouse to select the appropriate answer. The possible scale points were labeled (1) “past,” (2) “both past and present,” (3) “present,” (4) “both present and future,” and (5) “future.” Participants were told they could also choose 3 if they encountered an incongruent sentence with both past and future cues. For example, the participant could be presented with an incongruent sentence such as *heri cogitabo* “Yesterday I will think,” for which the correct answer was 3 and understood as the average of the items’ tenses (past [1] + future [5]/2 = 3). The correct answer for each trial, which Ellis and Sagarra (2011) referred to as the semidiem, is shown in **Table 1**. This task separately assessed the degree to which participants attended the adverb and verbal cues by determining the relative weight that learners put on adverbial and inflectional cues to time reference. For this reason, feedback was not provided.

The logic behind the experiment follows that of previous studies of learned attention and blocking (Ellis and Sagarra, 2010, 2011; Ellis et al., 2014). During Sentence Exposure, regardless of condition, all participants experience both the adverb and verbal cue together. If they pay equal attention to both cues during this phase then their judgment during the Comprehension test should be equally affected by both cues. However, if they are biased toward one cue or the other, it is expected that their judgment in the Comprehension test will be swayed toward the corresponding cue. Because the CC participants only saw the two cues together, their performance was expected to mirror how learners typically weigh these cues, which in the native speakers of English studied in Ellis and Sagarra (2010, 2011) was characterized by the overshadowing of morphological cues by the more salient and reliable adverbial cues.



Eye-Tracking

Eye-movement recordings were gathered using an ISCAN-ETL 400 eye-imaging system with a sampling rate of 60 Hz. The eye-tracking cameras were mounted on headgear. Before the start of Pretraining (or Sentence Exposure for the CC and VS participants), the participants' gaze was calibrated using a six-point calibration sequence. This sequence was again repeated for all participants before starting Comprehension testing. Stimuli were presented in E-Prime and were positioned within a screen area of 640×480 pixels. In the Sentence Exposure phase, the left stimulus (STIML) was centered at coordinates (x, y) 94, 99, and the right stimulus (STIMR) was positioned at coordinates 454, 99. For Comprehension testing, STIML and STIMR were positioned at 109, 108 and 505, 108, respectively. Participants' fixations were analyzed using ILAB (Version 3.6.4), an open-source program developed for the analysis of eye-movement recordings (Gitelman, 2002) through the MATLAB software platform (Version 7.12.0.635) (MathWorks Inc, 2011). For each condition, fixations were analyzed from 600 ms after the start of Sentence Exposure and Comprehension testing trials (coinciding with the end of the presentation of a fixation cross at the center of the screen) until the end of each trial (coinciding with participant response). Region of interest (ROI) analyses were calculated using two positions (left and right) at the upper-most part of the screen. Both ROIs had a height of 200 pixels and a width of 250 pixels; the ROI for STIML was centralized at 175, 103 pixels and the ROI for STIMR, at 465, 103 pixels. These relatively large ROIs reflect our simple setup, which involved merely a chin rest and forehead bar to stabilize the participant's head position. In some cases, for individual subjects it was necessary to edit coordinates for both ROIs to adjust for drift. Fixation analyses were run using the default ILAB fixation velocity/distance calculation parameters, with fixations determined according to degree of movement (horizontal 1.02° ; vertical 1.09°) and a minimum duration of 100 ms. Eye-movement analysis was done blind to stimulus content: the random order of stimulus presentation for each participant entailed that right and left fixation durations were assigned as verb and adverb fixation durations only in subsequent statistical analysis on the basis of trial number.

RESULTS

Behavioral Data

Verb Pretraining Data

Mean performance in the first quarter of Verb Pretraining was 79%. By the fourth quarter, mean performance was 93% (nine participants attained 100%, five 89%, and one 56%) demonstrating that the amount of training in Phase 1 was at an appropriate level.

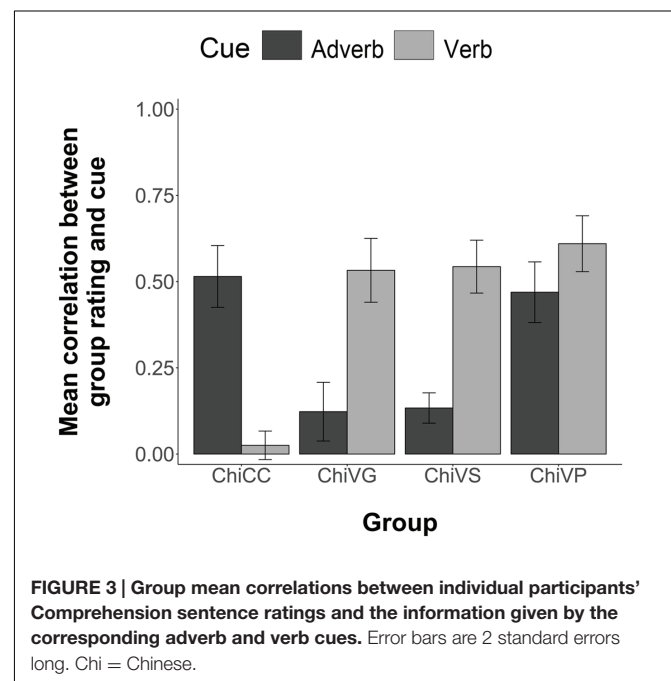
Sentence Exposure Data

Mean performance in the first quarter of Sentence Exposure was 60% for the CC group, 62% for the VG group, 49% for the VS group, and 74% for the VP group: the prior experience of VP participants gave them an advantage in the first quarter compared to the other groups. However, performance evened out for all

groups by the end of the phase. Mean performance in the final quarter was 82% for the CC group, 84% for the VG group, 73% for the VS group, and 89% for the VP group. A one-way ANOVA on these final quarter scores did not reveal a significant group effect, $F(3,63) = 1.55$, $p = 0.21$.

Comprehension Data

For each participant, we computed the Pearson correlation between the temporal ratings they provided for each of the 54 two item stimuli in the comprehension phase and the information given in each sentence by the corresponding adverb and verb cues. This correlation thus shows the degree to which each participant is biased by the verb and adverb cues. **Figure 3** illustrates the group mean correlations. Following Corey et al. (1998), when averaging or performing inferential statistics on the correlation coefficients, we first transformed the r values to z values, then performed the statistics, and then reverse transformed to report the values. Participants in the four groups differed in their cue use. Chinese CC participants were more influenced by the adverb, $M = 0.51$, 95% CI = $[0.34, 0.69]$ than the verb, $M = 0.03$, 95% CI = $[-0.06, 0.11]$. Chinese Verb grammar participants were more influenced by the verb, $M = 0.53$, 95% CI = $[0.35, 0.71]$, than by the adverb, $M = 0.13$, 95% CI = $[-0.04, 0.29]$. Chinese VS participants were more influenced by the verb, $M = 0.54$, 95% CI = $[0.39, 0.69]$, than by the adverb, $M = 0.13$, 95% CI = $[0.04, 0.22]$. Likewise, Chinese VP participants were more influenced by the verb, $M = 0.61$, 95% CI = $[0.45, 0.77]$, but relative to the other FFI groups, maintained some sensitivity toward the adverb cue, $M = 0.47$, 95% CI = $[0.30, 0.64]$. The one VP participant who attained less than 88% in Phase 1 pretraining showed little influence of verb bias in later comprehension ($r = 0.11$) compared to adverb bias ($r = 0.88$).



An ANOVA (4 Groups \times 2 Cues, with subjects nested within groups) revealed an overall effect of group, $F(3,63) = 3.83$, $p = 0.01$; and a significant group by cue interaction, $F(3,63) = 7.80$, $p < 0.001$. Individual ANOVAs (2 Groups \times 2 Cues) of each FFI group against the CC were conducted using Bonferroni adjusted alpha levels of 0.017 per test (0.05/3). The results yielded a significant interaction of group and cue for the CC group versus the VG group, $F(1,35) = 18.73$, $p < 0.001$; for the CC group versus the VS group, $F(1,32) = 25.84$, $p < 0.001$; and for the CC group versus the VP group, $F(1,32) = 8.51$, $p = 0.006$. All FFI treatments therefore increased sensitivity to the verb cue.

Figure 4 shows the reliability of these patterns across individual group members. Most CC individuals were predominantly influenced by the adverb cue, whereas most VS and VG participants were more influenced by the verb cue. Verb pretraining participants were more scattered: most showed greater sensitivity to the verb, though there were some who lay close to the 45° diagonal, suggesting that they were more evenly affected by both cues.

Eye-Tracking Data

Sentence Exposure

Figure 5 and Table 2 show the group mean fixation duration of these participants as they were studying the adverb and verb cues during exposure to the Latin sentences. Figure 5A shows the total fixation duration on these cues. Figure 5B shows these data as the proportion of the total fixations on each trial. The pattern in these Figures is clear, all groups looked at the verb more than the adverb, but it was the three FFI groups that did so to a greater extent. Individual ANOVAs on the total fixations (2 Groups \times 2 Cues, with subjects nested within groups) were conducted using Bonferroni adjusted alpha levels of 0.017 per test (0.05/3). The

results revealed a significant group by cue interactions for the VG group versus the CC group, $F(1,29) = 6.86$, $p = 0.014$; for the VS group versus the CC group, $F(1,28) = 17.71$, $p < 0.001$; but the interaction marginally failed to reach significance for the VP group versus the CC group, $F(1,28) = 4.06$, $p = 0.05$. VG and VS therefore paid more attention than the CC group to the verb cue during processing.

Correlations between Attention to Cue in Sentence Exposure and Subsequent Cue Comprehension

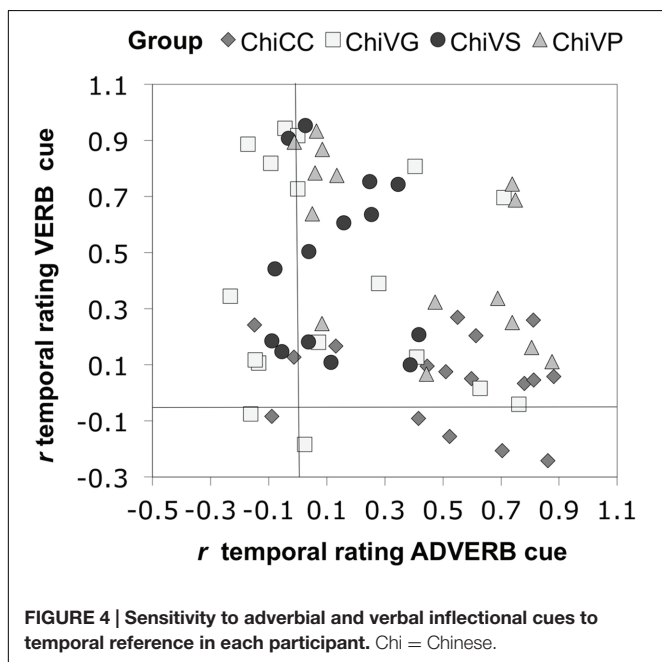
Pearson correlations investigating the relations between attention in the Sentence Exposure phase and comprehension ability in the Comprehension Phase across all the participants and groups of Experiment 1 show that the proportion of fixation time spent on the adverb during Sentence Exposure correlates significantly with later adverbial bias in Comprehension ($r = 0.50$, $p < 0.001$). Likewise, proportion of fixation time spent on the verb during Sentence Exposure correlates significantly with later verb bias in Comprehension ($r = 0.45$, $p < 0.001$).

Sentence Exposure Eye-Tracking Over Trials

Although the random order of stimuli was different for each participant, we can determine the degree to which the participants attended to the verb and adverb cues over trials. Figure 6A shows the total fixation on each cue by trial of experience in all L1 Chinese groups. It can be seen that CC participants initially spent more time looking at the verb, but interest in this cue waned over trials and more attention was paid to the adverbial cue. Participants in the three FFI conditions, however, maintained a steady attentional preference for the verb cue. These patterns are clearer in Figure 6B, which plots the proportion of fixation time on each trial spent on the adverb and verb cues, respectively.

DISCUSSION

The behavioral results of Experiment 1 show that under CC, adverbs were better attended than verb inflections. This finding replicates that of Ellis and Sagarra (2010, 2011) and Cintrón-Valentín and Ellis (2015). In the linguistic input, adverbial cues are more salient, simple and reliable cues compared to the verb-tense inflections. Furthermore, the adult language learners' prior use of adverb temporal reference in their Chinese L1 could have resulted in long-term blocking. In contrast to the CC treatment, training on the isolated verb cue under the VP condition reversed this bias, resulting in a better use of the verb cue during comprehension. This finding also replicates that of Ellis and Sagarra (2010, 2011) and Cintrón-Valentín and Ellis (2015), showing short-term learned attention effects, where prior learning of an isolated cue during pretraining shifts learners' attention to that cue in subsequent testing. In the two other FFI conditions, VG, where learners were first exposed to a short instructional sequence on how Latin verb-tense morphology works in Latin, and VS, where the verb inflections were made more salient by means of textual enhancement manipulations during exposure, participants were better able to use the verb



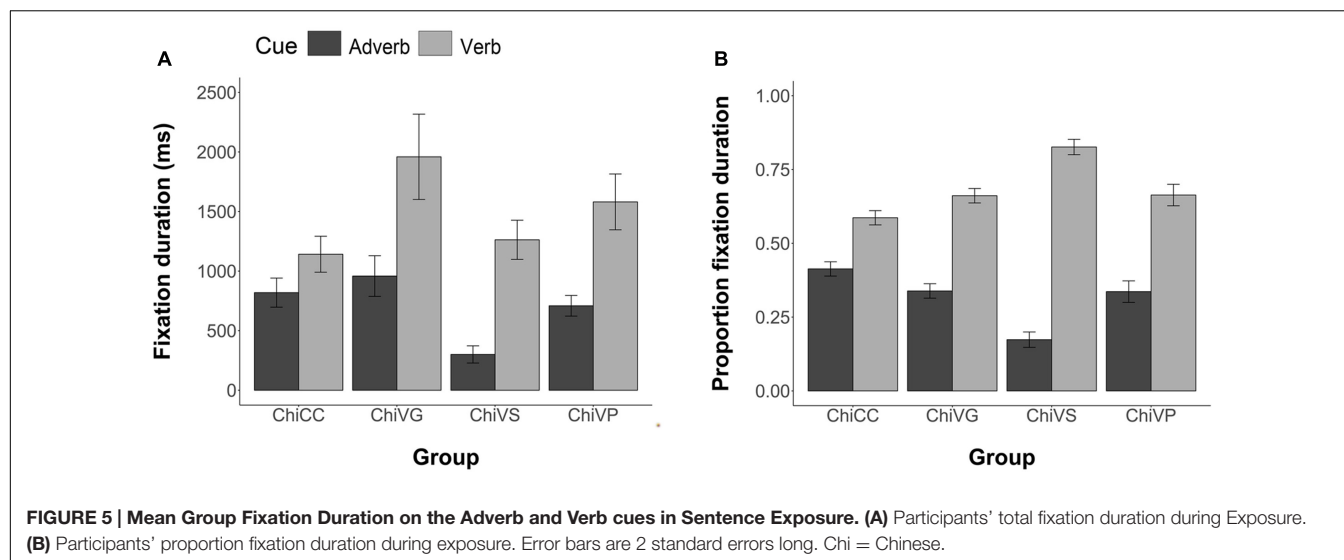


TABLE 2 | Mean participant fixations on the adverb and verb cues by the four groups of Experiment 1.

Group	Cue	Mean	95% CI
Mean total fixation duration (ms.)			
Control	Adverb	819.2	[402, 1236]
	Verb	1141.6	[628, 1656]
Verb grammar	Adverb	958.7	[340, 1577]
	Verb	1959.2	[662, 3256]
Verb salience	Adverb	300.3	[23, 578]
	Verb	1262.7	[631, 1894]
Verb pretraining	Adverb	708.8	[375, 1041]
	Verb	1581.1	[681, 2480]
Mean proportion fixation time			
Control	Adverb	0.41	[0.33, 0.50]
	Verb	0.59	[0.50, 0.67]
Verb grammar	Adverb	0.34	[0.25, 0.43]
	Verb	0.66	[0.58, 0.75]
Verb salience	Adverb	0.17	[0.07, 0.27]
	Verb	0.83	[0.73, 0.93]
Verb pretraining	Adverb	0.34	[0.20, 0.48]
	Verb	0.66	[0.52, 0.80]

cues in comprehension relative to the adverb cue than those in the CC condition. Of the three FFI conditions, VP resulted in more balanced acquisition of both verbal and adverbial cues.

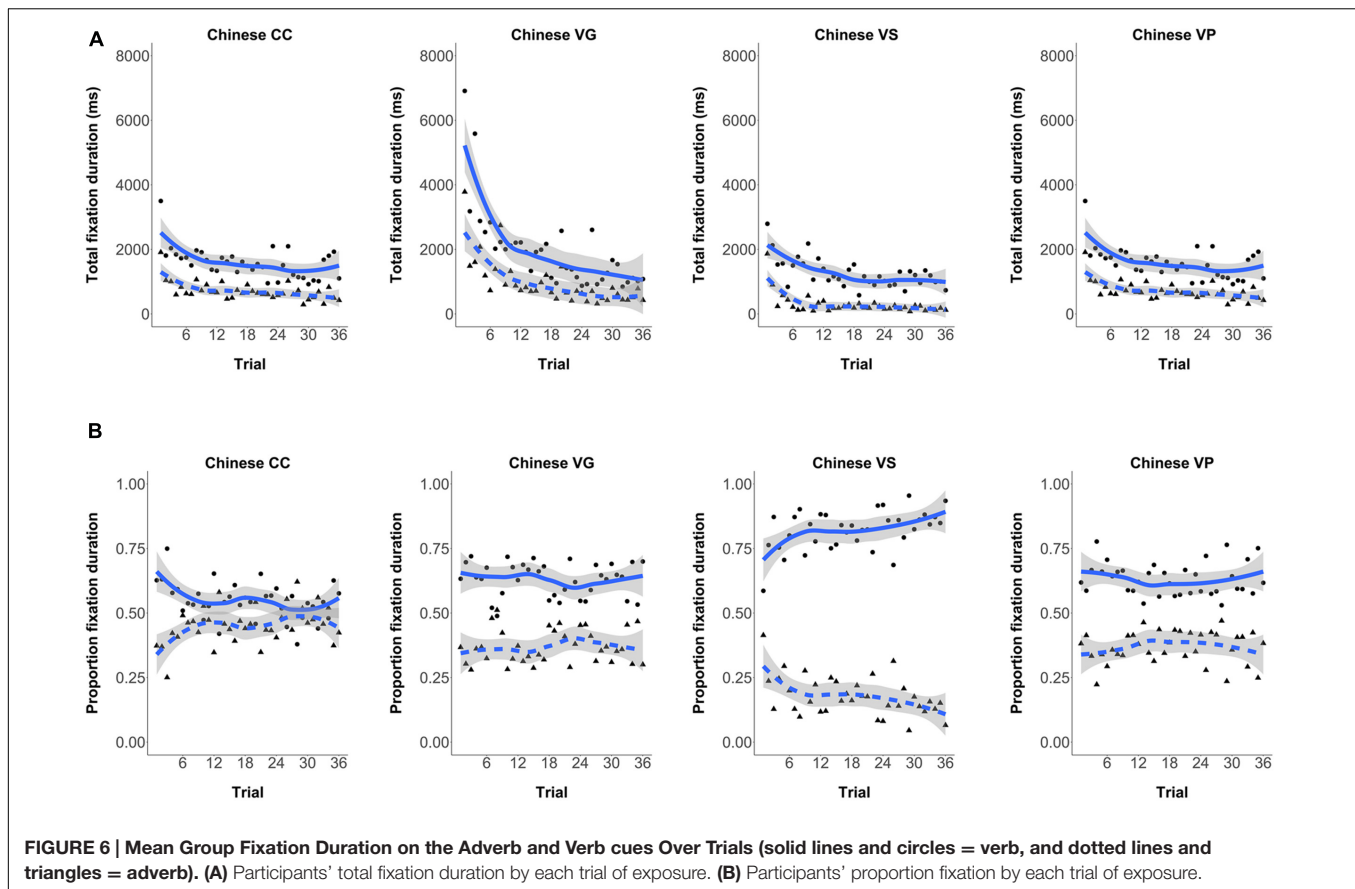
The eye-tracking data show how these FFI treatments affected attention to cues in the input processing. All participants looked at the verbs more than they did the adverb during sentence exposure. However, participants in the VG and VS conditions fixated upon the verbs significantly more during input processing than did Control participants. The VP group, however, did not differ significantly from the control group, although the same numerical trend was evident. The correlation analyses suggest that the relative amount participants spent processing the verb/adverb cues during exposure determined cue usage in

subsequent comprehension testing. The trial-by-trial analyses illustrated in **Figure 6**, show that CC participants initially spend more time looking at the verb, however, participants rapidly lose interest in the verb cue across trials and more attention is paid to the adverbial cue. One possible interpretation is that learners initially first fixate more on the verb + inflection because it is the longer word form, however, over trials they come to realize that the adverb is the simpler and more reliable cue, and as a result they shift their attention to it. The FFI participants on the other hand – for whom the verb forms or their functions were made more salient – pay more attention to verb from the start of language exposure, and this focus persists, leading to subsequent attention and use of this cue.

EXPERIMENT 2

Introduction

Our previous studies examining the effects of learned attention and blocking and the effects of FFI in overcoming learned attentional biases in L2 acquisition Ellis and Sagarra (2010, 2011) and Ellis et al. (2014) have focused on the learning of Latin only through the visual modality. As described in the introduction, spoken and written language are very different mediums. Whereas readers have the advantage of being able to control the amount and speed at which they process visual input, the fleeting nature of spoken language does not afford listeners the same advantage. These differences could well-affect the degree to which different language forms are salient in the input and thus control the degree to which they are attended, perceived, processed, and learned. Indeed Leow (2015, p. 122), in reviewing the relevance for instruction of this work on learned attention, explicitly asks for a potential replication study which addresses the issue of whether the findings can be extrapolated to the aural mode. This experiment therefore aims to replicate and extend previous work by comparing the attentional processes of L1 speakers of English in control (CC), VG, VS, and VP conditions



who learn from aural input with those whose input experience is visual.

Participants

Participants were 200 individuals recruited from a major university in the USA. They were volunteers who participated as part of an undergraduate psychology course requirement ($n = 182$) or were paid \$10 for their participation ($n = 18$). Inclusion criteria required participants to be native English speakers who had not learned Latin or Italian previously. They could know Spanish but could not have been raised bilingually before the age of 6 years. They were randomly assigned to one of eight conditions regarding instruction and modality of presentation. Those who received Aural presentation only were split into CCA (Control Condition Aural), $n = 25$ (15 females and 10 males), age range 17–45 years ($M = 21.56$); VGA (Verb Grammar Aural), $n = 25$ (16 females and 9 males), age range 18–22 years ($M = 18.84$); VSA (Verb Saliency Aural), $n = 25$ (22 females and 3 males), age range 17–20 years ($M = 18.36$); and VPA (Verb Pretraining Aural), $n = 25$ (15 females and 10 males), age range 18–20 years ($M = 18.44$). Participants who received instruction in the Visual modality only were split into CCV, $n = 25$ (18 females and 7 males), age range 18–21 years ($M = 18.40$); VGV (Verb Grammar Visual), $n = 25$ (8 females and 17 males), age range 17–22 years ($M = 18.68$); VSV (Verb Saliency Visual), $n = 25$ (9 females and 16 males),

age range 18–20 years ($M = 18.68$); and VPV (Verb Pretraining Visual), $n = 25$ (14 females and 11 males), age range 18–21 years ($M = 18.52$).

Procedure

The experiment was programmed in *PsychoPy* (Peirce, 2007) and consisted of the same phases as presented in Experiment 1 (see **Table 3** for detailed procedure). However, the stimulus set for Experiment 2 was more complex than that of Experiment 1. In Experiment 1 participants were presented with one verb stem, *cogit-*, which was combined with all appropriate past, present and future inflections, whereas in Experiment 2, participants were presented with four different verb stems and their appropriate past, present and future inflections.

Pretraining

Participants in the VP group were first pretrained on verb inflections and determined that each made reference to either present or past time. On each trial, Visual participants saw, or Aural participants heard, one of the four verb stems (*cant-*, *flea-*, *nat-*, *pugn-*) combining an inflection referencing the past (*-avi* *-avisti*, *-avit*) or present (*-o*, *-as*, *-at*). Participants were additionally presented with a picture of a stick figure that represented the action of the verb (see **Figure 7**). They were asked to select the Latin verb's temporal (past/present) reference from an on-screen menu. Feedback was provided on their responses.

