IMPLICATIONS OF PSYCHO-COMPUTATIONAL MODELLING FOR MORPHOLOGICAL THEORY

EDITED BY: Vito Pirrelli, Harald Baayen, Walter Daelemans and Robert Malouf

PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-541-6
DOI 10.3389/978-2-88976-541-6

About Frontiers

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

Frontiers Journal Series

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

Dedication to Quality

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

What are Frontiers Research Topics?

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

IMPLICATIONS OF PSYCHO-COMPUTATIONAL MODELLING FOR MORPHOLOGICAL THEORY

Topic Editors:
Vito Pirrelli, National Research Council (CNR), Italy
Harald Baayen, University of Tübingen, Germany
Walter Daelemans, University of Antwerp, Belgium
Robert Malouf, San Diego State University, United States

Citation: Pirrelli, V., Baayen, H., Daelemans, W., Malouf, R., eds. (2022). Implications of Psycho-computational Modelling for Morphological Theory.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-541-6

Table of Contents

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fpsyg.2021.688908

Origins of Dissociations in the English Past Tense: A Synthetic Brain Imaging Model

Gert Westermann and Samuel Jones*

Department of Psychology, Lancaster University, Lancaster, United Kingdom

Brain imaging studies of English past tense inflection have found dissociations between regular and irregular verbs, but no coherent picture has emerged to explain how these dissociations arise. Here we use synthetic brain imaging on a neural network model to provide a mechanistic account of the origins of such dissociations. The model suggests that dissociations between regional activation patterns in verb inflection emerge in an adult processing system that has been shaped through experience-dependent structural brain development. Although these dissociations appear to be between regular and irregular verbs, they arise in the model from a combination of statistical properties including frequency, relationships to other verbs, and phonological complexity, without a causal role for regularity or semantics. These results are consistent with the notion that all inflections are produced in a single associative mechanism. The model generates predictions about the patterning of active brain regions for different verbs that can be tested in future imaging studies.

Keywords: English past tense, connectionist modeling, synthetic brain imaging, experience-dependent brain development, verb inflection, verb morphology, neuroconstructivism

OPEN ACCESS

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
LouAnn Gerken,
University of Arizona, United States
Valeria Caruso,
University of Michigan, United States

****Correspondence:***
Gert Westermann
g.westermann@lancaster.ac.uk

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

Received: *31 March 2021*
Accepted: *31 May 2021*
Published: *02 July 2021*

Citation:
Westermann G and Jones S (2021) Origins of Dissociations in the English Past Tense: A Synthetic Brain Imaging Model. Front. Psychol. 12:688908. doi: 10.3389/fpsyg.2021.688908

INTRODUCTION

The English past tense has, over the past 35 years, taken center stage in the debate on the nature of language and cognitive processing. This is because the past tense is a prototypical "rules-and-exceptions" system, with regular verbs that form their past tense by adding *–ed* to the stem (e.g., *look-looked*), and irregular verbs with past tense forms that range from no change (*hit-hit*) and vowel changes with or without suffixation (*sleep-slept, sing-sang*) to completely idiosyncratic forms (*go-went*). The main question around which this debate has revolved is whether there are separate processing mechanisms for regular and irregular verbs, or if they can be accounted for in a system that produces both regular and irregular forms through a single associative mechanism. This question is important because it has wider implications, for example, for the rule-like nature of grammar (*is rule-like behavior evidence for an underlying mental rule or can it be explained through associative processes?*) and for the question of whether behavioral dissociations imply that the language system has a modular architecture. These questions touch on the very nature of language and cognitive processing, and the English past tense has therefore been called the "drosophila of language processing" (Pinker, 1994): a model system in which such questions can be studied in detail.

Two dominant theories of the nature of inflection processing have emerged. One view, the dual-mechanism or words-and-rules theory (e.g., Pinker, 1991, 1997, 1999; Marcus et al., 1995; Ullman et al., 1997; Clahsen, 1999; Pinker and Ullman, 2002; Ullman, 2004) holds that the processing differences between regular and irregular forms that have been observed in many studies are caused by distinct, qualitatively different underlying mechanisms: A mental symbolic rule for regular forms, and associative storage in the mental lexicon for irregular forms. According to this view grammatical differences are psychologically real in that the mental grammar is used directly in language processing (Clahsen, 1999), so that language processing separates into an associative mental lexicon and a rule-based system (i.e., words-and-rules).

An alternative view argues that all past tense forms are processed in a single associative system in which overlapping representations for regular and irregular forms compete for processing resources (e.g., Bybee and Slobin, 1982; Rumelhart and McClelland, 1986; MacWhinney and Leinbach, 1991; Plunkett and Marchman, 1991, 1993; Marchman, 1993; Joanisse and Seidenberg, 1999; Plunkett and Juola, 1999; McClelland and Patterson, 2002; Westermann and Plunkett, 2007; Westermann and Ruh, 2012; Engelmann et al., 2019). This view is closely tied to implemented connectionist neural network models that have simulated how graded dissociations between different verbs can arise without recourse to modularity and qualitatively different processes. In these systems, apparent dissociations between regular and irregular forms emerge on the basis of the different statistical properties of verbs, such as frequency, phonological complexity, similar sounding verbs with a similar sounding past tense form (i.e., "friends;" e.g., *sing* and *ring*), similar sounding verbs with a different sounding past tense form (i.e., "enemies;" e.g., *sing* and *bring*), or due to reliance on semantic vs. phonological factors.

A large amount of empirical and computational work has aimed to provide evidence for each view [for an overview, see McClelland and Patterson (2002), Pinker and Ullman (2002), and Westermann and Ruh (2012)]. While much of this research has focused on behavioral data from language acquisition and studies involving adults with and without brain damage, a number of brain imaging studies have also revealed brain regions involved in processing different verb inflections. These studies have found differences in neural activation patterns when participants inflected regular and irregular verbs, evidence cited by some researchers as support for a dual mechanism system in which the rule component and the associative mental lexicon are located in different brain regions [e.g., Jaeger et al., 1996; Lavric et al., 2001; Dhond et al., 2003; Sahin et al., 2006; Oh et al., 2011; Bakker et al., 2013; for an overview see Leminen et al. (2019)]. For example, in a seminal study by Jaeger et al. (1996) using positron-emission tomography (PET), participants were asked to generate past tense forms of visually presented monosyllabic verb stems. Jaeger et al. (1996) predicted that the left frontal lobe should be involved in regular processing due to its role in grammatical processing. Likewise, inflection of irregulars was predicted to involve posterior temporal or parietal activity as an index of memory retrieval. Results showed that although many brain regions were activated equally by all verbs, production of regulars selectively activated left dorsolateral prefrontal cortex and left anterior cingulate cortex. Irregulars, meanwhile, prompted higher overall activation and involved occipital visual processing areas. These systematic differences between both verb types were interpreted by the authors as strong evidence for the dual-mechanism account of inflection. Similar claims were made by Lavric et al. (2001) in an ERP study of covert past tense production. These authors found differences between regular and irregular past tense forms in a time window from 288 to 321 ms after visual presentation of the verb stem, and source localization indicated higher activation during this time window for regulars in right prefrontal and temporal areas and higher activation for irregulars in the left temporal area and the anterior cingulate cortex.

In another study using magnetoencephalography (MEG), Dhond et al. (2003) asked participants to covertly generate past tense forms of visually presented verb stems. Dhond et al. also found that generation of regulars and irregulars activated many brain areas in common, but that processing of regulars led to greater activation in left inferior prefrontal areas (Broca's area), and processing of irregulars preferentially activated left occipitotemporal cortex as well as right dorsolateral prefrontal cortex. These results were interpreted as indicating that regulars activated rule-based grammar regions and irregulars activated areas involved in the associative retrieval of forms, corresponding directly to the dual-mechanism theory. Different results were found in an fMRI study of covert past tense and plural production (Sahin et al., 2006) in which Broca's area was activated equally by regular and irregular verbs. Irregulars activated the anterior cingulate and supplementary motor area more than regulars, whereas regulars led to greater activation in some subcortical structures. Overall there was greater activation for irregulars. These results were interpreted within the dual-mechanism framework by suggesting that activation differences between regulars and irregulars were evidence for separate mechanisms and, therefore, against a single mechanism of inflection. Specifically, it was argued that Broca's area was involved in inflection processing, and that greater activation associated with irregulars indicated blocking of the application of the rule by a retrieved irregular form.

However, the results of these and other studies have been controversial. One problem is that specific methodological choices can strongly affect results. For example, because of the low temporal resolution of PET, Jaeger et al. (1996) used a block design in which all regular verbs and all irregular verbs were presented together. However, this design introduces the confound that participants could develop response strategies for regular but not for irregular verbs, suggesting that differences between both verb types should be found independently of the nature of the underlying processing mechanisms (Seidenberg and Hoeffner, 1998). Furthermore there have been inconsistencies between studies in the brain areas that were activated by different verbs [see also Table 1 in Desai et al. (2006)]. For example, Broca's area was activated selectively by regulars in one study (Dhond et al., 2003) which led the authors to argue that it is responsible for rule-based processing, but it was active equally for regular and irregular verbs in another (Sahin et al., 2006). Likewise, greater activation of the anterior cingulate cortex was found for regulars in one study (Jaeger et al., 1996) and for irregulars in others (Lavric et al., 2001; Sahin et al., 2006).

Several other imaging studies have investigated the possibility that the observed activation differences between regular and irregular verbs are due to the different statistical properties of verbs and not to separate underlying mechanisms. For example, an fMRI study in which participants covertly produced the past tense of auditorily presented stems (Joanisse and Seidenberg, 2005) found that regulars and irregulars activated common areas in both hemispheres, but that regulars, as well as irregulars that were phonologically similar to regulars (e.g., *burnt*, *slept*), additionally activated the inferior frontal gyrus bilaterally. In this study, irregulars did not activate any area more than regulars. Dissociations between verbs were thus argued to arise from the phonological properties of verbs instead of their regularity. In a similar fMRI study, Desai et al. (2006) also found widespread overlapping activation, including in Broca's area, for all verbs, and greater activation for regulars in the left dorsal superior temporal gyrus, involving the primary auditory areas and the planum temporale. This study also found regions of greater activation for irregulars compared with regulars (inferior frontal, precentral cortex and parietal cortex bilaterally). When the authors matched a subset of their verb set for phonological complexity of the past tense form, they found that no regions were activated more for regulars than for irregulars. Desai et al. (2006) explained the widespread activation of brain regions for irregular verbs in terms of higher demands on attention, working memory, and response selection for generating the past tense forms of these verbs. The fact that both regular and irregular production activated Broca's area was seen as contradicting the dual-mechanism account which assumes that regular, but nor irregular forms are generated through a mental grammar instantiated in Broca's area (Ullman et al., 1997). Greater activation in auditory areas for regulars was explained with regular forms being phonologically more complex than irregular forms (Burzio, 2002; Bird et al., 2003). Therefore, despite double dissociations between regular and irregular verbs these results were interpreted as evidence for a single-mechanism view of inflection processing.

In summary, previous imaging studies, despite each reporting single or double dissociations between regular and irregular verbs, have not provided a coherent picture of the brain areas involved in processing the English past tense: First, the activated regions for specific verb types differed considerably between studies; and second, the nature of the dissociations differed between studies. One study (Joanisse and Seidenberg, 2005) reported activation of distinct brain regions for regulars but not irregulars, another (Desai et al., 2006) reported the opposite pattern with distinct regions active for irregulars but not for regulars when verbs were matched phonologically, and other studies (Jaeger et al., 1996; Dhond et al., 2003; Sahin et al., 2006; Oh et al., 2011) reported a double dissociation with some regions more active for regulars and others more active for irregulars (although these regions differed in each case). These inconsistent patterns of activation have made it difficult to sufficiently constrain the theories of inflection for (or against) which they were meant to provide evidence. For example, involvement of Broca's area in the inflection of both regular and irregular verbs has been claimed to provide evidence both for (Sahin et al., 2006) and against (Desai et al., 2006) dual-mechanism views of inflection.

One possible explanation for the inconsistency in observed activation patterns in the discussed neuroimaging studies is that statistical factors and not grammatical class determine how a verb is processed, and that these factors differed between the specific verb stimuli used in existing studies. In each study, regular and irregular verbs were matched on certain factors, but the choice of factors had little theoretical foundation and differed greatly between studies. Jaeger et al. (1996) matched stem and past tense frequencies (albeit based on a word list that did not distinguish between nouns and verbs and therefore overestimated regular stem frequencies), Lavric et al. (2001) and Dhond et al. (2003) matched word frequency and letter length, Sahin et al. (2006) matched past and stem cluster frequency and syllable length and aimed for phonological similarity, Oh et al. (2011) matched phonological complexity and past tense frequency, and Joanisse and Seidenberg (2005) matched past tense frequency, imageability, and concreteness. The most careful matching was done by Desai et al. (2006), with past tense frequency, friend-enemy ratio, stem letter length, and stem and past tense syllable length all taken into account, in addition to a sub-group of verbs being further matched on number of phonemes and past tense syllable structure. However, which of these factors affect processing, and in what way, remains an open question. It is therefore also unclear whether a processing system that is sensitive to the statistical properties of verbs would give rise to the observed dissociations in active brain regions.

One approach to answering these questions is to consider how the adult language processing system is shaped through development. Adult psycholinguistics traditionally pays little heed to the mechanisms of language development although a better understanding of developmental trajectories could inform the nature of the adult processing system. Taking this perspective, in this paper we train an artificial neural network model on English past tense inflection (Westermann and Ruh, 2012), adopting a neuroconstructivist developmental process in which the architecture of the adult inflection processing system emerges through an interaction between experience-dependent structural development and experiences with verbs that have specific statistical properties. We then use what has been called "synthetic brain imaging" (e.g., Arbib et al., 1994; Tagamets and Horwitz, 1998, Cangelosi and Parisi, 2004, Horwitz et al., 1999; Arbib et al., 2000; Thomas et al., 2012) to analyze activation patterns across different parts of the model and show that such a system displays visible processing differences between regular and irregular verbs without relying on built-in dissociable processing modules. Finally, we investigate which statistical properties account for the observed dissociations, generating predictions for behavioral and imaging studies.

The computational model used in the current paper was developed by Westermann and Ruh (2012) for modeling behavioral aspects of the acquisition and adult processing of the English past tense. This model displayed a realistic acquisition profile, adult-like non-word generalization, and selective breakdown after damage to parts of the network. The model is based on the neuroconstructivist framework (Quartz and Sejnowski, 1997; Mareschal et al., 2007; Westermann et al., 2007), which stresses the importance of experience-dependent structural brain development in shaping an adult processing system that is specifically adapted to the learning task. There is overwhelming evidence that experience shapes the brain during cognitive development (e.g., Quartz and Sejnowski, 1997; Quartz, 1999; Casey et al., 2000, 2005; Johnson, 2001; Johnson and Munakata, 2005; Nelson et al., 2006; Mareschal et al., 2007; Westermann et al., 2007; Stiles, 2009; Bick and Nelson, 2017), and that differences in adult brain structures can at least partly be explained by a developmental process by which the brain adapts to the specific aspects of the tasks being learned. For example, one study involving Chinese-speaking adults and English-speaking adults living in the United States reported specific differences in the size of frontal, temporal and parietal cortical substrates between these groups (Kochunov et al., 2003). This structural difference was interpreted by the authors as an outcome of the different orthographic, phonetic and semantic characteristics of Chinese and English, which impacted experience-dependent brain development. Likewise, structural brain changes have been observed when learning a second language [see Li et al. (2014), for a review] and for bilinguals [see Bialystok (2017), for review]. For example, native Japanese speakers trained on learning English words for 16 weeks showed an increased density of gray and white matter in the right IFG, but a control group did not show these changes (Hosoda et al., 2013). Anatomical change in this study correlated positively with the participants' knowledge of English vocabulary. Other studies have begun to address systematic cross-linguistic variation in the neural structures supporting language processing (Chen et al., 2009; Mei et al., 2015) and more broadly have asked how specific experiences affect brain organization in members of different cultures (Park and Huang, 2010). Here, along similar lines, we argue that the specific processing demands of the English inflection system will lead to brain structures that are adapted to these demands through experience-dependent development. From this perspective, the specific dissociations between brain activation patterns observed in the adult language system are the outcome of experience-dependent structural development under the task demands of learning verb inflections with their characteristic distribution and statistical properties. This view is in contrast to a modular view of language processing according to which functionally specialized modules implement qualitatively different mental processes (e.g., Pinker, 1994).

Translating these ideas into a computational model, the artificial neural network developed by Westermann and Ruh (2012) integrates structural changes that mimic, on an abstract level, the experience-dependent development of cortical regions through childhood, allowing for the adaptation of its neural circuits to the specific demands of learning to inflect a large set of English verbs. It should be noted, however, that this relatively simple model, despite integrating aspects of neural development, is not a computational neuroscience model that aims to account for the formation of biological synapses or the internal processes of biological neurons, or to mimic the specific aspects of experience-dependent brain development. Connectionist models are usually conceptualized as higher-level models that are based on an abstract and simplified view of neural processing in the brain (e.g., Rumelhart, 1989): Interactions between simple processing units to generate complex behavior; learning of associations by adapting the efficacy of transmission between processing units; and the ability to extract statistical structures from the environment. As high-level models, units ("neurons") in connectionist models are not assumed to correspond to biological neurons on a one-to-one basis but instead to large ensembles of biological neurons (e.g., O'Reilly and Munakata, 2000). Nevertheless, the model is grounded in the assumptions that first, task-driven structural adaptation during learning qualitatively changes the learning process compared with learning in a fixed structure (Quartz, 1993; Quartz and Sejnowski, 1997; Westermann and Ruh, 2012; Westermann, 2016), and second, that it shapes the functional structure of the final system (Mareschal et al., 2007; Shultz et al., 2007) so that the adult system can best be understood as an outcome of such a structural developmental process. In building the model, therefore, we were interested in a principled mechanistic account of how a processing system that is sensitive to the statistical properties of verbs to which it is exposed while undergoing a structural developmental process, gives rise to the dissociations between activation patterns for different verbs that are observed in adult neuroimaging studies.

Synthetic brain imaging (Arbib et al., 1994, 2000; Tagamets and Horwitz, 1998; Horwitz et al., 1999) applies the idea of brain imaging—comparing brain region activation profiles between different test conditions to gain insights into underlying processing mechanisms—to artificial neural networks. In a structured neural network, different stimuli will generate specific activation patterns in different network components and, like in brain imaging, these patterns can be compared between conditions. Although synthetic brain imaging is still in initial stages of exploration, several results have been reported in modeling language processing. For example, one study showed how differential activation patterns for nouns and verbs arose in evolved agent-based networks (Cangelosi and Parisi, 2004): Whereas nouns activated preferentially sensory processing areas of the networks, verbs activated multisensory integration areas more broadly. These activation patterns were compared with brain imaging data showing that nouns activate more posterior brain areas whereas verbs also activate anterior motor areas. In another study, synthetic brain imaging was used in a model of sentence comprehension (Just et al., 1999) and accounted for fMRI data on brain regions involved in processing sentences of different complexities. A third study used synthetic brain imaging and lesioning to investigate whether impairment after brain damage and neuroimaging predict the same patterns of functional specialization (Thomas et al., 2012).

In deciding how to measure the synthetic analog to the fMRI BOLD signal it is important to consider which aspect of neural processing is reflected by BOLD. Current understanding is that the BOLD signal in fMRI does not measure neural activity (i.e., spike potentials) but rather the local field potential (LFP), which reflects the summation of post-synaptic potentials (Logothetis et al., 2001; Norris, 2006). This view would indicate that the closest correlate in a connectionist model to the BOLD signal is the incoming activation into a group of units, that is, the activation flowing through a pathway to this set of units. While it is beyond the scope of this report to explain in detail the many different activation patterns that have been observed in neuroimaging studies of the past tense, we aim to show how differential activation patterns can be generated in a single-mechanism system that is shaped through interactions between the statistical structure of the environment and experience-dependent brain development. In doing so we will account for some empirical results in detail and generate predictions for future neuroimaging studies.

There are a number of reasons why synthetic brain imaging using neural networks can inform theory building and help generate predictions for assessment in studies using real brain imaging. First, in neural networks, the experimenter has full control over the studied process. In imaging studies of inflection processing, the large number of active brain areas suggests that it is difficult to find a baseline condition that differs from the experimental condition only in the inflection process. For example, Desai et al. (2006) reported that the baseline task of reading verbs activated some brain regions that were not active when the verbs were inflected. In a model of verb inflection that takes a verb stem as input and produces its past tense as output, the inflection process can be isolated effectively. It is therefore not necessary to establish a baseline condition (such as reading a verb without inflecting it) and subtracting this baseline activation from the observed activation patterns in the inflection task. Second, a computational model allows for the precise analysis of what factors affect differential activation of network components in a much larger set of verbs than those typically used in neuroimaging studies, where small sets of verbs have to be matched for statistical factors. Third, the language experience that has shaped the computational model toward its final structure is precisely known and is under the control of the modeler. This allows for a better characterization of the statistical factors that underpin emerging dissociations.

MATERIALS AND METHODS

The Model

The neuroconstructivist neural network model (NCM; Westermann and Ruh, 2012) (**Figure 1**) starts out with a minimal architecture in which the input and output layers are fully connected. In a process of experience-dependent structural development, the hidden layer gradually expands to enable the past tense inflection task to be learned. The "adult" architecture of the model is therefore an outcome of, and optimally adapted to, the specific learning task.

In (Westermann and Ruh, 2012), the NCM was presented with phonological representations of verb stems and had the task of producing the corresponding past tense forms. Hidden layer units had a Gaussian (i.e., bell-shaped) activation function. Units of this type become active for a subset of similar-sounding verbs, forming a receptive field for a region of the phonological input space. Gaussian units are activated when an input (i.e., a verb stem) falls within their receptive field, and the closer the input is to the center of the receptive field the higher is the activation of the unit. In the NCM, lateral inhibition in the hidden layer was simulated by suppressing activation of all but the most active hidden unit. The position of the receptive field of this unit was adjusted at each presentation to move a small step toward the position of the current input. Receptive field sizes were also adapted to increase for fields that responded to a range of different verbs. Each hidden unit kept a local error counter to which the network's output error was added when the hidden unit was active.

The model attempted to learn the inflection task in the initial minimal architecture, and structural change occurred when the current structure no longer allowed for improvement in performance: When the average error over 10,000 verbs was no lower than for the previous 30,000 verbs, three new hidden unit receptive fields were inserted at the position of the existing hidden unit which had the highest local error, and their weights to the output layer were initialized randomly. In this process, a hidden unit whose activation leads to a high output error will become the preferred location for the insertion of new units. Because a high local error is usually caused by one hidden unit being responsible for too many input patterns with conflicting input-output transformations (e.g., *sink-sank* and *blink-blinked*), the insertion of additional resources led to a more fine-grained covering of the input space in those areas where similar sounding verbs have different past tense forms. As a consequence, the hidden units were "quasi-localist" [see Westermann and Ruh (2012)]: different units became responsive to between 1 and 136 verbs, with the degree of granularity an outcome of the task demands of the past tense inflection task. Note that this is different from purely localist "lexical entry units" (e.g., Joanisse and Seidenberg, 1999) where each unit is activated by exactly one verb.

Regressive events in the model were implemented by pruning hidden units that were not activated for 30,000 verb presentations [For further details of the implementation see Westermann and Ruh (2012)]. Together, these mechanisms led to a process in which the structure of the model—number, size and location of hidden unit receptive fields as well as the weight patterns in both the direct and indirect pathways—was a direct outcome of the experience with the environment of the English past tense, with its different verbs with specific inflections, phonological properties, similarity clusters, and frequencies. This way of developing the model is in contrast both to the more common static models in which only the weights but not the model structure are adapted, and to models in which change proceeds along a maturational timetable independent from environmental input (e.g., Elman, 1993). Indeed, whereas the developing model was shown to account for a wide range of data on acquisition, adult generalization, and selective impairment after brain damage, an equivalent static model did not account for many of these data (Westermann and Ruh, 2012).

This "neuroconstructivist" type of model also corresponds most closely to current views of experience dependent brain development in which new abilities become manifest in developing brain structures that are adapted to the demands of

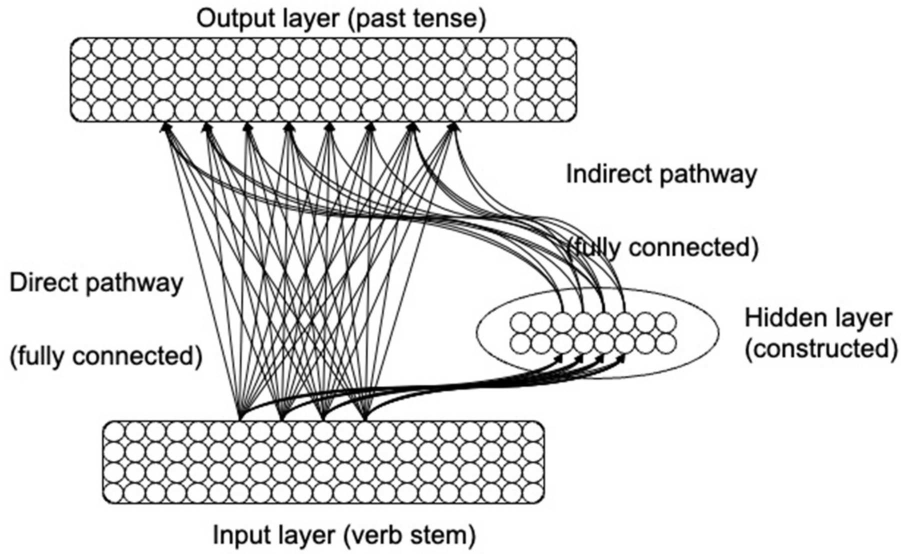

FIGURE 1 | The architecture of the neuroconstructivist past tense model.

a specific ability [see also Shultz et al. (2007)]. In the past tense model, more structure (i.e., hidden units and their connections) was allocated for forms that were "harder" to learn because of the statistical properties of the verb set.

Corpus

The NCM was trained on a set of 1,271 mono- and bisyllabic English verbs extracted from the CELEX database [Baayen et al., 1995; full training details are provided in Westermann and Ruh (2012)]. Of these verbs, 111 (i.e., 8.73% of types, 46.00% of tokens) were irregular. During training, verbs were drawn from this corpus on the basis of their past tense frequencies. The phonemes of each verb were inserted into a consonant-vowel template of the form xCCCVCC for each syllable [where x indicates if the syllable was stressed (1) or not (0)]. Individual phonemes were encoded by phonetic feature vectors, following the binary version of the PatPho coding scheme (Li and MacWhinney, 2002) which requires six features per vowel and seven features per consonant. The presence or absence of a feature was encoded by a value of 1 or −1, respectively, and all features for an empty phoneme slot were set to 0. The stem of a verb was encoded by 84 bits and the past tense form had an additional VC suffix (13 bits).

Training

Five networks were trained on 20 m verb tokens each. Verbs were presented randomly according to their past tense frequencies. Weights were updated after the presentation of each verb (online learning) using the perceptron learning rule (Rosenblatt, 1958). For earlier work on this model see Westermann and Ruh (2009).

Synthetic Brain Imaging Analysis

Synthetic brain imaging (SBI) in the models was performed by measuring the activation flowing through the direct (input-output) and indirect (hidden-output) pathways for each verb. Activation in the direct pathway was computed as the summed absolute activation flowing through the input-output connections:

$$\sum_{o}\sum_{i}|w_{oi}a_i|$$

where o are output units, i input units, w_{oi} the weight of the connection between input unit i and output unit o, and a_i the activation of input unit i. Likewise, activation in the indirect pathway was computed as the absolute activation flowing through the hidden-output connections as:

$$\sum_{o}|w_{oh}a_h|$$

where w_{oh} is the connection weight from the active hidden unit h to output unit o and a_h the activation of hidden unit h. Total activation was computed as the sum of the activation in the two pathways.

RESULTS

All models reached 100% accuracy on average after exposure to 16.8 million verbs, with an average number of 361 hidden units (range = 355–370). Since performance across networks was highly comparable, detailed results from a randomly sampled network will be reported unless otherwise specified.