TABLE 3 | The design of Phases 1–3 of Experiment 2.

Pretraining (Phase 1) (+ feedback)	Sentence Exposure (Phase 2) (+ feedback) 48 (24 × 2) randomized trials	Comprehension Test (Phase 3) (– feedback) 48 randomized trials	
		Stimulus	Semidiem
Control group	Present		
No pretraining	<i>hodie cantas</i>	<i>verb-o/as/at cras</i>	3
Verb Pretraining group	<i>hodie fleat</i>	<i>verb-avi/visti/vit cras</i>	1
(36 randomized trials)	<i>hodie pugnās</i>	<i>verb-abo/ bis/ bit cras</i>	5
<i>nato</i> “I swim”	<i>hodie nato</i>	<i>hodie verb-o/as/at</i>	3
<i>cantas</i> “You sing”	<i>cantas hodie</i>	<i>hodie verb-avi/visti/vit</i>	2
<i>pugnās</i> “You fight”	<i>fleat hodie</i>	<i>hodie verb-abo/bis/bit</i>	4
<i>fleat</i> “He/She cries”	<i>pugnās hodie</i>	<i>heri verb-o/as/at</i>	2
<i>cantavi</i> “I sang”	<i>nato hodie</i>	<i>heri verb-avi/visti/vit</i>	1
<i>natavisi</i> “You swam”	Past	<i>heri verb-abo/bis/bit</i>	3
<i>pugnāvit</i> “He/She fought”	<i>heri cantavi</i>	<i>cras verb-o/as/at</i>	4
<i>fleāvit</i> “He/She cried”	<i>heri fleāvi</i>	<i>cras verb-avi/visti/vit</i>	3
Verb Saliency group	<i>heri fleāvit</i>	<i>cras verb-abo/bis/bit</i>	5
No pretraining	<i>heri pugnāvit</i>	<i>verb-o/as/at hodie</i>	3
Verb Grammar group	<i>heri natavisi</i>	<i>verb-avi/visti/vit hodie</i>	2
Brief Grammar Lesson	<i>cantavi heri</i>	<i>verb-abo/bis/bit hodie</i>	4
See Figure 1	<i>fleāvi heri</i>	<i>verb-o/as/at heri</i>	2
	<i>fleāvit heri</i>	<i>verb-avi/visti/vit heri</i>	1
	<i>pugnāvit heri</i>	<i>verb-abo/bis/ bit heri</i>	3
	<i>natavisi heri</i>		
	Future		
	<i>cras cantabit</i>		
	<i>cras pugnābis</i>		
	<i>cras natabo</i>		
	<i>cantabit cras</i>		
	<i>pugnābis cras</i>		
	<i>natabo cras</i>		

The rating scale for the Comprehension Test ranged from 1 (past) to 5 (future). The correct answer for each trial is shown in the semidiem column.

In this phase they were not asked about the verb meaning, thus their understanding was focused upon the morphological tense reference.

Pretraining for the VG participants involved a brief lesson on Latin verb-tense morphology using similar slides to those shown in Figure 1, except that the Latin verb *amare* was used as an example in slides 2 and 3. Regardless of modality of language exposure these slides were presented visually.

Sentence Exposure

In Sentence Exposure, Visual participants saw, Aural participants heard, 24 different sentence combinations, which appropriately combined the adverb with a verb stem (see Table 3). While the sentence was exposed, participants saw onscreen a picture of a stick figure which appropriately represented the action the verb was referencing. Again, in this phase they were not asked to make any judgments regarding the picture they were shown. After each sentence, participants were asked to identify whether the sentences referred to the past, present, or future, responding via a visual menu presented on the computer screen. The sample of stimuli selected for presentation in Sentence Exposure ensured that each verb root was (1) presented once in each tense, and (2) appropriately combined one of the agreement markers for each tense. The Sentence Exposure procedure was identical for the CC, VG, and VP groups. For VS participants only, the stimuli were

either textually or aurally enhanced to increase their saliency, so for Visual presentation the verbal inflections were highlighted in bold and red, and for Aural presentation the verb inflections were spoken emphatically. Feedback was given for both correct and incorrect choices. For correct answers, the word “correct” would appear on the screen, whereas for incorrect answers, participants would see the word “wrong” accompanied by the correct answer (e.g., “Wrong – [*heri cantavit*] is [past]”).

Comprehension Test

Here, participants were presented with randomized verb-adverb/adverb-verb combinations as well as a selection of single word items. The single word items were verbs (*canto*, *fleat*, *natat*, *pugnās*, *cantavi*, *fleāvit*, *natavit*, *pugnāvisi*, *cantabit*, *fleābis*, *natabo*, *pugnābit*), half of which had been previously presented in the same inflection during Sentence Exposure. For the randomized verb-adverb/adverb-verb combinations, similar to Experiment 1, participants experienced sentences that were familiar to them from the previous task, but also combinations in which the verb and adverb were incongruent in their time reference. Here, participants saw six congruent combinations they had previously experienced during Sentence Exposure (*heri fleāvi*, *heri pugnāvit*, *heri natavisi*, *hodie cantas*, *hodie nato*, *cras pugnābis*) as well as six new congruent combinations they had not seen before (*heri natavi*, *heri cantavit*, *hodie fleo*,

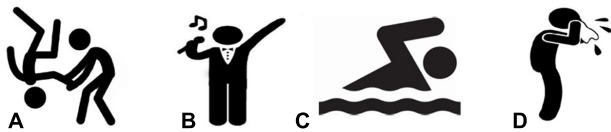


FIGURE 7 | Pictures presented during Verb Pretraining, Exposure and Comprehension testing. During Pretraining and Exposure each picture was presented in the presence of the corresponding Latin verb (either aurally or visually, depending on the condition), where **(A)** corresponded to any form of the verb *pugno* (representing the verb *to fight* in English); **(B)** corresponded to any form of the verb *canto* (representing the verb *to sing* in English); **(C)** corresponded to any form of the verb *nato* (representing the verb *to swim* in English); and **(D)** corresponded to any form of *fleo* (representing the verb *to cry* in English). During the Comprehension testing phase, a pane where all four pictures were present was shown in each trial, again in the presence of the corresponding Latin verb.

hodie fleas, cras cantabis, cras pugnabo). For the trials involving incongruent combinations, each of the verbs used for the congruent combinations were combined with all possible adverb forms. Overall this led to a total of 12 single-word items and 36 two-word combinations. In each of the trials, participants were additionally presented with a four-picture pane menu, where they saw the four pictures they had been previously presented with during Sentence Exposure. The position of the pictures was counterbalanced in the pane.

On each trial, participants were asked to make two judgments. The first judgment was whether the word string referred to the past, present, or future on a five-point scale. The possible scale points were the same as in Experiment 1. For the second judgment, participants were asked to select the picture that best represented the word or phrase they were presented with. This judgment tested how well they had processed the meaning of the verbs to which they had been exposed. Feedback was not provided.

RESULTS

Visual Modality

Verb Pretraining Data

Mean performance in the first quarter of Verb Pretraining was 63%. By the fourth quarter, mean performance was 86%, demonstrating acceptable completion of Phase 1.

Sentence Exposure Data

Mean performance in the first quarter of Sentence Exposure was 61% for the CCV group, 84% for the VGV group, 68% for the VSV group, and 82% for the VPV group. Both the VGV and the VPV groups were at an advantage in the first quarter compared to the other groups. However, performance evened out for all groups by the end of the phase. Mean performance in the final quarter was 96% for the CCV group, 98% for the VGV group, 93% for the VSV group, and 95% for the VPV group. A one-way ANOVA on these final quarter scores did not reveal a significant group effect, $F(3,96) = 1.71$, $p = 0.17$.

Comprehension Data

Perception of Time Cues

Each participant's temporal rating responses for the strings in Comprehension testing were correlated with the information provided by the verb cue and the information separately provided by the adverb cue to determine the degree to which each participant was biased by each cue type. Pearson correlations between each participant's temporal rating responses and the information provided by the verb and adverb cues separately are illustrated in **Figure 8A**. CCV participants were more influenced by the adverb, $M = 0.89$, 95% CI = [0.82, 0.96] than the verb, $M = 0.02$, 95% CI = [-0.03, 0.07]. VGV participants were more influenced by the verb, $M = 0.78$, 95% CI = [0.59, 0.97], than by the adverb, $M = 0.03$, 95% CI = [-0.07, 0.12]. VSV participants were more influenced by the verb, $M = 0.73$, 95% CI = [0.59,

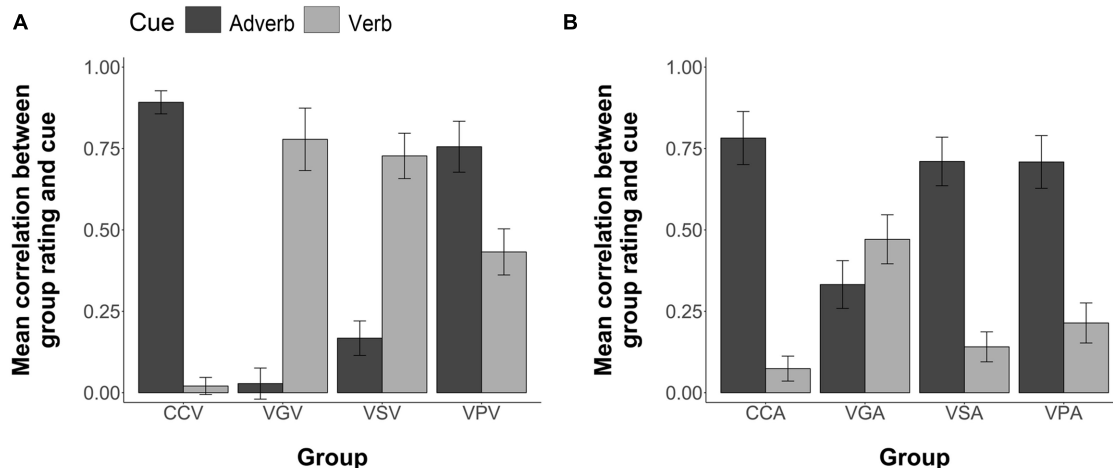


FIGURE 8 | Group mean correlations between individual participants' Comprehension sentence ratings and the information given by the corresponding adverb and verb cues. (A) Group mean correlations for the visual modality treatments **(B)** Group mean correlations for the aural modality treatments. Error bars are 2 standard errors long.

0.86], than by the adverb, $M = 0.17$, 95% CI = [0.06, 0.27]. Contrary to the other FFI conditions VPV participants were more influenced by the adverb, $M = 0.81$, 95% CI = [0.60, 0.91], but showed some sensitivity toward the verb cue, $M = 0.43$, 95% CI = [0.29, 0.57].

If we compare the comprehension data for the VPV participants with that of the Chinese VP participants in Experiment 1 (verb: $M = 0.61$, 95% CI = [0.45, 0.77]; adverb: $M = 0.47$, 95% CI = [0.30, 0.64], the pattern is quite different, with the VPV participants showing a greater degree of sensitivity toward the adverb cue relative to the verb cue. Although the VPV participants showed an increase in their sensitivity toward the verb cue when, compared to the CCV group (as confirmed by our analysis of variance below), it seems that the greater complexity of the stimulus set in Experiment 2 had an impact on the learners' attentional focus, and thus on the degree of sensitivity they showed toward the verbal morphological cues during comprehension.

An ANOVA (4 Groups \times 2 Cues, with subjects nested within groups) revealed a significant group by cue interaction, $F(3,96) = 7.80$, $p < 0.001$. As in Experiment 1, individual ANOVAs (2 Groups \times 2 Cues) of each FFI group against the CC were conducted using Bonferroni adjusted alpha levels of 0.017 per test (0.05/3). The results yielded a significant interaction of group and cue for the CCV group versus the VGV group, $F(1,48) = 82.19$, $p < 0.001$; for the CCV group versus the VSV group, $F(1,48) = 73.72$, $p < 0.001$; and for the CCV group versus the VPV group, $F(1,48) = 8.68$, $p = 0.004$. These results replicate those of Experiment 1, where the FFI conditions increased sensitivity to the verb cue.

Figure 9A shows the reliability of these patterns across individual group members. For the groups in the visual modality, most CCV participants were influenced by the adverb cue, whereas most VSV and VGV participants were more influenced

by the verb cue. VPV participants were more scattered: most showed greater sensitivity to the adverb, though there were some who showed greater sensitivity toward the verb, and one participant who lay close to the 45° diagonal, suggesting that they were more evenly affected by both cues.

To determine if the effects of FFI on cue use during comprehension testing differed based upon the nature of the items, that is, whether they were trained (i.e., previously presented during Sentence Exposure) or generalization items (i.e., only presented during Comprehension Testing) we ran a three-way ANOVA (4 Groups \times 2 item type \times 2 cues). The analyses revealed a non-significant effect of item type $F(1,96) = 0.19$, *n.s.*, a statistically significant group by cue interaction, $F(3,96) = 28.56$, $p < 0.001$, but no significant three-way interaction between item type (Trained or Generalization), group, and cue use, $F(3,96) = 0.69$, $p = 0.56$. Thus, participants performed at a similar level, regardless of item type (see Figure 10).

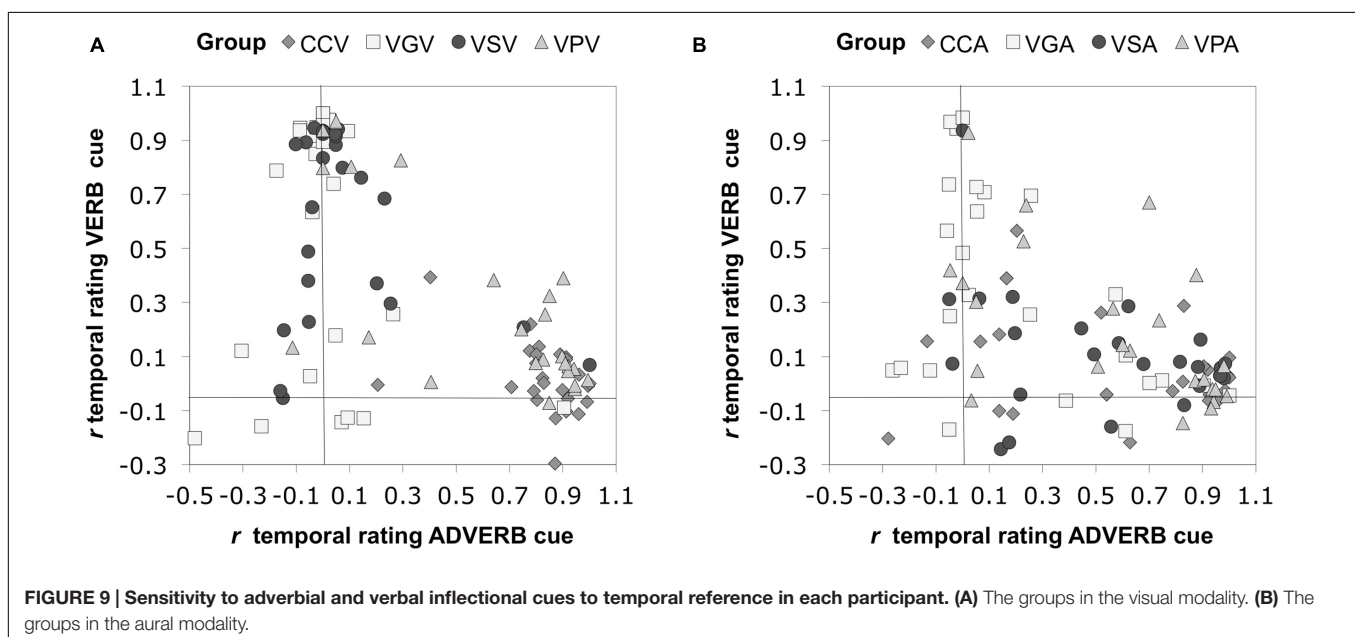
Perception of Verb Meaning

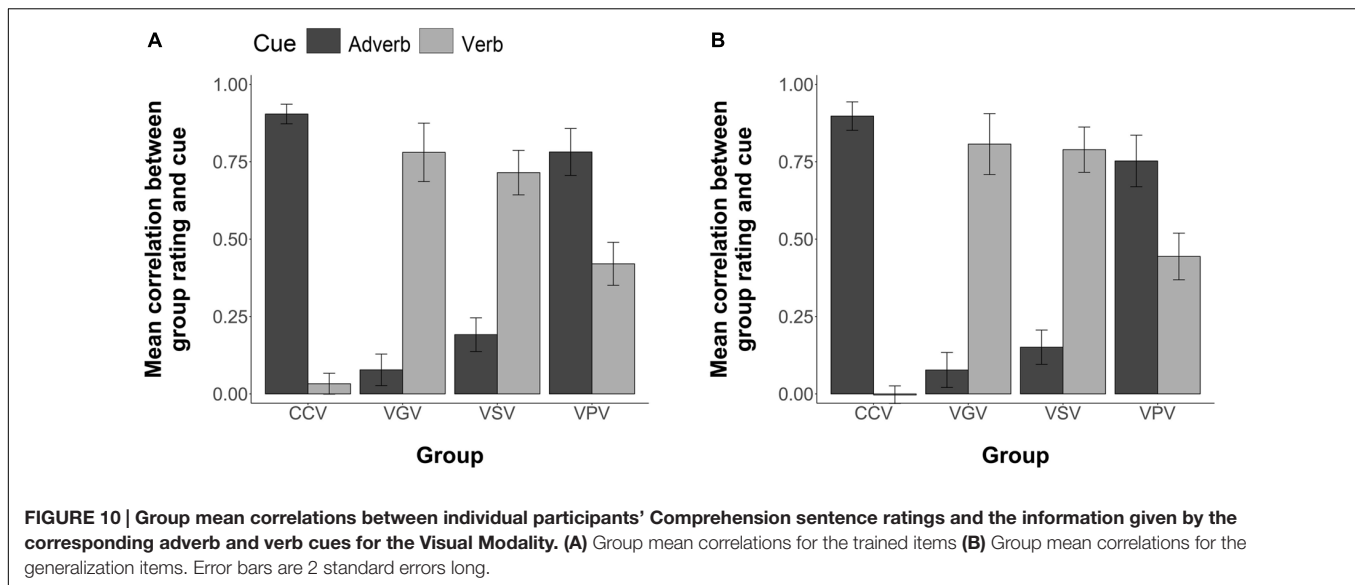
Mean accuracy scores for verb meaning was 0.82 for the CCV group, 0.80 for the VGV group, 0.76 for the VSV group, and 0.91 for the VPV group. A one-way ANOVA on each of the conditions' mean accuracy scores for the picture ratings showed a non-significant effect of group, $F(3,96) = 2.59$, $p = 0.06$. *Post hoc* Tukey HSD tests demonstrated just one significant pairwise group difference: between the VPV group, and the VSV group, $p = 0.04$, 95% CI = [0.004, 0.30].

Aural Modality

Verb Pretraining Data

Mean performance in the first quarter of Verb Pretraining was 66%. By the fourth quarter, mean performance was 80%, demonstrating acceptable completion of Phase 1.





Sentence Exposure Data

Mean performance in the first quarter of Sentence Exposure was 49% for the CCA group, 50% for the VGA group, 54% for the VSA group, and 67% for the VPA group. The pretraining on the verb allowed the VPA participants to be at an advantage in the first quarter compared to the other groups, and this advantage also persisted. Mean performance in the final quarter was 78% for the CCA group, 68% for the VGA group, 77% for the VSA group, and 87% for the VPA group. A one-way ANOVA on these final quarter scores revealed a significant group effect, $F(3,96) = 3.12$, $p = 0.03$. *Post hoc* Tukey HSD tests demonstrated one significant pairwise group difference: between the VPA group, and the VGA group, $p = 0.02$, 95% CI = [0.02, 0.35].

Comprehension Data

Perception of Time Cues

Pearson correlations between each participant's temporal rating responses and the information provided by the verb and adverb cues separately are illustrated in **Figure 8B**. CCA participants were more influenced by the adverb, $M = 0.78$, 95% CI = [0.62, 0.94] than the verb, $M = 0.07$, 95% CI = [-0.001, 0.15]. VGA participants were more influenced by the verb, $M = 0.47$, 95% CI = [0.32, 0.62], than by the adverb, $M = 0.33$, 95% CI = [0.19, 0.48]. Contrary to the VGA condition, VSA participants were more influenced by the adverb, $M = 0.71$, 95% CI = [0.56, 0.86], than by the verb, $M = 0.14$, 95% CI = [0.05, 0.23]. Likewise, VPA participants were more influenced by the adverb, $M = 0.71$, 95% CI = [0.55, 0.99], but showed some sensitivity toward the verb cue, $M = 0.21$, 95% CI = [0.09, 0.33].

The general patterns observed here were reliable across individual group members. **Figure 9B** shows the aural modality data. Again, most CCA individuals were predominantly influenced by the adverb cue. VGA participants showed greater sensitivity toward the verb, whereas VSA participants were predominantly influenced by the adverb. Similar to those in the visual modality, VPA participants were more scattered: most

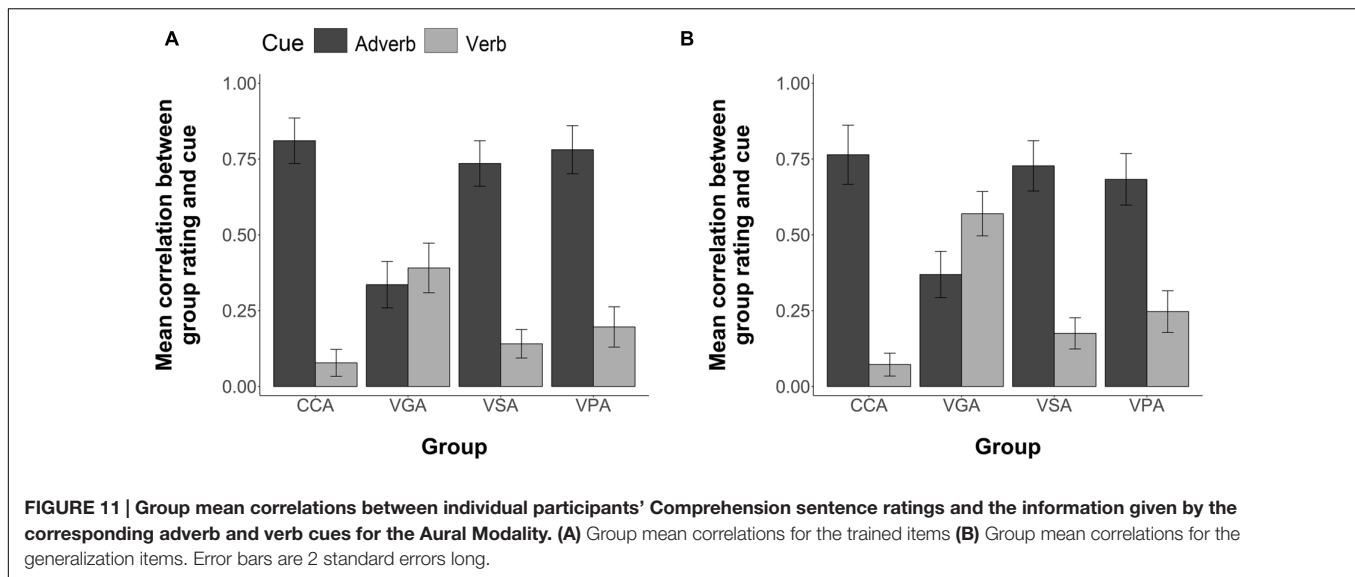
showed greater sensitivity toward the adverb, a small group of participants showed greater sensitivity toward the verb, and one participant lay close to the 45° diagonal, suggesting that they were more evenly affected by both cues.

An ANOVA (4 Groups \times 2 Cues, with subjects nested within groups) revealed a significant group by cue interaction, $F(3,96) = 5.53$, $p = 0.002$. Individual ANOVAs (2 Groups \times 2 Cues) of each FFI group against the CCA group yielded a significant interaction of group and cue for the CCA group versus the VGA group, $F(1,48) = 13.09$, $p < 0.001$; a significant main effect of cue for the CCA group versus the VSA group, $F(1,48) = 35.32$, $p < 0.001$, but no significant group by cue interaction, $F(1,48) = 0.63$, $p = 0.43$; and a significant main effect of cue for the CCA group versus the VPA group, $F(1,48) = 30.15$, $p < 0.001$, but no significant group by cue interaction, $F(1,48) = 1.07$, $p = 0.31$. Contrary to the results of Experiment 1 and those for Visual presentation described in 3.4.1, it seems only the VGA group increased sensitivity to the verb cue.

As for the visual modality, to determine if the effects of FFI on cue use during comprehension testing differed based upon the nature of the items (i.e., Trained or Generalization) we ran a three-way ANOVA (4 Groups \times 2 item type \times 2 cues). The analyses revealed a non-significant effect of item type $F(1,96) = 0.0002$, *n.s.*, a statistically significant item type by cue interaction, $F(1,96) = 8.04$, $p = 0.006$, but no significant three-way interaction between item type (Trained or Generalization), group, and cue use, $F(3,96) = 0.70$, $p = 0.55$. Thus, similar to the visual modality, participants in the aural modality performed at a similar level, regardless of item type (see **Figure 11**).

Perception of Verb Meaning

Mean accuracy scores for verb meaning was 0.59 for the CCA group, 0.56 for the VGA group, 0.63 for the VSA group, and 0.80 for the VPA group. A one-way ANOVA on each of the conditions' mean accuracy scores for the picture ratings showed a significant effect of group, $F(3,96) = 5.35$, $p = 0.002$. *Post*



hoc Tukey HSD tests demonstrated three significant pairwise group differences: between the VPA group and the CCA group, $p = 0.01$; between the VPA group and the VGA group, $p = 0.002$; and between the VPA group and the VSA group, $M = 0.62$, $p = 0.05$.

Modality by FFI Interactions

To determine if the effects of FFI on cue use during comprehension testing differed across modality of presentation, we ran a three-way ANOVA (4 Groups \times 2 modalities \times 2 cues). The analyses revealed a statistically significant three-way interaction between modality of input presentation (Aural or Visual), group, and cue use, $F(3,384) = 9.85$, $p < 0.001$.

Inspection of **Figure 8** shows the major loci of this interaction. Under CC, participants process the adverb and pay little or no heed to the verb. This is so for CCA and CCV. Pretraining on the verb in VP allows them a little better use of this cue, especially in VPV, but still it is overshadowed by the adverb. Making the verbal inflections salient during sentence exposure with visual presentation VSV allows learners to attend and learn to use verb morphology. But this is absolutely not so with auditory presentation VSA. Grammar instruction, however, does allow learners to make use of the morphological cues, both from auditory presentation, and particularly from visual exposure.

DISCUSSION

The findings for the Visual conditions follow that of the prior learned attention studies. The CCV group showed greater sensitivity toward the adverb than the verb cue. The VPV treatment allows participants to show sensitivity toward the verb cue, while also showing sensitivity to use of adverbs. As in Experiment 1, and in Cintrón-Valentín and Ellis (2015), both VG and VS treatments in the visual modality

shifted learners' attention to the verb cue in subsequent testing.

The general behavioral data for the Aural conditions show consistency in that learners in the CCA group focus more on the adverb than the verb cues. However, the other findings are in contrast to these patterns following Visual exposure. Of the three FFI conditions, only VGA produced a shift in their attention toward the verb cue. VSA participants' performance was similar to that of the CCA participants, showing more sensitivity toward the adverb relative to the verb cue. Although VPA participants showed an increase in their verb sensitivity, when compared to that of the CCA group, their attention was greater toward the adverb cue than to the verb cue. We will discuss these disparities below.

GENERAL DISCUSSION

Experiment 1 tested whether under normal conditions of exposure (CC), the effects of physical salience and learned attentional biases toward adverbial cues, prejudice the acquisition of verbal tense morphology, as indexed in participants' relative reliance on these cues in subsequent language comprehension (RQ1). The results of Experiment 1 lend support to the idea that the limited attainment of adult second and foreign language learners follows general principles of associative learning and cognition wherein saliency and attention are key factors (Ellis, 2006a,b). Under normal conditions of language exposure (CC), adverbs were better processed than verb inflections. We interpret this phenomenon, a standard finding in SLA research (e.g., Schmidt, 1984; Bardovi-Harlig, 1992, 2000; Noyau et al., 1995; Van Patten, 1996, 2006; Clahsen and Felser, 2006), as relating firstly to the relative saliency, simplicity and reliability of adverb cues which render them more learnable when compared to verb-tense morphology, and secondly, to adult language learners' prior knowledge of the use of adverb temporal reference in their L1 which results in the long-term blocking of these forms. This

is apparent in adult language learners' difficulty in learning morphology compared to child learners, and in studies such as Ellis and Sagarra (2010, Experiment 2) and Ellis and Sagarra (2011, Experiments 2 and 3) which demonstrate long-term language transfer effects whereby the nature of learners' first language (+/- verb tense morphology) biased the acquisition of morphological versus lexical cues to temporal reference in Latin. First language speakers of Chinese (no tense morphology) were less able than first language speakers of Spanish or Russian (rich morphology) to acquire inflectional cues from the same language experience.

Experiment 1 also tested whether early experience of morphological cues to temporal reference, through each of the FFI treatments VG, VS, and VP, counteract the effects of physical saliency and learned attentional biases, as indexed by participants' relative reliance on these cues in subsequent language comprehension (RQ2). The behavioral data demonstrated that all FFI interventions resulted in better attention to and use of the verbal inflectional cues. Participants in the VG group were initially provided with declarative statements about morphological function but still had to put this knowledge to use in subsequent phases. VS learners were introduced to the verbal cues during the exposure phase but still had to determine their function. Both of these treatments resulted in participants' attending these cues and using them over adverbial cues. In contrast, the verb pretraining in VP, where learners had to process the Latin verb forms for meaning in English, resulted in a more balanced acquisition of both verbal and adverbial cues. We believe that this is because, having learned to some extent how to use the morphology, they were next able to consider the role of adverbs too. This interpretation is consonant with other findings in the literature, where in the early stages of learning, when learners are confronted with multiple cues to interpretation, they typically focus upon one cue at a time. As they reduce errors of estimation regarding the outcome or interpretation of the cue, they then consider the role of the other cues (MacWhinney, 2001).

The results of the VG condition are consonant with prior findings in the literature on FFI (see, for instance, the meta-analyses of effects of type of instruction by Norris and Ortega, 2000; Spada and Tomita, 2010; Goo et al., 2015), which suggest that instructional conditions involving a focus on the rules underlying specific L2 structures generally lead to large advantages in the acquisition of target forms. In terms of the results obtained for the VS condition, as we explained in the introduction, reviews of the effectiveness of textual enhancement (TE) have yielded conflicting findings (Han et al., 2008; Lee and Huang, 2008), largely due to a wide variety of methodological differences. However, one pattern that seems to apply is that the provision of compound enhancement, that is, "TE in combination with attention-getting strategies such as corrective feedback" (Han et al., 2008, p. 609) tends to be more effective in encouraging noticing and subsequent processing than simple enhancement. The sentence exposure phase for our VS participants involved exactly this – visual saliency and corrective feedback.

Two additional research questions in Experiment 1 concerned (i) whether early experience of morphological cues to temporal reference lead to biases in subsequent overt perceptual attention (as indexed by number of fixations) during Sentence Exposure, where there are both adverbial and morphological cues to the same interpretation (RQ3), and (ii) whether any bias in overt attention to input cues in turn lead to subsequent attentional biases to the adverbial or morphological cues in subsequent language comprehension (RQ4). The eye-tracking data from Experiments 1 show how the FFI treatments affected attention to cues during input processing. All participants fixated upon the verbs significantly more than they did the adverb. As shown by the group by cue interactions, participants in the VG and VS conditions attended the inflections more than the control participants, as did the VP participants although in the latter case the interaction failed to reach significance. Additionally, the correlation analyses showed that the relative amount participants spent processing the verb/adverb cues during language exposure determined their cue usage in subsequent comprehension.

The trial-by-trial analyses of **Figure 6** show that although control participants initially spent more time looking at the verb, their interest in this cue waned over trials, with more attention being paid to the adverbial cue. One interpretation is that CC learners initially first fixated more on the verb + inflection because it is the longer more salient form. Initially, over the first 20 trials or so, they tried to induce the system of how the inflections signal temporality, but realizing that the adverb is the simpler and more reliable cue, they eventually shifted their attention to it. The FFI groups on the other hand – who were initially made aware of the verb forms or their functions, (1) by having the verb-inflections explained during pretraining (VG), (2) by having the inflections made more salient by textual enhancement during exposure (VS), (3) or by being pretrained on the verbal cues in non-redundant situations (VP) – paid more attention to the verbs from the start of the exposure phase, and this persisted through the end of the trials. Overall, the eye movement findings in the current study replicate with Chinese L1 speakers those of Cintrón-Valentín and Ellis (2015) with English L1 learners.

Experiment 2 investigated the effects of modality of language exposure upon the saliency and consequent processing of linguistic form, and on the ways in which different types of FFI interact with these very different mediums. The fleeting nature of spoken language does not afford listeners the control of scrutiny of input as does visual presentation, and these differences could well-affect the degree to which forms are salient in the input. This experiment therefore compared the attentional processes of L1 speakers of English in control (CC), VG, VS, and VP conditions who were exposed to aural input with those whose input experience was visual. Consonant with prior studies, it showed that participants under control conditions (either CCV or CCA) showed greater sensitivity toward the adverb than the verb cue.

Two specific research questions in Experiment 2 concerned (i) whether each of the FFI treatments VG, VS, and VP, counteract

the effects of physical saliency and learned attentional biases, as indexed by participants' relative reliance on these cues in subsequent language comprehension (RQ5), and (ii) whether each of the FFI treatments VG, VS, and VP are equally effective in reattuning learners' attention to the non-salient morphological cues through visual and auditory modalities of exposure (RQ6). Regarding RQ5, all forms of FFI were effective in increasing attention to verbal morphology in the visual modality, although VP resulted in balanced attention to both cues. However, it was generally the case that attending to the morphological cues was considerably more difficult under aural than under visual presentation. Only grammar instruction (VG) was successful in reattuning learners' attention to the non-salient morphological cues in both modalities. This instruction allowed learners to become aware of forms and their patterning of function prior to sentence exposure, and their subsequent processing of these cues in the input promoted their use during comprehension testing. Also relating to RQ6, a major difference was seen in the effectiveness of morphological saliency-raising in the two modalities. Increasing the saliency of the verb morphology was very effective with visual exposure. Textually enhancing the morphology promoted attention to and analysis of these cues during sentence exposure, and this in turn resulted in subsequent use of these cues even when no longer made salient. In contrast, although emphasized pronunciation of these cues led to their use during sentence exposure in the aural group, this did not result in sustained use of these cues once the emphasis was removed. With aural presentation, the emphasized inflectional forms were attended but not analyzed, so that abrupt removal of the emphasis removed the cues themselves.

There are many limitations to this study. Concerns include the small range of constructions being taught, the short-term nature of the experiment, the experimental environment and lack of ecological validity, the lack of long-term delayed testing, and the small range of outcome measures. The latter is a particular worry. As pointed out by one reviewer, attention/processing in this experiment is assessed through a considered comprehension temporal rating task, which likely taps predominantly explicit knowledge. Future research could well-incorporate a battery of measures ranging in their implicitness/explicitness. Meta-analyses of effects of instruction demonstrate that effectiveness varies as a result of explicitness of measure. Much also remains to be done particularly with regard to assessing the transfer of visual language experience to aural competence and vice-versa. We are currently comparing the effects of instruction under aural, visual, and bimodal conditions.

The findings in this study reinforce and extend prior studies in second language instruction. Specifically, in the absence of instruction, learners tend to ignore non-salient features in the input, such as verb morphology. FFI can increase the saliency of inflections and other commonly ignored features by (i) explicitly identifying the forms and their functions, as in VG, (ii) by having the inflections made more salient by textual enhancement, as in VS, or (iii) by introducing the verb alone in a non-redundant context, as in VP. These are the type of techniques that help learners attend to verb morphology, and broadly, they do so to the same extent in the visual modality. However,

our results demonstrate that the effectiveness of different types of FFI techniques in enhancing the saliency and processing of these forms can vary as a function of modality of input presentation (i.e., aural or visual). Here, VG was effective across modalities, but VS was only advantageous in the visual modality. These findings should be considered in the design of optimal L2 instruction programs. Here, we only examined one specific type of construction, that of verb-tense morphology. We do not believe that the findings in this study will necessarily be true for all linguistic constructions. As the literature on FFI shows, different forms will require different levels of explicitness and explanation (Long, 2006, chap. 5; Spada and Tomita, 2010; Tolentino and Tokowicz, 2014).

Taken together, these findings demonstrate a range of effects of saliency in L2 acquisition. Morphological forms are less well-attended than lexical forms. We believe that this reflects a combination of their relative psychophysical slightness in comparison to lexical cues, as well as effects of learned attention and blocking. There are effects of modality – attention to morphological cues is more effective from visual than from aural input. There are effects of instruction – form-focusing techniques such as grammar instruction, verb pretraining, or enhancing the saliency of forms through typological or prosodic enhancement, can increase attention to these forms and increase processing. There are interactions between the effectiveness of FFI and modality such that typological saliency enhancement from visual input is effective, while prosodic enhancement from aural input is not. Finally, brief EGI prior to language exposure is an effective means of raising the saliency of otherwise ignored cues and turning input into intake. In learning a second language, some attention to form is necessary, and the forms that need to be attended are often the least salient in the input. Successful L2 acquisition rests on attention-focusing manipulations which raise the significance of otherwise non-salient cues.

AUTHOR CONTRIBUTIONS

The Abstract, Introduction and Discussion Sections were all contributed to equally by both authors. The methods section was written up by MC-V and edited by NE. The results were analyzed/written up by MC-V and edited by NE.

FUNDING

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE 1256260. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

ACKNOWLEDGMENT

We thank Yiran Xu for helping design, pilot, score, and administer parts of these experiments.

REFERENCES

- Bardovi-Harlig, K. (1992). The use of adverbials and natural order in the development of temporal expression. *Int. Rev. Appl. Linguist.* 30, 299–320. doi: 10.1515/iral.1992.30.4.299
- Bardovi-Harlig, K. (2000). *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Oxford: Blackwell.
- Beckett, S. (1954). *Waiting for Godot*. New York, NY: Grove Press.
- Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.
- Bybee, J. (2000). *Mechanisms of Change in Grammaticalization: The Role of Frequency*. Albuquerque, NM: University of New Mexico.
- Cintrón-Valentín, M., and Ellis, N. C. (2015). Exploring the interface: explicit focus-on-form instruction and learned attentional biases in L2 latin. *Stud. Second Lang. Acquis.* 37, 197–235. doi: 10.1017/S0272263115000029
- Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/S0142716406060024
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477
- Corder, S. P. (1967). The significance of learners' errors. *Int. Rev. Appl. Linguist.* 5, 161–169. doi: 10.1515/iral.1967.5.1.4.161
- Corey, D. M., Dunlap, W. P., and Burke, M. J. (1998). Averaging correlations: expected values and biases in combined pearson rs and fisher's z transformations. *J. Gen. Psychol.* 125, 245–261. doi: 10.1080/00221309809595548
- Cutler, A., and Carter, D. M. (1987). The predominance of strong initial syllables in the english vocabulary. *Comput. Speech Lang.* 2, 133–142. doi: 10.1016/0885-2308(87)90004-0
- DeKeyser, R. M. (1995). Learning second language grammar rules. *Stud. Second Lang. Acquis.* 17, 379–410. doi: 10.1017/S027226310001425X
- Doughty, C. J., and Long, M. H. (2003). Optimal psycholinguistic environments for distance foreign language learning. *Lang. Learn. Technol.* 7, 50–80.
- Doughty, C., and Williams, J. (1998). *Focus on Form in Classroom Second Language Acquisition*. New York, NY: Cambridge University Press.
- Ellis, N. C. (1993). Rules and instances in foreign language learning: interactions of explicit and implicit knowledge. *Eur. J. Cogn. Psychol.* 5, 289–318. doi: 10.1080/09541449308520120
- Ellis, N. C. (ed.) (1994). *Implicit and Explicit Learning of Languages*. San Diego, CA: Academic Press.
- Ellis, N. C. (1996). Sequencing in SLA: phonological memory, chunking, and points of order. *Stud. Second Lang. Acquis.* 18, 91–126. doi: 10.1017/S0272263100014698
- Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024
- Ellis, N. C. (2005). At the interface: dynamic interactions of explicit and implicit language knowledge. *Stud. Second Lang. Acquis.* 27, 305–352. doi: 10.1017/S027226310505014X
- Ellis, N. C. (2006a). "The associative-cognitive CREED," in *Theories in Second Language Acquisition: An Introduction*, eds B. V. Patten and J. Williams (Mahwah, NJ: Lawrence Erlbaum Associates), 77–96.
- Ellis, N. C. (2006b). Selective attention and transfer phenomena in SLA: contingency, cue competition, saliency, interference, overshadowing, blocking, and perceptual learning. *Appl. Linguist.* 27, 1–31. doi: 10.1093/applin/aml015
- Ellis, N. C., Hafeez, K., Martin, K. I., Chen, L., Boland, J., and Sagarra, N. (2014). An eye-tracking study of learned attention in second language acquisition. *Appl. Psycholinguist.* 35, 547–579. doi: 10.1080/17470218.2014.961934
- Ellis, N. C., and Sagarra, N. (2010). The bounds of adult language acquisition: blocking and learned attention. *Stud. Second Lang. Acquis.* 32, 553–580. doi: 10.1017/S0272263110000264
- Ellis, N. C., and Sagarra, N. (2011). Learned attention in adult language acquisition: a replication and generalization study and meta-analysis. *Stud. Second Lang. Acquis.* 33, 589–624. doi: 10.1017/S0272263111000325
- Ellis, R. (2006). Current issues in the teaching of grammar: an SLA perspective. *TESOL Q.* 40, 83–107. doi: 10.2307/40264512
- Ellis, R. (2012). *Form-Focused Instruction and Second Language Learning Language Teaching Research and Pedagogy*. Hoboken, NJ: Wiley-Blackwell, 271–306.
- Elman, J. L. (1990). Finding structure in time. *Cogn. Sci.* 14, 179–211. doi: 10.1207/s15516709cog1402_1
- Gibson, J. J. (1977). *The Theory of Affordances*. Madison, WI: Hildale.
- Gitelman, D. R. (2002). ILAB: a program for post experimental eye movement analysis. *Behav. Res. Methods Instrum. Comput.* 34, 605–612. doi: 10.3758/BF03195488
- Goldschneider, J. M., and DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: a meta-analysis of multiple determinants. *Lang. Learn.* 51, 1–50. doi: 10.1111/1467-9922.00147
- Goo, J., Granena, G., Yilmaz, Y., and Novella, M. (2015). "Implicit and explicit instruction in L2 learning," in *Implicit and Explicit Learning of Languages*, ed. P. Rebuschat (Amsterdam: Benjamins), 443–482.
- Han, Z., Park, E. S., and Combs, C. (2008). Textual enhancement of input: issues and possibilities. *Appl. Linguist.* 29, 597–618. doi: 10.1093/applin/aml010
- Herron, D., and Bates, E. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *J. Mem. Lang.* 37, 217–239. doi: 10.1006/jmla.1997.2514
- James, W. (1890a). *The Principles of Psychology*, Vol. 1. New York, NY: Dover.
- James, W. (1890b). *The Principles of Psychology*, Vol. 2. New York, NY: Holt.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Amsterdam: Benjamins), 229–254.
- Kamin, L. J. (1969). "Predictability, surprise, attention, and conditioning," in *Punishment and Aversive Behavior*, eds B. A. Campbell and R. M. Church (New York, NY: Appleton-Century-Crofts), 276–296.
- Kruschke, J. K. (2006). "Learned Attention," in *Paper Presented at the Fifth International Conference on Development and Learning*, (Bloomington, IN: Indiana University).
- Kruschke, J. K., and Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychon. Bull. Rev.* 7, 636–645. doi: 10.3758/BF03213001
- Lee, S.-K., and Huang, H.-T. (2008). Visual input enhancement and grammar learning: a meta-analytic review. *Stud. Second Lang. Acquis.* 30, 307–331. doi: 10.1017/S0272263108080479
- Leow, R. P. (2015). The roles of attention and (un)awareness in SLA: conceptual replication of N. C. Ellis & Sagarra (2010a) and Leung & Williams (2012). *Lang. Teach.* 48, 117–129.
- Li, C. N., and Thompson, S. A. (1987). "Chinese," in *The World's Major Languages*, ed. B. Comrie (New York, NY: Oxford University Press), 811–833.
- Long, M. H. (1991). "Focus on form: a design feature in language teaching methodology," in *Foreign Language Research in Cross-Cultural Perspective*, eds K. de Bot, R. Ginsberg, and C. Kramsch (Amsterdam: John Benjamins Publishing), 39–52.
- Long, M. H. (2006). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.
- Lyster, R., and Ranta, L. (1997). Corrective feedback and learner uptake. *Stud. Second Lang. Acquis.* 19, 37–66. doi: 10.1017/S0272263197001034
- Mackintosh, N. J. (1975). A theory of attention: variations in the associability of stimuli with reinforcement. *Psychol. Rev.* 82, 276–298. doi: 10.1037/a0032254
- MacWhinney, B. (2001). "The competition model: The input, the context, and the brain," in *Cognition and Second Language Instruction*, ed. P. Robinson (New York: Cambridge University Press), 69–90.
- MathWorks Inc (2011). *Matlab Version 7.12.0.635 (R2011a)*. Natick, MA: MathWorks.
- Meisel, J. (1987). "Reference to past events and actions in the development of natural second language acquisition," in *First and Second Language Acquisition*, ed. C. Pfaff (New York, NY: Newbury House), 206–224.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi: 10.1037/h0043158
- Norris, J. M., and Ortega, L. (2000). Effectiveness of L2 instruction: a research synthesis and quantitative meta-analysis. *Lang. Learn.* 50, 417–528. doi: 10.1111/0023-8333.00136
- Noyau, C., Klein, W., and Dietrich, R. (1995). *Acquisition of Temporality in a Second Language*. Amsterdam: John Benjamins.
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *J. Neurosci. Methods* 162, 8–13. doi: 10.1016/j.jneumeth.2006.11.017
- Rebuschat, P. (ed.) (2015). *Implicit and Explicit Learning of Language*. Amsterdam: John Benjamins.

- Rescorla, R. A., and Wagner, A. R. (1972). "A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Theory and Research*, eds A. H. Black and W. F. Prokasy (New York, NY: Appleton-Century-Crofts), 64–99.
- Roberts, L., and Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Stud. Second Lang. Acquis.* 35, 213–235. doi: 10.1017/S0272263112000861
- Schmidt, R. (1984). The strengths and limitations of acquisition: a case study of an untutored language learner. *Lang. Learn. Commun.* 3, 1–16.
- Schmidt, R. (2001). "Attention," in *Cognition and Second Language Instruction*, ed. P. Robinson (Cambridge: Cambridge University Press), 3–32.
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh, PA: Psychology Software Tools Inc.
- Shanks, D. R. (1995). *The Psychology of Associative Learning*. New York, NY: Cambridge University Press.
- Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: theoretical bases. *Stud. Second Lang. Acquis.* 15, 165–179. doi: 10.1017/S0272263100011943
- Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84, 127.
- Simon, H. A. (1957). *Models of Man: Social and Rational*. New York, NY: Wiley and Sons.
- Spada, N. (1997). Form-focussed instruction and second language acquisition: a review of classroom and laboratory research. *Lang. Teach.* 30, 73–87. doi: 10.1017/S0261444800012799
- Spada, N., and Tomita, Y. (2010). Interactions between type of instruction and type of language feature: a meta-analysis. *Lang. Learn.* 60, 263. doi: 10.1111/j.1467-9922.2010.00562.x
- Terrell, T. (1991). The role of grammar instruction in a communicative approach. *Mod. Lang. J.* 75, 52–63. doi: 10.1111/j.1540-4781.1991.tb01083.x
- Tolentino, L. C., and Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Lang. Learn.* 64, 279–309. doi: 10.1111/lang.12048
- Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognit. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5
- Van Patten, B. (1990). Attending to form and content in the input. *Stud. Second Lang. Acquis.* 12, 287–301. doi: 10.1186/s12889-016-3013-1
- Van Patten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition*. New York, NY: Ablex.
- Van Patten, B. (2006). "Input processing," in *Theories in Second Language Acquisition: An Introduction*, eds B. Van Patten and J. Williams (Mahwah, NJ: Lawrence Erlbaum), 115–136.
- Van Patten, B., and Oikennon, D. (1996). Explanation versus structured input processing instruction. *Stud. Second Lang. Acquis.* 18, 495–510. doi: 10.1017/S0272263100015394
- Wills, A. J. (2005). *New Directions in Human Associative Learning*. Mahwah, NJ: Erlbaum.
- Winke, P. M., Godfroid, A., and Gass, S. M. (2013). Introduction to the special issue. *Stud. Second Lang. Acquis.* 35, 205–212. doi: 10.1017/S027226311200085X
- Wong, W. (2001). Modality and attention to meaning and form in the input. *Stud. Second Lang. Acquis.* 23, 345–368. doi: 10.1037/xlm0000088
- Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.
- Zuraw, K. (2003). "Probability in language change," in *Probabilistic Linguistics*, eds R. Bod, J. Hay, and S. Jannedy (Cambridge, MA: MIT Press), 139–176.