Emerging Double Dissociation Between Regulars and Irregulars

Figure 2 shows a longitudinal developmental SBI activation profile of the two network pathways. Early in development each

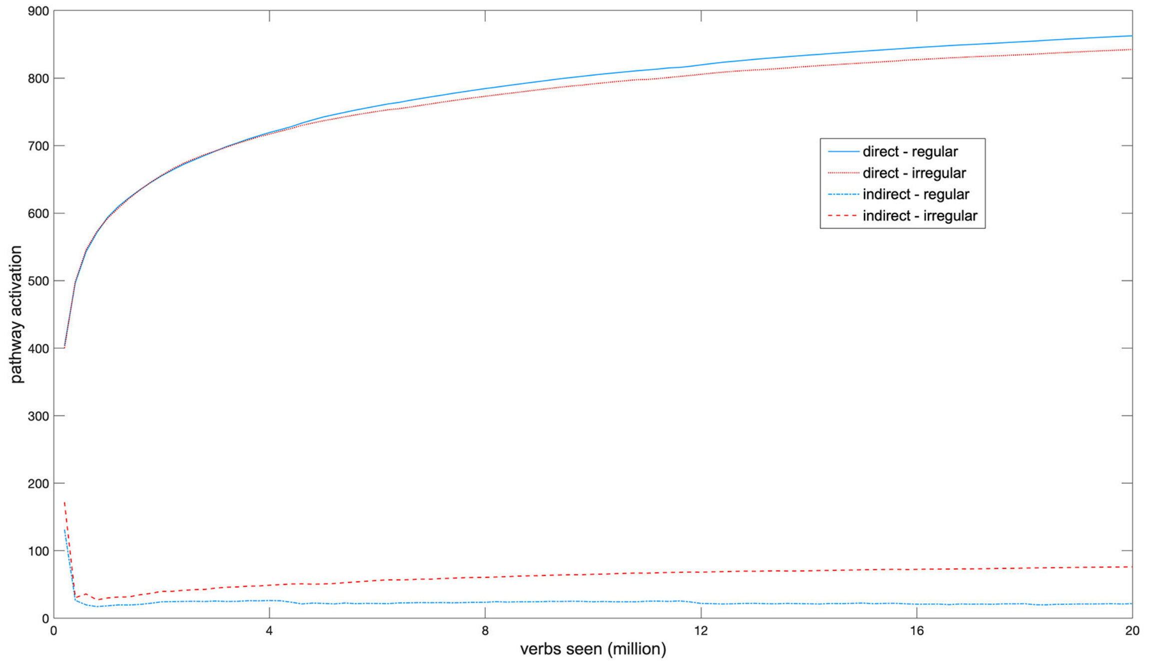

FIGURE 2 | Development of the activation profiles of regular and irregular verbs in both network pathways.

pathway was activated equally strongly by regular and irregular verbs. With development, activation in the direct pathway increased and separated between regular and irregular verbs, with regulars producing on average higher activation in this pathway than irregulars (mean activation at the end of training by regulars: $M = 864.3$, $SD = 111.9$; by irregulars: $M = 844.1$, $SD = 91.5$; Mann-Whitney U-test, $z = -2.19$, $p = 0.029$, mean rank for regulars = 642.97; and for irregulars = 563.21). Overall activation in the indirect pathway initially decreased because activation in this pathway interfered with learning the task due to the insufficient number of hidden units. Throughout the rest of development, mean activation differences between regular and irregular verbs then continued to increase (mean activation at the end of training by irregulars: $M = 75.5$, $SD = 59.1$; by regulars: $M = 21.4$, $SD = 13.8$; Mann-Whitney U-test, $z = -14.534$, $p < 0.001$, mean rank for regulars = 589.7, and for irregulars = 1119.7). This double dissociation between regular and irregular verbs emerged in the model without any functional pre-specification of either pathway and without explicit encoding of regularity, solely on the basis of the different task demands of producing the past tenses of different verbs.

Although to our knowledge there have not yet been developmental brain imaging studies of verb inflection, a developmental fMRI study of word production (Brown et al., 2005) found an increase in activation in some cortical areas and a decrease in others across age, with significant differences between age groups even when overt task performance was equal. The network for which results are displayed in **Figure 2** reached 100% correct performance in the inflection task after 16.0 million verb tokens, and it is interesting to observe that activation in the direct pathway likewise continued to increase after this point without a change in overt performance.

The fact that a double dissociation in regional activation patterns between regular and irregular verbs emerges in the NCM contradicts the argument that differential activation of brain regions for each verb type necessarily indicates an underlying qualitative processing difference between regular and irregular forms (e.g., Jaeger et al., 1996; Beretta et al., 2003; Sahin et al., 2006). In the "adult" NCM all past tense forms are generated through a single associative mechanism, but dissociations arise on the basis of statistical and distributional differences between verbs that have become manifest in the network's architecture during development. The developing hidden layer enables the model to allocate additional processing resources for verbs whose inflections are hard to learn in the direct pathway alone, as structure is added to this layer when learning no longer improves. The fact that the indirect pathway is activated more by irregular verbs is compatible with the "ease-of-processing" account of functional specialization in past tense processing (Westermann and Ruh, 2012). This account states that on average, irregular forms are harder to learn and process than regulars. An irregular is harder to process than a regular, however, not by virtue of its irregularity, which is a grammatical property of an individual verb, but instead as the result of a combination of statistical and distributional factors such as relative frequency and numbers of friends and enemies, which are statistical properties that arise from the verb corpus as a whole (Westermann and Ruh, 2012). The origin of the emergent dissociations is, therefore, the differential ease of processing of verbs and not their grammatical class.

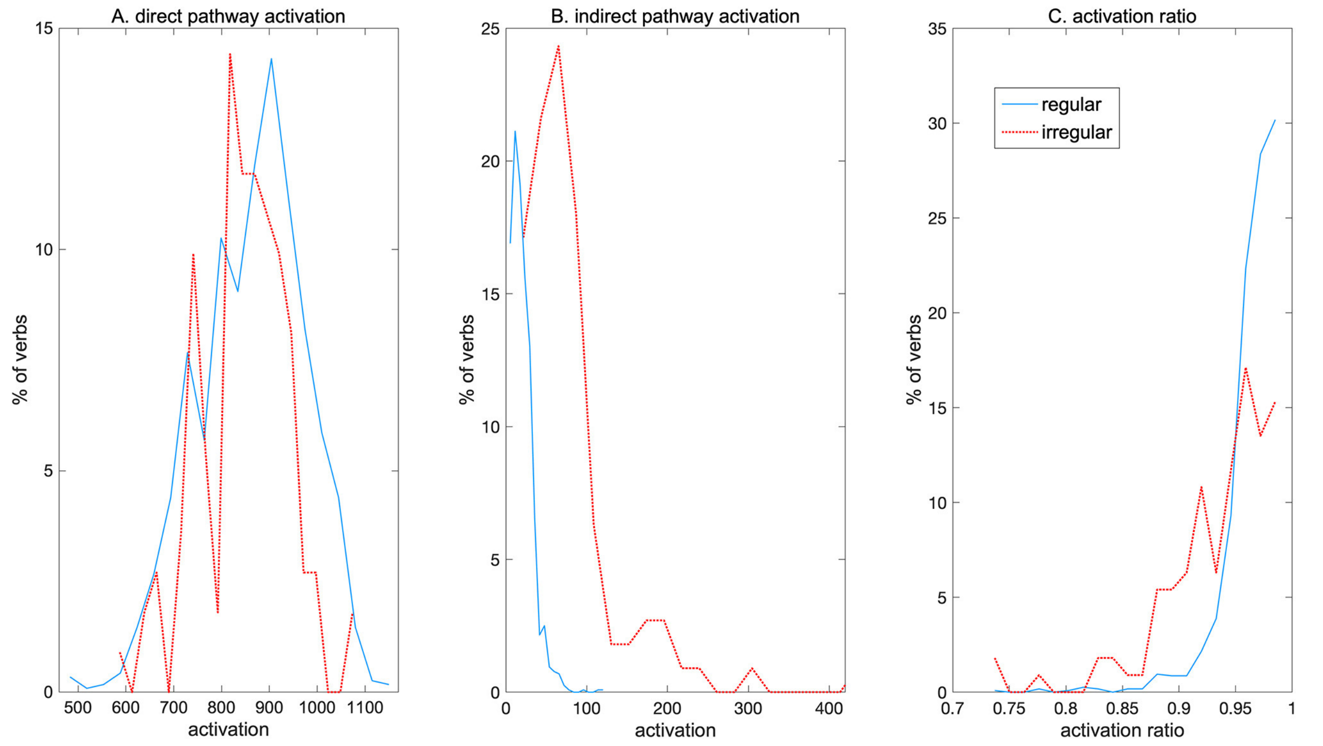

FIGURE 3 | Distribution of path activations by regular and irregular verbs. **(A)** Direct pathway activation. **(B)** Indirect pathway activation. **(C)** Activation ratio.

Double Dissociations Between Mean Activation Values Mask Distributional Differences

More detailed analysis of the NCM further revealed that, despite the observed double dissociation between regular and irregular verbs, each verb activated both pathways, albeit to different degrees. **Figure 3** shows the distribution of regular and irregular verbs activating each pathway. In the direct pathway (**Figure 3A**) the spread of activations is similar for regular and irregular verbs, with the highest activations resulting from regulars. In contrast, in the indirect pathway a higher proportion of irregulars than regulars were strongly activated, and most regulars only led to weak activation in this pathway (**Figure 3B**). **Figure 3C** shows the *activation ratio* which was computed as:

$$\frac{\textit{direct pathway activation}}{\textit{direct pathway activation} + \textit{indirect pathway activation}}$$

where activations in both pathways were scaled to a maximum value of 1. A ratio of >0.5 indicates that a specific verb activates the direct pathway relatively more than the indirect pathway. The figure shows that regulars as well as irregulars activated the direct pathway more than the indirect pathway, but regulars tended to have a higher activation ratio than irregulars. Nevertheless, some irregulars as well were produced almost solely through the direct pathway, with activation ratio near 1.0. These results indicate that although there is an apparent global specialization of the direct pathway for regular verbs and of the indirect pathway for irregulars as revealed by the observed double dissociation, this is an outcome of complex overlapping activation patterns for individual regular and irregular forms throughout the network.

Greater Regular Activation Is Due to Greater Phonological Complexity of Regulars

We further modeled more specific results from Desai et al.'s (2006) fMRI study. Desai et al. (2006) argued that a higher activation for regular verbs in some cortical regions was the consequence of the higher phonological complexity of the regular verbs used in their experiment. To test this claim they analyzed a subset of their verbs in which regulars and irregulars were matched for phonological complexity. As predicted, they found that for this matched set there was no brain region more active for regular verbs. We simulated this result by comparing the activation profiles in the model for all verbs with those for the matched subset of Desai et al. (2006). Of the 80 verbs in the subset, two irregulars were not in the network's training corpus (*break*, *cost*). These word's matched regular partners (*stay*, *guess*) were also removed from the test set, and the NCM was tested on the remaining 76 matched verbs. For this matched subset, as in Desai et al.'s (2006) study, no area was now more active for regulars. Whereas with the full verb set the average direct pathway activation was higher for regulars than for irregulars (see section Emerging Double Dissociation Between Regulars and

Irregulars), the matched subset showed the opposite pattern, with irregulars ($M = 853.2$, $SD = 99.1$) now on average activating the direct pathway more than regulars ($M = 798.8$, $SD = 92.8$; Mann-Whitney U-test, $z = -2.68$, $p = 0.007$, mean rank for regulars = 31.71; and for irregulars = 45.29). As in the full set of verbs, indirect pathway activation was higher for irregulars in the matched subset (irregular activation $M = 58.9$, $SD = 31.3$; regular activation $M = 31.0$, $SD = 20.0$; Mann-Whitney U-test, $z = -4.21$, $p < 0.001$, mean rank for regulars: 27.84, and for irregulars: 49.16).

Although Desai et al. (2006) found that with the phonologically matched subset some areas were activated more for irregulars than had been for the non-matched set (i.e., the precentral gyrus and left anterior cingulate gyrus), the region previously more active for regulars later showed no difference between regulars and irregulars. While in the NCM this area (i.e., the direct pathway) was now more active for irregulars than for regulars, the model accounted for Desai et al.'s (2006) main result of the disappearance of higher activation for regulars within the processing system when phonological complexity was controlled.

The NCM further provided a more general evaluation of the role of phonological complexity in observed regular-irregular dissociations. Whereas in experimental neuroimaging the effect of phonological complexity can only be controlled for by using matched subsets of verbs, in the NCM the same can be achieved by dividing the total activation in the direct pathway by the number of active input units. This is because in the distributed phonological representation of verbs, higher phonological complexity, here defined as number of phonemes or number of syllables, corresponds to more input units being active. Dividing the direct pathway activation by the number of active input units therefore normalizes this activation (Note that this is not necessary for the indirect pathway because only one hidden unit is active for each verb). Whereas, non-normalized activation in the direct pathway was higher for regulars than for irregulars, activation normalized for complexity was conversely smaller for regulars ($M = 30.6$, $SD = 8.4$) than for irregulars ($M = 36.0$, $SD = 7.3$; Mann-Whitney U-test, $z = -6.82$, $p < 0.001$, mean rank for regulars = 614.28, and for irregulars = 862.95), providing further evidence that systematic differences in phonological complexity can lead to regular-irregular dissociations.

Origins of dissociations

What, then, are the origins of the dissociations found in neuroimaging studies? The "easiness" view of past tense processing suggests that different statistical characteristics of verbs affect their ease of processing, and hence their activation profile, irrespective of whether they are regular or irregular. By using synthetic brain imaging we are able to investigate precisely which statistical factors are involved, as the model is tested on a large set of verbs in which these factors vary considerably. To do this, we characterized each verb along a range of factors that were accessible to the model during training: Past tense frequency, presence of a stem final alveolar consonant, phonological complexity, and number of friends and enemies within the training corpus (Note that a "friend" was defined as a verb with the same stem rime and the same past tense rime, e.g., *sing-sang* and *ring-rang*, and an "enemy" was defined as a verb with the same stem rime but with different past tense rime, e.g., *sing-sang* and *bring-brought*).

TABLE 1 | Correlations between the statistical properties of verbs and their activation ratio.

Correlation with activation ratio	Past tense frequency	Friends	Enemies	Phonological complexity
r	−0.696	0.226	−0.434	0.245

All correlations p < 0.001.

Table 1 shows that frequency, friend/enemy measures, and complexity correlate significantly with activation ratio (i.e., direct activation divided by total activation; see Formula 3). These correlations indicate that phonologically complex, low-frequency verbs with an advantageous neighborhood (i.e., many friends, few enemies) tend to activate the direct pathway relatively more strongly (i.e., lead to a higher activation ratio), while the indirect pathway is activated relatively more for frequent verbs with an unfavorable neighborhood.

To examine which statistical factors contributed to activation differences in each pathway, we entered these factors as independent variables into multi-level regression models across all networks, with pathway activation as the dependent variable and verb and network as random effects. All predictors were zero centered and scaled to $SD = 1$. Indirect pathway activation was most strongly predicted by past tense frequency ($\beta = 16.77$), enemies ($\beta = 5.64$), phonological complexity ($\beta = -3.64$), and friends ($\beta = -2.86$; all $p < 0.001$), with this model explaining 50% of the variance in pathway activation ($R^2 = 0.498$). For direct pathway activation, the only significant contributing factors were phonological complexity ($\beta = 38.78$) and friends ($\beta = 26.43$; both $p < 0.001$), with this model accounting for only 19% of the variance in pathway activation ($R^2 = 0.189$).

The picture emerging from these results is, therefore, slightly more complex than directly linking activation in the indirect pathway with low ease of processing. Although indirect pathway activation is predicted by a high number of enemies and low number of friends—both factors that would be expected to make processing harder—it is also predicted by high frequency, which would be expected to make processing easier. The reason for this counterintuitive result is that verbs that are frequently encountered by the model will lead to the accumulation of many small errors on hidden units (instead of fewer but larger errors for harder verbs) so that in the experience-dependent development of the network's structure new units will also be inserted in those regions of the input space. Importantly, the neuroconstructivist view of past tense processing therefore predicts that the same brain regions should be shared by the processing of frequent and hard verbs.

Typical and Non-typical Regulars and Irregulars

Given these results we used the activation ratio to establish "typical" and "non-typical" regular and irregular verbs from the imaging perspective. Typical regulars were regulars with a high activation ratio. The 10 most typical monosyllabic regulars according to this measure, given the specific training set of our model, were *nail, nurse, roar, hail, hiss, slice, frost, dawn, roast,* and *rate*. Note that these are not the most frequent regulars because frequent verbs also highly activated the indirect pathway, leading to a lower activation ratio. The 10 least typical monosyllabic regulars, that is, those regulars with the lowest activation ratio, were *ask, look, roam, mask, add, try, soil, dry, hum,* and *use.* Interestingly, several of these verbs would normally be regarded as prototypical regulars because of their high frequencies. The synthetic imaging results presented here, however, predict that in brain imaging studies they might actually activate similar regions to irregulars.

The 10 irregular verbs with the most typical irregular activation pattern, that is, a low activation ratio, were *say, see, think, stand, bring, do, go, make, get,* and *speak.* The 10 least typical irregulars according to this measure were *shrink, spin, sweep, flee, deal, creep, thrust, kneel, ride,* and *quit. Five* of these 10 verbs (*sweep, flee, deal, creep,* and *kneel*) are pseudo-regulars which add [t] or [d] to their past tense and, according to Joanisse and Seidenberg (2005), should be expected to cluster with regular verbs in their activation profile. In line with this claim, these verbs showed "regular-like" activation patterns in the NCM.

Analysis From a Dual-Mechanism Perspective

Although the NCM shows that regional double dissociations between regular and irregular verbs can emerge solely on the basis of the statistical properties of different verbs in a single processing mechanism that is shaped by experience-developmental structural development, and that the grammatical property of regularity plays no role in causing these dissociations, in brain imaging studies the underlying mechanisms remain unknown and are hypothesized on the basis of observed data. Thus, when dissociations between regular and irregular verbs are observed in an empirical study, researchers adopting a dual-mechanism framework explain these data in terms of separate processing mechanisms (Jaeger et al., 1996; Lavric et al., 2001; Beretta et al., 2003; Dhond et al., 2003; Sahin et al., 2006; Oh et al., 2011), with regions more active for regulars hypothesized to be responsible for the application of grammatical rules, such as regular inflection, and regions more active for irregulars indicating the retrieval of full forms from the mental lexicon located in this region. At the core of such dual-mechanism interpretations lies the assumption that grammatical class (i.e., regularity) forms the basis of observed dissociations.

To mimic this inferential process from data to hypothesized mechanism, we analyzed the activation differences in the model from a dual-mechanism perspective, which would assume that regularity itself is a predictor of the observed dissociations. We performed further multi-level regression analyses for pathway activation, with verb and network as random effects, and with regularity added to the inventory of independent variables first modeled in section Origins of Dissociations (each zero centered and scaled, $SD = 1$). Results were again highly significant, with past tense frequency ($\beta = 15.35$), regularity ($\beta = -36.28$), complexity ($\beta = -4.65$), and friends ($\beta = -1.68$; all $p < 0.001$) predicting indirect pathway activation ($R^2 = 0.557$). This model accounted for $\sim$6% more variance than the model without regularity as a factor, and this increase was significant ($p < 0.001$). Including regularity as a predictor in the regression model of direct pathway activation did not lead to a statistically significant improvement in model fit, and regularity did not predict activation ($p = 0.638$).

Although the NCM is a single-mechanism model, the results for indirect pathway activation correspond to the predictions made by the revised version of the dual-mechanism theory for lexical retrieval (Pinker and Ullman, 2002). This revised theory predicts lexical retrieval not only for irregulars but also for high frequency regulars because high frequency forms are more likely to be memorized than low frequency ones, and for regulars with low friend-enemy ratios because they are more likely to be attracted to irregular enemies. From a dual-mechanism perspective, activation patterns like those observed in the model would therefore be taken as backing for this theory, despite being caused by a very different underlying mechanism. This result highlights the benefit of computational modeling: when we collect empirical data we do not know the mechanism that generates them but we infer from the data to a potential underlying mechanism. When we construct a model we know the mechanism and we see how this mechanisms generates empirical data. In the past tense debate, empirically observed dissociations between regulars and irregulars have often been hypothesized as arising from two separate underlying mechanisms. The NCM shows that such dissociations, down to a level of detail that has previously served as refinement of the dual-mechanisms theory, arise in a single-mechanism system on the basis of statistical properties of verbs together with experience-dependent structural development. By designing the model we know that whether a verb is regular or not is not encoded in the training data and is therefore not accessible to the network. The fact that regularity nevertheless emerges as a significant predictor for indirect pathway activation is a consequence of two factors: on the one hand regularity correlates highly with several of the measures that lead to specialization of the network pathways (**Table 2**). Regular verbs tend to be less frequent, have more friends and fewer enemies than irregulars, and are phonologically less complex. As discussed above, learning high frequency verbs with few friends and many enemies leads to the allocation of hidden units in the indirect pathway, and thus to higher indirect pathway activation in the adult model. Therefore, regular verbs, which show the opposite profile, will on average have fewer dedicated hidden units and thus a lower activation of the indirect pathway, with these forms activating the direct pathway more.

Nevertheless, this cannot be the sole explanation of the significant effect for regularity, because if regularity was entirely predictable from these factors the hierarchical regression should not show a significant improvement when regularity is added. Instead, the explanation lies in the fact that associative learning mechanisms make use of all cues that facilitate learning of a mapping. In the case of the English past tense, these cues do not only lie in the distributional characteristics of verbs available to

TABLE 2 | Correlations between regularity and the statistical properties of verbs in the training data.

Correlation with regularity	Past tense frequency	Friends	Enemies	Phonological complexity
r	−0.363	0.176	−0.7	−0.122

All correlations p < 0.001.

the model indirectly through the training schedule, but also in the phonological characteristics of verbs, which are available directly as inputs. The second factor explaining why regularity emerges as a significant predictor of activation in the indirect pathway is that the mapping between stem and past tense for regulars is easy to learn in the direct pathway because regular verbs have the highest relative and absolute type frequency. To form their past tense, in English more verbs preserve their stem and add *–ed* than undergo any other transformation. Therefore, additional structure in the hidden pathway is not necessary for learning this transformation. As such, regularity-specific activation patterns in the model arise out of a combination between structural and distributional environmental cues together with the model's experience-dependent developmental process.

Mapping the Model to the Brain

A question that can be addressed in SBI studies is which brain areas give rise to specific behaviors. Indeed, one way in which SBI has been applied is in modeling the internal functioning of, and the interactions between, specific brain areas that are known to be involved in a task. For example, the model developed by Cangelosi and Parisi (2004) contained a sensory layer and a sensory/proprioceptive layer, and results from processing nouns and verbs were linked to previous fMRI results showing that nouns activate more sensory areas and verbs activated motor areas. Similarly, Horwitz et al. (1999) presented a biologically plausible large scale neural model of the interactions between specific brain areas, and used this model to account for cerebral blood flow data from PET studies. Nevertheless, of course, even "biologically plausible" models are far away from the biological substrate of the brain in terms of detail and number of interacting regions. Any link between activated regions in a higher-level model and the brain can therefore only serve as a suggestion that should be verified in subsequent neuroimaging studies. As such, speculating on such links can be beneficial for generating predictions about which stimuli might activate which brain areas preferentially.

Can a mapping between model and brain areas also be done on the basis of the present past tense model? One difficulty is that by necessity, the model is simple and the brain is complex, making any such attempted link seem tenuous. On the other hand, though, as the model clearly develops specialized processing pathways, it might seem a missed opportunity not to at least speculate how the model's pathways might map onto brain structures. A second difficulty in attempting such a mapping is that, unlike in the models described above, the data from past tense imaging experiments are anything but clear. As discussed above, studies have differed greatly in the areas that were found to be involved in regular and irregular processing. Furthermore, most imaging studies were not principally concerned with testing whether specific brain areas were involved in inflection processing, but instead investigated whether regular and irregular verbs activated different brain areas in principle (to provide evidence for dual-mechanism accounts of inflection) or whether regions for regulars and irregulars overlapped and dissociations were based on phonological and semantic factors (as evidence for single-mechanism accounts). When differences were found they were typically explained in a *post-hoc* manner. In one study favoring a dual-mechanism interpretation, for example, Dhond et al. (2003) noted that the left fusiform area, which was activated more by irregulars, has been implicated in lexico-iconic or word-form encoding and early lexical access, whereas Broca's area, which was in this study activated more by regular verbs, plays a role in rule-based past-tense formation, grammar, and syntactic parsing. Likewise adopting a dual-mechanism interpretation Sahin et al. (2006) and Jaeger et al. (1996) found equal activation of Broca's area for regulars and irregulars, and therefore attributed a general role in inflection processing to this region. Stronger activation for irregulars in the anterior cingulate and supplementary motor areas was in Sahin et al.'s (2006) study attributed to irregular verbs blocking the application of regular inflection, which is a central feature of dual mechanism accounts. Few studies include predictions about the specific areas involved in past tense processing. For example, Joanisse and Seidenberg (2005) hypothesized that, overall, activation should be distributed over areas responsible for phonological processing such as the inferior frontal gyrus (IFG), including Broca's area, and areas involved in semantic processing, particularly the posterior temporal lobe. These areas have also been shown in brain-damaged patients to lead to dissociations between verb types (Patterson et al., 2001; Bird et al., 2003).

The IFG and posterior temporal lobe are also a subset of the areas involved in inflection processing in Desai et al.'s (2006) study, which had the most carefully matched verb set. Given that studies with brain damaged patients also point toward these areas as being involved in inflection, we can hypothesize that the pathways in our model map to these areas. Specifically, the direct pathway in our model might reflect the functioning of the IFG, and the indirect pathway might reflect the functioning of the posterior temporal lobe. However, our model does not suggest that these are phonological and semantic areas respectively, as suggested by Joanisse and Seidenberg (2005). Instead, both reflect phonological processes, with the IFG processing direct mappings based on distributed phonological information, and posterior temporal areas providing more localist word representations that complement these distributed representations in the IFG. There has been a controversy on whether semantic representations are causal in enabling the formation of irregular past tense forms or whether semantic and irregular past tense representations are merely collocated in the IFG.

Deficits in irregular inflection are often associated with semantic impairments in Alzheimer's disease, semantic dementia and herpes simplex encephalitis (HSE) (Ullman et al., 1997; Patterson et al., 2001; Tyler et al., 2002) as a consequence to damage to the temporal lobes. Nevertheless, the association does not seem to be absolute, as would be expected if semantic representations formed the basis for irregular inflection. Several studies have reported cases in which patients with semantic deficits did not show disproportionate problems with semantic inflection (Tyler et al., 2004; Miozzo and Gordon, 2005) and others found patients with no semantic deficits but problems with irregular inflection (Miozzo, 2003). A longitudinal study of two semantic dementia patients (Bright et al., 2008) reported that the early stages of dementia were associated with semantic deficits, but that language deficits occurred later when brain atrophy was more widespread. We believe that these results raise the intriguing possibility that semantic and irregular verb representations are closely but not causally associated, because both constitute idiosyncratic representations. There is nothing in the sound of a word that signifies its meaning, and there is also nothing in the sound of a verb that predicts its irregular past tense. Both have to be learned, and it is possible that idiosyncratic information connected with words is stored in the posterior temporal areas of the brain. Computational modeling work supports this view. A well-known model of past tense impairment after brain damage (Joanisse and Seidenberg, 1999) included a set of localist units for each verb, and damage to these localist units led to greater irregular impairment. However, while these units were labeled as "semantic" in the model, there was nothing to connect them to the meaning of words; their main role was to encode idiosyncratic information. Likewise, in our neuroconstructivist model we did not include "semantic" representations (in the sense that the meaning of verbs was not encoded) because we found that a model without semantics (albeit with the ability to encode idiosyncratic information without recourse to semantics) accounted for a wide range of behavioral data in past tense tasks, spanning acquisition, adult processing and impairment after brain damage, and that assuming a causal role for semantics was therefore not necessary.

The proposed mapping of the pathways in our model to brain areas is corroborated by results from selectively damaging the model's pathways (Westermann and Ruh, 2012), where damage to the indirect pathway selectively affected irregular verbs and damage to the directed pathway affected all verbs, albeit regulars to a greater degree. These results correspond to patients with brain damage to left temporal areas and the IFG, respectively. Further evidence linking the direct pathway to the IFG comes from the fact that in Joanisse and Seidenberg's (2005) fMRI study, "pseudo-regulars" (such as *burnt*) clustered with regulars in the IFG, and the same was true in the model where several pseudo-regulars had activation ratios comparable with regular verbs. Finally, locating the direct pathway in the left IFG can provide insight into why some studies found equal activation for all verbs in Broca's area (Desai et al., 2006; Sahin et al., 2006), while others found greater activation for regular verbs in this region (Dhond et al., 2003). In our model, the direct pathway is strongly activated by all verbs, but slightly more by regulars. Whether activation differences in this path are found therefore depends on the precise choice of verbs. In accord with Joanisse and Seidenberg (2005), the model predicts that activation in the IFG is associated with phonological processing rather than with regularity: In our regression analysis of direct pathway activation, phonological complexity was the strongest predictor, but regularity was not a predictor.

DISCUSSION

The simulations described in this paper show how dissociations between brain activation patterns in inflection tasks can arise from a single associative mechanism together with experience-dependent structural development. The argument that dissociations between verbs reflect ease of processing has been made previously with respect to imaging studies (Seidenberg and Arnoldussen, 2003), but the present model provides a mechanistic account of how these dissociations can arise and a precise characterization of their underlying factors. Importantly, the model predicts that frequency acts in the opposite direction to ease of processing, and that hard-to-process verbs should generate similar activation patterns to high frequency verbs because dedicated structure in the developing system is allocated to both.