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.

Copyright © 2016 Cintrón-Valentín and Ellis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Social Salience Discriminates Learnability of Contextual Cues in an Artificial Language

Péter Rácz^{1,2*}, Jennifer B. Hay^{2,3} and Janet B. Pierrehumbert^{2,4,5}

¹ Department of Archeology and Anthropology, University of Bristol, Bristol, UK, ² New Zealand Institute of Language Brain and Behaviour, University of Canterbury, Christchurch, New Zealand, ³ Department of Linguistics, University of Canterbury, Christchurch, New Zealand, ⁴ Oxford e-Research Centre, University of Oxford, Oxford, UK, ⁵ Department of Linguistics, Northwestern University, Evanston, IL, USA

We investigate the learning of contextual meaning by adults in an artificial language. Contextual meaning here refers to the non-denotative contextual information that speakers attach to a linguistic construction. Through a series of short games, played online, we test how well adults can learn different contextual meanings for a word-formation pattern in an artificial language. We look at whether learning contextual meanings depends on the social salience of the context, whether our players interpret these contexts generally, and whether the learned meaning is generalized to new words. Our results show that adults are capable of learning contextual meaning if the context is socially salient, coherent, and interpretable. Once a contextual meaning is recognized, it is readily generalized to related forms and contexts.

Keywords: salience, language variation and change, experimental linguistics, morphology, indexicality, sociolinguistics, artificial language learning

OPEN ACCESS

Edited by:

Adriana Hanulíková,
University of Freiburg, Germany

Reviewed by:

Shiri Lev-ari,
Max Planck Institute for
Psycholinguistics (MPG), Netherlands
Lynne Nygaard,
Emory University, USA

*Correspondence:

Péter Rácz
peter.racz@bristol.ac.uk

Specialty section:

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

Received: 29 April 2016

Accepted: 09 January 2017

Published: 30 January 2017

Citation:

Rácz P, Hay JB and Pierrehumbert JB
(2017) Social Salience Discriminates
Learnability of Contextual Cues in an
Artificial Language.
Front. Psychol. 8:51.
doi: 10.3389/fpsyg.2017.00051

1. INTRODUCTION

Studies of sociolinguistic variation show that people are able to associate linguistic patterns with a wide array of non-linguistic contexts (see e.g., Hay and Drager, 2007; Drager, 2010). What remains unclear is how these associations are learned, and whether learners discriminate these contexts in some structured manner. This learning problem is central in the sense that it sheds light on both the way contextual linguistic variation is structured and the way adults acquire it during their lifetime.

Of particular interest in this paper is the degree to which learners may attend differently to different types of non-linguistic context. Does the social salience of the non-linguistic context affect success in associative learning?

Context is interpreted in a number of ways in the relevant literature. For sociolinguists, the non-linguistic context is very broad. It includes the addressee and the discourse situation, as well as the speaker's attitudes and ideologies, which are conjoined to give social meaning to a given utterance (Eckert, 2008). For psycholinguists and psychologists, the context can also encompass higher-level situational attributes. In a given experiment, however, it can have a more specific interpretation, such as the visual field (Chun and Jiang, 1998) or the speaker (Kraljic et al., 2008b).

Salience is also interpreted in a number of ways, even within linguistics (Rácz, 2013). The core meaning is bottom-up and perceptual (a salient entity differs from its environment). We contrast this meaning of salience with *social salience*, a top-down, phenomenological concept, which encompasses the observer's background knowledge on the relevance of various aspects of the interaction (hence the term "social"). We will use the concept of salience to differentiate

non-linguistic contexts that are equally complex in structural terms but are used to different degrees in anchoring linguistic variation. We use it as a general, neutral term for this distinction between contexts.

In this paper we introduce an experimental paradigm that facilitates investigation into the contextual learning of morphological patterns. Using this paradigm, we then conduct a series of six experiments that together demonstrate a very significant effect of social salience upon contextual morphological learning.

1.1. Denotative and Social Meaning

Linguistic constructions have denotative meaning and social meaning. Broadly speaking, the former is the concept that the construction denotes, while the latter is the social-cultural context of its use.

Denotative meaning does depend on the context—the denotative meanings of even common concrete nouns can vary with the topic of discussion and the use of metaphor; *bug* means something different in discussions of gardening and of computer programming. However, the social and topical dimensions of word choice are only moderately correlated (Altmann et al., 2011).

Social meaning—for example, information conveyed about who is speaking, who is being addressed, and the nature of their relationship—is more indirect and more variable than denotative meaning (Preston, 1996; Labov, 2001; Foulkes and Docherty, 2006; Silverstein, 2009). Both Labov and Silverstein note that awareness to social connotations can vary from having explicit stereotypes to no interpretation for social *indicator* variables that correlate with specific contexts in a way that is not acknowledged by the speakers.

In addition to dialect (Wells, 1982) and social group (Eckert, 2000), robust factors that influence social variation in language include age, gender, and sexual orientation (Labov, 2001; Pierrehumbert et al., 2004; Tagliamonte and Roeder, 2009). An important aspect of the social context is the addressee, or the speaker's relationship with the addressee (within its context). Linguistic accommodation to the addressee is a well-researched phenomenon (Coupland et al., 1991; Soliz and Giles, 2014). Certain languages, like Djirbal, develop lexical sets that are used with addressees belonging in specific kinship groups (Dixon, 1980).

Social meaning also varies in even more nuanced ways, as speakers dynamically use social-contextual information to take stances and negotiate in social interactions—different linguistic choices reflect the individual's linguistic experience and construction of social identity (Milroy, 1980; Eckert, 2000).

Listeners can, in turn, use such patterns to infer speaker characteristics or to adapt to different speakers when processing speech (for review of relevant literature see Hay and Drager, 2007; Foulkes and Hay, 2015). Some words are statistically associated with older speakers and others with younger speakers; sensitivity to these associations can be displayed through varied ease of psycholinguistic processing without any explicit awareness of age-based patterns on the individuals' part (Walker and Hay, 2011). Listeners are able to associate different speaker

personae with different combinations of morphological and phonological variables, based on fine-grained patterns in the ambient language (Campbell-Kibler, 2011). Individuals can also shift their categories of speech sounds based on cues of the broad cultural context, even if these cues are only peripherally present (Hay and Drager, 2010).

These examples show that social meaning can be attached to linguistic constructions that are not specific words or phrases—it can be generalized across linguistic patterns. It can apply to contexts of various levels of abstractness and expand to new contexts. It encompasses a wide variety of linguistic detail, from phonetics to word choice (Säily and Suomela, 2009; German et al., 2013). It relies on some contextual differences more heavily than on others, and speakers use it in a complex and subtle way to express intent and create a public persona (Eckert, 2012).

While numerous sociolinguistic and anthropological studies have revealed the importance of social meanings in language, less is known about how they are learned. Our understanding of social cognition in language leaves many unanswered questions about what details are noticed and remembered (Silverstein, 2009), as well as about what factors support generalization to new forms or new situations (Pierrehumbert, 2006). Foulkes (2010, p. 6) laments the lack of understanding of learning and storage of social meaning, stating that: “it now seems uncontroversial to conclude that social information is retained in memory alongside linguistic knowledge. Questions remain, however, over what sorts of social information are learned and stored, where and how they are stored in relation to linguistic information, and how social information affects linguistic processing.”

1.2. The Role of the Context in Learning Linguistic Categories

If context is important, how do we learn to use it?

The contextual learning literature that is most relevant to this paper focusses particularly on the role of the broader extra-linguistic context—above and beyond the referent—in learning a linguistic category. Three important findings emerge from it.

First, consistency across contexts aids recall: a category is remembered more accurately in the context in which it was originally learned. In word-learning tasks, words are retrieved more accurately if recall occurs in the same location as training. Godden and Baddeley (1975), for example, show that words learned underwater are more accurately retrieved underwater. This relates to more general work on memory retrieval, where it has been shown many times that consistency of contextual information between encoding and retrieval leads to increased recall. Smith and Vela (2001) review literature showing that “people tend to be aware of their surroundings even when memorizing something. As such, environmental features are typically encoded along with the to-be-remembered material.” Recurrent information will also invoke the context it was acquired in. Models of category learning and memory retrieval (Ratcliff, 1978; Grossberg, 1987) operate using notions of context-specificity.

Certain memories activate specific contexts in which they were learned.

Broad experimental work in psychology discusses the role of contextual cues in category learning (Chun and Jiang, 1998; Goujon et al., 2015). This work shows that visual decision tasks speed up if the trial (with a given visual context) is shown repeatedly, despite the fact that participants are unable to identify the contexts afterwards, suggesting that the effect of the context on learning can be implicit. Qian et al. (2014) show that in a “whack-a-mole” type game, players are faster at predicting the location of the mole if the location is probabilistically cued by moving background images that the player is not overtly oriented to. Gómez (2002) note that a consistent structure is learned better across multiple contexts. Observing the same pattern across multiple speakers improves learning as well (Rost and McMurray, 2009). Individuals use contextual memory to aid recall and prediction. Lleras and Von Mühlenen (2004) revisit Chun’s paradigm, and their results indicate that the success of contextual cues depends on whether participants are focussing on the task in a narrow sense (presumably discarding the context) or are trying to take a holistic approach.

Second, categories learned in one context can generalize to another, similar context. Van der Zande et al. (2014) show that phonetic categories that shift due to exposure to a speaker retain this shift even when listening to another speaker. Maye et al. (2008) extend this generalization to a new accent: their participants are able to use a context-specific vowel adaptation mechanism to process phonetic variation coming from a speaker and then, in turn, re-use the adapted vowel categories when encountering a similar speaker. Kraljic and Samuel (2006) show that a phonetic category distinction is generalized to a new linguistic context and also to a new speaker. That is, even if participants are trained on the distinction in one speaker voice, they carry over the distinction to another voice.

Third, listeners do not treat all available information the same way. Kraljic et al. (2008b) find that, while we first learn all phonetic detail as characteristic of a given speaker, we are later able to re-assess this knowledge and discard contextual variation that is based on an arbitrary idiosyncrasy of the speaker (such as talking with a pen in the mouth). Kraljic et al. (2008a) show that phonetic variation is processed differently if it is due to a consistent idiosyncrasy of a speaker (like a speech impediment) than if it represents a dialectal contextual allophone. Leung and Williams (2012) show that a distinction based on the animacy of referents is learned much more easily and generalized more readily than a distinction based on size differences between referents. Similar results have been found even for purely phonological contexts. Becker et al. (2011) find that Turkish speakers apply some statistical regularities in the Turkish lexicon, but not others, in a forced-choice wugs task; they conclude that speakers distinguish accidental from well-grounded statistical generalizations.

We are able to associate linguistic categories with non-linguistic contexts, even if these contexts are fairly arbitrary. We can extend this knowledge to similar but different contexts as well. And yet, we do not rely on all differences the same

way—we distinguish information that is *relevant* in the context from information that is *accidental* to it.

1.3. Weighing the Social Saliency of Contexts

The amount of detail observed in sociolinguistic variation (Hay and Drager, 2007), coupled with memory models (Nosofsky, 1988), suggests that language users are able to construct social meaning based on a vast number of contexts. Some assume, however, that human memory is too restricted for this. Therefore, at least some of the information may be discarded if it cannot be used to make generalizations and if it taxes resources overtly (for the debate, cf. Gluck and Myers, 2001; Denton et al., 2008; Baayen et al., 2013).

How do we choose between useful and irrelevant contextual information? While the statistical co-occurrence of contexts and patterns is important, this is not the complete story. Selective attention guides which details are more important in processing information (Itti et al., 1998). Variation can be interpreted differently depending on its source (Kraljic et al., 2008a).

Relevance in turn derives from complicated assumptions about how the world works: that some speaker differences are consistent and others are haphazard, and that some contexts are more informative of language variation than others. When these include assumptions about what sorts of people are members of the same group, they rest on social constructs or categories. What information is grouped together and what is discarded both play a role in structuring social-contextual language variation.

Many experimental studies have explored the associations between familiar social groups and accentual features. Relatively few studies have investigated learning that involves novel social groups, or learning of socio-linguistic cues other than ones at the phonetic level. This may be because it is a daunting task to set up scenarios in which the relationship between the social context and the linguistic pattern is transparent and well-controlled. However, there are several noteworthy studies. Work by Docherty et al. (2013) and Langstrof (2014) find that people can associate familiar dialectal variables with arbitrary “tribes” in a laboratory setting. Roberts (2008) shows that people are able to come up with morphological markers in a nonce language in order to demarcate in-group and out-group membership in a laboratory setting. Beckner et al. (2016) find that participants shift their linguistic patterns to accommodate to a group of human peers but not to a group of humanoid robots. The Beckner study shows extension of the accommodation pattern to new words that are similar in form to previously encountered ones. These studies do not look at extension or generalization to different speakers.

The term *relevance* often implies conscious decision making on which contexts to consider and which to discard. Work on phonetic learning, however, suggests that we discriminate contexts largely implicitly. We will use the term *social saliency* to compare the “usefulness” of a context for linguistic learning and—as a consequence—people’s ability to rely on it.

1.4. Individual Variation in Contextual Learning

As in any learning task, people's success rates will vary in contextual learning. Work on language variation and change provides an important body of evidence on how people learn linguistic patterns that are associated with non-linguistic contexts. Labov (2001) shows that a new sociolinguistic variant does not diffuse uniformly through the population. Typically, there are community leaders of language change, who are chiefly responsible for spreading an innovative variant in the community. Later work distinguishes components of this process, all of which are relevant here. First, computational models have led to the conclusion that an innovative variant succeeds only if it carries a positive social weight, something which of course depends on learning a social association for the variant in the first place. (cf. e.g., Baxter et al., 2009; Fagyal et al., 2010). However, this positive weight does not need to be present in the minds of everyone in the community, but only in the minds of a critical group of early adopters—people who take up and use the new variant before other people do (Pierrehumbert et al., 2014). Experimental evidence for the existence of linguistic early adopters is found in Schumacher et al. (2014), an artificial language learning experiment in which some participants adopted an unexpected number-marking system far more than others.

These results indicate that individual variation in contextual learning is far from being a footnote example of differences in task completion—it is significant for patterns of language variation and change.

The source of such individual differences, then, becomes crucial, but it is relatively unclear. Some individuals are likely better at recognizing or remembering contextual patterns than others. For verbal and non-verbal statistical learning tasks, Siegelman and Frost (2015) show that individual performances in a verbal or a non-verbal task are not strongly correlated with different measures of intelligence and cognitive capacity. In fact, Siegelman and Frost find little correlation across performance in different statistical learning tasks. The analysis of Lleras and Von Mühlenen (2004) of individual participant behavior in their learning experiments (adopted from the work of Chun and Jiang, 1998) indicates that participant success in a task depends on the strategy adopted by the participant. This high-level, complex decision is unlikely to be derived from any single cognitive or linguistic measure. However, some systematic effects have been identified. Vocabulary size is a good predictor of how easily new word forms are learned in children (Henderson et al., 2015 and therein). Henderson et al. (2015) do not find an effect in adults, but in a related pseudo-word rating study with more statistical power, Needle et al. (2015) do find that high-vocabulary adults have a better ability to decompose nonce compounds such as *angstroof*. Brooks et al. (2016) shows that learning and generalizing an L2 morphological pattern can be partially predicted by measures of non-verbal intelligence and statistical learning ability.

Ramscar et al. (2014) argue that the life-long accumulation of experience affects performance in psycholinguistic tasks. Older

people have more prior experience, and so work with a denser cue space in verbal tasks. In an explicit learning task with feedback, Metcalfe et al. (2015) find that older participants perform better, especially on unfamiliar items. Event-related potentials for the older participants indicate better ability to focus attention on feedback.

In short, participant accuracy is highly variable in learning tasks and this variation derives from a complex set of cognitive differences. However, two studies (Ramscar et al., 2014; Metcalfe et al., 2015) point to age as an interesting factor. Prior experience may affect performance on psycholinguistic tasks, either via richer mental representations gained through experience, or though better proficiency in allocating attention. Recruiting participants on Amazon Mechanical Turk provided us with a participant pool of diverse ages that makes it possible to assess this factor.

Diverse sources of evidence indicate that language use relies on contextual cues, and that speakers evaluate these cues both implicitly (based on saliency and statistical co-occurrence) and more explicitly (based on social saliency). How this behavior is learned is less clear, but the learning process and the individual differences manifested in it are both very relevant to the study of language variation and change.

2. AIMS

In this paper, we report on a series of experiments that build on previous results in context-specific category learning. In the experiments, participants have to learn linguistic patterns that depend on the context. The context can be linguistic—the choice of a suffix depends on the shape of the stem. It can be non-linguistic—the choice of a suffix depends on the conversation partner. Conversation partners can differ across various dimensions. Both the patterns and the contextual differences are more abstract than those explored in studies of phonetic adaptation such as Kraljic and Samuel (2006).

The linguistic patterns we look at are morpho-phonological. They are suffixation patterns in a simple artificial language that mark the diminutive or the plural. These are both transparent, iconic relations that also show considerable variation in English and other languages. The specific linguistic contextual pattern we use (the suffix vowel should match the stem vowel) is not found in English, so participants must learn it. Their success in doing so provides the baseline condition for the experiment, serving to validate the paradigm and shed light on the strengths of the effects found for the various social factors.

The social dimension we focus on is socially robustly interpretable, the gender of the conversation partner. We contrast it with a dimension that has similar visual prominence but lacks its social saliency, the spatial orientation of the conversation partner. We chose to explore the gender distinction because it is a very robust sociolinguistic marker. Children as young as 6 months, for example, preferentially match sex-cued voices and faces (Walker-Andrews et al., 1991). Sex and gender have a complex effect on the use of social meaning in general (Milroy and Milroy, 1993; Cheshire, 2002). Our experiments build on

each other to provide a solid foundation for the salience of this distinction, by showing that it holds up across differing amounts of exposure, types of extensions, or types of linguistic patterns.

We ask the following questions:

1. Given a morpho-phonological pattern, how quickly and well are participants able to learn its association with a linguistic context? Are they able to generalize the pattern to new words? How does generalization to new words compare to words seen in training?
2. How well can participants associate a morpho-phonological pattern with a social context: conversation partner gender? Are they able to generalize the pattern both to new words and to new instances of the appropriate context?
3. Are all types of social association equally learnable and generalizable? Or is an incidental social property (spatial orientation) processed differently from a more static conversation partner characteristic (gender)?
4. How do individuals vary in learning contextual associations for linguistic patterns?
5. Are older participants more successful, or less successful, at learning such social contextual associations?

As we will see, for the morpho-phonological patterns we look at, learning is possible for both linguistic and non-linguistic contexts. For non-linguistic contexts, participants are more successful in learning an association with a robust, salient context, conversation partner gender. This context is interpreted broadly—gender is recognized as the defining dimension. The morphological pattern is recognized and extended to previously unseen words after training. Older participants are better learners in our data.

We find these results with more types of conversation partners (such as children and adults) and with two distinct morphological patterns, the diminutive and the plural. These results indicate that the salience of conversation partner gender *vis-à-vis* spatial orientation is a broad and general phenomenon.

3. OVERVIEW OF EXPERIMENTS

Our experiments use a training-test paradigm based on a simplified version of adaptive tracking. The adaptive tracking paradigm described in Leek (2001). It was previously adapted to linguistic research by Schumacher et al. (2014). We discuss the paradigm in detail in the Methods Section of Experiment I.

Experiment I trains participants on a morphological pattern, presented visually. They see picture pairings, with a large and a small version of the same entity. The large version is named. They have to choose the name of the small version, which is a suffixed form of the name of the large version. There are two suffixes, and the correct one includes a vowel that matches the stem vowel. We find that participants learn to consider this context easily and also extend the pattern to new items.

In Experiment II the same morphological pattern is presented, again with different conversation partners. This time the pattern depends on the conversation partner. There are two conversation partners—a male and a female—and both are presented in two ways visually, in side view and in front view. One group of

participants is trained with the morphological pattern depending on conversation partner identity, cued by gender (answers for conversation partner A or B pattern together) The other group is trained with the pattern depending on conversation partner spatial orientation (answers pattern together according to whether the conversation partner is presented in front view or side view).

We find that an association of the pattern with the identity of the conversation partner is easier to learn than an association of the pattern with the spatial orientation of the conversational partner. This result resembles the findings by Kraljic et al. (2008b) on learning of incidental vs. characteristic patterns of phonetic variation. The morphological pattern is interpreted broadly—in the test session, it is extended to items not seen in the training session.

Experiment III expands the scope of contextual learning to examine whether participants generalize on the basis of conversation partner gender. Gender is one of the most widely discussed predictors of sociolinguistic variation. Morpho-phonological and lexical variation that depends on gender is not restricted to languages like Dyirbal. Languages like French and German use different adjective conjugations depending on the referent, including first and second person referents in the discourse, while stochastic differences for gendered language use have been found in English as well (Hay and Walker, 2013). We find that gender is a better cue than spatial orientation. Naming is readily extended to new conversation partners that fit this context (i.e., as another female or male conversation partner).

Experiment IV focuses on the way participants rely on the denotative and the contextual aspect of the naming pattern. The general layout is similar to experiments I–III. However, the test phase is different. Instead of a right and a wrong answer, they are forced to choose between one answer that is correct in its denotative aspect but wrong in its contextual aspect and another one that is set up the other way around. We find that if the contextual cue is conversation partner *gender*, participants have split preferences between the denotation and contextual cue. With a spatial cue, they overwhelmingly prefer the denotative aspect.

In experiments V and VI we extend the paradigm to investigate a new morphological pattern—the plural—and investigate the effect of a radically increased training set size. We find that learning the plural is similar to learning the diminutive, though the results are not conclusive. Increased training improves participant accuracy in test.

4. EXPERIMENT I

Experiment I establishes our experimental paradigm and investigates the role of the linguistic context in learning within this paradigm.

In Experiment I, participants learn a morphological pattern that is sensitive to the linguistic context. It is a vowel harmony or partial reduplication pattern (common in the world's languages, though not present in English): the vowel of the suffix has to match the vowel of the suffixed stem. The version of

adaptive tracking presented here was used successfully by Schumacher et al. (2014), who also recruited participants on Amazon Mechanical Turk. Our design, however, differs in its overall theme, as well as in the amount of training participants receive.

We explain our design in detail in Section 4.2 and address changes to it in subsequent Methods sections.

4.1. Participants

The experiment was hosted on Amazon Mechanical Turk (AMT). 47 people participated in the experiment. 22 are women, 25 men. All are native speakers of American English. We base this claim on the fact that all participants had IP addresses from the United States and self-identified as native speakers (those who did not were excluded from the results). The mean age is 31 years, with a standard deviation of 8.5. Participants were paid three dollars upon completion of the task.

For each of our experiments, we used Amazon Mechanical Turk worker IDs to exclude participants who had taken part in any of the other experiments. US worker IDs are independently verified by Amazon, making it very difficult for the same person to operate multiple accounts.

As in all following experiments, we used training speed to remove outliers (cf. below). For each across-subject condition, we removed the 2.5% of participants who took the most trials to finish training—i.e., the slowest ones. We filtered participants within the across-subject conditions, since we expect conditions to vary in length. In Experiment I, which has one across-subject condition, we removed 2 participants and report data from 45 participants. We return to the training phase and discuss our exclusion criteria in detail in the Results section of Experiment I.

For each experiment, we do not report the precise ratio of AMT workers who picked up the task vs. workers who finished it, since an online task can be interrupted for various reasons, including connection issues, disruptions, etc. On the whole, about 5% of the workers who started these experiments did not complete them. This, in our experience, is not an excessively high number for an online experiment.

By using Amazon Mechanical Turk, a burgeoning forum for psycholinguistic research (Munro et al., 2010), we were able to recruit a large number of participants in a short span of time. Amazon Mechanical Turk is especially fit for our experiment, which has a “game-ified,” button-input design. The game format allows for immersion of the participants, and increases the likelihood that they pay attention to the task—otherwise they cannot finish it. *Gamification* has been increasing in popularity in data collection in recent years (Von Ahn, 2006) and it has been used successfully in linguistic experiments as well (Fedzechkina et al., 2012; Schumacher et al., 2014). Relying on Amazon Mechanical Turk allowed us to run substantial numbers of subjects, so as to be able to see important differences across conditions in how likely our various predictions are borne out. Crump et al. (2013) show that more complex laboratory tasks on category learning can be replicated using subjects on AMT. However, AMT subjects are, overall, less successful learners than laboratory participants, possibly because they are less focused

and attentive when participating from their homes, without the presence of an experimenter.

Experiment I tests for a main effect of a single across-subject condition while Experiment II has two across-subject conditions and tests for interactions as well. This is why the latter has twice as many participants as the former. The same logic was applied to subsequent experiments. This is an economical use of participants, but does have the restriction that power to estimate participant-level effects (here, gender and age) will vary across experiments. We return to this issue in Section 8, in which we estimate these effects on a merged dataset.

This and the following experiments reported in the paper have been overviewed and approved by the Institutional Review Board of Northwestern University and the Human Ethics Committee of the University of Canterbury. During the time of data collection, the experimenters were not affiliated with any other institutions.

4.2. Methods

In Experiment I, participants play a computer game in which they have to help a bird flying roof to roof to return to its nest. The game consists of a training phase, followed by a test phase. The targets are presented in the following way. For a given *target*, the participant sees a *conversation partner* who shows a *query picture* to the main character, the bird, along with the *prompt*, the *name* of the depicted object. The bird responds with a *response picture* and *two possible names*. The participant has to choose one of them. The response picture is always the *diminutive* version of the conversation partner's picture (depicted as a small or juvenile version of the conversation partner's picture). This implies that one of the two possible names is the correct name of the diminutive of the query picture. A *target* is the combination of an item (a query-response picture pair) with a conversation partner. **Figure 1** shows the general layout with examples of the phases and the mechanics. Stimuli are *visual* only.

During training, targets are presented in a random order to each participant, and the participant has to give a correct answer for every target in order to move to the next target. If they give an incorrect response, they have to return to the previous target. The test phase does not use adaptive tracking. Here, targets are presented, again, in random order. The targets in the test phase include both the training items and previously unseen items. No feedback is given. Training takes place during the in-game day, test during the in-game night. Participants are also told when they enter the test phase. The way training relies on simplified adaptive tracking guarantees that each participant has responded to each stimulus correctly at least once before moving to the test phase. Unlike training protocols with a fixed number of trials, it provides an opportunity for participants who find the task difficult to improve by training for longer.

The name of the query picture is a nonce word with a CVC structure. These are drawn from a set of 12 syllables (cf. **Table 1**). Half of the name syllables have the <e> vowel, the other half the <a> vowel. The two possible names of the response picture are the name of the query picture plus one of two suffixes—*pek* and *pak*. These suffix syllables were selected such that one had <e> and one had <a>. The correct response always matches

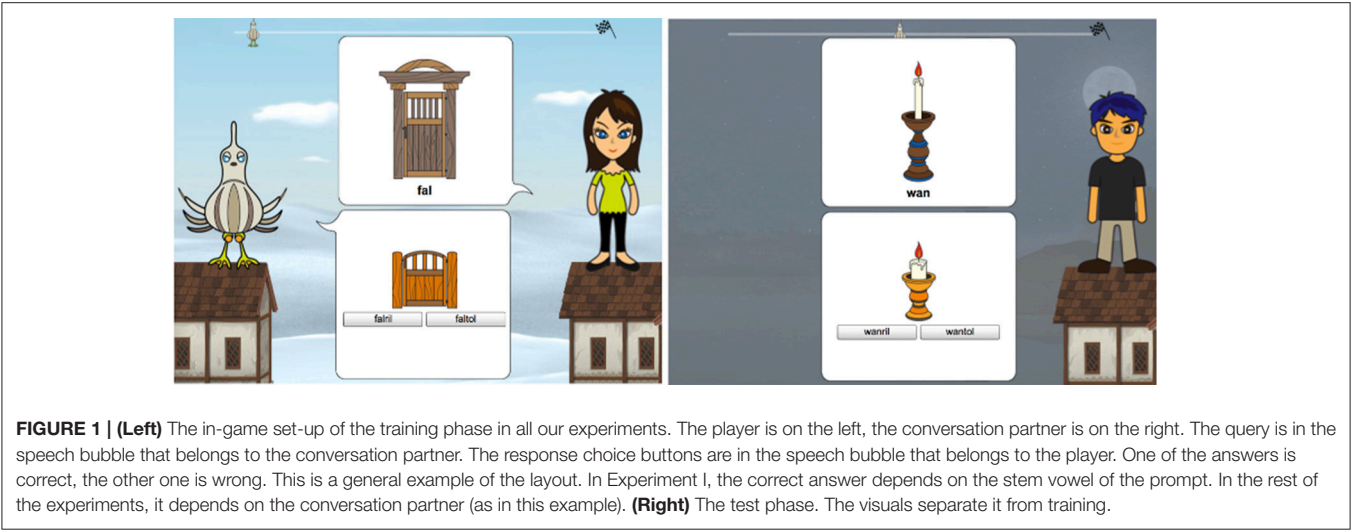


TABLE 1 | Stimuli set, Experiment I.

fek	ran	wek	fal
pel	ral	tek	rak
tas	fan	wen	fes

the vowel of the prompt. This echoes vowel harmony or partial reduplication systems commonly found in natural languages. Participants encounter six items in training and these six items plus six new items in test. These are items that do not occur in training.

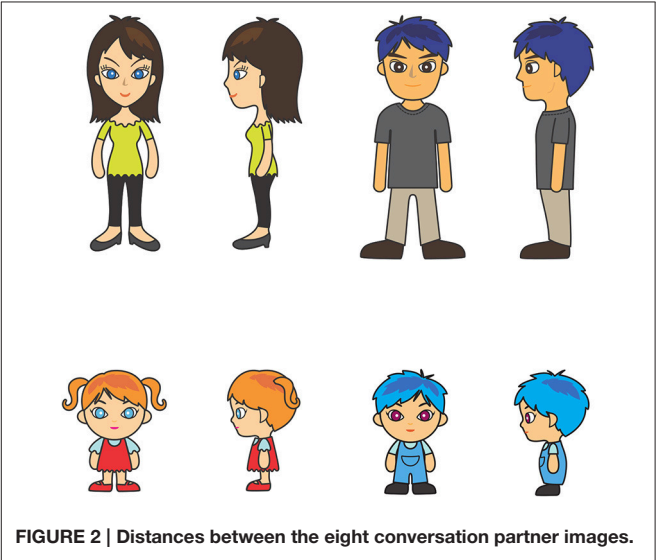
We designed the stimuli with the following principles in mind: (i) the syllables should be distinctive; (ii) they should consist of a small set of frequent letters; (iii) they should be easy to pronounce for our participants, who are American English speakers; (iv) the consonant clusters in the two-syllable words should cue English word boundaries in a uniform manner. These are somewhat competing requirements but our aim was to provide a relatively optimal set that balances all of these considerations.

The names of the individual objects are randomly assigned for each participant, using the set of twelve syllables in Table 1. Six occur in training and then also in the test. Six occur only in test.

In all our experiments, participants come across four conversation partner images during the game.

The images we use for our conversation partners in experiments I–VI can be seen in Figure 2. We will refer to them in the paper using the labels *woman*, *man*, *girl*, and *boy*. Each figure has two perspectives, *front* and *side*, giving us 8 conversation partner images in total.

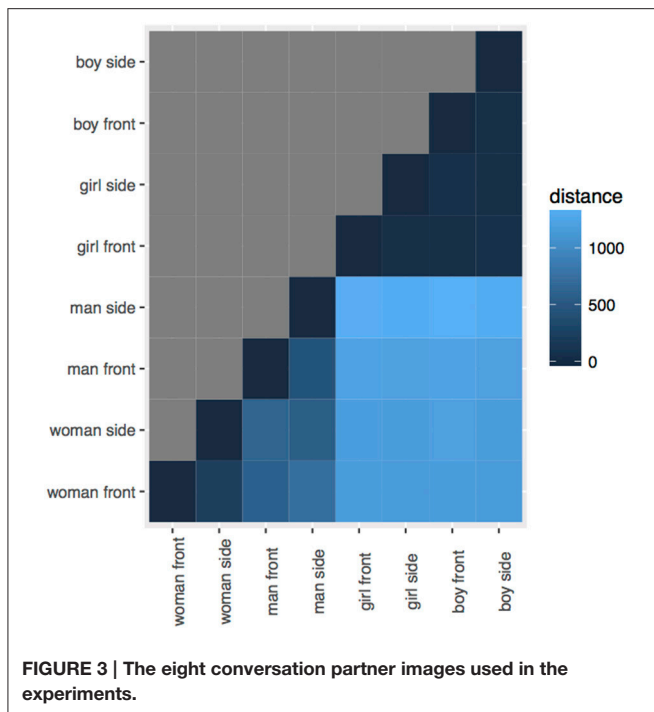
The particular images we used were designed to be matching in many respects, while still appearing visibly different according to gender and/or view. It is difficult to assess the degree to which we were successful with this aim, as human raters completing an explicit similarity rating task would be unable to avoid bringing their social knowledge to bear. However, in order to attempt an objective test that there were not strong visual differences between the different dimensions, we computed the



Levenshtein distance between uniformly binned histograms of the grayscale versions of the images using Matlab (Mathworks, 2016). Histogram comparison is a common method in image processing (Pele and Werman, 2010).

The Levenshtein distance between the images is roughly similar. For the adult images, there is a slightly larger distance between the woman and the man than between the front and side views for either the woman or the man. But for the child images, the order of the distances is interwoven between the grouping factors, and the largest difference is between the boy front and side view, and the smallest difference is between the girl and the boy front views. (Figure 3 is a tile plot that shows image distances. Darker hue means smaller distance).

This visual similarity metric patterns differently for the adult and child figures, but—as we will report in the following sections—none of our experiments reveal any difference in whether participants were trained on the adult or the child images



(c.f. experiments III; V). It therefore seems very unlikely that patterns relating to image similarity are driving the behaviors that we observe.

The four conversation partner images in Experiment I are the *woman* and *man* figures in **Figure 2**, viewed from the *front* and the *side*. All items occurred with all conversation partners, giving a minimum of 24 trials in training and 48 trials in test. Who the conversation partner is has no bearing on the correct name selection in training, since the latter only depends on vowel quality. Linguistic context is relevant, non-linguistic context is irrelevant. The conversation partners will become relevant in Experiment II. The “comic book” setup of the experiment allowed us to freely combine text with conversation partner images in all experiments.

In our experiments, the visual field of the experiment (the window in which it takes place on the user’s computer screen) is the non-linguistic context. The words that occur in this visual field (written in the latin alphabet) constitute the linguistic context. We use the visual display to manipulate a classic sociolinguistic factor, the *addressee*.

4.3. Participant Instructions

The experiments are designed to create a setting for linguistic or socio-linguistic learning that is controlled, yet still somewhat naturalistic. The task itself is made explicit, but the potential cues for the correct answers are not. Participants need to work out which cues to attend to from the potential cues available. These include the orthographic shapes of the word forms, the item pictures, and the conversation partner pictures (since the protagonist and the background are held constant). This makes the task harder. But it also makes it analogous to problems we encounter in language use, an issue that has received considerable

attention in the literature of contextual language learning (cf. Yu and Smith, 2007).

Participants receive written instructions at the beginning of the game. They are told that the bird is the protagonist (“our hero”), and that they need to help our hero return to its nest by flying from roof to roof. The hero will meet people who stand on the roofs and ask questions. The hero needs to answer the questions correctly in order to proceed. The questions are explained: the person names an object and shows our hero a smaller version of the object as well. Our hero has to guess what they would call the small object. It is explained that a second phase follows this first phase. In this phase, participants need to guess the names given to small objects, just like they did in the first part. They are asked to try to remember what the right answers were in the first part, and guess the right answer based on that.

4.4. Hypotheses

We hypothesized that participants would learn the association of the morpho-phonological pattern with the linguistic context and generalize it to new items. Based on related studies of phonetic learning, such as the study of an indexical allophonic pattern reported in German et al. (2013), we also predicted higher accuracy for test items seen in training than for unseen test items. We also evaluated age and gender as potential predictors of performance.

4.5. Results

Overall, the results show that many participants succeeded in learning and generalizing. This outcome is reflected in the time course of training and in accuracy in the test phase.

Since the length of training depends on the participant’s success at the learning task, training length is a good indicator of task difficulty. It is also a good indicator of participant attention and ability.

We use trial counts to express training duration. While individual trials vary in duration (there is no time limit on trial length, that is, people can spend as much time as they want on their decision), they do so to a modest degree (in Experiment I, mean (m) = 16 seconds, standard deviation (sd) = 12 seconds).

We prefer trial count to duration in time because the latter can be affected by user computer problems, server lag, and participant behavior (taking a break, answering the phone, etc.).

In similar experiments the accepted norm is removing participants who are 2 or 2.5 standard deviations outside the overall mean. This method has its problems, as shown by Leys et al. (2013), who recommend mean absolute deviation instead. For our data, neither the standard deviation threshold, nor the mean absolute deviation threshold are applicable. We decided to use a percentage threshold since trial length in our experiments is not normally distributed, making standard deviation a poor measure of the distribution of participant trial count. The distribution starts at 24 (the minimum possible number of trials) and has a long right tail. A method of outlier removal that relies on the 2 standard deviations threshold would remove about 5% of the participants from the experiment. Our method removes the slowest 2.5%. A participant cannot finish too quickly, and

so the distribution of training trial counts has no left tail. We remove outliers to safeguard against participants with very poor attention.

In all experiments, we filtered participants within the across-subject conditions because we expected these to vary in length. We used a quantile threshold to remove participants in the right tail of the training trial count distribution. For every condition, we establish the 0.975 quantile threshold of the distribution of training trial counts. We exclude participants over this threshold. The number of participants removed for each experimental condition ranges between 1 and 2. Outliers for the separate conditions add up to the sum of outliers for each experiment.

In Experiment I, 2 out of 47 participants are over the 97.5% threshold.

Participants finish training much faster than a player would by chance ($m = 43$, $sd = 18$). Individual variation for trial counts is large. Participants recruited through Amazon Mechanical Turk vary more in their behavior than would the college students recruited for a typical lab experiment.

Experiment I has one across-subject condition. Training speed in this condition is only informative inasmuch as it is, on average, much shorter than what we expect if participants were guessing. This shows that some form of learning is taking place in training¹.

Accuracy in test depends on whether the item was seen in training. **Figure 4** is a bean plot of participant accuracy in the test phase, grouped by whether the item was seen in training.

¹ Due to the use of the simplified adaptive tracking paradigm, a player who guesses randomly on all training targets would need 518 trials, on the average, to finish training. The high number is due to the fact that, in the adaptive tracking paradigm, an incorrect answer returns the previous trial, and an incorrect answer to that throws the participant back even further.

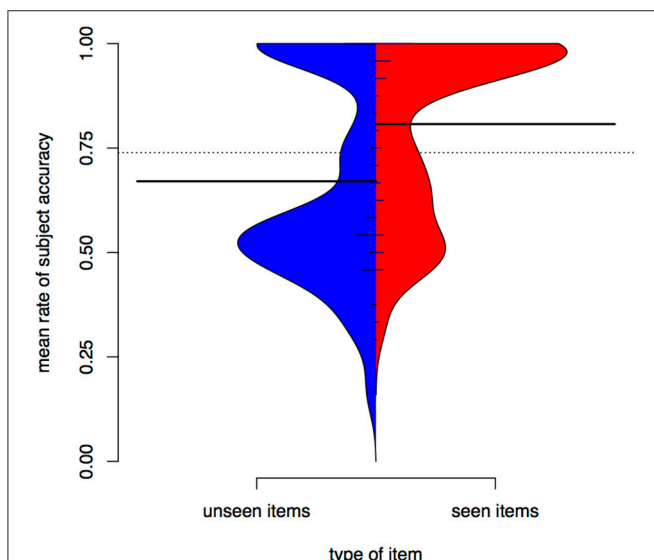


FIGURE 4 | Distributions of participant responses on previously seen and unseen test items, Experiment I. Black horizontal bars show the mean accuracy for each set of items. The dotted line shows the overall mean. Small horizontal lines show individual values; longer if multiple individuals have the same average.

The bean plot shows the distributions of participant responses along the y axis. Mean accuracy is higher for items seen in training (seen items, right) than for items participants only encountered in test (unseen items, left). This is indicated by the long black horizontal bars. The *distribution* of mean subject accuracy rates, however, is also revealing. For items not seen in training, we see a very clear bimodal distribution, with most subject means centred either around .5 or near 1 (up to 1, actually, since it is impossible to have a *higher* accuracy rate than 1). A person whose accuracy is around 0.5 in a task that involves binary choices is effectively guessing. A person whose accuracy is 1 has done a perfect job. For items seen in training, there also appears to be a bimodal distribution, but the total mass of the upper mode is greater and more participants perform at accuracies around 0.6 to 0.7.

We used the R statistical computing environment for our analyses (R Core Team, 2016). We created our plots using ggplot (Wickham, 2009).

We stepwise fit a binomial mixed-effects regression model on the test data, using response to individual items (*correct* or *not correct*) as an outcome variable and *presence in training* and participant *age* and *gender* as predictors, with a participant grouping factor (random intercept) (Gelman and Hill, 2006; Bates et al., 2012). We used a random intercept for participants to account for participant-specific differences in variation. Since object-name pairings are generated on the fly, these are different for each participant. As a consequence, we did not need to model item-level variation (e.g., with an item random intercept), making our models computationally more effective.

For each regression model in this paper, we started with a fully specified model including all interactions and removed non-significant predictors one by one, testing for model fit using analysis of variance and the Akaike information criterion (AIC). Where a combined model was too complex we fit interactions of participant-level predictors (age and gender) and experimental conditions (cue type, item presence in training, etc.) separately. We only report the best model, which means that we exclude predictors that were not significant.

Eight out of 498 participants (across the 6 experiments) did not disclose their age in the pre-test survey. When we tested age as a predictor, we re-fit models excluding these missing data and performed analysis of variance checks on these models to inform model selection. Models excluding participants with missing data were consistent with models fit on full data. For models for which age was justified as a predictor, the reported models exclude the few participants for which we have no age data. This model selection process assures that (i) we use all the available information in our models and that (ii) participant-level and experiment-level factors, along with their interactions, are tested in each experiment.

The best model for the test phase of Experiment I can be seen in **Table 2**. The model includes participant age. For Experiment I, all participants reported their age, and so no participants needed to be excluded on this basis.

The model shows that participants are more likely to pick the correct suffix in test if they have seen the item in training. Age is a significant predictor—older participants are more likely to give

TABLE 2 | Best model summary for Experiment I.

Formula: correct ~ item in training + age + (1 + item in training participant)				
	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	−2.17	1.17	−1.85	0.06
Item in training = TRUE	1.14	0.33	3.49	0.00
Age	0.11	0.04	3.05	0.00

correct answers. This effect is not strong compared to presence in training, but it is robust and remains even if we remove margin values.

4.6. Discussion

The results of Experiment I confirm that, within the current design, many people are able to accurately learn a morpho-phonological pattern they were trained on. They are able to ascertain the triggering linguistic context and choose the appropriate answer. They are also able to generalize this pattern to items not seen during training. This remains true despite the relatively low number of training items with which the cue was presented. However, performance is somewhat better for test items previously seen in training.

Test behavior follows training behavior closely. Participants who finish training earlier are more likely to have a high accuracy in test.

Figure 4 shows that average participant accuracy in test has a bimodal distribution. Those participants whose accuracy is below the overall mean (at 0.74) have means clustered around 0.5 (equivalent to chance), while participants above the overall mean have means clustered toward 1 (equivalent to a perfect score). Based on this difference we can divide participants into “good learners” and “poor learners.” This grouping is supported by the training data. If we compare training length for the “good learner” participants (those with mean above the overall mean in test) and the “poor learners,” we find that the former finish training faster (as supported by a Wilcoxon rank sum test, $W = 56$, $p < 0.001$). Note that there are 19 good learners and 26 poor learners, suggesting that the task is relatively hard (with a “passing rate” of 42%). These results may be compared to those in Becker et al. (2011), an experimental study of nonce words in Turkish in which vocalic cues to an alternation were found to be less learnable than a consonantal cue. Our results also show that many participants have difficulties learning a vocalic cue to an alternation.

The mean trial count of the “good learner” group in training is 32. That of the “poor learner” group is 51. Recall that training has 24 unique trials. If a participant does not find out the key to success in training (the stem vowel), and keeps guessing, but remembers every single guess and identifies it correctly afterwards, they will need about 36 trials to finish training on the average (since they have a 50% chance of guessing right in the first place, and only need to repeat half of the trials). If they keep guessing, they need 518 trials on the average. The mean of the “poor learner” group is clearly between these values,

suggesting that some rote learning did take place for this group (no participant needed 518 trials to finish), but it was not entirely efficient.

The “good learner”/“poor learner” distinction is *post-hoc*. Although we expected individual variability in learning, we did not hypothesize beforehand that listeners would fall into two clusters, with rather few “intermediate learners” falling between the “poor” and “good” learners. One possible interpretation is that the good learners are people who became consciously aware of the relevant cue. Conscious learning, also described in the research literature as “explicit learning” is generally faster than unconscious, or implicit learning (Goujon et al., 2015). While the good learners recognize the contextual pattern and simply apply it to all new items, some poor learners seem to perform rote learning as they repeatedly see training items. They do learn the correct suffixed form for some specific items, as evidenced by the greater number of participants who perform above chance in seen items, but they are not successful in generalizing to new items. It is also possible, of course, that this distribution does not relate to an explicit/implicit learning distinction at all, but rather reflects the distribution of individual learner characteristics in our data-set. Brooks et al. (2016), for example, have shown that morphological learning and generalization varies across individuals, in a way that correlates with measures of non-verbal intelligence and general statistical learning abilities. What follows is that if key individual learner characteristics are bimodal, then the learner outcomes would also be bimodal.

The results of Experiment I give us indications on how participants proceed through a learning task based on a cue associated with a linguistic context. The decisive point is whether a participant learned the pattern, and if this does not happen in training, participants will mostly guess in test. A sizeable group, but still a minority, learned the general cue association pattern. In Experiment II, we look at a similar task that uses a non-linguistic context.

5. EXPERIMENT II

In this experiment, the cue is no longer related to the name used by the conversation partner—rather, it is the conversation partner itself.

5.1. Participants

One hundred and five participants were recruited through Amazon Mechanical Turk. 51 are women, 54 men. Mean participant age is 34 years, with a standard deviation of 9.62. 54 participants were assigned to the *view* condition, 51 to the socially relevant *gender* condition. Four participants were excluded for not following the instructions properly. Four participants were removed based on training speed. We report data from the remaining 97 participants. All participants are native speakers of American English. Each person was paid three dollars upon completion of the task.

5.2. Methods

Experiment II modifies Experiment I in one major way. The correct response no longer depends on the vowel of the

prompt. Rather, it depends on the conversation partner, who was irrelevant in Experiment I. A non-linguistic contextual cue replaces the linguistic cue in learning a morpho-phonological pattern. The non-linguistic contextual cue is relatively basic. It is either who the conversation partner is or what physical orientation they have compared to the protagonist.

Experiment II has the same conversation partners as Experiment I, who, again, can each be seen in two different ways. This creates two groupings. One grouping, *gender*, is the identity of the conversation partner—who is either male or female. The other, *view*, is the spatial orientation of the conversation partner. The aim of this design is to teach naming patterns in conjunction with the *contextual cue* provided by the grouping. The images used can be seen in **Figure 2**. Learning the “view” cue requires participants to notice that changes in language use are correlated with changes in the direction the partner is facing. Learning the “gender” cue requires the participants to notice that changes in language use are correlated with changes in the speaker.

Who your conversation partner is has a huge effect on linguistic category learning. Listeners are able to keep track of information coming from two different speakers, adapt to new speakers, and recognize the difference between across-speaker and within-speaker variation and weigh them differently (Horton and Gerrig, 2002, 2005; Kraljic et al., 2008a). Perceived speaker gender is an especially robust cue (Johnson et al., 1999).