Together these results raise a number of important points. First, dissociations in activation patterns like those observed in the model have often been described as being between regular and irregular verbs, and have been taken as evidence for the existence of qualitatively distinct mechanisms (i.e., rule application and lexical retrieval) in the inflection of these verbs (Jaeger et al., 1996; Bergida et al., 1998; Lavric et al., 2001; Dhond et al., 2003; Sahin et al., 2006). The fact that the same dissociations emerge in the model on the basis of a single processing mechanism considerably weakens this argument. Whereas separate processing mechanisms would result in observable dissociations, the reverse implication is not true: Separate mechanisms are not *necessary* to obtain dissociations [see also Plaut (1995)]. Second, although superficially the emerging dissociations appear to be between regular and irregular verbs, their true nature is better described as a grading between low-frequency, phonologically complex verbs with many phonological friends and few enemies on the one hand, and high-frequency verbs with many enemies and few friends on the other. Although these statistical factors correlate with regularity, characterizing the dissociations observed as being between regulars and irregulars is a *post-hoc* abstraction of the actual underlying mechanisms. When this abstraction is used as an explanation of the underlying processes, as in dual mechanism approaches, a lot of the empirical data, such as the gradation of dissociations and the effects of phonology, friends, and enemies, cannot be captured. Third, regional activation patterns in imaging studies are likely to be a complex function of the statistical and phonological properties of the verbs used in a specific study. All imaging studies have taken this possibility into account and controlled for various properties. However, the selection of properties controlled for has generally not been systematic or based on evident theoretical considerations. The results presented here suggest that interactions between verb frequency, phonological complexity, and numbers of friends and enemies are the main factors affecting regional activation differences. Fourth, these results indicate that typical (i.e., high frequency) regulars and irregulars might not in fact activate different brain regions. Instead, according to the model, all frequent verbs share activated regions, and dissociations between regulars and irregulars will primarily be found among low frequency verbs. Finally, some previous approaches have also based explanations for dissociations on a differential involvement of semantics in the generation of regular and irregular forms (e.g., Joanisse and Seidenberg, 1999; Patterson et al., 2001). In this view, regular inflections rely on phonological representations, whereas irregular inflections are based on the semantic representations of verbs. Without precluding the possibility that semantic and irregular processing might be linked and correlated, the present model, which does not contain semantic representations, suggest that semantic and irregular impairments might correlate because they both refer to idiosyncratic information about verbs that cannot be directly retrieved from their phonological form.

In line with our argument that experience-dependent brain development shapes the adult cognitive architecture, the performance of the NCM model is an outcome of its experience with the learning environment. While we have made specific predictions about what are typical and atypical regular and irregular verbs in terms of brain activation patterns, as well as about the statistical factors predicting activation patterns in the model pathways, this point must serve as a caveat because it is not clear how closely the statistics of our verb set reflect those of real-world language learners. For example, the frequency statistics in our corpus are extracted from the CELEX database (Baayen et al., 1995) and are not derived from parental input to a child. Likewise, with respect to modeling, we made decisions about what statistical factors to consider in the first place. These choices were guided by two considerations. First, the factors must be available to the model. Therefore we did not include, for example, imageability as a factor because the model does not contain semantic representations. Second, a factor must be available for the majority of verbs. This precluded our use of age of acquisition norms, which are only available for a relatively small subset of verbs. Given these caveats, a worthwhile avenue for future research will be to investigate how variation in the input comes to be reflected in variation in the model's architecture and performance, and in how far model performance is robust to input variation.

In a similar vein, our model effectively isolates the past tense inflection process from the rest of cognitive processing. On the one hand this is a valuable abstraction because it allows for a precise investigation into the factors affecting activation patterns in this task alone. On the other hand, it is possible that different inflectional paradigms such as noun plurals or even inflections across languages known to multilinguals affect each other. While computational models exist that have simultaneously learned multiple inflections in a single system (Plunkett and Juola, 1999) there has been no systematic investigation of how these paradigms affect each other. Likewise, although we argued that the semantics of verbs are not causally linked to irregular inflection (Westermann and Ruh, 2012), omitting semantic representations from the model does not allow it to distinguish between homophones (e.g., *ring* and *wring*) and consequently these were excluded from the training data.

Although the model provides a precise account of the origins of different activation patterns in synthetic brain imaging, it is nevertheless possible that verbs might dissociate differently depending on the experimental paradigm, because representations in different areas for the same verb might be redundant. As discussed, the model predicts that in imaging studies inflecting frequent and hard verbs activate the same brain regions. It is, however, possible that in behavioral paradigms such as lexical decision tasks, frequent and hard verbs might dissociate as interactions between processing regions can differ with specific task demands even when the same regions are involved in processing both. This is because in a neuroconstructivist system that structurally develops on the basis of experience with the environment, a brain area that is activated by a certain process need not be necessary for this process because such a system would involve a degree of redundancy. For example, although high frequency verbs in the model activated the indirect pathway more than low frequency verbs, this does not mean that the indirect pathway is *necessary* for the production of high frequency past tense forms. Instead, their production might be possible based on the direct pathway alone, with indirect pathway representations being redundant. This would also indicate that one could expect differences between results from brain imaging and from behavioral studies with brain damaged patients (Price and Friston, 2002; Thomas et al., 2012). Damage to a certain brain area would therefore affect forms that activate this area in different ways, depending on whether the area is redundant for the processing of a specific form or not. For example, as described above, in lesioning the NCM to simulate selective impairment after brain damage (Westermann and Ruh, 2012), even when the indirect pathway was completely lesioned performance on regulars remained virtually unimpaired, indicating that the direct pathway is sufficient for producing all regular forms, despite, as reported here, regulars also activating the indirect pathway in synthetic brain imaging.

As a more general contribution, the model presented here highlights the importance of computational modeling in understanding the mechanisms of cognitive processing. As shown in the regression analyses, depending on the adopted theoretical perspective different explanations can be derived from the same observed dissociations. Under the assumption that the grammatical class (i.e., regular or irregular) of a verb is accessible to the model and analyzing the observed activation patterns from this perspective, the results would be taken as evidence for a dual-mechanism view of inflection processing. In studying the brain, this top-down approach from observed data to potential underlying mechanisms is the only possible approach. In a computational model, however, the mechanisms of processing are known and we can observe what data is generated through these known mechanisms. Using this bottom-up approach we know that regularity is not one of the factors accessible to the model, and that all inflections are based on a single mechanism that operates in a structured processing system. Likewise, the model has no access to semantic representations, with all inflections based on phonological information alone. Finding that dissociations between regular and irregular verbs nevertheless emerge under these constraints disconfirms the claim that such dissociations are evidence against a single-mechanism explanation and necessitate a dual-process system. However, these results also weaken the argument of prior single-mechanism accounts that semantic representations play a causal role in the inflection of irregular verbs. Computational modeling provides a detailed alternative explanation to these views by quantifying the interactions between statistical verb properties that give rise to the observed dissociations, and by providing a mechanism by which the structure of the environment comes to be reflected in the structure of the processing system through neuroconstrucivist development. Computational modeling is therefore an important approach in the gathering of converging evidence for theories of inflection processing, and for theories of cognitive processing in general.

Finally, together with previous work incorporating the NCM (Westermann and Ruh, 2012), which accounted for empirical data from past tense acquisition, adult generalization, and impaired processing after brain damage, we believe that our modeling of brain imaging data in the current paper illustrates how neuroconstructivist computational modeling can overcome one point of criticism sometimes levied against models in this domain—that each individual model is tailored specifically to account for a single phenomenon (Pinker and Ullman, 2003)—in providing a principled account of past tense processing by explaining existing data as well as generating predictions for future research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: Open Science Framework, https://osf.io/ejs4m/.

AUTHOR CONTRIBUTIONS

GW developed the study hypothesis, ran the simulations, and drafted the paper. GW and SJ performed the data analysis, discussed the results, and finalized the manuscript for submission. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Economic and Social Research Council International Centre for Language and Communicative Development (LuCiD) [ES/S007113/1 and ES/L008955/1].

ACKNOWLEDGMENTS

We thank Nicolas Ruh for help with the analysis scripts. An earlier version of part of this work was presented in Westermann and Ruh (2009).

REFERENCES

Arbib, M. A., Billard, A., Iacoboni, M., and Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw.* 13, 975–997. doi: 10.1016/S0893-6080(00)00070-8

Arbib, M. A., Bischoff, A., Fagg, A. H., and Grafton, S. T. (1994). Synthetic PET: analyzing large-scale properties of neural networks. *Hum. Brain Map.* 2, 225–233. doi: 10.1002/hbm.460020404

Baayen, H., Piepenbrock, R., and van Rijn, H. (1995). *The CELEX Lexical Database (CD-ROM): Linguistic Data Consortium*. Philadelphia, PA: University of Pennsylvania.

Bakker, I., MacGregor, L. J., Pulvermüller, F., and Shtyrov, Y. (2013). Past tense in the brain's time: neurophysiological evidence for dual-route processing of past-tense verbs. *NeuroImage* 71, 187–195. doi: 10.1016/j.neuroimage.2012.12.065

Beretta, A., Campbell, C., Carr, T. H., Huang, J., Schmitt, L. M., Christianson, K., et al. (2003). An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain Lang.* 85, 67–92. doi: 10.1016/S0093-934X(02)00560-6

Bergida, R., O'Craven, K., Savoy, R. L., and Ullman, M. T. (1998). "fMRI double dissociations in frontal and temporal regions between regular and irregular past tense production," in *Paper Presented at the 5th Annual Meeting of the Cognitive Neuroscience Society*. San Francisco, CA.

Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychol. Bullet.* 143, 233–262. doi: 10.1037/bul0000099

Bick, J., and Nelson, C. A. (2017). Early experience and brain development. *WIREs Cogn. Sci.* 8:e1387. doi: 10.1002/wcs.1387

Bird, H., Lambon-Ralph, M. A., Seidenberg, M. S., McClelland, J. L., and Patterson, K. (2003). Deficits in phonology and past-tense morphology: what's the connection? *J. Memory Lang.* 48, 502–526. doi: 10.1016/S0749-596X(02)00538-7

Bright, P., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2008). Longitudinal studies of semantic dementia: the relationship between structural and functional changes over time. *Neuropsychologia* 46, 2177–2188. doi: 10.1016/j.neuropsychologia.2008.02.019

Brown, T. T., Lugar, H. M., Coalson, R. S., Miezin, F. M., Petersen, S. E., and Schlaggar, B. L. (2005). Developmental changes in human cerebral functional organization for word generation. *Cereb. Cortex* 15, 275–290. doi: 10.1093/cercor/bhh129

Burzio, L. (2002). Missing players: phonology and the past-tense debate. *Lingua* 112, 157–199. doi: 10.1016/S0024-3841(01)00041-9

Bybee, J. L., and Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language* 58, 265–289. doi: 10.1353/lan.1982.0021

Cangelosi, A., and Parisi, D. (2004). The processing of verbs and nouns in neural networks: insights from synthetic brain imaging. *Brain Lang.* 89, 401–408. doi: 10.1016/S0093-934X(03)00353-5

Casey, B. J., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol. Psychol.* 54, 241–257. doi: 10.1016/S0301-0511(00)00058-2

Casey, B. J., Tottenham, N., Liston, C., and Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* 9, 104–110. doi: 10.1016/j.tics.2005.01.011

Chen, C., Xue, G., Mei, L., Chen, C., Dong, Q., and Joan, Y. C. (2009). Cultural neurolinguistics. *Progr. Brain Res.* 178, 159–171. doi: 10.1016/S0079-6123(09)17811-1

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behav. Brain Sci.* 22, 991–1013. doi: 10.1017/S0140525X99002228

Desai, R., Conant, L. L., Waldron, E., and Binder, J. R. (2006). FMRI of past tense processing: the effects of phonological complexity and task difficulty. *J. Cogn. Neurosci.* 18, 278–297. doi: 10.1162/jocn.2006.18.2.278

Dhond, R. P., Marinkovic, K., Dale, A. M., Witzel, T., and Halgren, E. (2003). Spatiotemporal maps of past-tense verb inflection. *Neuroimage* 19, 91–100. doi: 10.1016/S1053-8119(03)00047-8

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99. doi: 10.1016/0010-0277(93)90058-4

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., et al. (2019). How the input shapes the acquisition of verb morphology: elicited production and computational modelling in two highly inflected languages. *Cognit. Psychol.* 110, 30–69. doi: 10.1016/j.cogpsych.2019.02.001

Horwitz, B., Tagamets, M. A., and McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends Cogn. Sci.* 3, 91–98. doi: 10.1016/S1364-6613(99)01282-6

Hosoda, C., Tanaka, K., Nariai, T., Honda, M., and Hanakawa, T. (2013). Dynamic neural network reorganization associated with second language vocabulary acquisition: a multimodal imaging study. *J. Neurosci.* 33, 13663–13672. doi: 10.1523/JNEUROSCI.0410-13.2013

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., VanValin, R. D., Murphy, B. W., and Khalak, H. G. (1996). A positron emission tomographic study of regular and irregular verb morphology in English. *Language* 72, 451–497. doi: 10.2307/416276

Joanisse, M. F., and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: a connectionist model. *Proc. Natl. Acad. Sci. U. S. A.* 96, 7592–7597. doi: 10.1073/pnas.96.13.7592

Joanisse, M. F., and Seidenberg, M. S. (2005). Imaging the past: neural activation in frontal and temporal regions during regular and irregular past-tense processing. *Cogn. Affect. Behav. Neurosci.* 5, 282–296. doi: 10.3758/CABN.5.3.282

Johnson, M. H. (2001). Functional brain development in humans. *Nat. Rev. Neurosci.* 2, 475–483. doi: 10.1038/35081509

Johnson, M. H., and Munakata, Y. (2005). Processes of change in brain and cognitive development. *Trends Cogn. Sci.* 9, 152–158. doi: 10.1016/j.tics.2005.01.009

Just, M. A., Carpenter, P. A., and Varma, S. (1999). Computational modeling of high-level cognition and brain function. *Human Brain Mapp.* 8, 128–136. doi: 10.1002/(SICI)1097-0193(1999)8:2/3<128::AID-HBM10>3.0.CO;2-G

Kochunov, P., Fox, P., Lancaster, J., Tan, L. H., Amunts, K., Zilles, K., et al. (2003). Localized morphological brain differences between English-speaking Caucasians and Chinese-speaking Asians: new evidence of anatomical plasticity. *NeuroReport* 14, 961–964. doi: 10.1097/00001756-200305230-00011

Lavric, A., Pizzagalli, D., Forstmeier, S., and Rippon, G. (2001). A double-dissociation of English past-tense production revealed by event-related potentials and low-resolution electromagnetic tomography (LORETA). *Clin. Neurophysiol.* 112, 1833–1849. doi: 10.1016/S1388-2457(01)00615-0

Leminen, A., Smolka, E., Duñabeitia, J. A., and Pliatsikas, C. (2019). Morphological processing in the brain: the good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116, 4–44. doi: 10.1016/j.cortex.2018.08.016

Li, P., Legault, J., and Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: anatomical changes in the human brain. *Cortex* 58, 301–324. doi: 10.1016/j.cortex.2014.05.001

Li, P., and MacWhinney, B. (2002). PatPho: a phonological pattern generator for neural networks. *Behav. Res. Method. Instr. Comp.* 34, 408–415. doi: 10.3758/BF03195469

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005

MacWhinney, B., and Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition* 40, 121–157. doi: 10.1016/0010-0277(91)90048-9

Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *J. Cogn. Neurosci.* 5, 215–234. doi: 10.1162/jocn.1993.5.2.215

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: the exception that proves the rule. *Cogn. Psychol.* 29, 189–256. doi: 10.1006/cogp.1995.1015

Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M. W., Thomas, M., and Westermann, G. (2007). *Neuroconstructivism: How the Brain Constructs Cognition*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198529910.001.0001

McClelland, J. L., and Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *Trends Cogn. Sci.* 6, 465–472. doi: 10.1016/S1364-6613(02)01993-9

Mei, L., Xue, G., Lu, Z.-L., He, Q., Wei, M., Zhang, M., et al. (2015). Native language experience shapes neural basis of addressed and assembled phonologies. *NeuroImage* 114, 38–48. doi: 10.1016/j.neuroimage.2015.03.075

Miozzo, M. (2003). On the processing of regular and irregular forms of verbs and nouns: evidence from neuropsychology. *Cognition* 87, 101–127. doi: 10.1016/s0010-0277(02)00200-7

Miozzo, M., and Gordon, P. (2005). Facts, events, and inflection: when language and memory dissociate. *J. Cognit. Neurosci.* 17, 1074–1086. doi: 10.1162/0898929054475163

Nelson, C. A., Moulson, M. C., and Richmond, J. (2006). How does neuroscience inform the study of cognitive development? *Hum. Dev.* 49, 260–272. doi: 10.1159/000095579

Norris, D. G. (2006). Principles of magnetic resonance assessment of brain function. *J. Magnet. Resonance Imag.* 23, 794–807. doi: 10.1002/jmri.20587

Oh, T. M., Tan, K. L., Ng, P., Berne, Y. I., and Graham, S. (2011). The past tense debate: is phonological complexity the key to the puzzle? *NeuroImage* 57, 271–280. doi: 10.1016/j.neuroimage.2011.04.008

O'Reilly, R. C., and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/2014.001.0001

Park, D. C., and Huang, C.-M. (2010). Culture wires the brain. *Perspect. Psychol. Sci.* 5, 391–400. doi: 10.1177/1745691610374591

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., and McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia* 39, 709–724. doi: 10.1016/S0028-3932(01)00008-2

Pinker, S. (1991). Rules of language. *Science* 253, 530–535. doi: 10.1126/science.1857983

Pinker, S. (1994). *The Language Instinct.* New York, NY: Morrow. doi: 10.1037/e412952005-009

Pinker, S. (1997). Words and rules in the human brain. *Nature* 387, 547–548. doi: 10.1038/42347

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. New York, NY: Basic Books.

Pinker, S., and Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn. Sci.* 6, 456–463. doi: 10.1016/S1364-6613(02)01990-3

Pinker, S., and Ullman, M. T. (2003). Beyond one model per phenomenon. *Trends Cogn. Sci.* 7, 108–109. doi: 10.1016/S1364-6613(03)00021-4

Plaut, D. C. (1995). Double dissociation without modularity - evidence from connectionist neuropsychology. *J. Clin. Exp. Neuropsychol.* 17, 291–321. doi: 10.1080/01688639508405124

Plunkett, K., and Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cogn. Sci.* 23, 463–490. doi: 10.1207/s15516709cog2304_4

Plunkett, K., and Marchman, V. A. (1991). *U*-shaped learning and frequency-effects in a multilayered perceptron - implications for child language-acquisition. *Cognition* 38, 43–102. doi: 10.1016/0010-0277(91)90022-V

Plunkett, K., and Marchman, V. A. (1993). From rote learning to system building - acquiring verb morphology in children and connectionist nets. *Cognition* 48, 21–69. doi: 10.1016/0010-0277(93)90057-3

Price, C. J., and Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends Cogn. Sci.* 6, 416–421. doi: 10.1016/S1364-6613(02)01976-9

Quartz, S. R. (1993). Neural networks, nativism, and the plausibility of constructivism. *Cognition* 48, 223–242. doi: 10.1016/0010-0277(93)90041-S

Quartz, S. R. (1999). The constructivist brain. *Trends Cogn. Sci.* 3, 48–57. doi: 10.1016/S1364-6613(98)01270-4

Quartz, S. R., and Sejnowski, T. J. (1997). The neural basis of cognitive development: a constructivist manifesto. *Behav. Brain Sci.* 20, 537–596. doi: 10.1017/S0140525X97001581

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* 65, 386–408. doi: 10.1037/h0042519

Rumelhart, D., and McClelland, J. L. (1986). "On learning the past tense of English verbs: implicit rules or parallel distributed processing?" in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds J. L. McClelland, D. Rumelhart, and the PDP Research Group (Cambridge, MA: MIT Press), 216–271. doi: 10.7551/mitpress/5236.001.0001

Rumelhart, D. E. (1989). "The architecture of mind: a connectionist approach," in *Foundations of Cognitive Science,* ed M. I. Posner (Cambridge, MA: MIT Press), 133–159.

Sahin, N. T., Pinker, S., and Halgren, E. (2006). Abstract grammatical processing of nouns and verbs in broca's area: evidence from FMRI. *Cortex* 42, 540–562. doi: 10.1016/S0010-9452(08)70394-0

Seidenberg, M. S., and Arnoldussen, A. (2003). The brain makes a distinction between hard and easy stimuli: comments on Beretta et al. *Brain Lang.* 85, 527–530. doi: 10.1016/S0093-934X(03)00065-8

Seidenberg, M. S., and Hoeffner, J. H. (1998). Evaluating behavioral and neuroimaging data on past tense processing. *Language* 74, 104–122. doi: 10.1353/lan.1998.0087

Shultz, T. R., Mysore, S. P., and Quartz, S. P. (2007). "Why let networks grow?" in *Neuroconstructivism: Perspectives and Prospects,* eds D. Mareschal, G. Westermann, S. Sirois, and M. H. Johnson (Oxford: Oxford University Press), 65–98. doi: 10.1093/acprof:oso/9780198529934.003.0004

Stiles, J. (2009). On genes, brains, and behavior: why should developmental psychologists care about brain development? *Child Dev. Perspectiv.* 3, 196–202. doi: 10.1111/j.1750-8606.2009.00106.x

Tagamets, M. A., and Horwitz, B. (1998). Integrating electrophysiological and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain imaging study. *Cereb. Cortex* 8, 310–320. doi: 10.1093/cercor/8.4.310

Thomas, M. S. C., Purser, H. R. M., Tomlinson, S., and Mareschal, D. (2012). Are imaging and lesioning convergent methods for assessing functional specialisation? Investigations using an artificial neural network. *Brain Cogn.* 78, 38–49. doi: 10.1016/j.bandc.2011.10.003

Tyler, L. K., de Mornay-Davies, P., Anokhina, R., Longworth, C., Randall, B., and Marslen-Wilson, W. D. (2002). Dissociations in processing past tense morphology: neuropathology and behavioral studies. *J. Cogn. Neurosci.* 14, 79–94. doi: 10.1162/089892902317205348

Tyler, L. K., Stamatakis, E. A., Jones, R. W., Bright, P., Acres, K., and Marslen-Wilson, W. D. (2004). Deficits for semantics and the irregular past tense: a causal relationship? J. *Cognit. Neurosci.* 16, 1159–1172. doi: 10.1162/0898929041920559

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition* 92, 231–270. doi: 10.1016/j.cognition.2003.10.008

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., et al. (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *J. Cogn. Neurosci.* 9, 266–276. doi: 10.1162/jocn.1997.9.2.266

Westermann, G. (2016). Experience-dependent brain development as a key to understanding the language system. *Top. Cogn. Sci.* 8, 446–458. doi: 10.1111/tops.12194

Westermann, G., Mareschal, D., Johnson, M. H., Sirois, S., Spratling, M., and Thomas, M. (2007). Neuroconstructivism. *Dev. Sci.* 10, 75–83. doi: 10.1111/j.1467-7687.2007.00567.x

Westermann, G., and Plunkett, K. (2007). Connectionist models of inflection processing. *Lingue e Linguaggio* 1, 291–311. doi: 10.1418/25655

Westermann, G., and Ruh, N. (2009). "Synthetic brain imaging of English past tense inflection," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, eds N. A. Taatgen and H. van Rijn (Austin, TX: Cognitive Science Society), 1364–1369. Available online at: https://escholarship.org/uc/item/5d13r54m

Westermann, G., and Ruh, N. (2012). A neuroconstructivist model of past tense development and processing. *Psychol. Rev.* 119, 649–667. doi: 10.1037/a0028258

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 02 August 2021
doi: 10.3389/fpsyg.2021.678712

Morpho-Phonetic Effects in Speech Production: Modeling the Acoustic Duration of English Derived Words With Linear Discriminative Learning

Simon David Stein and Ingo Plag*

English Language and Linguistics, Department of English and American Studies, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Recent evidence for the influence of morphological structure on the phonetic output goes unexplained by established models of speech production and by theories of the morphology-phonology interaction. Linear discriminative learning (LDL) is a recent computational approach in which such effects can be expected. We predict the acoustic duration of 4,530 English derivative tokens with the morphological functions DIS, NESS, LESS, ATION, and IZE in natural speech data by using predictors derived from a linear discriminative learning network. We find that the network is accurate in learning speech production and comprehension, and that the measures derived from it are successful in predicting duration. For example, words are lengthened when the semantic support of the word's predicted articulatory path is stronger. Importantly, differences between morphological categories emerge naturally from the network, even when no morphological information is provided. The results imply that morphological effects on duration can be explained without postulating theoretical units like the morpheme, and they provide further evidence that LDL is a promising alternative for modeling speech production.

Keywords: speech production, linear discriminative learning, acoustic duration, morphological theory, derivation, mental lexicon

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
Jessie S. Nixon,
University of Tübingen, Germany
Kevin Tang,
University of Florida, United States
Valeria Caruso,
University of Michigan, United States
Christina Manouilidou,
University of Ljubljana, Slovenia

****Correspondence:***
Simon David Stein
simon.stein@uni-duesseldorf.de

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

Received: *10 March 2021*
Accepted: *29 June 2021*
Published: *02 August 2021*

Citation:
Stein SD and Plag I (2021) Morpho-Phonetic Effects in Speech Production: Modeling the Acoustic Duration of English Derived Words With Linear Discriminative Learning. Front. Psychol. 12:678712. doi: 10.3389/fpsyg.2021.678712

INTRODUCTION

Recent findings in morpho-phonetic and psycholinguistic research have indicated that phonetic detail can vary by morphological structure. For example, the acoustic duration of English word-final [s] and [z] differs depending on morphological status and inflectional function (Plag et al., 2017, 2020; Seyfarth et al., 2017; Tomaschek et al., 2019). For derivation, too, studies have demonstrated effects of morphological structure on phonetic output. For example, morphological geminates in English differ in duration depending on morphological category and informativity (Ben Hedia and Plag, 2017; Ben Hedia, 2019), and phonetic reduction in various domains can depend on how easily speakers can decompose a complex word into its constituents (e.g., Hay, 2003, 2007; Plag and Ben Hedia, 2018; Zuraw et al., 2020).

These findings raise several problems at the theoretical level. The observation that phonetic detail varies systematically with morphological properties is unaccounted for by traditional and current models of the morphology-phonology interaction and of speech production (e.g., Chomsky and Halle, 1968; Kiparsky, 1982; Dell, 1986; Levelt et al., 1999; Roelofs and Ferreira, 2019; Turk and Shattuck-Hufnagel, 2020). This is because these models are either underspecified regarding the processing of complex words, or do not allow for post-lexical access of morphological information. For example, feed-forward models of the morphology-phonology interface (e.g., Kiparsky, 1982) assume that morphological brackets around constituents are "erased" in the process of passing on a word through morphological and phonological levels of processing. This means that no trace of morphological structure should be left at the level of phonetic realization. Similarly, established psycholinguistic models of speech production (e.g., Levelt et al., 1999) assume that morphological units select general phoneme templates which are then passed on to an articulator module to be realized phonetically. Again, no morphological information is encoded in these templates, meaning that no systematic differences between morphological properties are expected at the phonetic level.

Yet, morphological effects on the phonetic output have repeatedly been observed, which is incompatible with these assumptions. For example, the observation that complex words are more acoustically reduced when they are less decomposable into their constituents (Hay, 2003, 2007; Plag and Ben Hedia, 2018; Zuraw et al., 2020) seems to suggest that information about morphological boundaries must somehow still be present at the phonetic level. From the perspective of the speech production models and theories of the morphology-phonology interaction outlined above, such effects are unexpected, and the mechanisms behind them are unclear. To better explain the morphology-phonetics interaction at the theoretical level and to understand the patterning of durations in complex words from a new perspective, we need alternative approaches.

One such approach is to model phonetic detail based on the principles of discriminative learning (see, e.g., Ramscar and Yarlett, 2007; Ramscar et al., 2010; Baayen et al., 2011). Such an approach sees form-meaning relations not as compositional, but as discriminatory instead. That is, form-meaning relations are created in a system of *difference*, which distinguishes between features based on their similarity and dissimilarity and connects them to each other in a learning process. In discriminative approaches, "signs" in the semiotic sense of relations of form and meaning (de Saussure, 1916) are not fixed units. Discriminative models refrain from sub-lexical static representations such as morphemes or roots in the lexicon. Instead, speech comprehension and production are the result of a dynamic learning process where relations between form and meaning are constantly recalibrated based on the speaker's experience. How strong associations between given forms and meanings are in the system depends on how often specific forms occur together with specific meanings, and on how often they fail to occur together with others. Each time a speaker makes a new experience, i.e., encounters a form together with a specific meaning, all associations of forms and meanings in the system are updated to reflect this new state of learning. An association strength increases when a "cue" (such as a specific form) occurs together with an "outcome" (such as a specific meaning), and an association strength decreases when a cue does not occur with the outcome.

Such an approach has clear advantages if we are to explain the evidence that morphology directly affects phonetic realization. A discriminative learning model lacks a feed-forward architecture which divides speech processing into separate levels. It is an end-to-end model that goes directly from form to meaning and from meaning to form. This means that the loss of morphological information between levels, e.g., through bracket erasure or phoneme template selection, is no longer an issue. Moreover, discriminative learning refrains from postulating morphemes or phonemes as psychologically relevant units in the first place. This opens the way for interpreting acoustic differences from a new perspective. In a discriminative approach, differences between morphological functions are expected to emerge naturally from sublexical and contextual cues. If we can model systematic acoustic variation between morphological functions with measures derived from a discriminative network, it is possible to explain potential effects by its theoretical principles of learning and experience.

While discriminative approaches have already been used to model other morphological correlates, such as reaction time (e.g., Baayen et al., 2011), the question arises whether a discriminative approach is able to successfully predict phonetic variation. Recently, Tomaschek et al. (2019) employed naïve discriminative learning (NDL) to model the duration of English word-final [s] and [z] of different morphological status. The measures derived from their network were predictive and indicated that a higher certainty in producing a morphological function leads to lengthening. While Tomaschek et al. (2019) focused on inflection, it is necessary to also test how well discriminative approaches can deal with derivational morphology. The present paper aims to account for this gap.

Our study investigates the durational properties of derived words in English. We modeled word durations for 4,530 tokens with the derivational functions DIS, NESS, LESS, ATION, and IZE from the Audio BNC (Coleman et al., 2012), using multiple linear regression models and mixed-effects regression models. The crucial predictors in our models are measures derived from the computational framework of linear discriminative learning (Baayen et al., 2019b).

Linear discriminative learning (LDL) is a new variant of naïve discriminative learning. Like NDL, it is *discriminative* because its system of form-meaning relations is generated by discriminating between different forms and meanings instead of building them from compositional units. Like NDL, LDL is a system of *learning* because the association strengths between forms and meanings are continuously recalibrated in a process of experience. This learning is simple and interpretable because, in contrast to deep learning, it features just two layers, an input layer and an output layer, both of which are linguistically transparent. Unlike NDL, however, LDL is *linear* and no longer "naïve." Its networks are linear mappings between form matrices and meaning matrices (which serve as either the input layer or the output layer, respectively). In this approach, forms are represented by vectors, and meanings are also represented by vectors, similarly to approaches in distributional semantics. The idea is that if we can express both forms and meanings numerically, we can mathematically connect form and meaning. In LDL, the network is no longer naïve because where NDL represents word meanings with binary vectors, LDL uses real-valued vectors, taking into account that words cannot only be similar in form, but also in meaning. How this is implemented is explained further below in the section Materials and Methods.

Our aim in this study is, first, to investigate how well LDL can account for the durational variation in our data. Second, we investigate what the effects of the LDL-derived measures tell us about the mechanisms of speech production. How can we interpret potential effects conceptually? Third, as we are interested in exploring how these findings relate to morphological functions, we also investigate how the results differ depending on how much information the network has about these functions. For this purpose, we initially trained three different LDL networks, two of which contain explicit morphological information. The first network does not include any information about morphological category and treats all derivatives as idiosyncratic (the Idiosyncratic Network). The second network uses vectors that include semantic information about the derivative and about the morphological category it belongs to (the Morphology Network). The third network uses vectors that include semantic information about the base word (instead of the derivative) and about the morphological category (the Base Network).

We hypothesize that LDL-derived measures can successfully (i.e., significantly) predict derivative durations. If they do, the effects of LDL-derived measures should be interpretable with regards to speech production (for example, they should mirror the finding by Tomaschek et al. (2019) that higher certainty is associated with longer durations). Lastly, we explore whether there are differences between the networks that contain information about the morphological category a derivative belongs to and the network that does not contain such information.

To preview our results, three key findings emerge from the analysis. First, all LDL networks achieve high learning accuracy and the proportion of variance in duration explained by the LDL-derived predictors is comparable to that explained by traditional predictors. Second, the effects of LDL measures highlight important patterns of speech production. For example, they suggest that words are lengthened in speech production when the semantic support of the word's predicted articulatory path is stronger (i.e., when certainty is higher), mirroring the finding by Tomaschek et al. (2019). Third, we find that, even though we did not provide the Idiosyncratic Network with any information about the morphological category a word belongs to, these categories still emerge from the network. For instance, the different morphological categories are reflected in the distributions of the correlation strength of a word's predicted semantics with the semantics of its neighbors. This corresponds to what we would traditionally describe as the differences in semantic transparency between affix categories.

The remainder of this paper is structured as follows. The section Materials and Methods describes our methodology, illustrating the procedure of collecting the speech data (the section Speech Data), building the LDL networks (the section Linear Discriminative Learning), the variables used (the section Variables) and the modeling procedure (the section Modeling Word Durations). The section Results outlines our results, followed by a discussion and conclusion in the section Discussion and Conclusion.

MATERIALS AND METHODS

Our methodology consists of three main steps: first, retrieving the speech data for the durational measurements for the response variable, second, building the LDL networks to retrieve LDL-derived predictors of interest, and third, devising regression models to predict derivative durations from various predictors. All data, scripts and materials can be found at osf.io/jkncb.

Speech Data

The speech data was obtained from the Audio BNC (Coleman et al., 2012). This corpus consists of both monologues and dialogues from different speech genres of several British English varieties. It comes phonetically aligned by an automatic forced aligner. Containing about 7.5 million words, it is large enough to yield enough observations per derivational function. A corpus approach has the advantage that that we are not only able to analyze a lot of data, but also that the type of data is conversational speech. This enables us to investigate a more authentic process of language production than with carefully elicited speech. It has been argued (e.g., Tucker and Ernestus, 2016) that research on speech production in particular needs to shift its focus to spontaneous speech to be able to draw valid conclusions about language processing.

The morphological categories selected for investigation are DIS, NESS, LESS, ATION, and IZE. We use the term *morphological category* in the traditional sense, referring to words that share a particular morphologically expressed meaning. We do not use the term *morpheme* because it is usually employed to denote a minimal sign combining a form and a meaning (e.g., /-ləs/ "without," see, e.g., Plag and Balling, 2020). We use the term *function* to refer to the semantic or grammatical contribution of a particular affix or process. LDL does not assume any fixed relationship between form and meaning. Meanings are dynamically mapped onto a stream of forms (overlapping triphones in our case), but never defined as being tied to strings that we would traditionally describe as being "morphemic." The terms *function* and *category* better reflect the fact that in LDL, derived words might be grouped into categories sharing similar semantics or features (cf. Word and Paradigm Morphology) but are not "composed" of form-meaning building blocks (cf. morpheme-based morphology). LDL's lexomes are pointers to meanings only, not to forms.

The five categories DIS, NESS, LESS, ATION, and IZE were chosen, first, because they featured sufficient token counts in the Audio BNC and are attested in Baayen et al.'s (2019b) vector space (explained in the section Training Data). Second, they were chosen because they cover a wide spectrum of characteristics traditionally considered important for affix classification. For example, following Bauer et al. (2013) and Plag (2018), the affixes corresponding to those categories differ in their semantic transparency: *-ness*, *-less*, and *dis-* produce mostly transparent derivatives, whereas *-ize* and *-ation* are overall a little less transparent in comparison. They vary in the range of their meanings, from relatively narrow and clearly definable semantics (e.g., the privative meaning of *-less* or the negative meaning of *dis-*) to more varied semantics (e.g., *-ness* denoting abstract states, traits, or properties) to highly multifaceted semantics (*-ize* can have locative, ornative, causative, resultative, inchoative, performative, or similative meaning, *-ation* can denote events, states, locations, products or means). They also differ in their productivity, with *-ness* and *-less* being considered highly productive, and *-ize*, *-ation*, and *dis-* being somewhat less productive. Lastly, they also differ phonologically. While *-ness*, *-less*, and *dis-* are not (obligatorily) subject to phonological alternations and not involved in resyllabification processes, *-ize* and *-ation* can cause stress shifts and other phonological alternations within their bases, and resyllabification is commonplace.

We obtained speech data for these morphological categories by entering pertinent query strings into the web interface of the Audio BNC and extracting the resulting wordlist and associated recordings and textgrids. These query strings searched for all word tokens that begin or end in the orthographic and phonological representation of each of the investigated derivational function. We manually cleaned the datasets by excluding words which were monomorphemic (e.g., *bless*, *disk*, *station*), whose semantics or base were unclear (e.g., *harness*, *disrupt*, *dissertation*), or which were proper names or titles (e.g., *Guinness*, *Stenness*, *Stromness*).

Before starting the acoustic analysis, manual inspection of all items was necessary to exclude items that were not suitable for further analysis. This was done by visually and acoustically inspecting the items in the speech analysis software Praat (Boersma and Weenik, 2001). Items were excluded that fulfilled one or more of the following criteria: the textgrid was a duplicate or corrupted for technical reasons, the target word was not spoken or was inaudible due to background noise, the target word was interrupted by other acoustic material, laughing, or pauses, the target word was sung instead of spoken, the target word was not properly segmented or incorrectly aligned to the recording. In cases where the alignment did not seem satisfactory, we examined the word-initial boundary and the word-final boundary in order to decide whether to exclude the item. We considered an observation to be correctly aligned if none of these boundaries would have to be shifted to the left or right under application of the segmentation criteria in the pertinent phonetic literature (cf. Machač and Skarnitzl, 2009; Ladefoged and Johnson, 2011). Following Machač and Skarnitzl (2009), we considered the shape of the sound wave to be the most important cue, followed by the spectrogram, followed by listening.

In a final step, the dataset was reduced to only those words that were attested in the TASA corpus as well as in CELEX, and whose base was simplex (this step is explained in the section Training Data). The final dataset of derivatives that entered the models comprised 4,530 tokens and 363 types. **Table 1** gives an overview of the data in each morphological category. Further descriptive statistics of the datasets are provided in the **Supplementary Material**.

TABLE 1 | Overview of tokens and types per morphological category.

	DIS	NESS	LESS	ATION	IZE
Tokens	233	344	145	3,403	405
Types	35	49	31	209	39

Linear Discriminative Learning

Our aim is to predict the durational patterning in the 4,530-token dataset described above with measures derived from an LDL network. These measures can be calculated on the basis of a transformation matrix that maps a cue matrix *C* for forms onto a semantic matrix *S* for meanings (for comprehension), and the semantic matrix *S* onto the cue matrix *C* (for production). The basic building blocks used to construct the meaning dimensions in matrix *S* are referred to as *lexomes*. Lexomes are atomic units of meaning in an LDL network and serve as pointers to semantic vectors. In comprehension, they are also the "outcomes" in the *S* matrix, which are predicted from the "cues" in the *C* matrix. Lexomes can for example correspond to words (content lexomes, such as LEMON), but also to derivational or inflectional functions (function lexomes, such as NESS).

It is important to note that function lexomes correspond to morphological categories, but are not the same thing as morphemes. In LDL, morphological categories (like NESS) are coded as semantic vectors and are not units of form and meaning, but units of meaning only. How these lexomes and their vectors were obtained, how the matrices were constructed and how they were mapped onto each other is illustrated in the following sections.

Training Data

To construct a linear discriminative learning network, it is necessary to obtain semantic vectors that represent the words' meanings (this will be explained in more detail in the section Matrices for Form and Meaning). For this, we made use of the vectors generated by Baayen et al. (2019b) from the TASA corpus, who used an algorithm to predict words in each sentence of the corpus from other words in that sentence (this will be explained further below). To make sure that we can use these semantic vectors for our derivatives, we first reduced our speech data set from the Audio BNC to those derivatives that are attested in TASA (losing 352 words). In a second step, we used the CELEX lexical database (Baayen et al., 1995) to obtain phonological transcriptions for the words in our data set. These transcriptions are necessary for constructing the matrices. Since CELEX did not have transcriptions for all words, this step led to a slight reduction of our data set (losing 9 words). In a final step, we excluded all derivatives (49 words) whose bases were already complex, i.e., all derivatives that have more than one derivational function (e.g., *stabilization, specification, attractiveness, disclosure, disagreement*). One reason for excluding these derivatives is that it is currently not clear how to build their semantic vectors. Another reason is that multi-affixed words in corpora are comparatively infrequent. Too infrequent derivatives might require a corpus even bigger than TASA from which to construct reliable semantic vectors.

The resulting dataset contained 363 unique derivatives (i.e., types). This dataset consists of all derivatives from the Audio BNC that are also attested in TASA. One problem with this

TABLE 2 | Schematic examples of a cue matrix *C* (left) and a semantic matrix *S* (right) for the words *cat*, *happiness*, *walk*, and *lemon*.

Schematic example of a *C* matrix						Schematic example of an *S* matrix				
	#k{	**k{ t**	**{t#**	**#h{**	**h{p**		**CAT**	**HAPPINESS**	**WALK**	**LEMON**
k{t	1	1	1	0	0	k{t	0.000000	−6.24e-05	4.71e-05	−0.000138
h{pInIs	0	0	0	1	1	h{pInIs	−0.000110	0.0000000	0.000194	−2.20E-05
w$k	0	0	0	0	0	w$k	0.000304	−0.0002335	0.000000	−3.74E-05
lEm@n	0	0	0	0	0	lEm@n	−7.28e-05	−2.41e-07	−2.68e-05	0.00000

Note that for the triphones in the C matrix, word boundaries are also counted, represented by a hash (#). The DISC phonetic alphabet is used for computer-readable transcription (Burnage, 1990).

dataset is that it would be rather unrealistic as training data. This is because a speaker encounters far more than just a few hundred words during their lifetime, and not all these encountered words contain one of the five investigated morphological categories DIS, NESS, LESS, ATION, and IZE. We therefore decided to merge this dataset with all words in TASA that had already been coded in Baayen et al. (2019b) for derivational functions (function lexomes) and phonological transcriptions (4,880 more words). This dataset contained 897 derivatives with the 25 derivational function lexomes AGAIN, AGENT, DIS, EE, ENCE, FUL, IC, INSTRUMENT, ATION, ISH, IST, IVE, IZE, LESS, LY, MENT, MIS, NESS, NOT, ORDINAL, OUS, OUT, SUB, UNDO, and Y, as well as 3,983 monomorphemic words. Derivational functions were coded irrespective of variation in affix spelling. Most of these words are not attested in our speech data and therefore not of interest for the durational modeling, but including them makes the training itself more realistic.

The resulting 5,176 unique word forms were then used for the *C* matrix, and the 5,201 unique lexomes (comprising the vectors for the 5,176 content lexomes and the 25 derivational function lexomes) were used for the *S* matrix. The next section illustrates what these matrices are and how they are constructed.

Matrices for Form and Meaning

In an LDL network, features of a word are represented by a vector for this word in a multidimensional space. Each word has a vector that specifies its form features, and a vector that specifies its semantic features. We therefore need two matrices: a cue matrix *C* for the words' forms and a semantic matrix *S* for the words' meanings.

The cue matrix *C* contains in rows the words' phonological transcriptions, and in columns form indicators that are either present or absent in those words. As shown in Arnold et al. (2017) and Shafaei-Bajestan et al. (2020), it is possible to use real-valued features extracted directly from the speech signal instead of discrete features. In the present study, we use triphones as form indicators, following Baayen et al. (2019b). These triphones overlap and can be understood as proxies for transitions in the articulatory signal. Each cell in the matrix codes in a binary fashion (1 for present or 0 for absent) whether the respective triphone string (specified in the column) occurs in the phonological transcription of the word (specified in the row). An example of the layout of the *C* matrix is given in **Table 2** on the left-hand side. For the *C* matrix in this study, we used the 5,176 unique word forms mentioned in the section Training Data.

The semantic matrix *S* contains in its rows the words' phonological transcriptions, and in its columns the semantic dimensions, or lexomes, with which the words are associated. In the present study, these lexomes correspond to interpretable linguistic items, such as words and derivational functions. Each cell in the S matrix contains a real number, which represents the association strength of a word (specified in the row) to a lexome (specified in the column). As mentioned in the Introduction, this is an important difference of LDL compared to NDL, where word meanings are initially coded as binary-valued vectors similar to the cue matrix. LDL, on the other hand, starts out with real-valued association weights. An example of the layout of the *S* matrix is given in **Table 2** on the right-hand side. For the *S* matrix in this study, we used the 5,201 unique lexomes mentioned in the section Training Data.

Where do these association weights come from? In the present study, we used association weights that were generated from word co-occurrence in real language data. For this, Baayen et al. (2019b) trained an NDL network on the TASA corpus (Ivens and Koslin, 1991; Landauer et al., 1998). This NDL network operated on an established learning algorithm (Widrow and Hoff, 1960) that incrementally learns association strengths between lexomes. In such an approach, words in a sentence are predicted from the words in that sentence. While the network goes through the sentences in the corpus, the associations strengths of the lexomes with each other are continuously adjusted over time. As language learning is about learning which connections are relevant, the association strength of lexomes that often occur together will be strengthened. As discriminative learning is also about *un*learning connections which are irrelevant, similarly, the association strength of lexomes will be weakened each time they do not occur together. For the implementational and mathematical details of this procedure, as well as for the validation of the resulting semantic vector space, the reader is referred to Baayen et al. (2019b). Importantly for the present study, Baayen and colleagues included lexomes not only for words, but also for derivational functions corresponding to suffixes and prefixes. This enables us to build LDL networks that take into account morphological categories shared between derivatives (in addition to an LDL network that does not take these into account and treats all words as idiosyncratic, i.e., as having a unique semantics that is not related to the semantics of constituents below the word level).

The so-called *lexome-to-lexome matrix* resulting from this learning process is a vector space in which each lexome vector represents a certain association with the meanings of all other lexomes. According to the idea that "you shall know a word by the company it keeps" (Firth, 1957), each value in the vector of a lexome represents the association strength of this lexome to the meaning of another lexome in TASA. Following Baayen et al. (2019b), we used a version of their lexome-to-lexome matrix which was trimmed to about five thousand dimensions and whose main diagonal was set to zero.[1] From this lexome-to-lexome matrix, we extracted the vectors for our 5,201 unique lexomes (described in the section Training Data), which we then used for the *S* matrix.

For the present study, we built three different LDL networks: one which contains no information about the morphological category a derivative belongs to but treats all derivatives as idiosyncratic, one in which the vectors contain information about the derivative and about the morphological category it belongs to, and one in which the vectors contain information about the base of a derivative and about the morphological category it belongs to. For each of these networks we need a matrix *S* and a matrix *C*. We will refer to the matrices with idiosyncratic derivatives as matrix S_I and matrix C_I, to the matrices with information about the derivative and its morphological category as matrix S_M and matrix C_M, and to the matrices with information about the base and the morphological category as matrix S_B and matrix C_B. We will refer to the networks as a whole as the *Idiosyncratic Network*, the *Morphology Network*, and the *Base Network*, respectively.

The Idiosyncratic Network with matrices S_I and C_I considered only the semantic vector of the derivative lexome (e.g., only the vector for HAPPINESS, which can be represented as $\overrightarrow{happiness}$). This vector was taken as is from the lexome-to-lexome matrix and straightforwardly entered matrix S_I for each word. This way, the vector contains only idiosyncratic information, and no information about any shared morphological category.

The Morphology Network with matrices S_M and C_M made use of the semantic vector of the content lexome of the derivative (e.g., the vector for HAPPINESS, i.e., $\overrightarrow{happiness}$) and the semantic vector of the corresponding derivational function lexome (e.g., the vector for NESS, which can be represented as $\overrightarrow{NESS}$).[2] We took both these vectors from the lexome-to-lexome matrix, and the sum of these two vectors entered matrix S_M for each word. That is, the semantic vector associated with the word *happiness* was the sum of the vectors for HAPPINESS and NESS: $\overrightarrow{happiness} + \overrightarrow{NESS}$. This way, the resulting vector contains idiosyncratic information, but also information about the morphological category it shares with other derivatives. While it is also conceivable to add to the vector of NESS the vector of HAPPY (instead of HAPPINESS), taking HAPPINESS better reflects the fact that derived words most often still carry some idiosyncratic meaning, i.e., signify more than merely the sum of their parts. The combination of HAPPINESS and NESS, thus, takes into account the morphological category NESS that the word shares with other derivatives, but still acknowledges that English derivatives are not characterized by strictly compositional semantics.

The Base Network with matrices S_B and C_B uses the semantic vectors of the content lexomes of the bases of derived words and the vectors of the derivational function lexomes. That is, instead of adding the derivational lexome vector to the lexome vector of the derivative as in the Morphology Network, in the Base Network we add the derivational lexome vector to the content lexome vector of the derivative's base. For instance, the semantic vector associated with the word *happiness* in matrix S_B is the sum of the vectors for HAPPY and NESS: $\overrightarrow{happy} + \overrightarrow{NESS}$. This way, the resulting vector contains information about the morphological category it shares with other derivatives, like in the Morphology Network. But unlike the Morphology Network, it contains no idiosyncratic information at all. The meaning of complex words in the Base Network is assumed (against our better knowledge) to be strictly compositional. In principle, this property makes this network unattractive and less suitable for predicting word durations, but it can be fruitfully used to gain further insights into the differences between architectures.

We now have three matrices (for each morphological setup, respectively) of the layout shown in **Table 2**. We have the *C* matrix, containing information about form, and the *S* matrix, containing information about meaning. These matrices can now be mapped onto each other.

Comprehension and Production Mapping

In speech comprehension, a listener encounters a form and needs to arrive at the corresponding meaning. Therefore, for comprehension we calculate a transformation matrix *F* which maps the semantic matrix *S* onto the cue matrix *C*, so that

$$CF = S. \tag{1}$$

In speech production, on the other hand, a speaker starts out with a meaning and needs to find the right form to express this meaning. Therefore, for production we calculate a transformation matrix *G* which maps the cue matrix *C* onto the semantic matrix *S*, so that

$$SG = C. \tag{2}$$

Mathematically, the transformation matrices *F* and *G* can be calculated by multiplying the generalized inverse (Moore, 1920; Penrose, 1955) of *C* with *S* (for comprehension) and the generalized inverse of *S* with *C* (for production). The transformations are visually illustrated in **Figure 1**.

As soon as we have obtained the transformation matrices, we can use them to estimate what forms and meanings the network would predict. For this, we calculate the predicted matrices $\hat{S}$ and

[1] The main diagonal of a lexome-to-lexome matrix represents the association strengths of each word to itself. Each word occurring in a sentence naturally predicts itself very well to occur in that sentence, but this value is not very informative about the word's relation to other words. Baayen et al. (2019b) therefore argue that when the researcher is interested in semantic similarity, they should replace these values with zero values.

[2] Note that the form matrices C_I, C_M, and C_B are identical, as the networks only differ in their construction of semantic vectors, not of form vectors.

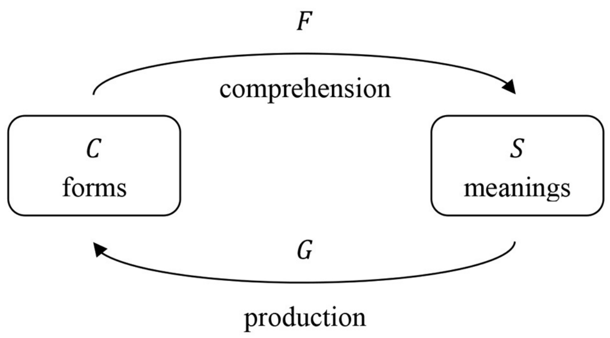

FIGURE 1 | Comprehension and production mapping, adapted from Baayen et al. (2019b). For comprehension, transformation matrix *F* transforms the cue matrix *C* into the semantic matrix *S*. For production, transformation matrix *G* transforms the semantic matrix *S* into the cue matrix *C*.

$\hat{C}$. For comprehension, we multiply the form matrix C with the transformation matrix F, i.e., we solve $\hat{S} = CF$. For production, we multiply the semantic matrix S with the transformation matrix G, i.e., we solve $\hat{C} = SG$. It is important to keep in mind that the mappings are simple linear transformations that are achieved by matrix multiplication (for an introduction in the context of LDL, see Baayen et al., 2019b). It is possible to think of the transformation matrices F and G like coefficients in linear regression, which try to approximate the target matrix but will not produce exactly the same values. This is true especially for large datasets like in the present study. The predicted matrices $\hat{S}$ and $\hat{C}$ are thus not exactly the same as the original matrices S and C.

We can also use the predicted matrices to evaluate model accuracy. To see how well the model predicts the semantics of an individual word in comprehension, we can multiply an observed form vector c from the cue matrix with the transformation matrix F to obtain a predicted semantic vector $\hat{s}$. We can then see how similar this predicted semantic vector $\hat{s}$ is to the target semantic vector s. For production, in turn, we can multiply an observed meaning vector s from the semantic matrix with the transformation matrix G to obtain the predicted form vector $\hat{c}$, which represents the estimated support for the triphones. We can then see how similar this predicted form vector $\hat{c}$ is to the target form vector c. If the correlation between the estimated vector and the targeted vector, i.e., between $\hat{s}$ and s or between $\hat{c}$ and c, respectively, is the highest among the correlations, a meaning or form is correctly recognized or produced. The overall percentage of correctly recognized meanings or forms is referred to as comprehension accuracy and production accuracy, respectively.

To obtain the mappings, we used the learn_comprehension() and learn_production() functions from the R package WpmWithLDL (Baayen et al., 2019a). Accuracy estimations were obtained with the functions accuracy_comprehension() and accuracy_production(). Finally, the measures of interest which we use to predict the durations were extracted from the networks with the help of the comprehension_measures() function and the production_measures() function. While we model word durations in the present study, which are the result of speech production, both speech production and speech comprehension mappings produce relevant measures for the analysis of production data. This is because the emergent structure of the learner's lexicon is determined both by the association of forms with meanings and of meanings with forms. In LDL, like in human learning, production and comprehension are inextricably linked to each other (see Baayen et al., 2019b for discussion). We will now describe the LDL-derived measures, as well as other used measures, in more detail.

Variables

As described above, many potentially useful LDL measures can be extracted automatically from the matrices by the package WpmWithLDL (Baayen et al., 2019a). However, some of the variables provided by this package capture similar things and are strongly correlated with each other. Careful variable selection, and sometimes adaptation, was therefore necessary. Further below we illustrate our selection and explain the conceptual dimensions we aim to capture with each variable.

Conceptually, it is desirable to not have any traditional linguistic covariates in the models that are not derived from the network, such as lexical frequencies, neighborhood densities, or bigram frequencies. It is important to build models instead which contain LDL-derived variables only. This is because, first, we are interested in how well an LDL network fares on its own in predicting speech production. Second, many traditional covariates bring along implicit assumptions that LDL does not want to make, such as the existence of discrete phonemic and morphemic units. Third, it is unclear how these traditional measures contribute to learning and processing. At the same time, however, the traditional measures might tap into properties of the linguistic signal that are picked up in a discriminative learning process. Hence, LDL measures often correlate with traditional measures.[3]

The models of interest therefore only include LDL-derived variables (described in the section LDL-Derived Predictor Variables), with one exception: the one important non-LDL variable that needs to be taken into account is SPEECH RATE. This is an influence that is beyond the control of the network.

In addition, we built models with just non-LDL variables (we describe these variables in the section Traditional Predictor Variables). This is to compare the explanatory power of the LDL-derived models with traditional models used in morpho-phonetic research.

Response Variable

Duration Difference

One important problem in analyzing spontaneous speech is that which words are spoken is uncontrolled for phonological and segmental makeup. This problem is particularly pertinent for the present study, as our datasets feature different affixes whose derivatives vary in word length. To mitigate potential durational differences that arise simply because of the number and type

[3]Correlation matrices and variable clustering trees for both LDL-derived variables and traditional variables are documented in the **Supplementary Material**.

of segments in each word, we refrained from using absolute observed duration as our response variable. Instead, we derived our duration measurement in the following way.

First, we measured the absolute acoustic duration of the word in milliseconds from the textgrid files with the help of scripts written in Python. Second, we calculated the mean duration of each segment in a large corpus (Walsh et al., 2013) and computed for each word the sum of the mean durations of its segments.[4] This sum of the mean segment durations is also known as "baseline duration," a measure which has been successfully used as a covariate in other corpus-based studies (e.g., Gahl et al., 2012; Caselli et al., 2016; Sóskuthy and Hay, 2017; Engemann and Plag, 2021). It would now be possible to subtract this baseline duration from the observed duration, giving us a new variable that represents only the difference in duration to what is expected based on segmental makeup. However, we found that this difference is not constant across longer and shorter words. Instead, the longer the word is on average, the smaller the difference between the baseline duration and the observed duration. In a third and final step, we therefore fitted a simple linear regression model predicting observed duration as a function of baseline duration. The residuals of this model represent our response variable. Using this method, we factor in the non-constant relationship between baseline duration and observed duration. We named this response variable DURATION DIFFERENCE, as it encodes the difference between the observed duration and a duration that is expected on the basis of the segmental makeup.

LDL-Derived Predictor Variables

Mean word SUPPORT

MEAN WORD SUPPORT is a measure that we introduce to capture how well-supported on average transitions from one triphone to the next are in the production of a word. Taken together, these transitions are referred to as an articulatory "path." MEAN WORD SUPPORT is calculated based on the variable PATH SUM from the package WpmWithLDL. PATH SUM refers to the summed semantic support that a given predicted articulatory path receives from its corresponding predicted semantic vector $\hat{s}$, i.e., the path from one triphone to the next in the predicted form of a word. This is illustrated in **Figure 2** with the toy example *lawless*. Each node in the path, i.e., each triphone, has a certain probability of being selected against all the other possible triphones when trying to produce a word based on its semantics. The maximum value per transition is therefore 1, i.e., a 100% probability of being selected. However, with longer words, there are also more transitions. For example, if a word's form is perfectly predicted across all triphone transitions, but there are five such transitions, PATH SUM would take the value 5. Thus, the problem with PATH SUM is that it increases not only with higher support, but also with increasing segmental length of words. This would not be ideal as a measure of semantic support when modeling durations, since durations naturally increase with longer words. The interpretation of PATH SUM as a measure for mere semantic support would be difficult. Therefore, we decided to divide each value of PATH SUM, i.e., each summed support of a word's path, by the number of path nodes in a word. This new variable MEAN WORD SUPPORT controls for path length and only reflects the average transition support in each word. MEAN WORD SUPPORT can be read as a metaphor for certainty. The higher the average transition probabilities in a word, the more certain the speaker is in pronouncing this word based on its semantics. Based on previous studies which have found higher certainty of various operationalizations to be associated with lengthening (Kuperman et al., 2007; Cohen, 2014, 2015; Tomaschek et al., 2019; Tucker et al., 2019), words with higher MEAN WORD SUPPORT can be expected to be longer in duration.