In contrast, conversation partner spatial orientation is much less salient as a social-indexical cue. People learn both deictic expressions (denoting spatial relations, such as “here” and “that”) and words with implicit spatial relations (such as “wide” or “tall”) easily, since these are frequent forms of every language. Variation between deictic expressions, however, does not typically carry social meaning.

Note that, if our participants in this task learn the association of the linguistic pattern with conversation partner, we have no way of knowing whether they are imputing a person-specific pattern, or a more general distinction based on person gender. As the most salient difference between the two partners is the gender difference, we here refer to the cue as a *gender* cue. Whether the learned cue is identity or gender can not be established from the design of Experiment II. However we will explicitly test the degree to which the learning to generalized to other speakers on the basis of gender in later experiments.

The game, then, has four conversation partner images. Each occurs once with each item in training. Again, targets are presented in a random order to each participant, and the participant has to give a correct answer for every target in order to move to the next target. In the test phase, targets are presented, again, in random order. No feedback is given. Training consists of six items, so it has $4 \times 6 = 24$ targets in total. The test contains these six items, and six items unseen in training, presented with each of the four conversation partners, so it has $4 \times 12 = 48$ targets in total.

The nonce language we used is similar to Experiment I, except that the vowel harmony pattern is absent. Instead, we used the five English vowel letters to make stimuli maximally distinct. The list of stimuli used in Experiment II can be seen in **Table 3**. The same principles guided stimuli selection as in

TABLE 3 | Stimuli set, Experiment II.

fek	rik	wuk	fal
pel	ril	tol	rul
wan	fen	wun	tas
fis	tos		

Experiment I. Since stem vowel is no longer relevant, we used the five English vowel letters to make the syllables more distinct. For each participant, two syllables are randomly selected as suffixes (marking conversation partner gender or spatial orientation, depending on the condition) while the rest are randomly assigned as item names.

There are four conversation partner images in the experiment, and two suffixes. Each suffix corresponds to two conversation partner images. The *across-subject* factor of Experiment I is the grouping of the conversation partner images. In the *gender* condition, the correct suffix (and, consequently, the correct response) is *cued* by the identity/gender of the conversation partner. In the *view* condition, the correct suffix (and so the correct response) is *cued* by the conversation partner's orientation (facing outwards or facing left).. The *within-subject* factor is whether a test item was seen in training.

5.3. Hypotheses

Experiment II looks at the association of a morpho-phonological pattern and a non-linguistic context. We had three hypotheses for Experiment II: (i) Participants would learn the diminutive pattern and extend it to new items in the *gender* condition (ii) learning and extension would be poorer in the *view* condition (iii) participants would be more likely to assign the correct pattern to items in the test phase if they have *seen* them in the training phase. We also evaluated participant age and gender as predictors of performance.

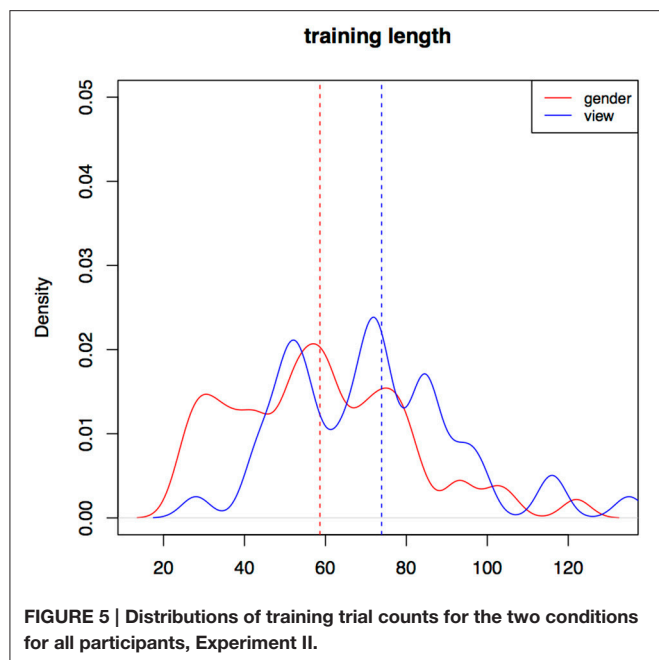
5.4. Results

We find that the pattern is indeed easier to learn with the *gender* condition. Unlike in Experiment I, item presence in training has no effect on response accuracy.

We use two measures of participant performance. In the training phase, we look at the number of trials it takes a participant to finish the experiment. This number provides information about the difficulty of learning in training and how much attention the participant pays to the task—this is why we use it as our main exclusion criterion. Participant responses in the test phase tell us how much they remember training and how easily they extend the pattern to new items and conversation partners.

Training takes longer (in terms of trial counts) in Experiment II ($m = 66$, $sd = 25$) than in Experiment I ($m = 42$, $sd = 18$) (a significant difference according to a Wilcoxon rank sum test, $W = 3450$, $p < 0.001$).

In Experiment II, participant training trial count is longer in the *view* condition ($m = 74$, $sd = 27$) than in the *gender* one ($m = 59$, $sd = 22$, $W = 1565$, $p < 0.01$). Training with the *gender* cue in Experiment II is still significantly longer than training in Experiment I. **Figure 5** is a kernel density plot of training trial



count for individual participants grouped by the two conditions. Mean trial count is shorter in both conditions than what we would expect for random behavior. Trial count is the number of trials it takes a participant to finish training. The smoothing bandwidth and the y axis are held constant for all density plots in this paper to aid comparison.

Figure 6 is a bean plot of participant responses, contrasting the *gender* condition and the *view* condition. For the *view* cue, most participants have a mean around 0.5—they are effectively guessing in test. In contrast, a sizeable proportion of participants has high accuracy for the *gender* cue. The bimodal structure of the *view* distribution strongly resembles the distribution of participant results in Experiment I for the unseen items.

We stepwise fit a binomial mixed-effects regression model on the test data, using response to individual items (*correct* or *not correct*) as an outcome variable and cue type (*gender* or *view*), *item presence in training*, and participant age and gender as predictors, with a participant random intercept. The summary for the best model can be seen in **Table 4**. Since age was a significant predictor, this model excludes 2 out of 97 participants who had no age data available. Model fitting in Experiment II is similar to Experiment I, we start with the most complex regression model and remove predictors one after the other until we reach the best fit. We test for all interactions of our terms.

The model shows that participants who are trained on the *gender* cue have much higher accuracy in test. Unlike in Experiment I, item presence in training is irrelevant—participant accuracy remains the same with previously seen and unseen items. Age is a significant predictor of test accuracy: older participants are more accurate.

5.5. Discussion

Cue type is a strong and independent predictor of test accuracy in Experiment II. Participants trained with the *gender* cue have

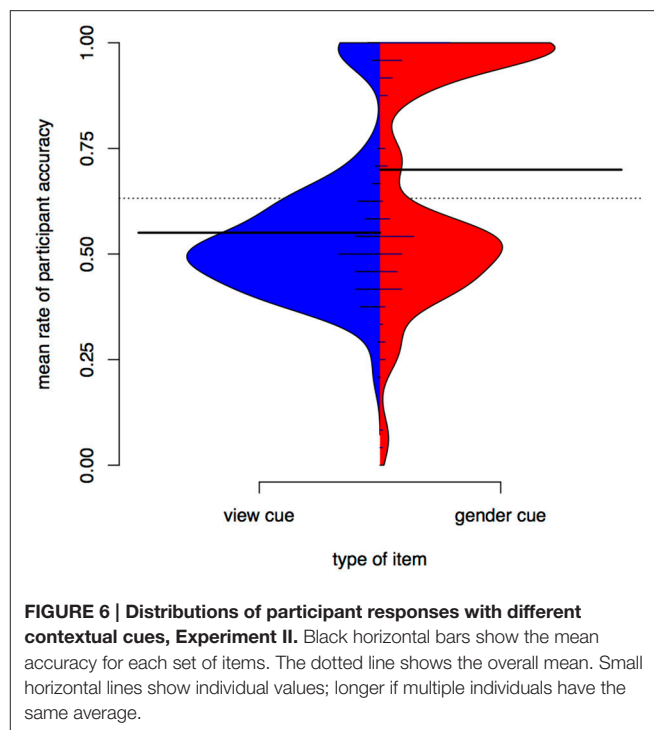


TABLE 4 | Best model summary for Experiment II.

Formula: $\text{correct} \sim \text{age} + \text{cue type} + (1 | \text{participant})$

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	-1.51	0.63	-2.40	0.02
Age	0.06	0.02	3.21	0.00
Cue type = gender	1.09	0.34	3.20	0.00

a much higher test accuracy, echoing results in the contextual learning of phonetic categories. Item presence in training does not affect test accuracy.

The Somers' Dxy Rank Correlation between test response accuracy and training trial count is modest (0.37). This is probably because participants show two types of behavior, much as in Experiment I. As we speculated above, some participants may have explicitly recognized the the context-pattern association, while others did not.

If we group participants with mean test accuracy above the overall mean as “good learners” and those below the overall mean as “poor learners,” we find that good learners finish training in significantly fewer trials ($W = 484, p < 0.001$).

If we look at the distribution of good learners across cue type, we find that most good learners are to be found in the *gender* condition (cf. **Table 5**). This tabulation supports the results of the regression analysis: the context-pattern association is easier to recognize for the *gender* cue than for the *view* cue.

When we compare Experiment II with Experiment I, we see that learning the non-linguistic cue is harder than learning the linguistic cue. As we note above, training takes longer. This

TABLE 5 | Good learners and poor learners across cue type, Experiment II.

	View cue	Gender cue
Good learner	5	25
Poor learner	40	27

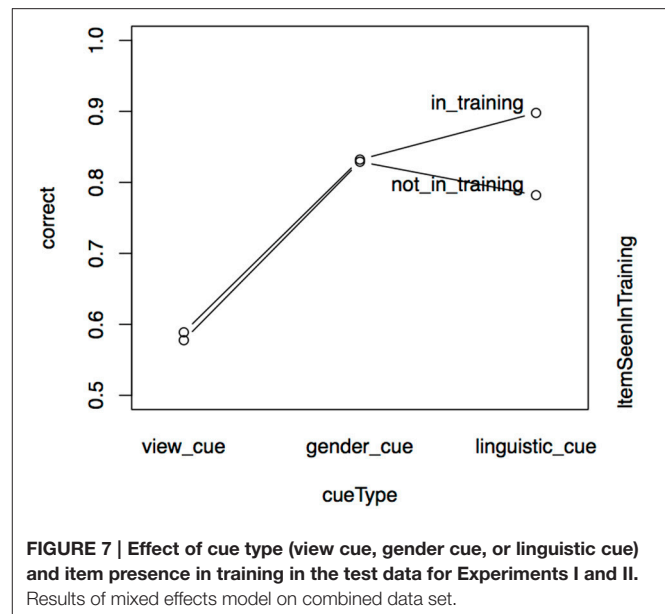
remains true if we compare the *gender* cue with the linguistic cue in Experiment I (learning the linguistic cue takes significantly fewer trials, $W = 639$, $p < 0.001$). An important difference between Experiment I and Experiment II, however, is that the linguistic cue is learned through exposure to a range of linguistic items, but the gender cue is learned through a contrast between just two people. A number of studies have shown that repetition and variability of context leads to improved learning (Gómez, 2002; Rost and McMurray, 2009). We cannot therefore directly compare the learnability of the linguistic and the social cue from these experiments alone.

Test accuracy for the *gender* cue in Experiment II is not worse overall than test accuracy in Experiment I. There is, however, an important difference in relation to item presence in training, which is significant in Experiment I but not in Experiment II. We merged the data from Experiments I and II and performed a binomial mixed-effects regression analysis, using the interaction of *item presence in training* and *cue type* (view, gender, linguistic) as predictors, with a participant random intercept. Effect sizes can be seen in **Figure 7**.

For the two contextual cues tested in Experiment II, item presence in training is not relevant. For the linguistic cue in Experiment I, participants are better at recalling names for items they have seen in training. This result could indicate that rote-based learning of items is relevant for the linguistic cue, but less so for the contextual cues (even for the *gender* cue, where a substantial amount of learning takes place). Note, however, that the role of the stem is different in the two experiments. In Experiment I, the participant must attend to the different stems in order to select the correct suffix. In contrast to Experiment I, the key to success in Experiment II is paying attention to the social context. The available responses always share the stem with the prompt word, which is irrelevant in relation to success on the task. This fact could explain the differing outcomes.

It remains clear that, in Experiment II, most learning happens with the socially salient, interpretable cue, *gender*—the identity of the conversation partner. The notion of social salience afforded by this experiment, however, is very narrow—it entails a distinction between two specific conversation partners (a woman and a man) as opposed to their position in space.

We have referred to the two cues as *view* vs. *gender*, assuming it is very likely that participants rely on the visible gender difference between the conversation partners in making their decisions. In Experiment II it is impossible to know whether participants are performing a categorization based on speaker gender, or simply associating the cue with the particular speakers. In Experiment III, we therefore continue exploring non-linguistic contexts, by more explicitly testing whether the associations learned in this type of experiment are extended to other partners, on the basis of conversation partner gender.



6. EXPERIMENT III

In this experiment, we look at whether participants generalize from the learning process we have seen in Experiment II, by extending the contextual cue to new conversation partners on the basis of gender.

6.1. Participants

The experiment was hosted on Amazon Mechanical Turk. 101 people participated in the experiment. 57 are women, 44 men. 50 are in the *gender* condition, 51 in the *view* one. Mean age is 32 years, with a standard deviation of 10.83. Four were excluded from the analysis based on training length. We report data from the remaining 97 participants. All are native speakers of American English. Participants were paid three dollars upon completion of the task.

6.2. Methods

Experiment III was designed to replicate the results of Experiment II, and in addition it investigates whether participants are able to generalize the contextual cue to new conversation partners in the test phase. It was identical to Experiment II except for the fact that Experiment III has eight conversation partners instead of four. Four conversation partners are present in training and test (just like in Experiment II) and four conversation partners are only present in test. Both previously seen and novel items are presented with previously seen and novel conversation partners in test, making the test twice as long as in Experiment II.

Experiment III uses all the conversation partner images in **Figure 2**. Conversation partners are grouped according to a *gender* attribute, as well as a *perspective* (*view*) one, their spatial orientation. We used an adult/child distinction for conversation partner images that are present in training vs. unique to the test. The reason for this is that we wanted to keep the two conversation

partner categories distinct visually. Some higher level knowledge is needed to realize that an adult and a child share the same gender. In contrast, two adult images of the same gender could have been matched based on visual similarity only.

The experiment has two across-subject factors. Half the participants have to learn the relevant cue (*gender*), and half of them the accidental cue (*view*). Also, half the participants are trained with children, and the other half with adults, creating four different training groups.

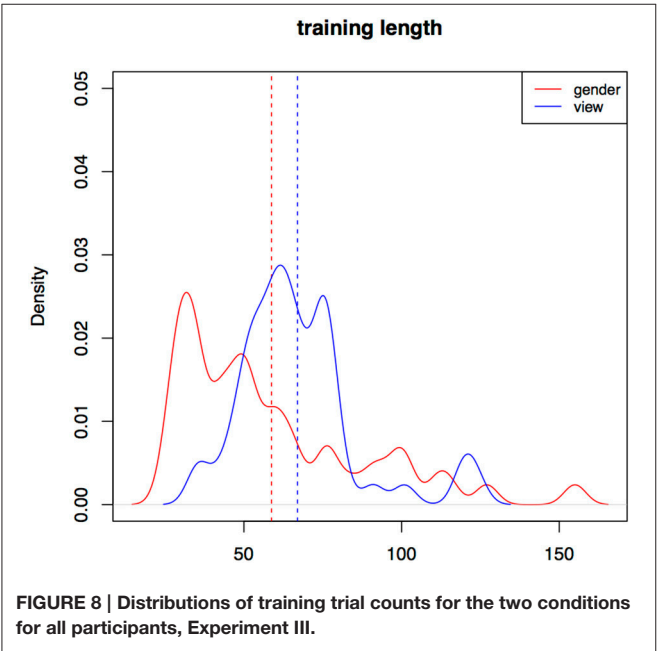
6.3. Hypotheses

We evaluated four hypotheses for Experiment III: (i) Participants would learn the diminutive pattern and extend it to new items and new conversation partners in the *gender* condition, (ii) The diminutive pattern would be easier to learn if it is associated with the *gender* cue than with the *view* cue (iii) Participants would be better able to assign the correct pattern to items in the test phase if they have seen them in the training phase, (iv) participants would be better able to assign the correct pattern to conversation partners that they had seen in the training phase. We also evaluated age and gender as predictors.

6.4. Results

Participants finish training much faster than a player would at random. On average, it takes participants longer to finish training in the *view* condition than in the *gender* condition (a significant difference according to a Wilcoxon rank sum test, $W = 1556$, $p < 0.01$). **Figure 8** is a density plot of training length for individual participants grouped by the two conditions. Training trial count in Experiment III is not significantly different from Experiment II.

Figure 9 shows a bean plot of participant test responses for the *gender* condition and for the *view* condition. Mean accuracy is much higher for the *gender* condition.



We stepwise fit a binomial mixed-effects regression model on the test data, using response to individual items (*correct* or *not correct*) as an outcome variable and the interaction of cue type (*gender* or *view*) and item *presence in training*, conversation partner *presence in training*, conversation partner *type* in training (*children* or *adults*), and participant age and gender as predictors, with a participant random intercept. Response accuracy is predicted by cue type. It does not depend on familiarity with items or conversation partners. Accuracy does not improve significantly with age, or differ by participant gender. The summary of the best model can be seen in **Table 6**.

6.5. Discussion

The results of Experiment III support the results of Experiment II, and show that the learning generalizes to other partners. Even when exposed to just one person in the training, participants extend this learning to others, on the basis of the person's *gender*.

While half the participants are trained with children and the other half with adults, this makes no difference in test accuracy. This further supports our assumption that the perceptual difference between our conversation partner images is far less relevant than their socially salient grouping characteristics.

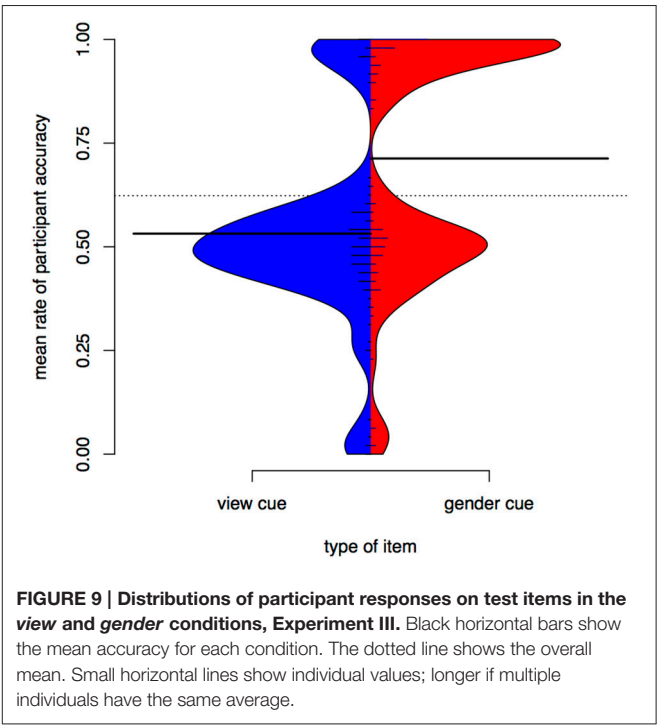


TABLE 6 | Best model summary, Experiment III.

Formula: $\text{correct} \sim \text{cue type} + (1 \mid \text{participant})$

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	0.28	0.31	0.90	0.37
Cue type = gender	1.68	0.44	3.81	0.00

As in the previous two experiments, participant mean test accuracy ratings show a clear bimodal distribution. We can group participants as good learners or poor learners, according to whether their test mean is above or below the overall mean. If we tabulate good learners across cue type, we find that the *gender* cue is easier to learn. This can be seen in **Table 7**.

The results of Experiment III are very similar to Experiment II. The main difference is that, in Experiment III, we have evidence that participants clearly rely on a more abstract context to establish generalizations. If they recognize conversation partner gender as the contextual cue, they are able to interpret it generally. They are able to learn this cue with adults and extend it to children and vice versa. This is comparable to the recognition of phonetic categories in stereotypical male and female voices. The huge difference is, however, that this distinction is both much more abstract (relying on a distinction in diminutive use) and simpler (a single difference in suffixes as opposed to a complex envelope of distinction between stereotypical male and female voices). This grants additional power to our socially salient distinction, which is generalized to differences between stereotypically male and female characters. This distinction, trained with only one instance of each gender, is straightforwardly generalized to a new instance (from a woman to a girl, etc.).

Now that we have established that learning based on just one person is extended to another person of the same gender in the test, this substantiates the choice of *gender* (rather than identity) as the most appropriate label to use for the person-based cue.

Note that the item presence in training is not a significant predictor of test accuracy in either Experiment II or III. This suggests that participants completely disregard the prompt word form and focus on the suffix and the associated context (if they focus on anything at all). The design of these experiments, however, does not allow us to explicitly test whether participants pay attention to the suffix vs. the stem and how this relates to training performance. Experiment IV addresses this question.

7. EXPERIMENT IV

In this experiment, we return to the learning process in Experiment II and look at the relative importance of our various cues by offering participants two test choices that are both “wrong,” in different ways.

7.1. Participants

The experiment was hosted on Amazon Mechanical Turk. 80 people participated in the experiment. 46 are women, 34 men. 40 are in the *gender* condition, 40 in the *view* one. Mean age is 32 years, with a standard deviation of 9.99. Two participants were excluded from the reported data based on training length. We

TABLE 7 | “Good” learners across cue type, Experiment III.

	View cue	Gender cue
Good learners	8	25
Poor learners	41	23

report data from the remaining 78 participants. All participants are native speakers of American English. Participants were paid three dollars upon completion of the task.

7.2. Methods

Experiment IV uses the adult *woman* and *man* conversation partners in *front* and *side* view.

For Experiment IV, as for Experiment II, context determines the correct response during the training phase. Each target has two possible responses. One has the suffix associated with the present context, the other has the suffix associated with the absent context. So, if the context is *gender* the participant must choose the response with the suffix that matches the gender of the conversation partner on screen. The stem of the two available responses is always the same, the name of the query, which is also visible on the screen.

The test phase of Experiment IV differs from Experiment II in two respects. First, during the test phase, participants are only exposed to previously seen items, no novel items are presented. And second, the query and the prompt name are no longer visible on the screen. One possible response for the target has the stem which is the name of the query of the target (as seen in training) but a context-inappropriate suffix (this is a choice present in the previous experiments). The other possible response has the correct suffix, but it has a stem that is not the name of the query of the target (as seen in training). Both answers are wrong (compared to training), but for different reasons. One has the correct prompt name, one the correct suffixation pattern, but neither has both. **Table 8** gives an example.

In the training phase, the stimuli were generated from the same pool as in Experiment III. For each participant, two syllables are assigned as suffixes. Six syllables are assigned as item names. In test, the “wrong conversation partner” answer was generated using the prompt name and the wrong suffix. The “wrong prompt name” was generated using a different, randomly assigned prompt name and the correct suffix. This means that the wrong stems were different for the same item across test trials.

Picking the response with the correct stem (the name of the query in training) but the wrong suffix (the one that belongs to the other cue) means that, during training, participants pay more attention to the entity they name than the context. Picking the response with the correct suffix (the one that belongs to the present cue) but the wrong stem (not the name of the query) means that, during training, participants pay more attention to the context than the entity they name. The naming task in

TABLE 8 | Example stimuli, Experiment II vs. Experiment IV.

	Training		Test	
Exp II	stem	tas	stem	tas
	stem+correct suffix	tasrul	stem+correct suffix	tasrul
	stem+incorrect suffix	taspel	stem+incorrect suffix	taspel
Exp IV	stem	tas	(stem not visible)	
	stem+correct suffix	tasrul	incorrect stem + correct suffix	fenrul
	stem+incorrect suffix	taspel	correct stem + incorrect suffix	taspel

Experiment II and III derive the name from the query image and the conversation partner, and the naming task in Experiment IV allows directly comparing their degree of relevance.

7.3. Hypotheses

We had two hypotheses for Experiment IV: (i) as in the previous experiments, participants would finish training faster in the *gender* condition than in the *view* condition. (ii) Participants would be more likely to focus on the suffix in the *gender* than in the *view* condition; as seen in experiments II–III, the *gender* cue contributes more to learning success, and hence it is likely easier to recognize and learn.