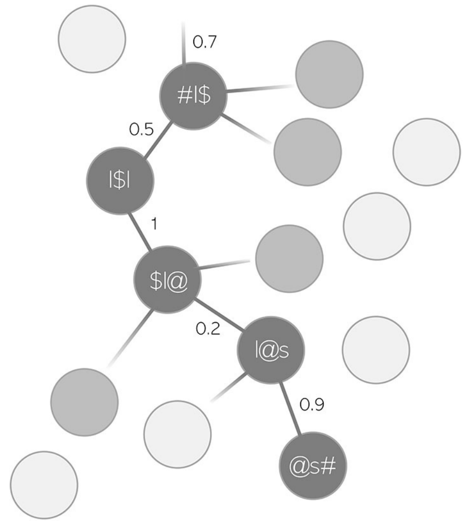

FIGURE 2 | Toy example of an articulatory path for the word *lawless*. Each connection between a triphone node is assigned a probability of being selected against other triphones.

Path Entropies

Like MEAN WORD SUPPORT, PATH ENTROPIES considers the transition probabilities between nodes in the path from one triphone to the next in the predicted form of a word. PATH ENTROPIES is the Shannon entropy calculated over the support that a given path in the predicted form vector $\hat{c}$ receives from its corresponding predicted semantic vector $\hat{s}$. Entropy is a measure of the uncertainty in the choice of one of several alternatives. Higher entropy generally means a larger number of possibilities of similar probabilities, in other words, less certainty. Similarly to MEAN WORD SUPPORT, this measure is thus related to certainty, albeit in a conceptually different way. The higher the entropy, the less certain the speaker is in producing a word, because there is not much informational value in the path support differences. Higher PATH ENTROPIES thus indicate more uncertainty. Based on the above-mentioned previous studies on certainty (Kuperman et al., 2007; Cohen, 2014, 2015; Tomaschek et al., 2019; Tucker et al., 2019), words with higher PATH ENTROPIES can thus be expected to be shorter.

[4]We used a different corpus than the Audio BNC for this task because the Audio BNC does not provide this information in an accessible and reliable form.

Semantic Vector Length

SEMANTIC VECTOR LENGTH refers to the L1 distance, also known as taxicab distance, Manhattan distance, or city-block distance, of $\hat{s}$. It thus measures the length of the predicted semantic vector by summing the vector's absolute values. We decided to use the L1 distance instead of the correlated L2 distance, as the former does not lose information by smoothing over the city-block distance. The longer the predicted semantic vector becomes, the stronger the links to other lexomes become. SEMANTIC VECTOR LENGTH can thus be understood as a measure of semantic activation diversity. It is the extent to which a given word predicts other words. As a result, it can also be understood as a measure of polysemy. The more semantic dimensions a speaker is active on for a word and the more other meanings the word can predict, the more collocational relations it has and the more varied and confusable the meanings of this word are (cf. Tucker et al., 2019, also cf. the notion of "sense uncertainty" in Filipović Durdević and Kostić, 2021). Following Tucker et al. (2019), words with higher activation diversity can be expected to be shorter: the speaker is more uncertain when more meanings are activated and therefore invests less energy in maintaining the signal.

Semantic Density

SEMANTIC DENSITY refers to the mean correlation of $\hat{s}$ with the semantic vectors of its top 8 neighbors' semantic vectors. A strong average correlation of the estimated semantic vector with the vectors of its neighbors means that the neighboring words are semantically very similar to the word in question. The higher the density, the more semantically similar these words are. SEMANTIC DENSITY applied to derived words is thus an important measure of semantic transparency: Words in a dissimilar neighborhood are idiosyncratic and their meaning is not predictable. Words in a semantically similar neighborhood are semantically transparent, i.e., mathematically shifted in the same direction. It is currently unclear whether one should expect a facilitatory or inhibitory effect of measures related to semantic transparency on duration. We explore this question in more detail in the discussion in the section Discussion and Conclusion.

Target Correlation

TARGET CORRELATION refers to the correlation between a word's predicted semantic vector $\hat{s}$ and the word's target semantic vector s. This is a measure for how accurate the network is in predicting meaning based on form. The closer the predicted meaning to the actual targeted meaning, the more successful is the model, and the better is the learner in making the correct connection between form and meaning. Being better in making the correct connection between form and meaning could be expected to have a facilitatory effect in both comprehension and production, i.e., in our case, to lead to shorter durations.

Traditional Predictor Variables

Speech Rate

SPEECH RATE is the only covariate in our LDL-derived models, and the only predictor that is not derived from the LDL networks. It is, of course, also used in the traditional models. The duration of a word is naturally influenced by how fast we speak. SPEECH RATE can be operationalized as the number of syllables a speaker produces in a given time interval (see, e.g., Pluymaekers et al., 2005b; Plag et al., 2017). In the window containing the target word plus 1 s before and 1 s after it, we divided the number of syllables by the duration of this window. This is a good compromise between a maximally local speech rate which just includes the adjacent segments, but allows the target item to have much influence, and a maximally global speech rate, which includes larger stretches of speech but is vulnerable to changing speech rates during this larger window. The number of syllables in the window and the duration of this window were extracted from the textgrids with a Python script. A higher speech rate (i.e., more syllables being produced within the window) should lead to shortening.

Word Frequency

WORD FREQUENCY has been shown to affect acoustic durations (and processing in general) in many different studies (for an overview, see, e.g., Baayen et al., 2016). Higher word frequency is expected to lead to shorter durations. We extracted the WORD FREQUENCY, i.e. the frequency of the derivative, from the *Corpus of Contemporary American English* (COCA, Davies, 2008), with the help of the corpus tool Coquery (Kunter, 2016). Derived words are often rare words (see, e.g., Plag et al., 1999). For this reason, very large corpora are necessary to obtain frequency values for derived words. We chose COCA because this corpus is much larger than the BNC, and therefore had a much higher chance of the words and their bases being sufficiently attested. We prioritized covering a bigger frequency range with more tokens. Following standard procedures, we log-transformed WORD FREQUENCY before it entered the models instead of using raw frequency. We added a constant of +1 to the variable in order to be able to take the log of the zero frequency of non-attested derivatives (cf. Howes and Solomon, 1951; Baayen, 2008).

Relative Frequency

RELATIVE FREQUENCY refers to the frequency of the base word relative to the frequency of its derivative from COCA (Davies, 2008), calculated by dividing BASE FREQUENCY by WORD FREQUENCY. It is a frequency-based measure for morphological decomposability. Morphological decomposability, or segmentability, has been found to affect duration in a number of studies (Hay, 2003, 2007; Pluymaekers et al., 2005b; Schuppler et al., 2012; Zimmerer et al., 2014; Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2018; Zuraw et al., 2020). The higher the relative frequency, the more decomposable the item is assumed to be. According to Hay (2001, 2003, 2007), more decomposable words should feature longer durations (although some studies have also found the opposite). We added a constant of +1 and log-transformed the variable.

Bigram frequency

BIGRAM FREQUENCY refers to the frequency of the target derivative occurring together with the word following it in the COCA (Davies, 2008). It has been found that the degree of acoustic reduction can be influenced by the predictability of the following context (see, e.g., Pluymaekers et al., 2005a; Bell et al., 2009; Torreira and Ernestus, 2009). It is thus expected that the higher the bigram frequency, the shorter the duration. We added a constant of +1 and log-transformed the variable.

Mean Biphone Probability

The variable BIPHONE PROBABILITY refers to the sum of all biphone probabilities (the likelihood of two phonemes occurring together in English) in a given target derivative. It has been found that segments are more likely to be reduced or deleted when they are highly probable given their context (see, e.g., Munson, 2001; Edwards et al., 2004; Turnbull, 2018; also see Hay, 2007 on transition legality effects on reduction). Thus, biphone probability can be expected to negatively correlate with duration: the more probable the biphones, the shorter the durations. Biphone probabilities were calculated by the Phonotactic Probability Calculator (Vitevitch and Luce, 2004). For this, we first manually translated the target derivatives' ASCII transcriptions of the Audio BNC into the coding referred to as Klattese, as this is the computer-readable transcription convention required by this calculator.

AFFIX

AFFIX is a categorical variable coding which affix category the derivative belongs to. This is to account for any potential idiosyncrasies in durations of affix categories.

Modeling Word Durations

Due to the distributional properties of the words in our dataset, we decided to fit both standard multiple linear regression models and mixed-effects regression models to the data. In our dataset, we have many types that are attested only once, which precludes the use of mixed-effects regression.[5] Having many single observations for one type involves the danger that certain word types may become too influential in the model. Mixed-effects regression, on the other hand, can prevent certain word types from being too influential in the model but necessitates the exclusion of items for which no repeated measurements are available. We decided to address this problem by fitting and documenting both types of model. All regression models were fitted in R (R Core Team, 2020), using the lme4 package (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2016) for the mixed models.

In the course of fitting the regression models, we trimmed the dataset by removing observations from the models whose residuals were more than 2.5 standard deviations away from the mean, which led to a satisfactory distribution of the residuals (see, e.g., Baayen and Milin, 2010). For the standard regression models, this resulted in a loss of 82 observations (1.8% of the data) for the model based on the Idiosyncratic Network, and 74 observations (1.6% of the data) for the models based on the Morphology Network and the Base Network.

[5]We provide plots illustrating the frequency distribution in our data in the **Supplementary Material**.

For the mixed models, we only included word types that occurred more than once (reducing our dataset from 363 to 261 types, or from 4,530 to 4,358 observations). The trimming procedure resulted in a loss of 71 observations (1.6% of the data) for the models based on the Idiosyncratic Network and the Base Network, and 70 observations (1.6% of the data) for the model based on the Morphology Network.

From our experience, LDL-derived variables are often strongly correlated with each other. As explained in the section Variables, we made sure to select variables that are not highly correlated and that had least conceptual overlap with each other, in terms of representing specific concepts such as certainty or semantic transparency. Still, we used variance inflation factors to test for possible multicollinearity of the remaining variables. All of the VIF values were smaller than 2, i.e., far below the critical value of 10 (Chatterjee and Hadi, 2006).

The initial models were fitted including all variables described in the section Variables. The models were then simplified according to the standard procedure of removing non-significant terms in a stepwise fashion. An interaction term or a covariate was eligible for elimination when it was non-significant at the 0.05 alpha level. Non-significant terms with the highest p-value were eliminated first, followed by terms with the next-highest p-value. This was repeated until only variables remained in the models that reached significance at the 0.05 alpha level.

RESULTS

General Comparison of the Networks

Network accuracy was generally satisfactory, with comprehension accuracies at 81, 82, and 83% for the Idiosyncratic Network, the Morphology Network, and the Base Network, respectively, and production accuracies at 99, 99, and 98%, respectively.

Before turning to the regression models that predict duration, let us compare the predicted semantic matrices $\hat{S}$ of the three networks. This can be done by calculating the correlation of each predicted semantic vector $\hat{s}$ from one network with its corresponding predicted semantic vector $\hat{s}$ from the other two networks, and then taking the mean of these correlations for all words. Comparing the semantic vectors $\hat{s_I}$ of the Idiosyncratic Network to the semantic vectors $\hat{s_M}$ from the Morphology Network, we find that they are on average very weakly correlated: the mean correlation between the vectors of the $\hat{S_I}$ matrix and the $\hat{S_M}$ matrix was $r = 0.08$. This means that the matrices are rather different. Likewise, the mean correlation between the vectors of the $\hat{S_I}$ matrix and the $\hat{S_B}$ matrix is weak ($r = 0.1$).

However, the mean correlation between the vectors of the $\hat{S_M}$ matrix and the $\hat{S_B}$ matrix is extremely high ($r = 0.9$). This indicates that it is probably the information about derivational function that differentiates the semantic vectors of the Idiosyncratic Network from the semantic vectors of the other two networks. Morphological category matters.

TABLE 3 | Final standard linear regression models reporting effects on duration difference with variables from the three networks.

	Idiosyncratic Network model			Morphology Network model			Base Network model		
	Estimate	SE		Estimate	SE		Estimate	SE	
Intercept	0.216901	0.026210	***	0.090708	0.025887	***	0.408246	0.029999	***
MEAN WORD SUPPORT	0.170726	0.023507	***	0.250262	0.020700	***	0.050723	0.012716	***
PATH ENTROPIES	−0.008688	0.002242	***	−0.008442	0.002309	***	−0.009342	0.002259	***
SEMANTIC DENSITY	−0.043545	0.008925	***	0.033868	0.012372	**	−0.093906	0.025844	***
SPEECH RATE	−0.058757	0.001148	***	−0.058602	0.001159	***	−0.058702	0.001171	***
N	4,448			4,456			4,456		
R^2 *adjusted*	0.3778			0.3742			0.3623		

Full models are documented in the ***Supplementary Material****. Significance codes: *** < 0.001, ** < 0.01, * < 0.05.*

TABLE 4 | Final mixed-effects regression models reporting effects on duration difference with variables from the three networks.

	Idiosyncratic Network model			Morphology Network model			Base Network model		
	Estimate	SE		Estimate	SE		Estimate	SE	
Intercept	1.328e-01	4.601e-02	**	2.146e-01	6.024e-02	***	2.595e-01	2.510e-02	***
MEAN WORD SUPPORT	2.722e-01	4.600e-02	***	2.535e-01	4.572e-02	***	1.211e-01	2.654e-02	***
PATH ENTROPIES	−1.173e-02	5.625e-03	*	−1.163e-02	5.633e-03	*			
SEMANTIC VECTOR LENGTH	−1.606e-02	6.860e-03	*	−3.294e-02	1.550e-02	*			
SPEECH RATE	−5.944e-02	1.116e-03	***	−5.937e-02	1.116e-03	***	−5.936e-02	1.117e-03	***
N	4,357			4,358			4,357		
R^2 *marginal*	0.3690016			0.3638608			0.3487138		
R^2 *conditional*	0.5198377			0.5168201			0.5200542		

Full models are documented in the ***Supplementary Material****. Significance codes: *** < 0.001, ** < 0.01, * < 0.05.*

Predicting Durations With LDL Variables

Let us now turn to the regression models predicting duration. **Tables 3**, **4** report the final models regressing duration difference against the LDL-derived variables and SPEECH RATE.

The model in **Table 3** reports the results of the standard regression models. As we can see, of the LDL-derived variables, MEAN WORD SUPPORT, SEMANTIC DENSITY, and PATH ENTROPIES significantly affect duration in the regression models of all three networks. In addition, SPEECH RATE is significant in all three models. The variables SEMANTIC VECTOR LENGTH and TARGET CORRELATION, on the other hand, did not reach significance and were therefore excluded from these final models.

The model in **Table 4** reports the results of the mixed models. These models are very similar to the standard regression models, with two important differences. The variables MEAN WORD SUPPORT and SPEECH RATE display the same effects as in the standard models. PATH ENTROPIES also displays the same effects for the Idiosyncratic Network and the Morphology Network (it was only marginally significant for the Base Network and therefore excluded). However, SEMANTIC DENSITY does not reach significance in the mixed models. Instead, there is a significant effect of SEMANTIC VECTOR LENGTH in the models derived from the Idiosyncratic Network and the Morphology Network, but not in the Base Network.

Before taking a look at the effects of individual variables, let us first examine how much variation is actually explained by the models. **Tables 3**, **4** show that for all three networks in both types of model, the R^2 of the fixed effects is between 0.36 and 0.37, i.e., about 36–37% of the variance in duration is explained by the predictors (the marginal R^2 of the mixed model for the Base Network is an exception, being slightly lower with about 35%). To put this number into perspective, we compared the explained variance of the LDL-derived models to that of a model containing predictor variables that are traditionally used in morpho-phonetic corpus studies of duration. We fitted a standard linear regression model and a mixed model including the traditional predictors from the section Traditional Predictor Variables. These variables were fitted to the response variable DURATION DIFFERENCE. Some observations were lost due to the same trimming procedure as explained in the section Modeling Word Durations (80 observations, or 1.8% of the data, for the standard model, and 74 observations, or 1.7% of the data, for the mixed model). For the sake of comparison of the explanatory power of individual predictors, we did not remove insignificant variables from the models. The models are summarized in **Table 5**. WORD FREQUENCY, RELATIVE FREQUENCY, and BIGRAM FREQUENCY were not significant in the models, while MEAN BIPHONE PROBABILITY, some levels of AFFIX, and SPEECH RATE were. We can see that about the same

TABLE 5 | Standard linear regression model and mixed-effects regression model reporting effects on duration difference with traditional, non-LDL predictors.

	Traditional standard regression model			Traditional mixed-effects model		
	Estimate	SE		Estimate	SE	
Intercept	3.888e-01	8.345e-03	***	4.159e-01	1.106e-02	***
WORD FREQUENCY	4.970e-08	3.764e-08		−2.608e-07	2.328e-07	
RELATIVE FREQUENCY	−2.136e-05	4.166e-05		−1.446e-05	8.931e-05	
BIGRAM FREQUENCY	−6.542e-07	6.293e-07		7.978e-07	6.382e-07	
MEAN BIPHONE PROBABILITY	−5.188e+00	8.872e-01	***	−7.167e+00	1.545e+00	***
AFFIX ATION						
DIS	8.145e-03	6.700e-03		−1.405e-03	1.438e-02	
IZE	−2.316e-02	5.251e-03	***	−1.491e-02	1.377e-02	
LESS	−5.749e-02	8.226e-03	***	−7.569e-02	1.524e-02	***
NESS	−5.473e-02	5.700e-03	***	−3.630e-02	1.295e-02	**
SPEECH RATE	−5.893e-02	1.163e-03	***	−5.986e-02	1.116e-03	***
N	4,450			4,354		
R^2 *adjusted/marginal*	0.3731			0.3705799		
R^2 *conditional*				0.5344904		

Full models and ANOVAs are documented in the ***Supplementary Material****. Significance codes: *** < 0.001, ** < 0.01, * < 0.05.*

TABLE 6 | Relative importance of variables in the models for the overall explained variance (marginal variance for mixed models).

	Relative importance metrics (lmg)							
	Idiosyncratic Network		Morphology Network		Base Network		Traditional model	
	lm	*lmer*	*lm*	*lmer*	*lm*	*lmer*	*lm*	*lmer*
MEAN WORD SUPPORT	0.0089	0.1649	0.0148	0.0956	0.0025	0.1641		
PATH ENTROPIES	0.0023	0.0031	0.0023	0.0017	0.0030			
SEMANTIC DENSITY	0.0067		0.0020		0.0014			
SEMANTIC VECTOR LENGTH		0.0064		0.0399				
SPEECH RATE	0.3605	0.1946	0.3556	0.2266	0.3559	0.1845	0.3561	0.2140
WORD FREQUENCY							0.0007	0.0065
RELATIVE FREQUENCY							0.0006	0.0044
BIGRAM FREQUENCY							0.0007	0.0034
MEAN BIPHONE PROBABILITY							0.0025	0.1178
AFFIX							0.0136	0.0246
Total variance explained	0.3778	0.3690	0.3742	0.3639	0.3623	0.3487	0.3731	0.3706

proportion of the variance is explained by the traditional models ($R^2 = 0.37$).

Partitioning how much each of the predictors contributes to the proportion of explained variance, using the lmg metric (Lindeman et al., 1980) from the relaimpo package (Grömping, 2006) and the calc.relip.mm function (Beresewicz, 2015) reveals that in both the traditional models and the LDL models, by far most of the variance is explained by speech rate (which alone explains about 35% of the total variance in the standard regression models and about 20% in the mixed models). This is shown in **Table 6**. The variables of interest MEAN WORD SUPPORT, PATH ENTROPIES, SEMANTIC DENSITY, and SEMANTIC VECTOR LENGTH are all comparable in their explanatory power to the categorical AFFIX variable and MEAN BIPHONE PROBABILITY, and often better than the three frequency measures WORD FREQUENCY, RELATIVE FREQUENCY, and BIGRAM FREQUENCY. While the small differences in the explained variance between the LDL-derived variables and the traditional variables after factoring out the contribution of SPEECH RATE are not large enough to truly say which set of variables is "better," they clearly show that they are in the same ballpark. We can thus say that LDL-derived variables can compete against traditional variables from morphophonetic studies.

We can now take a closer look at the effects of each of the variables. **Figure 3** (for the standard regression models) and **Figure 4** (for the mixed models) plot the effects of the LDL-derived variables and SPEECH RATE on duration. **Figure 5**

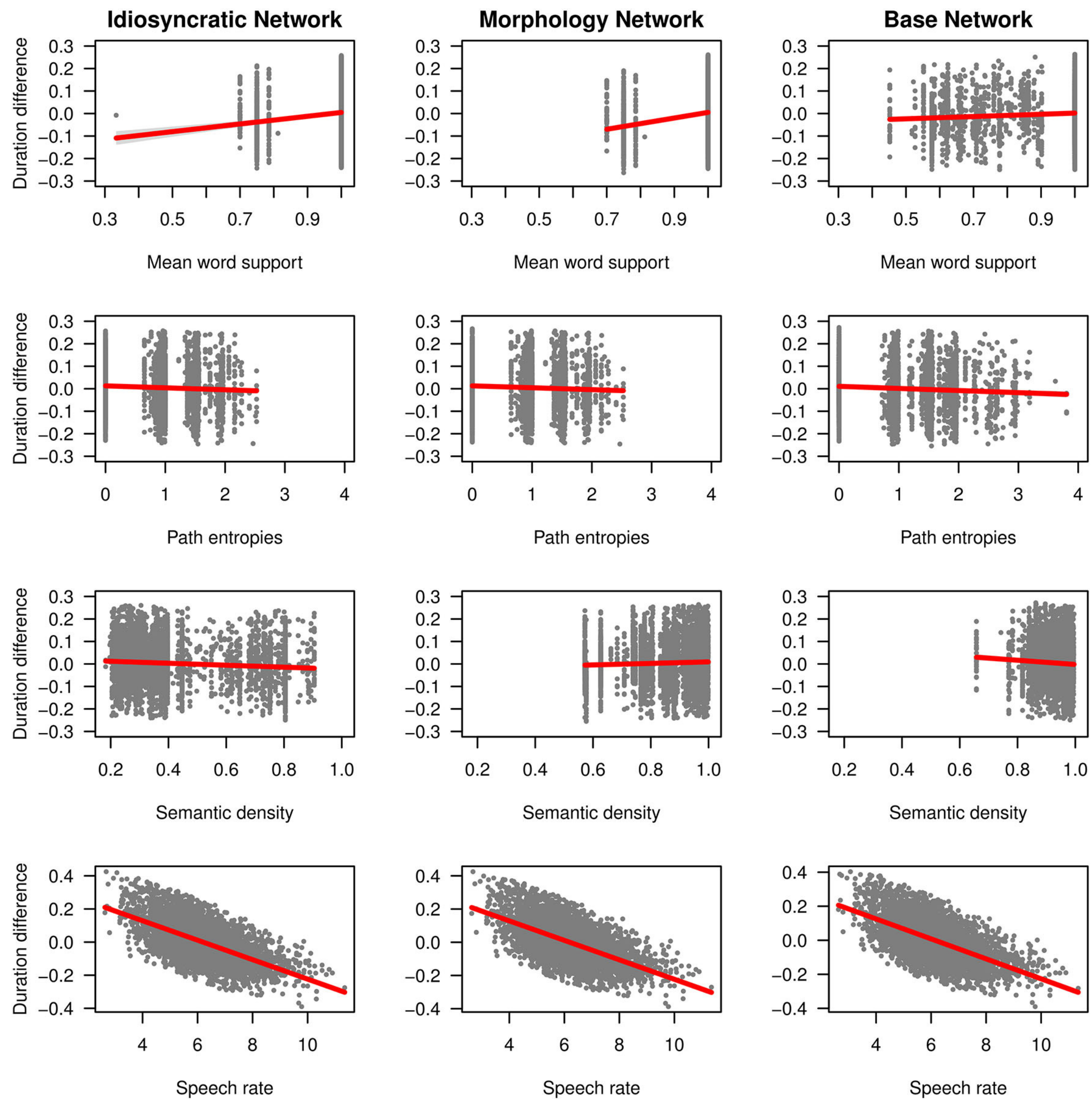

FIGURE 3 | Effects on duration difference in the standard linear regression models for the Idiosyncratic Network variables (left column), the Morphology Network variables (middle column) and the Base Network variables (right column).

displays the density distributions of the variables in all three networks. We will discuss the two variables relating to certainty in the articulatory path first (MEAN WORD SUPPORT and PATH ENTROPIES), followed by a discussion of the two variables relating to the semantic relations between words (SEMANTIC DENSITY and SEMANTIC VECTOR LENGTH). The covariate SPEECH RATE and the variable TARGET CORRELATION will not be further discussed, as SPEECH RATE behaves as expected (see the bottom rows of **Figures 3**, **4**) and TARGET CORRELATION was not significant in any of the models.

Mean Word Support and Path Entropies

As explained in the section LDL-Derived Predictor Variables, the two variables MEAN WORD SUPPORT and PATH ENTROPIES both reflect properties of the semantic support for the predicted articulatory path, and they both tap into articulatory certainty. Given that the way these variables are calculated, MEAN WORD SUPPORT is a measure of certainty, while PATH ENTROPIES is a measure of uncertainty, they should mirror each other by showing opposite effects on duration. We find that this is the case.

Let us start with MEAN WORD SUPPORT. This variable has a significant effect on duration difference in all models. We can see from the coefficients in **Tables 3**, **4** as well as from its positive slope in the top row of **Figures 3**, **4** that higher MEAN WORD SUPPORT is significantly associated with longer durations. The higher the average semantic support of a word's predicted triphone path, the longer this word is pronounced. This means that the more certain the speaker is in producing

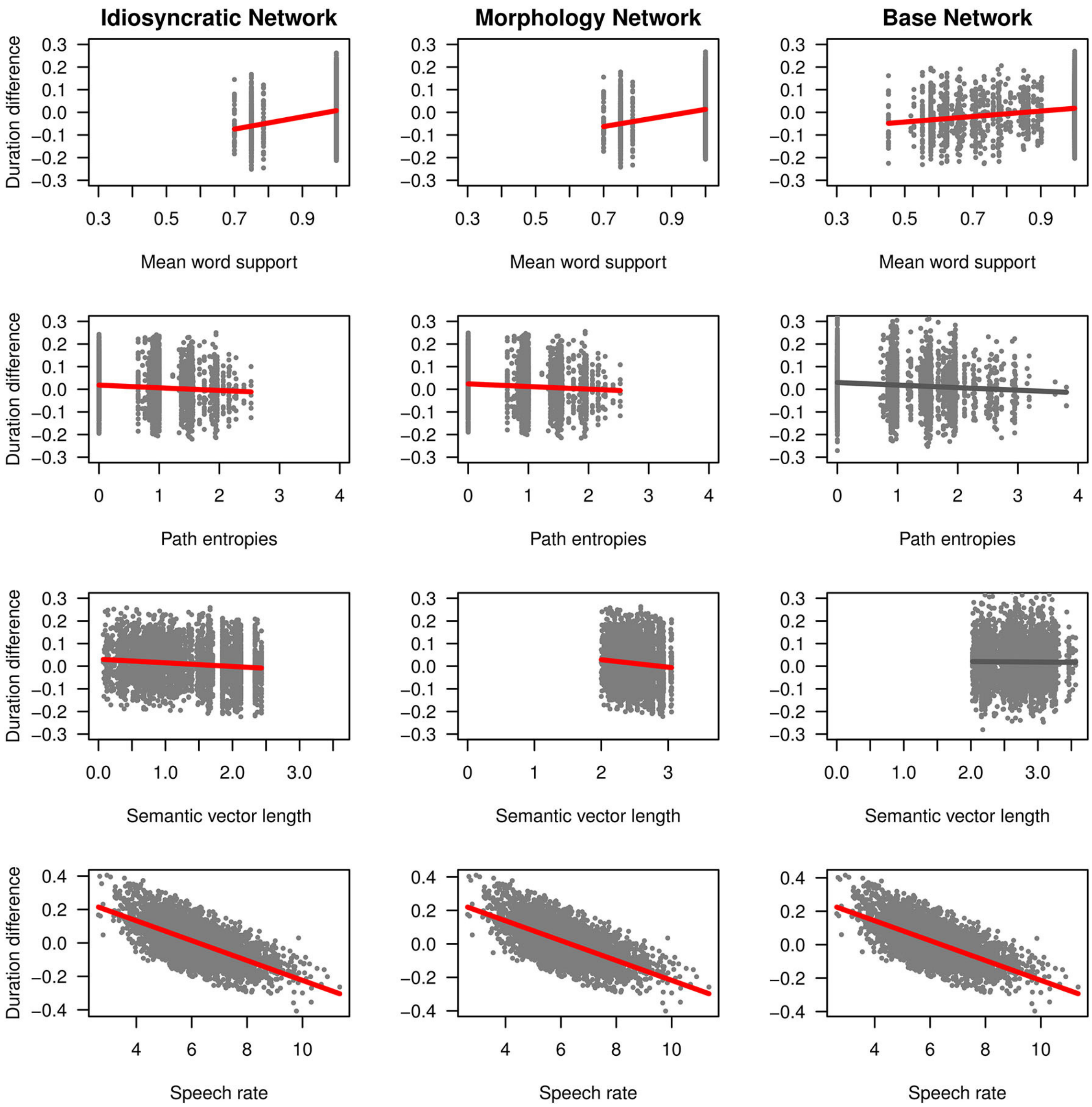

FIGURE 4 | Effects on duration difference in the mixed-effects regression models for the Idiosyncratic Network variables (left column), the Morphology Network variables (middle column) and the Base Network variables (right column). Red regression lines indicate significant effects from the final models, gray regression lines indicate non-significant effects from the initial models before the non-significant predictors were excluded.

the word, the more the articulation is durationally enhanced. In other words, more certainty is associated with lengthening. Interestingly, if we look at the distribution of MEAN WORD SUPPORT in the top row of **Figure 5**, we can see that mainly two derivational functions are responsible for this effect: Whereas the paths of IZE and ATION words are always very well-supported (as well as the paths of DIS in the Idiosyncratic Network and in the Morphology Network), paths of NESS and LESS words often feature weaker transition probabilities between triphones. The distributional differences of each of these two categories compared to the others are significant (Mann-Whitney, $p < 0.001$). This is true for all three networks. However, it is notable that the mean support of words is generally lower in the Base Network, especially for IZE, NESS, and LESS words. We will come back to these differences between morphological categories and between networks in the discussion.

If MEAN WORD SUPPORT indicates that with greater certainty, durations become longer, our next predictor PATH ENTROPIES should indicate that with greater uncertainty, durations become shorter. This is the case. Moving on to the second row in **Figures 3**, **4**, we can observe negative slopes for the effect of PATH ENTROPIES, which was significant in the models (marginally significant in the mixed model for the Base Network). The higher the Shannon entropy of the semantic support for the

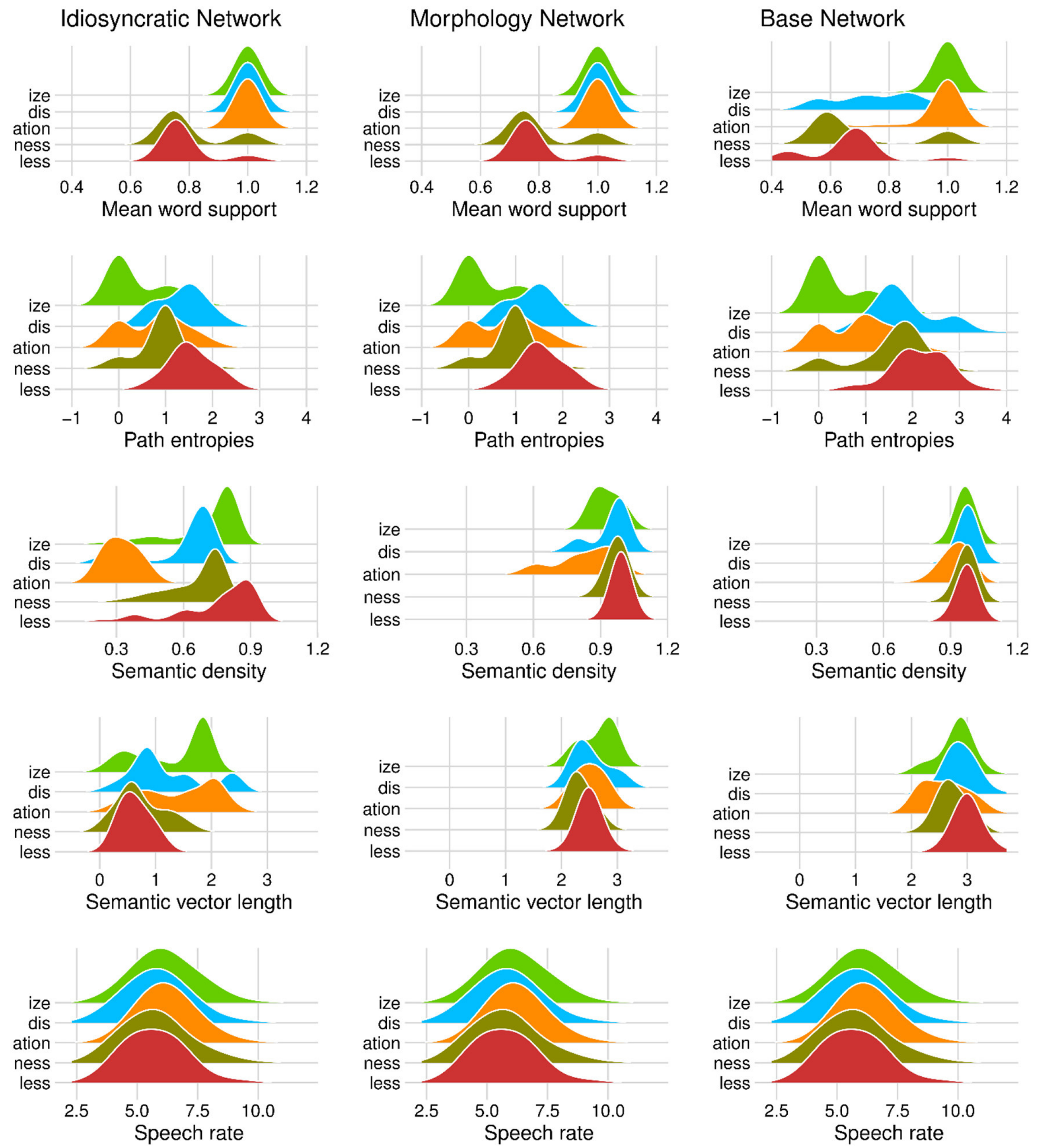

FIGURE 5 | Density distributions of variables by derivational function in the Idiosyncratic Network models (left column), the Morphology Network models (middle column), and the Base Network models (right column). Note that in the first two panels in the top tow, the density curves around 1.0 are calculated over a single value.

predicted articulatory paths becomes, i.e., the more variation of support there is in the system, the shorter the durations are. More uncertainty is associated with reduction. In other words, a speaker's lower certainty in production means the articulatory signal is less strengthened or less enhanced. Again, there are differences between morphological categories in all three networks. For example, words with IZE are characterized by more diverse and informative support values, while the other categories often feature more entropic supports across the paths, especially LESS and DIS. All differences in the distributions are significant at $p < 0.001$, except for the non-significant difference between LESS and DIS in the Idiosyncratic Network and the Morphology Network, and the difference between NESS and DIS in the Base Network.

Semantic Density and Semantic Vector Length

Let us now look at the two variables that capture the semantic relations to other words, SEMANTIC DENSITY and SEMANTIC VECTOR LENGTH.

SEMANTIC DENSITY is significant in the standard regression models, but did not reach significance in the mixed models. Its coefficients in **Table 3** show that while it has a negative effect on duration when derived from the Idiosyncratic Network and the Base Network, it has a positive effect on duration when derived from the Morphology Network. This is illustrated in the third row of **Figure 3**. For the Idiosyncratic Network and the Base Network, the stronger an estimated semantic vector correlates with its neighbors, the shorter becomes the duration of a word. For the Morphology Network, the stronger an estimated semantic vector correlates with the semantic vectors of its neighbors, the longer becomes the duration of a word. High-density words live in a space more semantically close to other words, i.e., they can be said to be less idiosyncratic and, due to their being derived words, more semantically transparent. Higher transparency can thus lead to both lengthening and shortening, depending on how the network is constructed.

Investigating the distribution of this variable, we observe that SEMANTIC DENSITY shows differences between the networks. The data points in **Figure 3** and the distributions in **Figure 4** show that density is lowest in the Idiosyncratic Network, higher in the Morphology Network, and highest in the Base Network. This means that density increases with the amount of morphological structure we encode in the networks. SEMANTIC DENSITY also shows differences between derivational functions. Especially in the Idiosyncratic Network, this difference is very pronounced. This is again illustrated in **Figure 5** (third row, first column). Words with LESS and IZE have particularly high densities, whereas densities are lower for DIS and NESS words, and lowest for ATION words. All of the distributions are significantly different from each other at $p < 0.001$. The fact that these morphological categories cluster so distinctly is particularly surprising, given that the Idiosyncratic Network was not provided with any information about these categories. We will return to the peculiar behavior of this variable in the discussion.

Turning to the second semantic variable, we can see that SEMANTIC DENSITY is replaced by SEMANTIC VECTOR LENGTH in the mixed models: SEMANTIC VECTOR LENGTH, while not significant in the standard regression models, reaches significance in the mixed models for the Idiosyncratic Network and the Morphology Network (**Table 4** and third row in **Figure 4**). When derived from these networks, SEMANTIC VECTOR LENGTH has a negative effect on duration. Recalling that this variable captures activation diversity, we can say that being active on more semantic dimensions as a speaker has a facilitatory effect in production. The more collocational relations a word has to other words and the more meanings are activated, the shorter it is pronounced.

Investigating the distribution of SEMANTIC VECTOR LENGTH (**Figure 5**, fourth row), we observe that the estimated semantic vectors are generally longer in the Morphology Network and the Base Network than in the Idiosyncratic Network. Not only are they longer on average, they also cluster more closely together in terms of their length: the L1 distance in the Morphology Network and the Base Network covers a range from about 2 to 3, while in the Idiosyncratic Network, it is spread out across a range from about 0 to 2.5. One reason for this may be purely mathematical: The vectors in the two networks with information about the morphological category can often be longer because the vector for the derivational function lexome is added to the vector of the derived word's content lexome. However, the vectors are not just generally longer in these networks, but the spread of the datapoints is also narrower. This indicates that the words cluster more closely together. Since SEMANTIC VECTOR LENGTH can represent activation diversity, this is expected: If words share a morphological function with other words, they become more similar to each other, hence are more likely to be semantically active when a member of their category is accessed. In the Idiosyncratic Network, words do not explicitly share a morphological category, hence members of a given category are not as likely to be co-activated. Again, the distributions show that vector lengths cluster differently depending on derivational function, meaning that different morphological categories are characterized by different degrees of semantic activation diversity.

It is interesting to note that when modeling durations, it is the Base Network that seems to behave differently from the other two networks, even though it shares with the Morphology Network its property of having information about morphological categories. The mixed model based on the variables from the Base Network is the least successful, as two predictors that are significant in the other networks (PATH ENTROPIES and SEMANTIC VECTOR LENGTH) do not reach significance in the Base Network. In the section Matrices for Form and Meaning, we have already discussed that the Base Network is conceptually unappealing and theoretically flawed, as it wrongly assumes that the meaning of a derived word is strictly composed of the meaning of its base word and the meaning of the affix. However, we now find that it also seems to perform less optimal in modeling durations. Importantly, it is surprising that the Base Network shows a facilitatory effect of SEMANTIC DENSITY similar to the Idiosyncratic Network, instead of behaving like the Morphology Network, i.e., showing an inhibitory effect. This is despite the fact that the distribution of SEMANTIC DENSITY is very similar in the Base Network and in the Morphology Network, but very different in the Idiosyncratic Network (see again **Figure 5**, third row). Moreover, it was the $\hat{S_M}$ matrix and the $\hat{S_B}$ matrix which are extremely highly correlated with each other (see the section General Comparison of the Networks) and not at all correlated with the $\hat{S_I}$ matrix.

Exploring the aberrant behavior of the Base Network further, we investigated the semantic space of the Base Network in more detail and found that the clustering of words in the semantic space is detrimental. This is exemplified in **Table 7**, which shows an extract from the list of closest semantic neighbors to words with DIS in the three networks. Quite expectedly, the Idiosyncratic Network features a lower number of DIS words as neighbors of target DIS words than the other two networks. And there are more neighbors featuring DIS in the Base Network than in the Morphology Network. This increase of the number of DIS words as neighbors across the three networks mirrors the increasing role of explicit morphological information encoded in these networks. There is an important

TABLE 7 | Extract from the closest semantic neighbors of DIS words in the three networks.

Word	Phones	Neighbors						
Idiosyncratic network								
disarm	dIs,m	m1d1	kInt	w{m	m{mb5	kr{Nkl	n5zl	bl{Nkll
disband	dIsb{nd	m1d1	kInt	bl{Nkll	w{m	m{mb5	kr{Nkl	pIpIn
discard	dIsk,d	dIs@r1	dIst1st	dIskrEdIt	dIsgr1s	dIskVmf@t	$l	dIs@b1
discharge	dIsJ,=	dIsl2k	dIsQnIst	dIstrVst	dIs@gri	dIskVmf@t	dIsgr1s	dIsk@ntEnt
disclose	dIskl5z	m1d1	kInt	m{mb5	w{m	bl{Nkll	n5zl	SIt
discount	dIsk6nt	dIsQnIst	dIskVmf@t	dIsgr1s	dIsk@ntEnt	dIstrVst	dIst1st	dIsg2z
discourse	dIsk$s	dIs@r1	dIst1st	dIskrEdIt	dIsgr1s	dIskVmf@t	dIsp{r@tI	dIslQ=
disease	dIziz	dIskVv@R	dIs@p7R	dIs$d@R	dIsJ,=	dIsl2k	dIsk6nt	dIs@gri
disgrace	dIsgr1s	dIst1st	dIskVmf@t	dIs@r1	dIskrEdIt	dIs@b1	dIslQ=	dIsp{r@tI
disguise	dIsg2z	dIskVmf@t	dIsgr1s	dIst1st	dIs@r1	dIsQnIst	dIsk@ntEnt	dIskrEdIt
dislike	dIsl2k	dIsQnIst	dIskVmf@t	dIsgr1s	dIstrVst	dIsk@ntEnt	dIst1st	dIsg2z
Morphology network								
disarm	dIs,m	dIsjun@tI	dIs5n	dIsb{nd	dIs@r1	dIskrEdIt	dIsp{r@tI	dIs@b1
disband	dIsb{nd	dIsjun@tI	dIs5n	dIs,m	dIs@r1	dIskrEdIt	dIs@b1	dIsp{r@tI
discard	dIsk,d	dIskVmf@t	dIsgr1s	dIst1st	dIsQnIst	dIs@r1	dIsk@ntEnt	dIslQ=
discharge	dIsJ,=	dIsl2k	dIsQnIst	dIstrVst	dIs@gri	dIskVmf@t	dIsgr1s	dIsk@ntEnt
disclose	dIskl5z	dIs@r1	dIs5n	dIs,m	dIskrEdIt	dIsjun@tI	dIsb{nd	dIsp{r@tI
discount	dIsk6nt	dIskVmf@t	dIsQnIst	dIsgr1s	dIsl2k	dIs@gri	dIstrVst	dIsg2z
discourse	dIsk$s	dIskVmf@t	dIsgr1s	dIst1st	dIsQnIst	dIsk@ntEnt	dIs@r1	dIsrIg,d
disease	dIziz	dIskVv@R	dIs@p7R	dIs$d@R	dIsJ,=	dIsl2k	dIsk6nt	dIs@gri
disgrace	dIsgr1s	dIst1st	dIskVmf@t	dIs@r1	dIskrEdIt	dIs@b1	dIslQ=	dIsp{r@tI
disguise	dIsg2z	dIskVmf@t	dIsgr1s	dIst1st	dIs@r1	dIsQnIst	dIsk@ntEnt	dIskrEdIt
dislike	dIsl2k	dIskVmf@t	dIsQnIst	dIsgr1s	dIstrVst	dIs@gri	dIsk@ntEnt	dIsg2z
Base network								
disarm	dIs,m	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIsl2k	dIs@bidj@ns	dIspl1s
disband	dIsb{nd	dIsg2z	dIsp{r@tI	dIs@r1	dIsgVst	dIsl2k	dIspl1s	dIs@bidj@ns
discard	dIsk,d	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIs@bidj@ns	dIsl2k	dIspl1s
discharge	dIsJ,=	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIsl2k	dIs@bidj@ns	dIsQnIst
disclose	dIskl5z	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIsl2k	dIs@bidj@ns	dIslQ=
discount	dIsk6nt	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIs@bidj@ns	dIsl2k	dIsQnIst
discourse	dIsk$s	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIspl1s	dIs@bidj@ns	dIsQnIst
disease	dIziz	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIs@bidj@ns	dIsl2k	dIsQnIst
disgrace	dIsgr1s	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIsl2k	dIs@bidj@ns	dIsQnIst
disguise	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIs@bidj@ns	dIsl2k	dIsQnIst	dIslQ=
dislike	dIsl2k	dIsg2z	dIsp{r@tI	dIsgVst	dIs@r1	dIs@bidj@ns	dIslQ=	dIsQnIst

difference, however, between the Morphology Network and the Base Network. While in the Morphology Network, the DIS neighbors consist of many different word types with DIS, in the Base Network these are very often exactly the same word types. A type analysis of the neighbors for all morphological categories in the three networks confirms this impression: **Figure 6** shows that the Base Network is characterized by the least diverse neighbor space of the three networks, and that this is true for every investigated morphological function. Given this behavior, it is thus no longer surprising that measures derived from the Base Network might behave strangely or not display effects. We conclude that the Base Network is not only theoretically the least appealing of the three networks, but that these problems also lead to an empirically unattractive model.

DISCUSSION AND CONCLUSION

This study set out to explore how morphological effects on the phonetic output, which have been frequently observed in the literature, can be explained. From the perspective of current speech production models and theories of the morphology-phonology interaction, such effects are unexpected, and the mechanisms behind them are unclear. Our study investigated whether we can successfully model the durations of English derivatives with a new psycho-computational approach, linear discriminative learning. We hypothesized that measures derived from an LDL network are predictive of duration. We also explored what insight their effects can give us into the mechanisms of speech production, and whether the measures derived from these networks differ in their predictive power

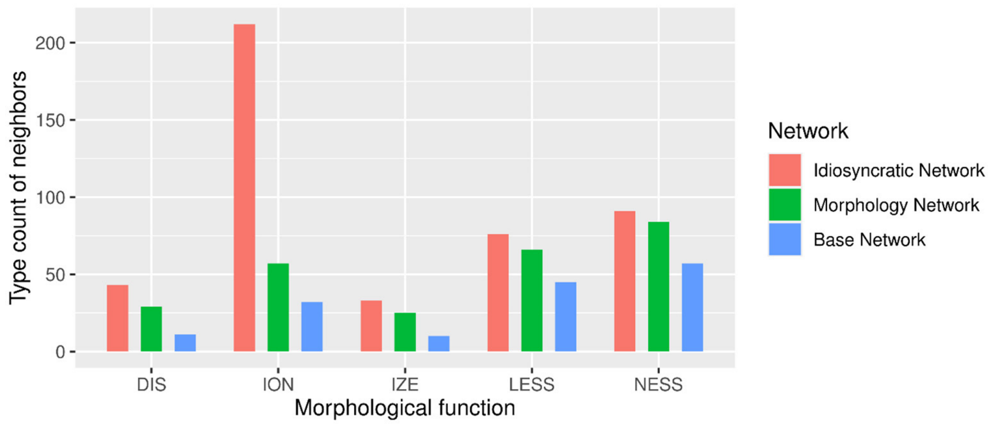

FIGURE 6 | Type count of top 8 neighbors by network and morphological function.

depending on the kind of information they have about morphological functions.

Our study demonstrated that LDL-derived variables can successfully predict derivative durations. The mean semantic support of a word's articulatory path, the entropy of a word's path supports, the mean correlation of a word's predicted semantics with the semantics of its neighbors, and the distance of the semantic vector in the semantic space all significantly affect duration. We have also shown that these measures explain a reasonable proportion of the durational variance, in the sense that their contribution to the explained variance is comparable to the contribution of traditional linguistic variables used in corpus studies of duration. The present study thus contributes to the growing literature that demonstrates that LDL is a promising alternative approach to speech production which can explain the variation in fine phonetic detail we find in different kinds of words, be they simplex, complex, or non-words (cf. Baayen et al., 2019b; Chuang et al., 2020).

Regarding the question what the effects of LDL-derived variables can tell us about speech production, we find that two important concepts relevant for production are the certainty in the association of form with meaning and the semantic relations of words to other words. The positive effects of MEAN WORD SUPPORT and the negative effects of PATH ENTROPIES on duration both indicate that generally, higher certainty in the association of form and meaning is associated with longer durations. The better an articulatory path is on average semantically supported, and the less these supports vary over the path, the more strengthened the articulation becomes. It is important to note that the metaphor of "certainty" which is ascribed to these measures can generate two opposing expectations, both of which are intuitive in their own way. On the one hand, it could be assumed that the more certain a speaker is, the more strengthened the signal will be, leading to longer durations. This may be because a speaker invests more energy in maintaining duration when they are certain, and less energy when they are uncertain, in order to not prolong a state of uncertainty (Tucker et al., 2019). On the other hand, it could be assumed that the more certain a speaker is, the more efficient they can articulate, leading to shorter durations. This may be because more certainty could enable a speaker to select the correct path more quickly. The present results provide support for the first interpretation rather than the second one.

This interpretation is in line with the findings for other measures that have been interpreted with reference to the concept of certainty. Tomaschek et al. (2019), for instance, found that with higher functional certainty, gauged by the support for a word's inflectional lexome and the word's overall baseline support, segment durations of different types of English final S are lengthened. Kuperman et al. (2007) found that with higher certainty, gauged by the paradigmatic support (probability) of Dutch compound interfixes, these interfixes are realized longer. Cohen (2014) found that higher certainty, gauged by the paradigmatic probability of English suffixes, is associated with phonetic enhancement, i.e., again with longer durations. Cohen (2015) found that higher paradigmatic support can also enhance Russian vowels. Tucker et al. (2019) found that with higher support for tense and regularity (more certainty), acoustic duration of stem vowels increases, and with greater activation diversity (more uncertainty), acoustic duration decreases. In sum, regarding the question whether certainty has an effect of enhancement or reduction, recent evidence—including the present study—points toward enhancement.

The significant effects of SEMANTIC DENSITY and of SEMANTIC VECTOR LENGTH indicate that a second relevant factor in the production of derivatives is the semantic relation of a word to other words. Starting with SEMANTIC DENSITY, depending on the architecture of the network, the average semantic similarity of a word's neighbors to this word can lead to both longer and shorter durations. If the network has information about the semantics of the morphological category of the derivative and of the derivative itself, higher densities are associated with longer durations. If the network has no such information and treats all words as idiosyncratic, or if the network has information about the morphological category and the semantics of the derivative's base word, higher densities are associated with shorter durations. In order to get a better understanding of this somewhat puzzling finding, three observations are helpful.

Let us first compare the Idiosyncratic Network and the Morphology Network. We can see in the data points in **Figure 3** as well as in the density plots in **Figure 5** that SEMANTIC DENSITY is distributed very differently when derived from the Idiosyncratic Network than when derived from the Morphology Network (both the model results as well as the distributions are plotted on the same x-axis scale, respectively, for easier comparison). For the Idiosyncratic Network, there are hardly any data points above 0.8 and the vast majority of data points have density values below 0.4. For the Morphology Network, on the other hand, the vast majority of data points show densities above 0.8. At the conceptual level this makes sense: We would expect words sharing the semantics of their morphological category to be closer to their neighboring words, i.e., to be more transparent and less idiosyncratic. This means that if the model has information about morphological categories, density should be generally higher. This is the case. In contrast, words in the Idiosyncratic Network are generally more dissimilar to each other because they do not share the semantic information that comes with belonging to a particular morphological category. This difference between the two networks is also illustrated by the example of DIS neighbors in **Table 7**, which shows that in the Morphology Network a larger proportion of nearest neighbors comes from the morphological category of the target word.

Returning to the relation between SEMANTIC DENSITY and duration, we can now see in **Figure 3** that the contradictory effects happen at different ends of the distribution. The negative effect found in the Idiosyncratic Network is carried by the low-density words, while the positive effect of semantic density on duration is carried by the high-density words. The positive effect of densities above 0.8 is even visible in the Idiosyncratic Network: the residuals in that range are clearly skewed toward higher durations. If we attempt an interpretation of the relation of SEMANTIC DENSITY and word duration across these two networks, we can say that the shortest durations are found in the middle of the semantic density range. Having many close semantic relatives slows down articulation, and so does having very few relatives.

What about SEMANTIC DENSITY in the Base Network? SEMANTIC DENSITY in this network is distributed similarly to the Morphology Network, yet the effect is similar to the Idiosyncratic Network, as it negatively affects duration. However, our exploration of the type diversity in the semantic space of the networks in the section Semantic Density and Semantic Vector Length has shown that the neighbors that are behind these semantic densities are not at all diverse in the Base Network. This was true to such an extent that for the DIS words, for example, the Base Network considered the same few words (especially *disguise, disparity, disgust, disarray*) to be the closest neighbors for the vast majority of the target words. We consider this clustering to be rather unrealistic. Most likely, it is the consequence of the questionable premises underlying this network architecture discussed earlier. Overall, we conclude that the effect of SEMANTIC DENSITY in this network is not interpretable.

The question remains how we can understand the opposite effects of SEMANTIC DENSITY in the Idiosyncratic Network and the Morphology Network. If our interpretation that SEMANTIC DENSITY captures semantic transparency is correct, we would expect higher densities to lead to longer durations. More transparent words should be more protected against phonetic reduction because they feature a stronger morphological "boundary," i.e., they are more decomposable. Such lengthening effects induced by supposed morphological boundaries have been observed in several studies (e.g., Hay, 2001, 2003, 2007; Plag and Ben Hedia, 2018). If we assume that the theoretical concept of a morphological boundary and the similarity of a word to its neighboring words capture the same underlying dimension of semantic transparency, we should still be able to replicate this effect. However, it is not entirely clear why a higher degree of semantic transparency would lead to lengthening. Given that a higher semantic transparency means that more words will be more strongly activated, we would rather expect durations to shorten. This is because semantic activation diversity has been found to be associated with reduction (Tucker et al., 2019). This reduction in speech production is mirrored in reaction time experiments that have found shorter reaction times with larger morphological family sizes (Schreuder and Baayen, 1997; Bertram et al., 2000). This family size effect has been interpreted as a semantic effect arising through activation spreading between morphologically related words. Interestingly, in the study by Bertram et al. (2000), the effect was restricted to transparent family members. This is an indication that the effect may not be as linear as standardly assumed.

A non-linear, *U*-shaped effect of transparency on reaction times was observed by Plag and Baayen (2009). These authors demonstrated that suffixes that are either very easily segmentable or hardly segmentable have lower processing costs (as gauged by shorter reaction times in lexical decision) than suffixes in the middle of the segmentability range. Plag and Baayen (2009) interpreted this as an effect of the opposing forces of storage and computation. Assuming that our high-density words are those that are easily segmentable, while our low-density words are the ones that are not segmentable, we can come up with the tentative interpretation that the short durations in the mid-range of density are a reflection of the higher processing costs incurred by the forms in the middle of the segmentability scale. One problem with this account is, however, that higher processing costs in lexical decision seem to be correlated with shorter durations in production, but with longer latencies in comprehension. This contradiction can only be solved if we know more about the specific processing differences between production and comprehension, or about the specific processing stages involved in lexical decision vs. freely generated conversational speech. We leave this issue to be explored in future studies.

The second variable capturing the semantic relation between words that this study has shown to be able to successfully predict duration is SEMANTIC VECTOR LENGTH. Compared to SEMANTIC DENSITY, the effect of SEMANTIC VECTOR LENGTH is more straightforward to interpret. A longer semantic vector, i.e., a higher activation diversity, is associated with shorter duration. Tucker et al. (2019) argue that the more semantic dimensions a speaker is active on for a word, the more confusable the meanings of this word are. When more meanings are activated and these meanings are more confusable, the speaker is more uncertain and therefore invests less energy in maintaining the signal. In this account, our finding that words with higher activation diversity are shorter is thus expected.

Let us now return to the role of morphological information in our networks. Importantly, our results for the two semantic variables show that differences between morphological categories can emerge even from the network without any information about derivational functions. For example, semantic density is significantly higher for words with the derivational functions NESS, LESS and DIS than for words with ATION. This is in accordance with traditional descriptions of the semantic transparency of affixes, which posit *-ness*, *-less*, and *dis-* as producing mostly transparent derivatives, while words with-*ation* are assumed to be less transparent (Bauer et al., 2013; Plag, 2018). Only IZE does not fit that pattern, as many IZE words are characterized by high densities but are considered about as transparent as *-ation* (however, *-ize* is considered to be more productive than *-ation*). Another interesting example of this is the distribution of SEMANTIC VECTOR LENGTH. The longer the vector of a word, the higher its semantic activation diversity becomes and the more collocational relations it has to other words, i.e., the more polysemous it is. The average vector length was highest for IZE and ATION words. This reflects traditional descriptions of *-ize* and *-ation* having highly multifaceted semantics (cf. the locative, ornative, causative, resultative, inchoative, performative or similative meaning of *-ize*, and the meanings of *-ation* denoting events, states, locations, products, or means; Bauer et al., 2013; Plag, 2018). The affixes *-less*, *dis-*, and to a lesser extent *-ness*, on the other hand, have comparatively clearer and narrower semantics. In sum, these differences between morphological categories in the Idiosyncratic Network demonstrate that LDL can discriminate derivational functions from sublexical and contextual cues alone.

Our results have implications for morphological theory and speech production models. First, the acoustic properties of morphologically complex words can be modeled successfully by implementing a discriminative learning approach. Traditional approaches were largely unable to accommodate effects of morphological structure on the phonetic output production (Chomsky and Halle, 1968; Kiparsky, 1982; Dell, 1986; Levelt et al., 1999; Roelofs and Ferreira, 2019; Turk and Shattuck-Hufnagel, 2020). Many theories of morphology-phonology interaction assume that morphological boundaries are erased in the process of passing morphemic units on to phonological processing. And many models of speech production assume an articulator module that realizes phonemic representations with pre-programmed gesture templates independently of morphemic status. These approaches lack explanations for the fact that a word's morphological structure or semantics can cause differences in articulatory gestures, as they do not allow for a direct morphology-phonetics interaction. In LDL, however, such interaction is expected and can be explained by its underlying theoretical principles of learning and experience.