7.4. Results

The overall training duration of Experiment IV is not significantly different from that of Experiment III or Experiment II. As in Experiment II, training in the *gender* condition is significantly shorter than in the *view* one ($W = 1023$, $p < 0.01$, using a Wilcoxon rank sum test).

In the test phase, overall, participants pick the answer containing the correct suffix significantly more often than the answer with the correct stem (the “original” name) (59% of the time).

During test, participants in the *view* condition pick the correct stem overwhelmingly more (76% of the time) than in the *gender* condition (41% of the time). In the *gender* condition, the correct suffix is preferred more often (59% of the time).

Figure 10 shows the degree to which participants pick the stem (1) or the suffix (0), that is, the preference for the *stem* with the *view* cue (left) and the *gender* cue (right).

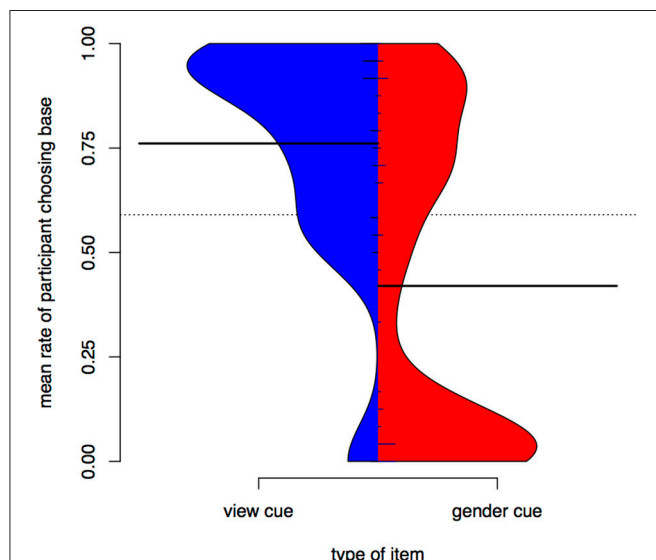


FIGURE 10 | Distributions of participants picking right stem plus wrong suffix (1) vs. wrong stem plus right suffix (0) in the two conditions, Experiment IV. Black horizontal bars show the mean rate of “stem” preference for each condition. The dotted line shows the overall mean. Small horizontal lines show individual values; longer if multiple individuals have the same average.

We stepwise fit a binomial mixed-effects regression model on the test results using “picked correct stem” (as opposed to “picked correct suffix”) as outcome variable and condition (*gender* or *view*) and participant age and gender as predictors, with a participant random intercept. The summary of the best model can be seen in **Table 9**. The only significant predictor is the condition, with the *view* cue leading to a stronger preference for the stem than the *gender* cue.

7.5. Discussion

In the *view* condition, participants overwhelmingly focus on the stem of the response, rather than the suffix. This suggests that, in the *gender* condition, the suffix is much easier to learn than in the *view* condition. This is also the outcome for Experiments II & III: accuracy for the *view* condition is not much higher than chance. Some participants learn the *view* cue, but many fewer than the *gender* cue.

The *gender* condition of Experiment IV is more interesting. The tight answer ratio (59 vs. 41%) indicates that participants can rely on either—there are “object” people and “people” people. This difference does not vary with participant age or gender. We can infer more about being an “object” or a “people” person as a learning strategy if we look at training performance for these two groups. **Figure 11** shows training trial counts for participants who overwhelmingly go for the stem or the suffix in test.

TABLE 9 | Best model summary, Experiment IV.

Formula: correct stem \sim cue type + (1 | participant)

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	1.86	0.44	4.27	0.00
Cue type = gender	−2.81	0.62	−4.54	0.00

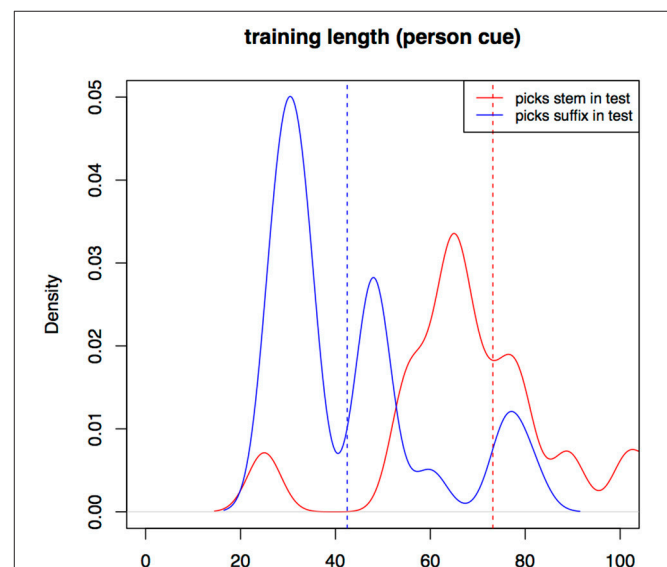


FIGURE 11 | Distributions of training trial counts for “object” and “people” participants, Experiment IV.

People who go on to pick the suffix in test are much faster to finish training than people who go on to pick the stem. We can interpret training and test performance in Experiment IV as results of either of two learning strategies. “Object” people focus on the stem, therefore take a while to finish the training, and overwhelmingly pick the stem in the test. “People” people focus on the suffix, finish training much faster, and pick the suffix in the test. We should note that “object” participants in the *gender* condition are as slow in training as participants in the *view* condition.

In the *gender* condition, as in the analogous conditions of experiments II–III, both the linguistic context and the non-linguistic context of the morphological pattern (the prompt name and the conversation partner’s gender/identity) are readily available. A group of participants are able to pin down the relevant factor in variation, namely, the conversation partner, and pick their responses accordingly. Others remain “distracted” by the prompt name. In the *view* condition, the non-linguistic context is barely if at all accessible—consequently, all participants focus on the linguistic context.

We use the word “focus” to refer both the participant’s attention (what part of the frame they pay attention to) and the participant’s weighing of the cues (how much importance they attribute to any cue; that is, any part of the frame that changes from trial to trial). These cannot be separated in our analysis, but likely together lead to the observed dichotomized participant behavior, dividing successful and poor learners in the experiment.

We now have strong reasons to believe that a robust and salient non-linguistic context is easier to learn than a less salient one. The generality of these findings, however, is somewhat compromised by the fact that we have thus far only looked at one morphological pattern, the diminutive, which is both highly variable in English and which has strong associations with gender in many languages (Jurafsky, 1993/2012). In order to make our findings more robust, we repeated Experiment III using the plural instead of the diminutive as the iconic relationship between prompts and targets. The main question was whether participant accuracy changes with visual stimuli cueing the plural replacing diminutive stimuli.

8. EXPERIMENT V

In this experiment, participants work with an artificial language that is based on a different iconic relationship, the plural instead of the diminutive.

8.1. Participants

The experiment was hosted on Amazon Mechanical Turk². 89 people participated in the experiment. 50 are women, 39 men. 46 are in the *gender* condition, 43 in the *view* one. Mean age is 37 years, with a standard deviation of 15.47. All are native speakers of American English. Participants were paid three dollars upon

²We initially ran 40 participants in this experiment. A reviewer pointed out that the lower participant count is problematic given that we compare this experiment to Experiment III. We have then run additional participants for Experiment V. Regression analysis shows no difference in the performance of the first and second batch in Experiment V.

completion of the task. We excluded 4 participants from the analysis based on training speed. We report data from the remaining 85 participants.

8.2. Methods

Experiment V is identical to Experiment III except for the prompt and target images. Experiment III, like all previous experiments, used a normal sized item and a diminutive item as the pair of pictures. The instructions told the participant to identify the name of the small item based on the larger item. In Experiment V, by contrast, each query picture displays an item and each target picture displays three of the same item. The instructions tell the participant to identify “the plural, the word for multiple instances of the same item.” Otherwise instructions are unchanged. The goal of Experiment V is to determine whether the results of Experiment III generalize to a morphological process (pluralization) that is highly general and productive in English and many other languages.

Experiment V uses all conversation partners in *front* and *side* view.

8.3. Hypotheses

Our hypothesis was that the patterns that we had previously observed would generalize beyond the particular case of the diminutive. We therefore evaluated the same hypotheses for Experiment V as for Experiment III. Based on the results of Experiment III, we expected that participants would learn the plural pattern and extend it to new items and new conversational partners. We expected learning to be more successful in the *gender* condition than in the *view* condition. We expected no advantage for seen items or partners, and we also looked for effects of participant age and gender.

8.4. Results

Here, we first look at Experiment V by itself and then together with Experiment III. Cue type has no effect on training speed in Experiment V. The mean rate of participant accuracy in test can be seen in **Figure 12**. We fit a mixed-effects binomial regression model on the test data using item and conversation partner presence in training, type of cue, and participant age and gender as predictors, with a participant random intercept. The model summary can be seen in **Table 10**. The only predictor that shows any effect is cue type. The effect size is above the level of statistical significance. This result is similar to what we see in Experiment III, even though the effect is weaker in test—and absent in training.

The only way to tell whether the experiments differ from each other significantly is to use statistical tests on the joint data from the two experiments.

Training in Experiment V does not differ significantly in length from training in Experiment III.

We merged the two datasets and stepwise fit a mixed-effects binomial regression model on the combined test data using item and conversation partner presence in training, type of cue, type of pattern (*diminutive* or *plural*), and participant age and gender as predictors, with a participant random intercept. The plural dataset patterns essentially the same as the diminutive dataset.

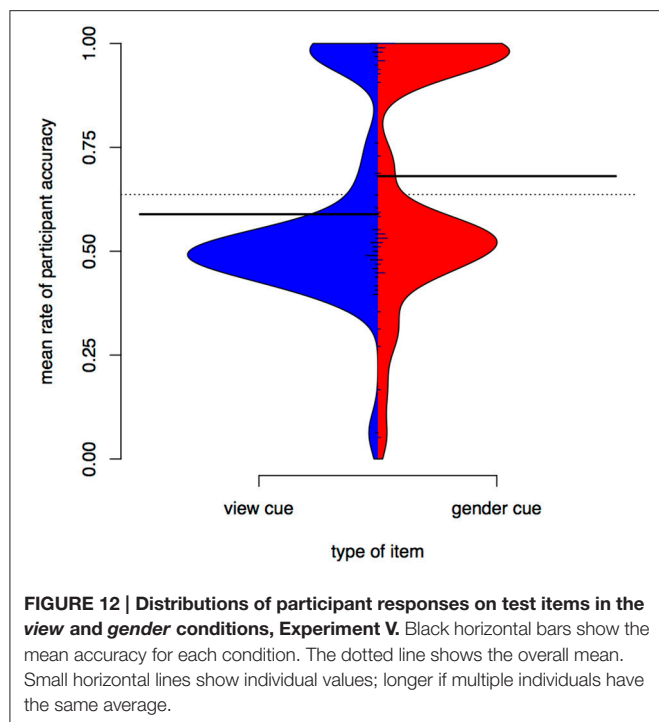


TABLE 10 | Best model summary, Experiment V.

Formula: correct ~ cue type + (1 | participant)

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	0.77	0.33	2.33	0.02
Cue type = gender	0.84	0.46	1.83	0.07

Cue type is a significant predictor of test accuracy. Participant age and gender and item presence in training are not significant. The type of pattern (diminutive/plural) does not affect test accuracy significantly. The summary of the best model of the merged test data can be seen in **Table 11**.

8.5. Discussion

Experiment V shows that the learning difference between the socially salient cue and the irrelevant cue persists when these cues are tied to a different morphological pattern, the plural. This adds further robustness to this distinction.

When we look at the experiment in itself, the effect of the gender cue is weaker than in other experiments, e.g., Experiment III. It is also above the generally accepted threshold of statistical significance. However, the statistical analysis of the two datasets together indicates that this difference is not statistically significant. The joint analysis gives us no ground to reject the null hypothesis that the plural does not differ from the diminutive. In general, finding significance levels for differences in statistical significance is difficult and would require a study considerably exceeding our scope at present. If future work establishes that socio-indexical associations for plural patterns are indeed more difficult to learn than for diminutive

TABLE 11 | Best model summary, Experiments III and V.

Formula: correct ~ cue type + (1 | participant)

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	0.50	0.23	2.22	0.03
Cue type = gender	1.29	0.32	4.03	0.00

patterns, the reasons for this difference would be of considerable interest. Potential factors could include adult differences in the adaptability of the derivational vs. the inflectional morphology, and pre-existing associations between the diminutive and social attributes of age, gender, or status (as described in Jurafsky, 1993/2012 as well as Kruisinga, 1942 cited by Bauer, 1997). For English, a further aspect is that the language has a number of competing diminutive suffixation patterns (such as *-ling*, *-ly*, *-ie*, etc), but only one broad, productive plural pattern.

In experiments II, III, and V, it is only with the salient cue that participants show a large degree of learning. However, only about half the participants exposed to the salient cue show high accuracy in test, while the other half of this group resorts to guessing, much like participants learning the non-salient cue. In Experiment IV we looked at learning strategies and proposed that, when both types of information are accessible, some participants will focus on the linguistic context (the prompt), and others at the non-linguistic context (the conversation partner). What remains unclear is whether participants make a by and large random choice at the beginning to focus on either context and then remain with it, or whether the effect of the non-linguistic context can be increased by expanding training. This is an especially relevant question given that item presence in training does not affect test accuracy, suggesting that the recognition of the relevant context is far more important than exposure to the specific training items. We look at this question in Experiment VI.

Another important question arising from our results across experiments I-V is the role of individual participant characteristics. We evaluated age and gender as individual predictors, with mixed results. These participant characteristics were not controlled in the participant recruitment procedure, and different experiments enrolled slightly different age and gender distributions.

In order to obtain more statistical power to look at these participant effects, we combined the test data for experiments II, III, and V, which have the same training size, the same test setup, and the same cue differences (*gender* and *view*). The pattern is either the diminutive or the plural. Those are also comparable. Each experiment has items in the test phase that were seen in training as well as new items. We fit a binomial mixed-effects regression model on the combined test data for participants with age available—273 participants. The outcome variable is response accuracy, the predictors are participant age and gender, as well as cue type and item presence in training, with a participant random intercept. The best model has age and cue type as significant predictors. The model summary can be seen in **Table 12**. Older

participants are more accurate overall, and responses to socially salient cues are much more likely to be correct. Cue type is a much stronger predictor than age. Participant gender is not a significant predictor. Similarly, inspection of the training data reveals a significant age effect, with older participants completing the training in significantly fewer trials.

The result shows that older participants are doing better with the socio-indexical learning. The effect, however, is not very strong.

9. EXPERIMENT VI

In this experiment, participants undergo an extended training phase. Extended training allows us to explore whether participants can modify their focus of attention based on feedback during training.

9.1. Participants

The experiment was hosted on Amazon Mechanical Turk. 100 people participated in the experiment. 55 are women, 45 men. 58 are in a *gender* condition. 42 participants are in the *view* condition. Mean age is 35 years, with a standard deviation of 10.98. Four participants were excluded based on training length. We report data for the remaining 96 participants. All participants are native speakers of American English. Participants were paid three dollars upon completion of the task.

9.2. Methods

Experiment VI is based on Experiments II and V. The morphological pattern is the plural, as in Experiment V. The extent of exposure is three times as great as in Experiment V: The plural pattern is trained with 18 (instead of 6) items and 4 conversation partners, and it is tested with these 18 items and 18 previously unseen items. There are no unseen conversational partners in the test phase, as in Experiment II. Since the focus is on the effect of familiarity with training items and since including new conversation partners as well would have prolonged the experiment to a large degree, we only included conversation partners seen in training in the test. We use the same conversation partners as in experiments I, II, and IV: the *woman* and the *man*. Our list of stimuli was expanded for Experiment VI (cf. Table 13). The main principle was to avoid adding syllables with consonant clusters which would upset the symmetry of the concatenated words. This was achieved by adding “ng,” the English consonant letter for the velar nasal, to the set of available syllable codas.

TABLE 12 | Best model summary, Experiments II, III, and V.

Formula: $\text{correct} \sim \text{age} + \text{cue type} + (1 | \text{participant})$

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	−0.51	0.43	−1.20	0.23
Cue type = gender	1.23	0.25	5.01	0.00
Age	0.03	0.01	2.45	0.01

Experiment VI uses the adult *woman* and *man* conversation partners in *front* and *side* view.

9.3. Hypotheses

Our hypotheses were based on the the results of Experiments II and V. We expected that participants would learn the contextual association more easily in the *gender* condition than in the *view* condition. We also expected that they would generalize to new items. In Experiment VI, we were also seeking a more in depth understanding of individual success rates. In addition to evaluating the effects of participant age and gender, we asked whether the lengthened training phase improves the success rate, compared to the previous experiments, and whether it affects the distribution of the good learners vs. poor learners.

9.4. Results

As in the previous experiment, we first analyzed the data from Experiment VI and then went on to compare it with Experiment II, which has a similar setup but shorter training.

Similar to most previous experiments, training takes longer with the non-salient cue (*view*) than with the salient cue (*gender*) ($W = 584, p < 0.001$).

The mean rate of participant accuracy in test can be seen in Figure 13.

We fit a regression model on the test phase of Experiment VI following the same procedure as in the previous experiments. The model summary can be seen in Table 14. The only significant predictor is cue type ($\beta = 2.45, se = 0.51, p < 0.001$).

We then compare the test data from Experiment VI to test data from Experiment II. Experiment II constitutes the best comparison since it also has no new conversation partners in test. The pattern type is the diminutive rather than the plural, but Experiment V provided little evidence that this would be a relevant dimension.

We merge the two test datasets to see whether test accuracy in Experiment VI (which has 18 training items) is better than in Experiment II (which has 6 training items), and whether this has any interaction with cue type (*gender* or *view*).

We stepwise fit a binomial mixed-effects regression model on the test data using response as an outcome variable and item presence in training, cue type, and participant age and gender as

TABLE 13 | Stimuli set, Experiment VI.

fos	ruk	wik	ril
fol	pil	fel	tos
fon	tang	rong	fok
tel	fek	rel	tas
feng	fong	ros	wis
wal	tal	rek	pek
pung	fus	tol	rik
wun	rak	ren	ral
tus	wus	rok	tok

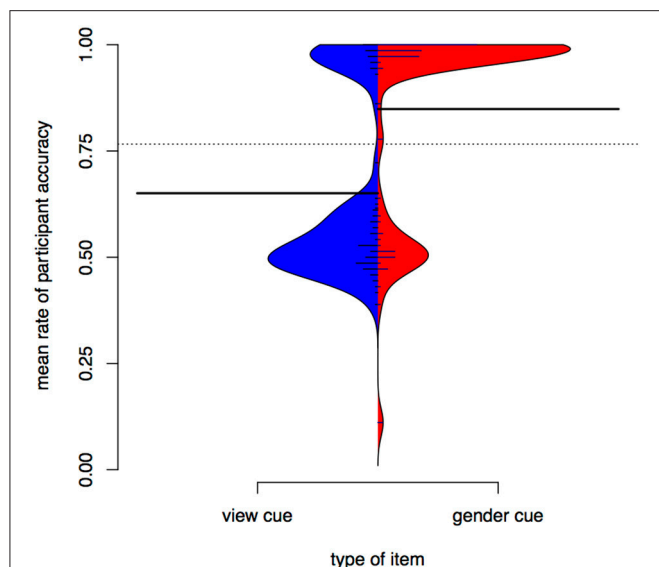


FIGURE 13 | Distributions of participant responses on test items in the *view* and *gender* conditions, Experiment VI. Black horizontal bars show the mean accuracy for each condition. The dotted line shows the overall mean. Small horizontal lines show individual values; longer if multiple individuals have the same average.

TABLE 14 | Best model summary, Experiment VI.

Formula: $\text{correct} \sim \text{cue type} + (1 | \text{participant})$

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	1.18	0.38	3.10	0.00
Cue type = gender	2.45	0.51	4.77	0.00

predictors, with a participant random intercept. The summary of the best model can be seen in **Table 15**³.

We see that cue type has a strong positive effect on test accuracy. Training length matters. Participants who go through longer training are more accurate in the test phase. However, this effect is mostly carried by the *gender* cue: longer training is beneficial to those who are trained with the gender distinction. The effect plot can be seen in **Figure 14**.

If we tabulate good learners and poor learners across conditions and compare the results to Experiment II (which is similar in structure but has a shorter training phase), we find that the ratio of good learners increases with increased training (cf. **Table 16**—there are more good learners in Experiment VI than in Experiment II).

9.5. Discussion

Based on the results of experiments I–IV, we hypothesized that two types of information are available to participants in this

³The way the experimental platform assigns participants to conditions has a slight random element, and one of the conditions has considerably fewer participants in it. In order to make sure that our results are robust, we re-fit our model on a subset of the test data with 39 participants in each condition. The main effects did not change considerably.

TABLE 15 | Best model summary, Experiments II and VI.

Formula: $\text{correct} \sim \text{training length} * \text{cue type} + (1 | \text{participant})$

	Estimate	Std. Error	z-value	Pr(> z)
(Intercept)	1.64	0.34	4.82	0.00
Training length = 18 items	1.94	0.46	4.24	0.00
Cue type = gender	−0.86	0.48	−1.78	0.08
Training length = 18 items:cue type = gender	−1.56	0.67	−2.34	0.02

experimental paradigm, the stem (the linguistic context) and the suffix (the non-linguistic context). If the non-linguistic context is salient, it is more readily available as a factor in variation. What remains unclear is whether participants then decide to focus on the stem or the suffix (resulting in less or more success in the experiment) at random or based on specific learning strategies. Experiment VI shows that if we increase the training set, more participants are able to determine the relevant cue for the linguistic pattern (the non-linguistic context). This indicates that, if participants adopt a learning strategy (e.g., focussing on the prompt picture or on the stem), some of them are able to update it based on evidence that it does not yield good results. Increasing the amount of evidence available by lengthening the training phase enables more participants to modify their strategy and succeed at the task.

Age has no effect on test accuracy for participants with extended training length. This indicates that training can overcome the age effect. It is true, however, that the effect of age in our experiments is not robust with this sample size. This means that we have to be very cautious in interpreting the lack of an effect here. Ultimately, the relationship of training length and age could only be tested with a larger sample, which is outside the scope of this paper.

10. SUMMARY

We have given a review of the literature to show that the non-linguistic context is extremely influential in learning linguistic constructions. Indeed, language use is shaped to a large degree by the social context. However, the link between contextual language learning and the observed structural complexity of social language use is far from completely understood.

We presented a series of artificial language learning experiments in which learning takes place in different contexts, which have different degrees of social-cognitive salience. The experiments were designed to investigate whether the relative social salience of contextual cues is relevant to learning a language pattern and whether this pattern is generalized to new lexical items and contexts. We hypothesized that participants would fare better at learning the link between the type of conversation partner and morphological pattern if the categorization of conversation partners was socially salient. We also assumed that this salient link would be generalized to new items and new conversation partners.

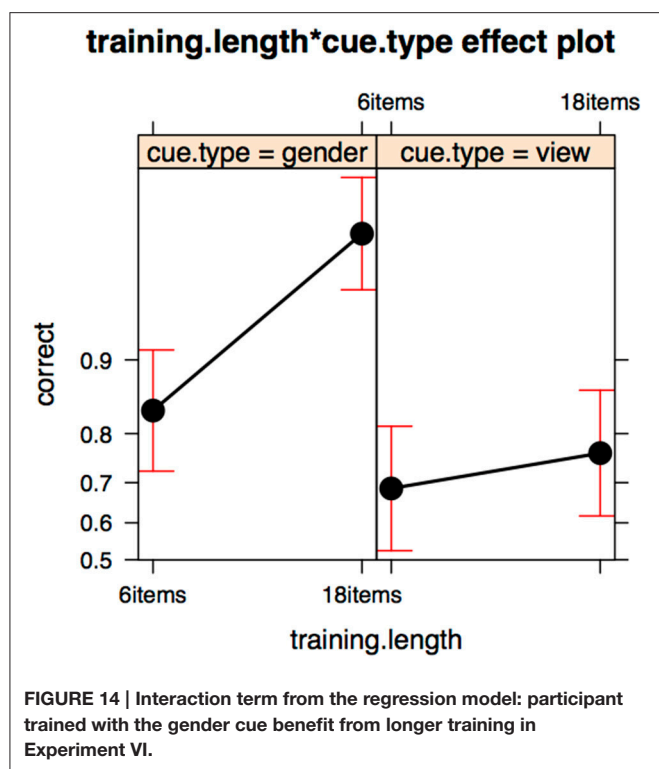


TABLE 16 | “Good” learners across cue type, Experiments II and VI.