Second, our implementations show that morphological functions can emerge as a by-product of a morpheme-free learning process. Morphology is possible without morphemes. Given the many problems with the morpheme as a theoretical construct (see, e.g., Baayen et al., 2019b), this is a welcome finding. Finding morphological effects on phonetic realization need not lead to the conclusion that these effects must originate from morphemes. They can also emerge in the mapping of forms and meanings that have no information on morphology at all (see, e.g., Baayen et al., 2011 et seq. for more examples of this). As Divjak (2019) puts it, "it is not because a phenomenon can be described in a certain way that the description is psychologically realistic, let alone real" (p. 247). Of course, the success of LDL in this study and others does not allow us to infer that there is no cognitive plausibility to these structural units at all. If LDL is modeling rather how children learn languages, adult speakers may learn differently once they have explicit knowledge of morphemic structure. Such structure might also be acquired after-the-fact, when a speaker has seen enough words to start seeing analogies, or after learning about this structure explicitly. The morpheme might be epiphenomenal rather than superfluous. However, LDL does demonstrate that such fixed units of form and meaning are at the very least not obligatory. The connection between form and meaning can be dynamic and relational, allowing morphological theory to reframe its semiotic legacy. In fact, it has been argued that since its discriminative underpinnings emphasize that language is a system of *différence*, discriminative learning elegantly carries the discipline back to its Saussurean heritage (Blevins, 2016).

There are several potential future directions for discriminative learning studies on the phonetics of derived words. First, it would be interesting to model the durations of more derivational functions in a larger dataset. Investigating more than the five morphological categories of the present study might reveal further important differences between these categories. Second, one issue that we would like to resolve in future studies concerns the response variable. In a corpus study of duration with different word types, it is essential to control for phonological length. This is why instead of duration, we decided to model duration difference, i.e., the residuals of a model regressing a word's absolute duration against the sum of its average segment durations. However, for an LDL implementation, this response variable is not optimal, since strictly speaking it still implicitly assumes segmental structure. It would be desirable to control for segmental makeup without actually having to refer to segments. Third, we think it could be fruitful to investigate how best to construct vectors for words with multiple derivational functions. This would enable us to gain more insight into the complex interplay of morphological categories. And, finally, we think that to further test how well LDL can predict durations when the semantics of derivatives are strictly compositional (like in the Base Network), one interesting avenue for future research would be to use vectors that already assume this compositionality when generating lexome-to-lexome vectors.[6] That is, while in the present study we used lexome vectors that Baayen et al. (2019b) generated by using the Widrow-Hoff algorithm to predict function lexomes in addition to content lexomes for words in a sentence, it is conceivable to use vectors generated by predicting function lexomes in addition to content lexomes for *bases* in a sentence. The lexome vector $\overrightarrow{happy}$, for instance, would then capture relations to contextual lexomes surrounding

[6]Thanks are due to Reviewer 2 for this suggestion.

the word *happiness* as well. Similarly, one could generate vectors of derived words to use in the Idiosyncratic Network that do not capture cues of any functional lexomes by refraining from coding them altogether in the training data. We leave this interesting option to be explored in future studies.

To summarize, this study modeled the acoustic duration of 4,530 English derivative tokens with the morphological functions DIS, NESS, LESS, ATION, and IZE in natural speech data, using predictors derived from a linear discriminative learning network. We have demonstrated that these measures can successfully predict derivative durations. They reveal that more semantic certainty in pronunciation is associated with acoustic enhancement, i.e., longer durations, which is consistent with previous studies of paradigmatic probability and semantic support measures. We have also shown that differences between morphological categories emerge from the network, even without explicitly providing the network with such information. This further strengthens the position of LDL as a promising theoretical alternative for speech production, and provides further evidence that morphology is possible without morphemes.

DATA AVAILABILITY STATEMENT

The data, scripts and materials used in this study can be found in an online repository at https://osf.io/jkncb.

AUTHOR CONTRIBUTIONS

IP and SS contributed to conception and design of the study. SS retrieved the data, performed the modeling and statistical analysis, and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft), Research Unit FOR2373 Spoken Morphology, project Morpho-Phonetic Variation in English, PL 151/7-2 and PL 151/8-2, which we gratefully acknowledge.

ACKNOWLEDGMENTS

We thank Yu-Ying Chuang and Harald Baayen for their advice on the intricacies of linear discriminative learning.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.678712/full#supplementary-material

REFERENCES

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLoS ONE* 12:e0174623. doi: 10.1371/journal.pone.0174623

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Chuang, Y.-Y., and Heitmeier, M. (2019a). *WpmWithLdl: Implementation of Word and Paradigm Morphology With Linear Discriminative Learning. R Package*. Available online at: http://www.sfs.uni-tuebingen.de/$\sim$hbaayen/software.html

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019b). The discriminative lexicon. A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019, 1–39. doi: 10.1155/2019/4895891

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28. doi: 10.21500/20112084.807

Baayen, R. H., Milin, P., Durdević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–481. doi: 10.1037/a0023851

Baayen, R. H., Milin, P., and Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology* 30, 1174–1220. doi: 10.1080/02687038.2016.1147767

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database*. Philadelphia, PA: Linguistic Data Consortium.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bauer, L., Lieber, R., and Plag, I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60, 92–111. doi: 10.1016/j.jml.2008.06.003

Ben Hedia, S. (2019). *Gemination and Degemination in English Affixation: Investigating the Interplay Between Morphology, Phonology and Phonetics*. Berlin: Language Science Press.

Ben Hedia, S., and Plag, I. (2017). Gemination and degemination in English prefixation. Phonetic evidence for morphological organization. *J. Phon.* 62, 34–49. doi: 10.1016/j.wocn.2017.02.002

Beresewicz, M. (2015). *calc.relip.mm: Variable Importance for Mixed Models*. Available online at: https://gist.github.com/BERENZ/e9b581a4b7160357934e

Bertram, R., Baayen, R. H., and Schreuder, R. (2000). Effects of family size for complex words. *J. Mem. Lang.* 42, 390–405. doi: 10.1006/jmla.1999.2681

Blevins, J. P. (2016). "The minimal sign," in *The Cambridge Handbook of Morphology*, eds A. Hippisley and G. Stump (Cambridge: Cambridge University Press), 50–69.

Boersma, P., and Weenik, D. J. M. (2001). *Praat: Doing Phonetics by Computer*. Available online at: http://www.praat.org/

Burnage, G. (1990). *CELEX: A Guide for Users*. Nijmegen: Centre for Lexical Information.

Caselli, N. K., Caselli, M. K., and Cohen-Goldberg, A. M. (2016). Inflected words in production. Evidence for a morphologically rich lexicon. *Q. J. Exp. Psychol.* 69, 432–454. doi: 10.1080/17470218.2015.1054847

Chatterjee, S., and Hadi, A. S. (2006). *Regression Analysis by Example, 4th Edn*. Hoboken, NJ: John Wiley & Sons.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY; Evanston, IL; London: Harper and Row.

Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2020). The processing of pseudoword form and meaning in production and comprehension. A computational modeling approach using linear discriminative learning. *Behav. Res. Methods [Preprint]*. doi: 10.3758/s13428-020-01356-w

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. Contextual and paradigmatic effects on morpheme pronunciation. *Morphology* 24, 291–323. doi: 10.1007/s11525-014-9243-y

Cohen, C. (2015). Context and paradigms. *Ment. Lex.* 10, 313–338. doi: 10.1075/ml.10.3.01coh

Coleman, J., Baghai-Ravary, L., Pybus, J., and Grau, S. (2012). *Audio BNC: The Audio Edition of the Spoken British National Corpus*. Available online at: http://www.phon.ox.ac.uk/AudioBNC

Davies, M. (2008). *The Corpus of Contemporary American English: 450 Million Words, 1990–Present*. Available online at: http://corpus.byu.edu/coca/

de Saussure, F. (1916). *Cours de Linguistique Générale.* Paris: Payot.

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

Divjak, D. (2019). *Frequency in Language: Memory, Attention and Learning.* Cambridge: Cambridge University Press.

Edwards, J., Beckman, M. E., and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *J. Speech Lang. Hearing Res.* 47, 421–436. doi: 10.1044/1092-4388(2004/034)

Engemann, U. M., and Plag, I. (2021). Phonetic reduction and paradigm uniformity effects in spontaneous speech. *Ment. Lex.* 16, 165–198. Available online at: https://benjamins.com/catalog/ml.20023.eng

Filipović Durdević, D., and Kostić, A. (2021). We probably sense sense probabilities. *Lang. Cogn. Neurosci.* 1–28. doi: 10.1080/23273798.2021.1909083

Firth, J. R. (1957). "A synopsis of linguistic theory, 1930–1955," in *Studies in Linguistic Analysis* (Oxford: Blackwell), 1–31.

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *J. Mem. Lang.* 66, 789–806. doi: 10.1016/j.jml.2011.11.006

Grömping, U. (2006). Relative importance for linear regression in R. The package relaimpo. *J. Stat. Softw.* 17, 1–27. doi: 10.18637/jss.v017.i01

Hay, J. (2001). Lexical frequency in morphology. Is everything relative? *Linguistics* 39, 1041–1070. doi: 10.1515/ling.2001.041

Hay, J. (2003). *Causes and Consequences of Word Structure.* New York, NY; London: Routledge.

Hay, J. (2007). "The phonetics of un," in *Lexical Creativity, Texts and Contexts*, eds J. Munat (Amsterdam; Philadelphia, PA: John Benjamins), 39–57.

Howes, D. H., and Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *J. Exp. Psychol.* 41, 401–410. doi: 10.1037/h0056020

Ivens, S. H., and Koslin, B. L. (1991). *Demands for Reading Literacy Require New Accountability Measures.* Brewster: Touchstone Applied Science Associates.

Kiparsky, P. (1982). "Lexical morphology and phonology," in *Linguistics in the Morning Calm: Selected Papers From SICOL*, ed I.-S. Yang (Seoul: Hanshin), 3–91.

Kunter, G. (2016). *Coquery: A Free Corpus Query Tool*. Available online at: www.coquery.org

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *J. Acoust. Soc. Am.* 121, 2261–2271. doi: 10.1121/1.2537393

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2016). *lmerTest: Tests in Linear Fixed Effects Models*. R Package. Available online at: https://cran.r-project.org/web/packages/lmerTest/index.html

Ladefoged, P., and Johnson, K. (2011). *A Course in Phonetics, 6th Edn*. Boston, MA: Wadsworth Cengage Learning.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes* 25, 259–284. doi: 10.1080/01638539809545028

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

Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis.* Glenview, IL; London: Scott, Foresman and Company.

Machač, P., and Skarnitzl, R. (2009). *Principles of Phonetic Segmentation.* Prague: Epocha Publishing House.

Moore, E. H. (1920). On the reciprocal of the general algebraic matrix. *Bull. New Ser. Am. Math. Soc.* 26, 394–395.

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *J. Speech Lang. Hearing Res.* 44, 778–792. doi: 10.1044/1092-4388(2001/061)

Penrose, R. (1955). A generalized inverse for matrices. *Math. Proc. Cambridge Philos. Soc.* 51, 406–413. doi: 10.1017/S0305004100030401

Plag, I. (2018). *Word-Formation in English, 2nd Edn*. Cambridge: Cambridge University Press.

Plag, I., and Baayen, R. H. (2009). Suffix ordering and morphological processing. *Language* 85, 109–152. doi: 10.1353/lan.0.0087

Plag, I., and Balling, L. W. (2020). "Derivational morphology: an integrative perspective on some fundamental questions," in *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*, eds V. Pirrelli, I. Plag, and W. U. Dressler (Berlin; Boston, MA: Mouton de Gruyter), 295–335.

Plag, I., and Ben Hedia, S. (2018). "The phonetics of newly derived words: testing the effect of morphological segmentability on affix duration," in *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*, eds S. Arndt-Lappe, A. Braun, C. Moulin, and E. Winter-Froemel (Berlin; New York, NY: Mouton de Gruyter), 93–116.

Plag, I., Dalton-Puffer, C., and Baayen, R. H. (1999). Morphological productivity across speech and writing. *English Lang. Linguist.* 3, 209–228. doi: 10.1017/S1360674399000222

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology. The acoustics of word-final S in English. *J. Linguist.* 53, 181–216. doi: 10.1017/S0022226715000183

Plag, I., Lohmann, A., Ben Hedia, S., and Zimmermann, J. (2020). "An S is an'S, or is it? Plural and genitive-plural are not homophonous," in *Complex Words: Advances in Morphology*, eds L. Körtvélyessy and P. Stekauer (Cambridge: Cambridge University Press), 260–292.

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005a). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146–159. doi: 10.1159/000090095

Pluymaekers, M., Ernestus, M., and Baayen, R. H. (2005b). Lexical frequency and acoustic reduction in spoken Dutch. *J. Acoust. Soc. Am.* 118, 2561–2569. doi: 10.1121/1.2011150

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* Available online at: http://www.R-project.org/

Ramscar, M., and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback. A new approach to the logical problem of language acquisition. *Cogn. Sci.* 31, 927–960. doi: 10.1080/03640210701703576

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957. doi: 10.1111/j.1551-6709.2009.01092.x

Roelofs, A., and Ferreira, V. S. (2019). "The architecture of speaking," in *Human Language: From Genes and Brains to Behavior*, ed P. Hagoort (Cambridge, MA: MIT Press), 35–50.

Schreuder, R., and Baayen, R. H. (1997). How complex simplex words can be. *J. Mem. Lang.* 37, 118–139. doi: 10.1006/jmla.1997.2510

Schuppler, B., van Dommelen, W. A., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/. Studies at the phonemic and sub-phonemic level. *J. Phon.* 40, 595–607. doi: 10.1016/j.wocn.2012.05.004

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. *Lang. Cogn. Neurosci.* 33, 32–49. doi: 10.1080/23273798.2017.1359634

Shafaei-Bajestan, E., Moradipour-Tari, M., Uhrig, P., and Baayen, R. H. (2020). LDL-AURIS: error-driven learning in modeling spoken word recognition. *PsyArXiv* 1–36. doi: 10.31234/osf.io/v6cu4

Sóskuthy, M., and Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166, 298–313. doi: 10.1016/j.cognition.2017.05.032

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Phonetic effects of morphology and context. Modeling the duration of word-final S in English with naïve discriminative learning. *J. Linguist.* 57, 1–39. doi: 10.31234/osf.io/4bmwg

Torreira, F., and Ernestus, M. (2009). "Probabilistic effects on French [t] duration," in Interspeech (Brighton), 448–451.

Tucker, B., Sims, M., and Baayen, R. H. (2019). Opposing forces on acoustic duration. Preprint submitted to Elsevier. *PsyArXiv [preprint]*. 1–38. doi: 10.31234/osf.io/jc97w

Tucker, B. V., and Ernestus, M. (2016). Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *Ment. Lex.* 11, 375–400. doi: 10.1075/ml.11.3.03tuc

Turk, A., and Shattuck-Hufnagel, S. (2020). *Speech Timing: Implications for Theories of Phonology, Speech Production, and Speech Motor Control*. New York, NY: Oxford University Press.

Turnbull, R. (2018). Patterns of probabilistic segment deletion/reduction in English and Japanese. *Linguist. Vanguard* 4, 1–14. doi: 10.1515/lingvan-2017-0033

Vitevitch, M. S., and Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instruments Comput.* 36, 481–487. doi: 10.3758/BF03195594

Walsh, L., Hay, J., Derek, B., Grant, L., King, J., Millar, P., et al. (2013). The UC QuakeBox Project. Creation of a community-focused research archive. *N. Zeal. Eng. J.* 27, 20–32. doi: 10.26021/2

Widrow, B., and Hoff, M. E. (1960). "Adaptive switching circuits," in *Periodical: IRE WESCON Convention Record Part IV*, 96–104. doi: 10.21236/AD0241531

Zimmerer, F., Scharinger, M., and Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *J. Phon.* 45, 64–75. doi: 10.1016/j.wocn.2014.03.006

Zuraw, K., Lin, I., Yang, M., and Peperkamp, S. (2020). Competition between whole-word and decomposed representations of English prefixed words. *Morphology*. 31, 201–237. doi: 10.1007/s11525-020-09354-6

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

The handling editor declared a past co-editorship with one of the authors IP.

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fpsyg.2021.680889

Durational Differences of Word-Final /s/ Emerge From the Lexicon: Modelling Morpho-Phonetic Effects in Pseudowords With Linear Discriminative Learning

Dominic Schmitz[1]*, *Ingo Plag*[1], *Dinah Baer-Henney*[2] *and Simon David Stein*[1]

[1] *English Language and Linguistics, Heinrich Heine University, Düsseldorf, Germany,* [2] *Linguistics and Information Science, Heinrich Heine University, Düsseldorf, Germany*

Recent research has shown that seemingly identical suffixes such as word-final /s/ in English show systematic differences in their phonetic realisations. Most recently, durational differences between different types of /s/ have been found to also hold for pseudowords: the duration of /s/ is longest in non-morphemic contexts, shorter with suffixes, and shortest in clitics. At the theoretical level such systematic differences are unexpected and unaccounted for in current theories of speech production. Following a recent approach, we implemented a linear discriminative learning network trained on real word data in order to predict the duration of word-final non-morphemic and plural /s/ in pseudowords using production data by a previous production study. It is demonstrated that the duration of word-final /s/ in pseudowords can be predicted by LDL networks trained on real word data. That is, duration of word-final /s/ in pseudowords can be predicted based on their relations to the lexicon.

Keywords: morphology, speech production, linear discriminative learning, computational modelling, pseudoword paradigm, subphonemic differences

OPEN ACCESS

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
LouAnn Gerken,
University of Arizona, United States
Yu-Ying Chuang,
University of Tübingen, Germany

Correspondence:
Dominic Schmitz
dominic.schmitz@uni-duesseldorf.de

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

Received: *15 March 2021*
Accepted: *24 June 2021*
Published: *09 August 2021*

Citation:
Schmitz D, Plag I, Baer-Henney D and Stein SD (2021) Durational Differences of Word-Final /s/ Emerge From the Lexicon: Modelling Morpho-Phonetic Effects in Pseudowords With Linear Discriminative Learning. Front. Psychol. 12:680889. doi: 10.3389/fpsyg.2021.680889

INTRODUCTION

Many studies on the acoustic properties of phonologically homophonous elements have shown unexpected effects of their morphological structure on their phonetic realisation. Such effects were shown for seemingly homophonous lexemes (Gahl, 2008; Drager, 2011), for free and bound variants of stems (Kemps et al., 2005a,b), and for prefixes (Ben Hedia and Plag, 2017; Ben Hedia, 2019).

For the level of individual segments, a number of studies have shown that the acoustic realisation of word-final /s/ and /z/ (henceforth S) in English depends on its morphological status and category. Corpus studies (Zimmermann, 2016; Plag et al., 2017) found that non-morphemic word-final S shows longest acoustic durations, followed by suffixes, which in turn are followed by clitics. Experimental studies (Walsh and Parker, 1983; Hsieh et al., 1999; Seyfarth et al., 2017; Plag et al., 2020) confirm durational differences between different types of S. However, their results are mostly not as clear as those by previous corpus studies. That is, only recently a study by Schmitz et al. (2020) on word-final S in pseudowords confirmed the pattern of durational differences found previously only in corpus studies.

Most importantly, none of the aforementioned studies on the matter of word-final S was able to explain found differences on a theoretical level. Traditional models of speech production come with the assumption of having no morphological information in phonetic processing (Levelt et al., 1999; Roelofs and Ferreira, 2019; Turk and Shattuck-Hufnagel, 2020), thus rendering an explanation on the basis of differing morphological categories improbable. Other accounts, e.g., standard feed-forward theories of morphology-phonology interaction (e.g., Chomsky and Halle, 1968; Kiparsky, 1982) or prosodic phonology (e.g., Booij, 1983; Selkirk, 1996; Goad, 1998, 2002), do not offer a satisfying explanation for such durational differences, either.

Only recently, Tomaschek et al. (2019) analysed durational differences between types of S by means of an implementation of naïve discriminative learning (Ramscar and Yarlett, 2007; Ramscar et al., 2010; Baayen et al., 2011). Their results indicate that the duration of a word-final S in English can be sufficiently approximated by considering the support for its morphological function from the word's sublexical and collocational properties.

This paper continues this line of evidence by making use of the computational model of linear discriminative learning (Baayen et al., 2019b; Chuang et al., 2020), the more advanced successor of naïve discriminative learning. We analyse the durational differences between non-morphemic and plural word-final /s/ found not in real words, but in pseudowords. By using nonce words, we want to rule out potentially confounding effects of the lexical and contextual properties of the individual utterances (e.g., Caselli et al., 2016). Making use of measures derived from this implementation of linear discriminative learning, the present study demonstrates that the effects found by Tomaschek et al. (2019) can be confirmed. Differences in phonetic duration emerge from differences in the strengths of associations between form and meaning.

We proceed as follows. The next section will give an overview on studies on the duration of word-final S, and possibilities and obstacles of theoretical accounts. Section "Introduction to LDL" introduces linear discriminative learning on a theoretical level, while Section "Combining Real Words and Pseudowords in an LDL Implementation" presents the implementation of linear discriminative learning used in the present study. The analysis and results of our study are given in Sections "Analysis" and "Results." A discussion of the obtained results and a conclusion follow in Section "Discussion."

WORD-FINAL /s/ AND ITS DURATION

A number of morphological categories can take the phonological form of /s/ in English, i.e., plural, genitive, genitive plural, third person singular, and the clitics of is, has, and us. In itself, there is nothing in the phonological form of these morphological categories that indicates systematic differences in realisation on the phonetic level between different S morphemes or a non-morphemic S. Yet, a number of studies report on durational differences between different types of S.

Corpus studies on word-final S in English find differences in duration between non-morphemic, suffix, and clitic variants. Zimmermann (2016) on New Zealand English, and Plag et al. (2017) and Tomaschek et al. (2019) on North American English find that non-morphemic S (as in grace, cheese, bus) shows longer durations than plural S and the clitic S of has and is, while plural S in turn shows longer durations than clitic S.

Turning to experimental studies, results are not as consistent. Walsh and Parker (1983) conducted a production experiment with three homophonous word pairs with all words ending in either a non-morphemic or morphemic word-final S. Tested in three different contexts, they find durational differences in two of them. They conclude that morphemic S in English is systematically lengthened by speakers (Walsh and Parker, 1983: 204). However, their conclusion relies on only a small number of 110 observations, a mixture of common and proper nouns as items, and lacks appropriate inferential statistical methods as well as an integration of covariates.

Hsieh et al. (1999) find that plural S is longer than third person singular S in child-directed speech. However, as their data was originally elicited for another study (Swanson and Leonard, 1994), half of all plural items occurred sentence-finally, while almost all third person singular items occurred sentence-medial. Thus, the durational differences found by Hsieh et al. (1999) may be attributed to effects of phrase-final lengthening (e.g., Klatt, 1976; Wightman et al., 1992) rather than to phonetic differences between different types of S.

In another study, Seyfarth et al. (2017) conducted a production experiment on word-final /s/ and /z/ in non-morphemic, plural, and third person singular contexts. Their results indicate that non-morphemic S is shorter than morphemic S. However, they do not find a difference between voiced and voiceless instances, even though previous studies confirm differences dependent on voicing (e.g., Plag et al., 2017). With only six items ending in /s/, but twenty items ending in /z/, it is questionable how meaningful their results on different types of S are.

Comparing affixes, Plag et al. (2020) find that plural and genitive plural S differ in duration. That is, in their study the genitive plural suffix shows a longer duration than the plural suffix.

Most recently, Schmitz et al. (2020) conducted a production experiment on pseudowords carrying either a non-morphemic, plural, or is- or has-clitic S. Their results are in line with those of aforementioned corpus studies. That is, non-morphemic S shows longest S durations, followed by plural S, which in turn is followed by clitic S, while there is no significant durational difference between the two clitics. An overview of the durational differences found in corpus and experimental studies is given in **Table 1**.

There is a noteworthy discrepancy between experimental results and the results based on conversational speech data. Results of corpus studies are in line with each other, but they might be flawed due to imbalanced data sets. Experimental studies, on the other hand, often rely on small data sets, and lack phonetic covariates, appropriate statistical methods, or a proper distinction of voiced and voiceless segments. Additionally, previous experimental studies rely on different experimental methods, making their results subject to their pertinent limitations. Another crucial difference between corpus and experimental studies is the use of homophones. While corpus

TABLE 1 | Overview of durational differences of word-final /s/ found in previous studies.

Study	Findings
Zimmermann, 2016; Plag et al., 2017; Tomaschek et al., 2019; Schmitz et al., 2020	Non-morphemic > plural > clitics
Walsh and Parker, 1983	Plural > non-morphemic
Hsieh et al., 1999	Plural > third person singular
Seyfarth et al., 2017	Plural > non-morphemic
Plag et al., 2020	Genitive plural > plural

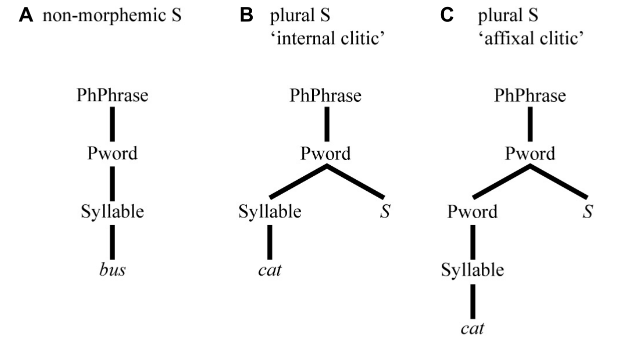

FIGURE 1 | Prosodic configuration for **(A)** non-morphemic and **(B,C)** plural S.

studies take into consideration all words, most experimental studies restrict their data to homophone pairs. This limitation to homophones and the competition between their representations might be a problem of its own as it is unclear how members of such homophone pairs may influence each other in speech production. Lastly, differences in results might also arise due to potentially confounding effects of the lexical properties and contextual effects of the items under investigation.

But even if the direction of durational differences between different types of S is not entirely clear yet, it appears that there are indeed durational differences of some sort. How is one to explain such differences? In standard feed-forward theories of morphology-phonology interaction (e.g., Chomsky and Halle, 1968; Kiparsky, 1982) all types of S, morphemic and non-morphemic, are treated in a similar way. For morphologically complex words, e.g., words ending in a morphological word-final S, a process named "bracket erasure" is said to remove any morphological information. Thus, leaving speech production with no information on the morphology of a complex word (e.g., the plural form cats), rendering its morphological information equal to that of a morphologically simple word ending in a non-morphemic word-final S (e.g., the singular form bus). In such a system, there is nothing that could account for realisational differences between phonologically identical forms of suffixes, clitics, and non-morphemic segments.

A similar distinction of lexical and post-lexical processing is also found in established theories of psycholinguistics. According to models of speech production (e.g., Levelt et al., 1999; Roelofs and Ferreira, 2019), morphemic types of word-final S do not differ in their realisation from non-morphemic instances of word-final S. For a plural form, e.g., cats, the lemma of the lexical concept CAT and a plural specification are retrieved. Then, during morphological encoding, the plural specification is mapped onto the base lemma, i.e., cat, and the plural suffix, < -s >. During phonological encoding, phonemes are selected for the corresponding morphemes, i.e., /k/, /æ/, /t/, and /s/. Finally, the phonemes are syllabified, resulting in a phonological word representation. Such phonological forms are then forwarded and used in speech production. Thus, no information on the morphological origin of particular segments is contained in the phonetic realisation, rendering an explanation on durational differences between types of S on morphological grounds improbable.

In prosodic phonology (e.g., Booij, 1983), differences in phonetic realisation may arise from the position of sounds in different configurations of prosodic constituency. For instance, different types of word-final S can be analysed as being integrated at different levels of the hierarchical prosodic configuration. In the case of word-final S, different levels co-determine differing degrees of integration of an S to the word it belongs to. Non-morphemic S, uncontroversially, is an integral part of the prosodic word itself (Selkirk, 1996), see (A) of **Figure 1**. For plural S, Goad (1998) analyses it as an "internal clitic", see (B), while Goad (2002) analyses it as an "affixal clitic", see (C).

Thus, the prosodic approach posits a structural prosodic difference between types of S. However, it is not so clear what particular phonetic effects these differences would predict. Most plausibly, a higher degree of integration would correlate with shorter durations, predicting shortest S durations in monomorphemic words. Yet, findings on S duration show the opposite (e.g., Zimmermann, 2016; Plag et al., 2017; Tomaschek et al., 2019; Schmitz et al., 2020), i.e., the duration of non-morphemic S is longest.

An alternative explanation for durational differences between different types of S can be found within the computational modelling framework of naïve discriminate learning (NDL; e.g., Ramscar and Yarlett, 2007; Ramscar et al., 2010; Baayen et al., 2011). NDL is based on simple but powerful principles of discriminative learning theory (Wagner and Rescorla, 1972; Rescorla, 1988), i.e., learning results from exposure to informative relations among events in the individual's environment. Such events are used to form associations between them, while the associations and their resulting representations are constantly updated on the basis of new experiences. Associations are built between features ("cues", e.g., biphones) and content lexemes or morphological functions ("outcomes", e.g., different types of S), which co-occur in events in which the individual is predicting outcomes from cues (Tomaschek et al., 2019: 11). Using the Rescorla-Wagner equations (Rescorla and Wagner, 1972; Wagner and Rescorla, 1972; Rescorla, 1988), relations between cues and outcomes are modelled. That is, the weight of an association, i.e., its strength, increases every time a cue and an outcome co-occur, while it decreases if a cue occurs without the outcome. The result of this process is a continuous recalibration of association strengths, which is a crucial part of discriminative learning.