	Experiment II (6 training items)		Experiment VI (18 training items)	
	View cue	Gender cue	View cue	Gender cue
Good learners	5	25	12	41
Poor learners	40	27	28	15

We found that participants learned the association of two morphological patterns (a diminutive suffix and a plural suffix) with conversation partner identity or gender, much as they learned the linguistic pattern that we used as a baseline (a cooccurrence constraint between the stem and suffix vowels).

Successful learners of the contextual association generalized well to new items. However, learning contextual associations was overall rather difficult for the participants. There were substantial individual differences in learning. As in the survey of statistical learning in various domains by Siegelman and Frost (2015), we found that people vary in their individual ability to learn from training data—some people have high accuracy, and others perform worse.

The test data distributions generally showed two distinct modes, one for “good learners” and one for “poor learners.” The adaptive tracking training enabled us to examine the differences between good learners and poor learners in more depth. Good learners finished the training phrase faster, suggesting that they identified and focused on the relevant cue better than poor learners did. However, even participants who were “good

learners” needed to make a number of mistakes to learn the pattern. The distributions of training trial counts for “good learners” reveal that training took good learners longer than would be needed for a player who plays ideally. This means that each of them had to make at least a few mistakes before they learned the pattern. With the lengthened training phrase in Experiment VI, a greater number of opportunities to notice the relevant cue had the result that a greater number of participants responded to failure by readjusting their focus, ultimately patterning as good learners in the test phase.

An important result of these experiments is the relative success for different non-linguistic contextual dimensions. Social saliency is very important. When the link between the conversation partner and the item appeared relatively accidental (side-facing, vs. front-facing), the association was very difficult to learn. When the link was socially coherent and interpretable (conversation partner gender), the learning task was considerably easier (Experiments III, V, and VI). Participants learn relatively easily, for example, that a particular adult female calls a small *fen* a *fenwun*, whereas a different person—a male—calls it a *fentas*. Participants orient early to the contextual cue of gender, and easily generalize this both to new items and to other conversation partners (Experiment III). This is true even when little evidence of generality is actually given. That is, exposure to just one female partner saying *fenwun* (in two views) leads to the hypothesis that all females would prefer *fenwun* to *fentas*.

Another aspect of learning is the competition between the linguistic context and non-linguistic context. In Experiment I, where participants need to focus on the linguistic context (the prompt name), familiar items (ones seen in training) are chosen more accurately in test. In Experiments II and III, where participants need to focus on the non-linguistic context (along with the suffix), this effect is absent. This remains true even for the “poor learner” participant group—those who did not seem to pin down the relevant contextual difference (conversation partner gender or spatial orientation).

In Experiment IV, we see that participants who focus on the suffix in test also finish training faster. This, in turn, supports the interpretation that the two types of information (stem and suffix) are competing with each other. Concentrating on the suffix is the key to success. Experiment V shows that this learning strategy is robust (applies for learning both the plural and the diminutive) while Experiment VI shows that the choice of stem or suffix as the main locus of attention is not fixed. With increased training, more participants figure out the relevant dimension and respond like “good learners” in the test phase.

These results can be compared with the results of Lleras and Von Mühlenen (2004). They find that, in a learning experiment, where contextual cues correlate with tasks, participants that employ an “active” searching strategy, and focus on the task itself, do not rely on contextual cues. The participants in our experiments follow an analogous pattern. Those who focus on the stem ignore social contextual cues. For the baseline Experiment I, the social contextual information is irrelevant and focusing on the stem would lead to success; but in the other experiments, focusing on the stem would cause the participant to overlook the information that is actually relevant to the task. The interesting

point is that *whether* they focus on the stem or the suffix depends on the kind of social contextual cue present. They are more likely to rely on the social contextual cue if it is salient.

Taken together, the results of these experiments provide solid evidence that adults are able to learn contextual meaning, and that they orient more toward contextual information that is socially salient and relevant than to contextual variation that appears accidental.

11. GENERAL DISCUSSION

The focus of this article is learning associations between a morphological pattern and a non-linguistic context. The main question is how the social salience of the context influences success in learning.

We contrasted the learning of two cues, one of which is socially salient (*gender*) and one of which is not (*view*), showing that the former is learned more easily than the latter. As we note in Section 3, the perceptual differences between the images are unlikely to affect their categorization.

Of course, the forced-choice paradigm has its limitations. When we say that participants were better at learning the *gender* cue, this needs to be interpreted within the context of the task. We do not know whether they learned it well enough to produce it unprompted, for example, nor do we know whether, outside of a forced-choice task, they would have preferred some unknown other response. The positive side of a forced-choice paradigm is that our results are easy to interpret statistically. But the results do open up further questions about how the results would pattern if a free-response paradigm was used. As we only contrast *gender* and *view*, a further question is whether these two conditions may vary on unknown dimensions other than salience. Further research using other images and other contrasts is therefore still required.

Social salience has a top-down effect. Prior experience teaches us that some differences are more important than others, and we pay more attention to these in linguistic categorization. The way we see the world, then, has a strong influence on our language use, resulting in the complex structures of indexicality discussed by sociolinguists on the population level. This article provides evidence that this influence is present on an individual level. The social salience of the images is likely to rely on more than "gender", but it remains the core manipulation that participants react to. The manipulation appears to provide a very reliable effect, despite the simplicity of the experimental paradigm.

Our results indicate that participants give more accurate answers when they recognize the relevant distinction (e.g., female/male in the *gender* condition). This is broadly analogous to explicit social stereotypes in terms of recognizing both the pattern and the context, as well as the connection between the two. At the same time, it can also be extended to cases in which either the context or the pattern is recognized and negotiated explicitly. The former is typical of most cases of social-indexical variation, in which we know our conversation partner's principal characteristics. The latter is typical of word patterns in particular, such as the choice between the formal and the informal second

person conjugation in French and German, and the dialect-specific vocabularies of English, German, French, and many other national languages.

Experiment I, as well as the combined analysis of Experiments II, III, and V, showed that older participants were more successful at both morpho-phonological and socio-indexical learning.

The age effect may arise because prior experience, increasing throughout the lifespan, has a beneficial effect on learning tasks, as proposed by Ramscar et al. (2014). As older participants were better at learning all associations presented (including the un-natural *view* association), this cannot be viewed as an effect of increased experience with socially relevant distinctions. Rather, it would have to be interpreted as an effect of increased general experience with learning socio-indexical and linguistic associations. The age effect may also arise in some way from the specific nature of our task. For example, if participants select the wrong answer, they get feedback. This feedback could provide them with information that are orienting to the wrong cue. There is some evidence that older participants make better use of feedback, especially in situations in which they are initially uncertain (Metcalfe et al., 2015).

It is important to note that the age range of our participants is restricted, when considered in the context of the literature on ageing. Only one of our 498 participants (with age data available) is over 70. The considerable literature on cognitive decline in ageing across a range of psychological tasks compares younger and older adults (usually over 70), with an assumption that speakers in the middle (i.e., 40–60) fall somewhere in between (Lachman, 2004). While the middle-aged group tends to be less studied, there are at least some studies which show improvement from younger adults to middle age, before declining again in older adults. Tasks where such an effect has been reported include everyday problem-solving (Thornton et al., 2013) and social problem-solving (D'Zurilla et al., 1998).

It is interesting to note that we see no age effect for the task with extended training (Experiment VI). This may suggest that, whatever the root of the age effect, additional practice can neutralize the benefits of increased age in this task.

Our results indicate a major role for social salience in the acquisition of contextual meaning in morphology. In our task, the contextual information is associated with the morphological pattern in the cognitive representation, and influences recall and generalization. Whether or not participants are overtly aware of the association, it is sufficient to nudge them in the correct direction in a two-way forced choice test. The fact that the gender-dependent association is learned better than an accidental association, and that performance slightly improves with age, reveals the role of prior knowledge and expectation about what aspects of the context may potentially be relevant. Many aspects of language vary according to the gender of the conversation partner, and in the participants' prior sociolinguistic learning, the gender of the conversation partner will have been relevant many times. Foulkes (2010) hypothesizes that some types of indexical properties should be more readily transmitted than others, based on the frequency with which they have been relevant in individuals' past experience. He identifies gender as one of the earliest learned socio-indexical associations. Children as young as

6 months, for example, preferentially match sex-cued voices and faces (Walker-Andrews et al., 1991). It is likely this considerable prior experience that facilitates a ready generalization across conversation partners.

12. CONCLUSION

Our paradigm demonstrates differences in adult learning of socially salient vs. accidental non-linguistic contextual cues. It also reveals a number of questions about the way we learn contextual associations of higher level linguistic structures. Does a varying non-linguistic context aid the learning of a linguistic pattern? Do we learn the diminutive more easily, for example, if we are exposed to more types of conversation partners who use it? What is the effect of attention to particular linguistic patterns and non-linguistic contexts? Does variance in a non-linguistic difference that we explicitly attend to aid language learning? Finally, amongst socially salient non-linguistic cues, are some easier to learn than others? Is it easier to learn the association of a linguistic pattern with gender, for example, than with age? These questions remain to be answered by follow-up research.

Our controlled laboratory experiments are, of course, still many worlds apart from the type of complex socio-contextual learning and generalization that occurs in every day interaction and language acquisition. However they do provide some first steps toward shedding some light on the complex cognitive mechanisms that must be at play in such learning. Whether an associative pattern is attended to, learned and recreated by a speaker will be affected by a range of factors—including who that

individual is, how socially salient the relevant context is, and how much the learner is exposed to that association. Our experiments have shown that individual variability in individual listeners, the saliency of socio-contextual associations, and differing patterns of exposure, likely all play some role in affecting socio-contextual learning in morphology.

AUTHOR CONTRIBUTIONS

PR designed and ran the experiments, performed statistical analysis, and wrote up the results. JH and JP contributed to the design, provided feedback on the experiments and analysis and contributed to writing up the manuscript.

FUNDING

This project was made possible through the support of a grant to Northwestern University from the John Templeton Foundation (Award ID 36617). The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation. JH was also supported by a Royal Society of New Zealand Rutherford Foundation Grant (Grant Number E5909).

ACKNOWLEDGMENTS

The authors would like to thank Chun Liang Chan, Viktória Papp, Alex Schumacher, Márton Sóskuthy, Clay Beckner, Patrick LaShell, Steven Franconeri, and Kayo Takasugi.

REFERENCES

- Altmann, E. G., Pierrehumbert, J. B., and Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLoS ONE* 6:e19009. doi: 10.1371/journal.pone.0019009
- Baayen, R. H., Hendrix, P., and Ramscar, M. (2013). Sidestepping the combinatorial explosion: an explanation of *n*-gram frequency effects based on naive discriminative learning. *Lang. Speech* 56, 329–347. doi: 10.1177/0023830913484896
- Bates, D., Maechler, M., and Bolker, B. (2012). *lme4: Linear Mixed-Effects Models using Eigen and R*. R Package.
- Bauer, L. (1997). Evaluative morphology: in search of universals. *Stud. Lang.* 21, 533–575. doi: 10.1075/sl.21.3.04bau
- Baxter, G. J., Blythe, R. A., Croft, W., and McKane, A. J. (2009). Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. *Lang. Variat. Change* 21, 257–296. doi: 10.1017/S095439450999010X
- Becker, M., Ketrez, N., and Nevins, A. (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87, 84–125. doi: 10.1353/lan.2011.0016
- Beckner, C., Rácz, P., Brandstetter, J., Hay, J., and Bartneck, C. (2016). Participants conform to humans but not to humanoid robots in an English past tense formation task. *J. Lang. Soc. Psychol.* 35, 158–179. doi: 10.1177/0261927X15584682
- Brooks, P. J., Kwoka, N., and Kempe, V. (2016). Distributional effects and individual differences in L2 morphology learning. *Lang. Learn.* doi: 10.1111/lang.12204. [Epub ahead of print].
- Campbell-Kibler, K. (2011). Intersecting variables and perceived sexual orientation in men. *Amer. Speech* 86, 52–68. doi: 10.1215/00031283-1277510
- Cheshire, J. (2002). "Sex and gender in variationist research," in *Handbook of Language Variation and Change*, eds J. K. Chambers, P. Trudgill, and N. Schilling (Hoboken NJ: Wiley Blackwell), 423–443.
- Chun, M. M., and Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn. Psychol.* 36, 28–71. doi: 10.1006/cogp.1998.0681
- Coupland, J., Coupland, N., and Giles, H. (eds.). (1991). "Accommodation theory: communication, context and consequences," in *Contexts of Accommodation* (Cambridge; Paris: Cambridge University Press, Éditions de la Maison des Sciences de l'homme), 1–68.
- Crump, M. J., McDonnell, J. V., and Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE* 8:e57410. doi: 10.1371/journal.pone.0057410
- Denton, S. E., Kruschke, J. K., and Erickson, M. A. (2008). Rule-based extrapolation: a continuing challenge for exemplar models. *Psychon. Bull. Rev.* 15, 780–786. doi: 10.3758/PBR.15.4.780
- Dixon, R. M. W. (1980). *The Languages of Australia*. Cambridge: Cambridge University Press.
- Docherty, G. J., Langstrof, C., and Foulkes, P. (2013). Listener evaluation of sociophonetic variability: probing constraints and capabilities. *Linguistics* 51, 355–380. doi: 10.1515/ling-2013-0014
- Drager, K. (2010). Sociophonetic variation in speech perception. *Lang. Linguist. Compass* 4, 473–480. doi: 10.1111/j.1749-818X.2010.00210.x
- D'Zurilla, T. J., Maydeu-Olivares, A., and Kant, G. L. (1998). Age and gender differences in social problem-solving ability. *Pers. indiv. Diff.* 25, 241–252.
- Eckert, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Hoboken, NJ: Wiley-Blackwell.
- Eckert, P. (2008). Variation and the indexical field. *J. Sociolinguist.* 12, 453–476. doi: 10.1111/j.1467-9841.2008.00374.x
- Eckert, P. (2012). Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Ann. Rev. Anthropol.* 41, 87–100. doi: 10.1146/annurev-anthro-092611-145828

- Fagyal, Z., Swarup, S., Escobar, A. M., Gasser, L., and Lakkaraju, K. (2010). Centers and peripheries: network roles in language change. *Lingua* 120, 2061–2079. doi: 10.1016/j.lingua.2010.02.001
- Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17897–17902. doi: 10.1016/j.lingua.2010.02.001
- Foulkes, P. (2010). Exploring social-indexical knowledge: a long past but a short history. *Lab. Phonol.* 1, 5–39. doi: 10.1515/labphon.2010.003
- Foulkes, P., and Docherty, G. (2006). The social life of phonetics and phonology. *J. Phonet.* 34, 409–438. doi: 10.1016/j.wocn.2005.08.002
- Foulkes, P., and Hay, J. B. (2015). “The emergence of sociophonetic structure,” in *The Handbook of Language Emergence*, eds B. MacWhinney and W. O’Grady (Hoboken, NJ: Wiley Blackwell), 292.
- Gelman, A., and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- German, J. S., Carlson, K., and Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *J. Phonet.* 41, 228–248. doi: 10.1016/j.wocn.2013.03.001
- Gluck, M. A., and Myers, C. E. (2001). *Gateway to Memory: An Introduction to Neural Network Modeling of the Hippocampus and Learning*. Cambridge, MA: MIT Press.
- Godden, D. R., and Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *Br. J. Psychol.* 66, 325–331. doi: 10.1111/j.2044-8295.1975.tb01468.x
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13, 431–436. doi: 10.1111/1467-9280.00476
- Goujon, A., Didierjean, A., and Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends Cogn. Sci.* 19, 524–533. doi: 10.1016/j.tics.2015.07.009
- Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cogn. Sci.* 11, 23–63. doi: 10.1111/j.1551-6708.1987.tb00862.x
- Hay, J., and Walker, A. (2013). “Skewed experience with words affects lexical access patterns,” in *Variation and Language Processing (VALP) 2* (Christchurch).
- Hay, J. B., and Drager, K. (2007). Sociophonetics. *Ann. Rev. Anthropol.* 36, 89–103. doi: 10.1146/annurev.anthro.34.081804.120633
- Hay, J. B., and Drager, K. (2010). Stuffed toys and speech perception. *Linguistics* 48, 865–892. doi: 10.1515/ling.2010.027
- Henderson, L., Devine, K. W., and Gaskell, M. (2015). When the daffodot flew to the intergalactic zoo: off-line consolidation is critical for word learning from stories. *Dev. Psychol.* 51, 406–417. doi: 10.1037/a0038786
- Horton, W. S., and Gerrig, R. J. (2002). Speakers’ experiences and audience design: knowing when and knowing how to adjust utterances to addressees. *J. Mem. Lang.* 47, 589–606. doi: 10.1016/S0749-596X(02)00019-0
- Horton, W. S., and Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition* 96, 127–142. doi: 10.1016/j.cognition.2004.07.001
- Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* 20, 1254–1259. doi: 10.1109/34.730558
- Johnson, K., Strand, E. A., and D’Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *J. Phonet.* 27, 359–384.
- Jurafsky, D. (1993/2012). “Universals in the semantics of the diminutive,” in *Annual Meeting of the Berkeley Linguistics Society*, Vol. 19 (Berkeley, CA). doi: 10.3765/bls.v19i1.1531
- Kraljic, T., Brennan, S. E., and Samuel, A. G. (2008a). Accommodating variation: dialects, idiolects, and speech processing. *Cognition* 107, 54–81. doi: 10.1016/j.cognition.2007.07.013
- Kraljic, T., and Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychon. Bull. Rev.* 13, 262–268. doi: 10.3758/BF03193841
- Kraljic, T., Samuel, A. G., and Brennan, S. E. (2008b). First impressions and last resorts. How listeners adjust to speaker variability. *Psychol. Sci.* 19, 332–338. doi: 10.1111/j.1467-9280.2008.02090.x
- Kruisinga, E. (1942). *Diminutieve en Affektieve Suffixen in de Germaanse Talen*. Groningen: Noord-Hollandsche Uitgevers Maatschappij.
- Labov, W. (2001). *Principles of Linguistic Change: Social Factors*. Hoboken, NJ: Wiley-Blackwell.
- Lachman, M. E. (2004). Development in midlife. *Annu. Rev. Psychol.* 55, 305–331. doi: 10.1146/annurev.psych.55.090902.141521
- Langstrof, C. (2014). *Sociophonetic Learning in L1 and L2*. Freiburg: Habilitation, University of Freiburg.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Percept. Psychophys.* 63, 1279–1292. doi: 10.3758/BF03194543
- Leung, J. H., and Williams, J. N. (2012). Constraints on implicit learning of grammatical form–meaning connections. *Lang. Learn.* 62, 634–662. doi: 10.1111/j.1467-9922.2011.00637.x
- Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013
- Lleras, A., and Von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. *Spat. Vis.* 17, 465–482. doi: 10.1163/1568568041920113
- Mathworks (2016). *MATLAB Version 9.1.0.441655 (R2016b)*. Natick, MA: The Mathworks, Inc.
- Maye, J., Aslin, R. N., and Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cogn. Sci.* 32, 543–562. doi: 10.1080/03640210802035357
- Metcalfe, J., Casal-Roscum, L., Radin, A., and Friedman, D. (2015). On teaching old dogs new tricks. *Psychol. Sci.* 26, 1833–1842. doi: 10.1177/0956797615597912
- Milroy, J., and Milroy, L. (1993). Mechanisms of change in urban dialects: the role of class, social network and gender. *Int. J. Appl. Linguist.* 3, 57–77. doi: 10.1111/j.1473-4192.1993.tb00043.x
- Milroy, L. (1980). *Language and Social Networks*. Oxford: Oxford University Press.
- Munro, R., Bethard, S., Kuperman, V., Lai, V. T., Melnick, R., Potts, C., et al. (2010). “Crowdsourcing and language studies: the new generation of linguistic data,” in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk* (Association for Computational Linguistics), 122–130.
- Needle, J. N., Pierrehumbert, J. B., and Hay, J. B. (2015). “Effects of pseudomorphology on the wordlikeness of pseudowords,” in *Architectures and Mechanisms for Language Processing*, 3–5 September. University of Malta Valletta (Los Angeles, CA).
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *J. Exp. Psychol. Learn. Mem. Cogn.* 14:700. doi: 10.1037/0278-7393.14.4.700
- Pele, O., and Werman, M. (2010). “The quadratic-chi histogram distance family,” in *European Conference on Computer Vision* (New York, NY: Springer), 749–762.
- Pierrehumbert, J. B. (2006). The next toolkit. *J. Phonet.* 34, 516–530. doi: 10.1016/j.wocn.2006.06.003
- Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., and Bailey, J. M. (2004). The influence of sexual orientation on vowel production. *J. Acoust. Soc. Am.* 116, 1905–1908. doi: 10.1121/1.1788729
- Pierrehumbert, J. B., Stonedahl, F., and Daland, R. (2014). A model of grassroots changes in linguistic systems. *arXiv preprint arXiv:1408.1985*.
- Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Lang. Awareness* 5, 40–74. doi: 10.1080/09658416.1996.9959890
- Qian, T., Jaeger, T., and Aslin, R. N. (2014). *Implicit Learning in a Non-stationary Environment: Knowing When It’s Groundhog Day*. MS, University of Rochester (Accessed March 17, 2014).
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rácz, P. (2013). *Salience in Sociolinguistics: A Quantitative Approach*. Berlin: De Gruyter - Mouton.
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Top. Cogn. Sci.* 6, 5–42. doi: 10.1111/tops.12078
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85:59. doi: 10.1037/0033-295X.85.2.59
- Roberts, G. (2008). Language and the free-rider problem: an experimental paradigm. *Biol. Theory* 2, 174–183. doi: 10.1162/biot.2008.3.2.174
- Rost, G. C., and McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Dev. Sci.* 12, 339–349. doi: 10.1111/j.1467-7687.2008.00786.x
- Säily, T., and Suomela, J. (2009). “Comparing type counts: the case of women, men and -ity in early English letters,” in *Corpus Linguistics: Refinements*

- and Reassessments, eds A. Renouf and A. Kehoe (Amsterdam: Rodopi), 87–109.
- Schumacher, R. A., Pierrehumbert, J. B., and LaShell, P. (2014). “Reconciling inconsistency in encoded morphological distinctions in an artificial language,” in *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Cognitive Science Society (Quebec City, QC).
- Siegelman, N., and Frost, R. (2015). Statistical learning as an individual ability: theoretical perspectives and empirical evidence. *J. Mem. Lang.* 81, 105–120. doi: 10.1016/j.jml.2015.02.001
- Silverstein, M. (2009). “The limits of awareness,” in *Linguistic Anthropology: A Reader*, Vol. 1, ed A. Duranti (Oxford: Blackwell), 382–401.
- Smith, S. M., and Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychon. Bull. Rev.* 8, 203–220. doi: 10.3758/BF03196157
- Soliz, J., and Giles, H. (2014). Relational and identity processes in communication: a contextual and meta-analytical review of communication accommodation theory. *Commun. Yearbook* 38, 106–143. doi: 10.1080/23808985.2014.11679160
- Tagliamonte, S. A., and Roeder, R. V. (2009). Variation in the English definite article: socio-historical linguistics in t’speech community. *J. Sociolinguist.* 13, 435–471. doi: 10.1111/j.1467-9841.2009.00418.x
- Thornton, W. L., Paterson, T. S., and Yeung, S. E. (2013). Age differences in everyday problem solving: the role of problem context. *Int. J. Behav. Dev.* 37, 13–20. doi: 10.1177/0165025412454028
- Van der Zande, P., Jesse, A., and Cutler, A. (2014). Cross-speaker generalisation in two phoneme-level perceptual adaptation processes. *J. Phonet.* 43, 38–46. doi: 10.1016/j.wocn.2014.01.003
- Von Ahn, L. (2006). Games with a purpose. *Computer* 39, 92–94. doi: 10.1109/MC.2006.196
- Walker, A., and Hay, J. B. (2011). Congruence between ‘word age’ and ‘voice age’ facilitates lexical access. *Lab. Phonol.* 2, 219–237. doi: 10.1515/labphon.2011.007
- Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., and Diaz, I. (1991). Infants’ bimodal perception of gender. *Ecol. Psychol.* 3, 55–75. doi: 10.1207/s15326969eco0302_1
- Wells, J. C. (1982). *Accents of English*, Vol. 1. Cambridge: Cambridge University Press.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Yu, C., and Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychol. Sci.* 18, 414–420. doi: 10.1111/j.1467-9280.2007.01915.x

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.

Copyright © 2017 Rácz, Hay and Pierrehumbert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