NDL has been used successfully to model various morphological phenomena, e.g., reaction times in studies on morphological processing (e.g., Baayen et al., 2011; Blevins et al., 2016; see Plag, 2018, chapter 7 for an introduction to NDL in morphological research). For word-final S, Tomaschek et al. (2019) reproduce the differences in duration found by Plag et al. (2017) by means of NDL measures. Their study shows that the duration of different types of S can be approximated by considering the support for these morphological functions from a word's sublexical and collocational properties. In the NDL network, all words and their diphones within a five word window centred on the target word that contained the S served as cues, and were associated with the morphological functions, which served as outcomes. Two main measurements from this network emerged as predictive for S duration. First, the so-called "activation" as a measure of an outcome's baseline activation, i.e., of how well an outcome is entrenched in the lexicon. Second, the so-called "activation diversity" as a measure to quantify the extent to which the cues in a given context also support other targets. Taken together, the following pattern for S duration emerges: When the uncertainty about a targeted outcome increases, i.e., the level of "activation" decreases and the level of "activation diversity" increases, the duration of S decreases. In other words: The stronger the support for a morphological function is, both from long-term entrenchment and short-term from the context, the longer its duration.

While NDL implementations apparently offer some form of explanation for different durations of different types of S, they also come with shortcomings and limitations. In NDL, a word's meaning is defined in terms of the presence or absence of an outcome, i.e., NDL "adopted a stark form of naive realism" (Baayen et al., 2019b: 4) just for computational reasons. That is, NDL takes into account that words tend to have similar forms, but ignores that words are also similar in meaning. Thus, Baayen et al. (2019b) introduced semantic vectors of reals replacing the binarily coded row vectors of the semantic matrix (see Section "The S Matrix: Semantic Vectors"), naming their new implementation linear discriminative learning (LDL) instead of naïve discriminative learning. Outcomes are no longer assumed to be independent, i.e., semantic similarities are now reflected, and networks are mathematically equivalent to linear mappings of matrices, i.e., vector spaces. It is the implementation of such linear discriminative learning that the present paper makes use of for analysing the duration of word-final types of S. Our paper explores whether measures derived from an LDL implementation are predictive of different types of S and their durations. In order to better understand the relation between traditional psycholinguistic variables (such as lexical frequencies, neighbourhood densities, bigram probabilities, morphological category etc.) and LDL measurements we also compare models that use measures derived from an LDL implementation with models that use traditional measures to predict S durations. Finally, we test whether measures derived from an LDL implementation render the specification of morphological structure proper (affix vs. no affix) as predictor variable for S duration unnecessary.

INTRODUCTION TO LDL

Overview

Linear discriminative learning as a computational model implements a discriminative view of learning. In contrast to deep learning models that have multiple hidden layers based on non-linear functions, LDL networks are very simple two-layer networks and are linguistically transparent and interpretable. In LDL, the mental lexicon consists of five high-dimensional numeric matrices, each of which represents a different subsystem: the visual matrix, retina; the auditory matrix, cochlea; the semantic matrix; the speech matrix, speaking; and the spelling matrix, typing. For the current implementation, the semantic and the speech matrix are most important.

With regard to the mappings between vectors, linear mappings are implemented. These mappings are estimated using the linear algebra of multivariate regression. Thus, each mapping is defined by a matrix A that transforms the row vectors in a matrix X into the row vectors of a matrix Y, i.e., $Y = XA$. Then, $A = X'Y$, where X' is the generalised inverse of X. We will return to the mapping of matrices in Section "Comprehension and Production," and refer the interested reader to Baayen et al. (2019b) for an introduction to the mathematical details, as well as to Milin et al. (2017) for a detailed discussion on the restrictions and possibilities of linear mappings.

Another important feature of LDL is its notion of lexomes, i.e., basic semantic units corresponding to words or morphological functions. As outlined in Chuang et al. (2020), lexomes fall into two groups: content lexomes, and inflectional and derivational lexomes. Content lexomes can be morphologically simple or complex forms, i.e., *cat* and *cats*. Inflectional lexomes represent inflectional functions, e.g., number, tense, and aspect. Derivational lexomes represent derivational functions, e.g., morphological categories such as -NESS, -LESS, or UN-. Each lexome is paired with a vector of the aforementioned five subsystems. That is, for the semantic matrix, each lexome is paired with a semantic vector, making each lexome a pointer to a semantic vector on the one hand (Milin et al., 2017), and a location in a high-dimensional space on the other hand. For monomorphemic words, the semantic vector is identical to the semantic vector of the corresponding lexome. That is, the semantic vector of the word *cat*, $\overrightarrow{cat}$, is identical to the vector of the lexome CAT. For complex words, the semantic vector is the sum of its corresponding lexome vectors. That is, the semantic vector of the word *cats*, $\overrightarrow{cats}$, is the sum of the semantic vectors of the lexomes CAT and PLURAL, $\overrightarrow{cat} + \overrightarrow{plural}$. The implementation of LDL and the matrices necessary for the present paper are introduced in the subsequent sections. Please refer to https://osf.io/zy7ar/?view_only=ef43a5caf6444270a56074027d7d6482 for the full documentation of the data set, the implementation in R (R Core Team, 2020), and the R script.

The S Matrix: Semantic Vectors

The semantic matrix S contains semantic vectors of word forms on basis of their corresponding lexomes. That is, the semantic vector $\overrightarrow{s}$ in S for a simplex word is identical to its corresponding lexome, while the semantic vector $\overrightarrow{s}$ in S for a cosmplex word is the sum of its corresponding lexomes, e.g., $\overrightarrow{apple} + \overrightarrow{plural}$ for *apples* (Baayen et al., 2019b). Semantic vectors of lexomes can be derived in different ways (e.g., Landauer and Dumais, 1997; Jones and Mewhort, 2007; Shaoul and Westbury, 2010; Mikolov et al., 2013).

The C Matrix: Form Vectors

The present study uses triphones to represent form, as previous studies (Milin et al., 2017; Baayen et al., 2019b; Chuang et al., 2020) have shown that triphones capture the variability of neighbouring phonological information well for English. Triphones are sequences of three phones within a word form. They overlap and can be understood as proxies for phonetic transitions. The cue matrix C encodes the forms of words in a binary fashion, giving information on which triphones are part of which word. This is illustrated in (1). In each word's individual form vector $\overrightarrow{c}$, the presence of a triphone is marked with 1, while the absence is marked with 0. The cue vectors of all words of a set of words constitute its C matrix. That is, each row in such a C matrix represents a word form, while the columns of the respective C matrix represent all triphones of its underlying word set.

Comprehension and Production

In LDL, comprehension refers to a model that has form vectors as input and semantic vectors as output. We illustrate the C matrix of a set of words with a toy lexicon containing the words *cat*, *bus*, and *eel* in (1). Here, the DISC keyboard phonetic alphabet (the "Distinct Single Character" representation introduced by Burnage, 1988) is used for triphone representation. Word boundaries are marked by the # symbol.

$$C = \begin{array}{c} \\ cat \\ bus \\ eel \end{array} \begin{array}{c} \begin{array}{cccccccc} \#k\{ & k\{t & \{t\# & \#bV & bVs & Vs\# & \#il & il\# \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}.$$

For the same toy lexicon, suppose that the semantic vectors for these three words are the row vectors of the following S matrix:

$$S = \begin{array}{c} \\ cat \\ bus \\ eel \end{array} \begin{array}{c} \begin{array}{ccc} cat & bus & eel \end{array} \\ \begin{pmatrix} 1.0 & 0.2 & 0.5 \\ 0.4 & 1.0 & 0.1 \\ 0.2 & 0.3 & 1.0 \end{pmatrix} \end{array}.$$

To map forms onto meanings we need transformation matrix F, such that

$$CF = S.$$

The transformation matrix F is straightforward to obtain. Let C' denote the Moore-Penrose generalised inverse[1] of C, available in R as the *ginv* function of the MASS package (Venables and Ripley, 2002). Then,

$$F = C'S.$$

For the toy lexicon example,

$$F = \begin{array}{c} \\ \#k\{ \\ k\{t \\ \{t\# \\ \#bV \\ bVs \\ Vs\# \\ \#il \\ il\# \end{array} \begin{array}{c} \begin{array}{ccc} cat & bus & eel \end{array} \\ \begin{pmatrix} 0.33 & 0.06 & 0.16 \\ 0.33 & 0.06 & 0.16 \\ 0.33 & 0.06 & 0.16 \\ 0.13 & 0.33 & 0.03 \\ 0.13 & 0.33 & 0.03 \\ 0.13 & 0.33 & 0.03 \\ 0.10 & 0.15 & 0.50 \\ 0.10 & 0.15 & 0.50 \end{pmatrix} \end{array},$$

with CF being exactly equal to S in this simple example. That is, taking form vectors as input for the prediction of semantic vectors as output, i.e., solving $\hat{S} = CF$, this toy example correctly predicts 100% of all (three) words' semantics, i.e., $\hat{s}_i = s_i$. In more complex cases, semantic vectors are only approximately identical, thus, for a word i and its predicted semantic vector $\hat{s}_i$, comprehension is successful if $\hat{s}_i$ shows the highest correlation with the targeted semantic vector s_i (Baayen et al., 2019b). Following this method, one can report the percentage of comprehension accuracy.

Production as modelled in in LDL takes semantic vectors as input and delivers form vectors as output. Using the same toy lexicon as before, we adapt its C matrix, i.e., we borrow the notation by Baayen et al. (2019b) and henceforth call it T as is contains the Targeted triphones. For production, the transformation matrix G is of interest. Similar to F for comprehension, it is straightforward to obtain. Let S' denote the Moore-Penrose generalised inverse of S. Then,

$$G = S'T.$$

Given G, one can then predict the triphone matrix $\hat{T}$ from the semantic matrix S by solving

$$\hat{T} = SG.$$

For our toy lexicon example, the G transformation matrix is

$$G = \begin{array}{c} \\ cat \\ bus \\ eel \end{array} \begin{array}{c} \begin{array}{cccccccc} \#k\{ & k\{t & \{t\# & \#bV & bVs & Vs\# & \#il & il\# \end{array} \\ \begin{pmatrix} 1.14 & 1.14 & 1.14 & -0.06 & -0.06 & -0.06 & -0.56 & -0.56 \\ -0.44 & -0.44 & -0.44 & 1.05 & 1.05 & 1.05 & 0.12 & 0.12 \\ -0.09 & -0.09 & -0.09 & -0.30 & -0.30 & -0.30 & 1.08 & 1.08 \end{pmatrix} \end{array}.$$

As this is a toy example, SG is identical to T. For more complex cases, $\hat{T}$ will not be virtually identical to T "but will be

[1]The inverse of a matrix needs not exist, rendering such a matrix a singular one. Most matrices used in LDL implementations are singular matrices. Thus, an approximation of the inverse must be used instead of an inverse itself. One such approximation is the Moore-Penrose generalized inverse (Moore, 1920; Penrose, 1955).

$$\mathrm{F} = \begin{pmatrix}0.5 & 1\\ 0.1 & 0.2\end{pmatrix}$$

$$\mathrm{C} = \begin{pmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{pmatrix} \; \underset{\longleftarrow}{\longrightarrow} \; \mathrm{S} = \begin{pmatrix}0.5 & 1\\ 0.1 & 0.2\\ 0.9 & 0.7\end{pmatrix}$$

$$\mathrm{G} = \begin{pmatrix}-1.22 & -0.24\\ 1.57 & 0.31\end{pmatrix}$$

FIGURE 2 | Illustration of mapping between *C* and *S* matrix via *F* (i.e., comprehension), and *S* and *C* matrix via *G* (i.e., production). In production, *C* is referred to as *T*.

an approximation of it that is optimal in the least squares sense" (Baayen et al., 2019b: 21). Triphones with strongest support are expected to be the triphones making up a word's form. As triphones are not ordered, it is also checked whether the sequence of phones can be constructed correctly. Both, checking triphone support and sequence, are conveniently done by the functions of the WpmWithLdl package (Baayen et al., 2019a). Following this method, one can report the percentage of production accuracy.

Figure 2 summarizes the mapping between form and meaning by the *F* and *G* transformation matrix for comprehension and production modelling.

COMBINING REAL WORDS AND PSEUDOWORDS IN AN LDL IMPLEMENTATION

The Semantics of Pseudowords

The present paper follows the implementational basics outlined in Section "Introduction to LDL." However, as we are interested in /s/ durations in pseudowords (and not in real words), there are a number of complications. The most important complication arises from the widely shared belief that pseudowords do not have meaning. So how can we map form and meaning with forms that have no meaning? In a recent study (Chuang et al., 2020) it was shown that the assumption that pseudowords are bare of meaning is most probably wrong. Due to their formal similarity with existing words, pseudowords resonate with the lexicon. As a result, they may in fact carry meaning. The authors demonstrate that quantitative measures gauging the semantic neighbourhoods of pseudowords predict reaction times of lexical decision and acoustic durations. The present study is inspired by these results and implements a similar architecture. To model resonance of pseudowords with the lexicon, both real words and pseudowords must be included in the networks. The following sections will detail the combined LDL implementation of real words and pseudowords.

Data Set: Real Words and Pseudowords

The pseudowords and their phonetic realisations that this paper is based on are taken from the study of word-final /s/ production by Schmitz et al. (2020). In their study, participants were given pictures of "alien creatures" and their respective names (which were the target pseudowords), a short explanation of a situation, and a question relevant to the situation which was to be answered aloud. For each participant, pairings of pictures and pseudowords were randomised. That is, each pseudoword was represented by different pictures across participants. By button-press, a question was given to elicit an answer with the pertinent type of S while the context slowly faded out. The fading out of the question forced participants not to rely on the reading-aloud of the given context. In total, 24 pairs of pseudowords were used in that study. Each pseudoword form can act as singular or plural noun, e.g., *glaits* is either realised as singular, i.e., *a glaits*, or as plural, i.e., *two glaits*. Additionally, some pseudowords show a number of different realisations by the participants in the experiment, e.g., *prups* is sometimes produced as /pɹʌps/, and sometimes it is produced as /pɹups/. Thus, not 48 (i.e., 2 × 24) but 78 different phonological forms are included in the pseudoword set. **Supplementary Table 1** gives an overview of all pseudowords and their phonological forms.

The second set of words contains real words and their phonetic realisations. Following Chuang et al. (2020) we extracted these words from the MALD corpus (Tucker et al., 2019a). While the MALD corpus contains 26,793 real words, only a subset of 8,285 words is used for a number of reasons. First, some 7,577 words in the corpus contain multiple affixes. As it was unclear how to handle such words, these were excluded. Second, only words for which we have semantic vectors could be used, leading to the exclusion of further 6,828 words. Third, only words with transcriptions available in the CELEX corpus (Baayen et al., 1995) were retained, i.e., there was no transcription available for 818 words. Fourth, 3,285 words showed ambiguities regarding their morphology, e.g., *walks* as a third person singular verb versus the plural of a noun. As huge numbers of words lead to extensive computation times, we decided to exclude such cases as well. The final set of real words contains 6,165 simple and 2,120 complex word forms.

Cue Matrices

As introduced in Section "The C Matrix: Form Vectors," cue matrices are coded in binary form, giving information on which triphones are part of which word. For the current implementation, two such cue matrices are created using the WpmWithLdl package's (Baayen et al., 2019a) *make_cue_matrix* function. First C_{rw}, the real word cue matrix, is created for the set of real words. Then, a second cue matrix, C_{pw}, is created for the set of pseudowords. However, C_{pw} is a lot smaller than C_{rw} as there are only 78 phonological forms for pseudowords, but more than 8,000 for real words. Thus, the C_{rw} is of dimension 8285 × 7610, while C_{pw} is of dimension 78 × 78. We will come back to this issue of differing dimensions in the next section.

Semantic Matrices

To introduce semantics, i.e., semantic vectors, for the present set of real words, a pre-built semantic matrix *A* from Baayen et al. (2019b) was used. These authors derived semantic vectors based on the TASA corpus (Ivens and Koslin, 1991). For this, words were parsed into their lexomes, i.e., inflected words were represented by their stem and sense-disambiguated labels for their respective inflectional functions. Ambiguous forms, e.g., *walks*, were disambiguated using part of speech tagging (Schmid, 1999). Derived words were assigned a lexome for their stem and a lexome for derivational function. Then, following Baayen et al., 2016 and Milin et al. (2017), Naïve Discriminative Learning (Baayen et al., 2011; Sering et al., 2019) was used to build semantic vectors. The Rescorla-Wagner update rule (Rescorla and Wagner, 1972; Wagner and Rescorla, 1972; Rescorla, 1988) was applied incrementally to the sentences of the TASA corpus. That is, for each sentence the algorithm was given the task to predict the lexomes in that sentence from all lexomes of that sentence. This resulted in a 23562×23562 weight matrix A. This matrix lists all lexomes as rows and columns. Thus, for a given lexome at row i, the association strengths of this lexome with all other lexomes as given as columns is contained. In this state of the A matrix, lexomes predict themselves. Thus, the diagonal of the A matrix is set to zero (see Baayen et al., 2019b, for a discussion on this procedure). Lastly, columns which mostly contain zeros, i.e., no information, and show small variances ($\sigma < 3.4 \times 10^{-8}$) are removed. The resulting A matrix is of dimension $23,562 \times 5030$. Following the method outlined in Section "The S Matrix: Semantic Vectors," a semantic matrix for real words S_{rw} can be constructed based on A. That is, the semantic vector $\overrightarrow{s}$ in S_{rw} for a simplex word is identical to its corresponding lexome, while the semantic vector $\overrightarrow{s}$ in S_{rw} for a complex word is the sum of its corresponding lexomes. That is, the semantic vector of *apple* is $\overrightarrow{apple}$, while the semantic vector of *apples* is the sum of the vectors of the lexomes APPLE and PLURAL, i.e., $\overrightarrow{apples} = \overrightarrow{apple} + \overrightarrow{plural}$. As a set of real words is used, S_{rw} contains only semantic vectors for this set of real words (instead of, e.g., all word forms of the TASA corpus). The final real word semantic matrix S_{rw} is of dimension 8285×5487.

While this procedure is rather straightforward, the creation of a pseudoword semantic matrix S_{pw} is not. Due to the nature of pseudowords, their lexomes are not contained within any corpus or our A matrix, for that matter. Instead, one can estimate a pseudoword's semantic content by utilising the semantic and phonological information on real words, i.e., their C and S matrix (Chuang et al., 2020). That is, the same transformation matrix F that is used for mapping real word cues onto predicted real word meanings (see Section "Comprehension and Production") can be used to map pseudoword cues onto their estimated semantics. That is, one must first solve

$$F = C'_{rw}S_{rw}$$

to obtain F. Then, one can make use of the pseudoword cue matrix C_{pw}, and estimate pseudoword semantics, as

$$S_{pw} = C_{pw}F,$$

with S_{pw} denoting the originally estimated semantic matrix for pseudowords. In this semantic matrix, pseudowords of identical segmental makeup show identical semantics as semantics are calculated only based on triphone occurrence, i.e., the semantics of $pleeps_{singular}$ is identical to the semantics of $pleeps_{plural}$. To differentiate between singular and plural pseudowords, the semantic vector of the PLURAL lexome is added to all plural pseudowords in the S matrix. Similarly, the semantic vectors of ALIEN and CREATURE are added to all pseudoword semantic vectors as participants in the original production experiment were told that pseudowords describe alien creatures. As explained in Section "Model B: LDL Measures and Affix Specification," the pairing of the pictures with pseudowords representing the alien creature was randomised during the experiment by Schmitz et al. (2020). A pertinent pseudoword thus only contains the semantics of "alien creature" as a constant part of its own semantics, while other factors such as appearance, e.g., colour, shape, or number of eyes, differ across participants. We can assume that in the course of the experiment, participants gradually came to realize that the looks of these alien creatures, i.e., colour, shape, etc., are not relevant to their label names. Thus participants were just aware of the fact that these are all alien creatures, without paying much attention to their individual features. Please see the aforementioned complementary material for a detailed implementation.

Comprehension and Production

Pseudoword comprehension and production are not computed and evaluated in isolation but in combination with real words, simulating a real person's lexicon in a pseudoword comprehension and production situation, respectively. For this, we created a cue matrix C_{comb} based on a combined set of words, containing all aforementioned real words and pseudowords. In total, 8440 word forms are part of this set of words. A combined semantic matrix S_{comb} is created by attaching S_{pw} to S_{rw}, and reordering its rows to reflect the same order of words as found in C_{comb}.

Then, using the functions of the WpmWithLdl package (Baayen et al., 2019a) in R, a comprehension model is trained and checked for accuracy. That is, taking form vectors as input for the prediction of semantic vectors of output, $\hat{S}_{comb} = C_{comb}F$ is solved. Comprehension is successfully modelled for a word i if its predicted semantic vector $\hat{s}_i$ is most highly correlated with its targeted semantic vector s_i. This is true for 74.41% of cases (i.e., 6,165 word forms) in our comprehension model. In total, 25.59% of cases (i.e., 2,120 word forms) are incorrectly predicted, with 1,912 simple and 208 complex word forms. None of the incorrectly predicted word forms is a pseudoword.

Similarly, a production model is trained and checked for accuracy using functions of the aforementioned R package. Thus, semantic vectors are provided as input to predict form vectors as output, i.e., to solve $\hat{T}_{comb} = S_{comb}G$. Production is successfully modelled for a word i if its predicted triphones are those triphones present in its targeted cue vector in the correct sequence (possible sequences of triphones will be referred to below as "paths"). This is true for 97.3% of cases (i.e., 8,061 word forms) in our production model. In total, 2.7% of cases (i.e., 224 word forms) are incorrectly predicted, with 98 simple and 126 complex word forms. None of the incorrectly predicted word forms is a pseudoword.

Measures

In order to explore the potential of different measures emerging from the network to predict phonetic duration, we extracted a whole range of measures, based on the measures introduced by the WpmWithLdl package (Baayen et al., 2019a) and by Chuang et al. (2020). Please see the **Supplementary Material** for exploratory analyses of individual measures.

In the following, we first describe the semantic measures before we turn to the phonetic measures.

L1NORM and L2NORM: The L1NORM is the sum of the absolute values of vector elements of a given word's predicted semantic vector $\hat{s}$, i.e., its city-block distance. The L2NORM is the square root of the sum of the squared values of a given word's predicted vector $\hat{s}$, i.e., its Euclidian distance. For both variables, higher values imply more strong links to many other lexomes. Thus, both measures may be interpreted as semantic activation diversity.

DENSITY: For DENSITY, the correlation values of a word's predicted semantic vector $\hat{s}$ and its eight nearest neighbours' semantic vectors $s_{n1} \cdots s_{n8}$ are taken into consideration. The mean of these eight correlation values describes DENSITY, with higher values indicating a denser semantic neighbourhood.

ALC: The Average Lexical Correlation, ALC, is the mean value of all correlation values of a pseudoword's estimated semantic vector as contained in S_{pw} with each of the real word semantic vectors as contained in S_{rw}. Higher ALC values indicate that a pseudoword's semantics are part of a denser semantic neighbourhood. Thus, ALC may be interpreted as a measure of semantic activation diversity for pseudowords.

EDNN: This variable describes the Euclidian Distance between a pseudoword's estimated semantic vector s and its Nearest semantic real word or pseudoword Neighbour. Thus, higher values indicate a larger distance to the nearest semantic neighbour. EDNN may be regarded as a measure of semantic neighbourhood density.

NNC: The Nearest Neighbour Correlation is computed by taking a pseudoword's estimated semantic vector as given in S_{pw} and checking it for the highest correlation value against all real word semantic vectors as given in S_{rw}. This highest correlation value is taken as NNC value. Thus, higher values indicate that a pseudoword is semantically close to a real word. Additionally, one can tell which real word a pseudoword's semantics are closest to. This measure may be interpreted as a measure of similarity between nonce and real words, indicating the co-activation of a real word when confronted with a pseudoword.

SUPPORT: This measure describes the amount of support the word-final triphone (i.e., fs#, ks#, ps#, ts#) obtains for each pseudoword. The value of SUPPORT is extracted from $\hat{T}$. Higher values of this variable indicate a higher semantic support for the word-final triphone which includes the segment of interest, i.e., word-final S.

PATH_COUNTS: PATH_COUNTS describes the number of paths, i.e., possible sequences of triphones, detected for the production of a word by the production model. PATH_COUNTS may be interpreted as a measure of phonological activation diversity, as higher values indicate the existence of multiple candidates (and thus paths) in production.

PATH_SUM: PATH_SUM describes the summed support of paths for a predicted form. PATH_SUM may be interpreted as a measure of phonological certainty, with higher values indicating a higher certainty in the candidate form.

PATH_ENTROPIES: PATH_ENTROPIES contains the Shannon entropy values which are calculated over the path supports of the predicted form in $\hat{T}$. Thus, PATH_ENTROPIES may be interpreted as a measure of phonological uncertainty, with higher values indicating a higher level of disorder, i.e., uncertainty.

ALDC: The Average Levenshtein Distance of all Candidate productions, ALDC, is the mean of all Levenshtein distances of a word and its candidate forms. That is, for a word with only one candidate form, the Levenshtein distance between that word and its candidate form is its ALDC. For words with multiple candidates, the mean of the individual Levenshtein distances between candidates and targeted form constitutes the ALDC. Thus, higher values indicate that a word's candidate forms are very different from the intended pronunciation. ALDC may be interpreted as a measure of phonological neighbourhood density as it takes into account real word neighbourhoods for pseudowords, i.e., large values indicate sparse real word neighbourhoods.

ANALYSIS

The data set by Schmitz et al. (2020) contains non-morphemic, plural, or clitic word-final S as final segment of a pseudoword. As our LDL implementation does not include information on clitics, we only consider durational data on non-morphemic and plural S for the present study. A subset of 666 data points remains, with 303 observations with non-morphemic S and 363 observations with plural S. Due to some variable pronunciations requiring triphones not included in our LDL implementation, 13 data points had to be excluded, resulting in a final data set with non-morphemic and plural S durations of 653 data points, i.e., 300 entries on non-morphemic S and 353 entries on plural S.

Covariates

Besides the aforementioned variables extracted and computed from the LDL implementation itself (see Section "Measures"), the following covariates, adopted from previous analyses of word-final S (e.g., Plag et al., 2017; Tomaschek et al., 2019; Schmitz et al., 2020), are included in the analysis. The main reason for this is to allow us to compare the performance of these predictors with the performance of LDL predictors. LDL measures often correlate with traditional measures (such as lexical frequencies, transitional probabilities, or neighborhood densities), but the traditional measures have no clear correlating mechanisms in learning or processing.

There are, however, also covariates that do not tap into lexical properties, but that control for other influences, such as speech rate, the speaker, gender, the order of stimuli in an experiment, etc. These will be referred to as "non-lexical covariates" and they will also be included in our regression models.

AFFIX: This binary variable indicates whether a word contains an affix, i.e., whether the pertinent pseudoword is a singular or plural form. It takes the value NM for pseudowords without affix, and PL for pseudowords with affix.

SPEAKINGRATE: Analysing durational data, speech rate is a self-evident variable to consider. As speech rate is no inherent part of any LDL measure, we calculated speaking rate as the number of syllables in an utterance divided by the duration of the utterance (e.g., Tomaschek et al., 2019; Schmitz et al., 2020). This was done automatically using a script in Praat (de Jong and Wempe, 2008; Boersma and Weenink, 2019).

BASEDURLOG: Base duration was taken as a more local measure of speech rate (e.g., Plag et al., 2017, 2020; Schmitz et al., 2020). Here, the term "base" refers to the string of segments preceding the word-final S, for both non-morphemic and morphemic pseudowords. Base duration was then log-transformed to achieve a closer to normal distribution.

PAUSEBIN: To account for final-lengthening effects, stretches of silence between the offset of the word-final S and the onset of the following word were measured. Silence of 50 ms and above was considered as pause (Lee and Oh, 1999; Krivokapić, 2007). In order to make sure that closures of following plosives were not mistaken for pauses, their average closure duration (see Yao, 2007) was subtracted of the pertinent measured silence. Following the results by Schmitz et al. (2020), pause information was included as binary variable with the values PAUSE / NO PAUSE.

DISC: As some pseudowords were produced with multiple pronunciations, their transcription was incorporated as a categorical variable. This variable is called DISC after the DISC keyboard phonetic alphabet (Burnage, 1988).

BIPHONEPROBSUMBIN: The summed biphone probability for each pseudoword and its phonological variants is included as the binary variable BIPHONEPROBSUMBIN. It was calculated using the Phonotactic Probability Calculator (Vitevitch and Luce, 2004). The rationale for this variable is that more probable biphones should lead to shorter durations (e.g., Schmitz et al., 2020).

LIST & SLIDENUMBER: To account for priming effects, the list number (1–12) and the point of occurrence during the original experiment by Schmitz et al. (2020) are included.

PREC: To account for potential effects of the consonant preceding the word-final S (Umeda, 1977), it is included as PREC variable (similar to e.g., Tomaschek et al., 2019).

BIPHONEPROB: The probability of the final biphones /fs/, /ks/, /ps/ and /ts/ in monomorphemic words is included as covariate to account for potential effects of phonotactics (see Schmitz et al., 2020, for a detailed explanation).

FOLTYPE: As the segment following the word-final S is no part of the individual pseudoword, it is also not considered in LDL measures. Thus, the covariate FOLTYPE is introduced (similar to e.g., Tomaschek et al., 2019), coding the following segment by its segmental class (i.e., approximant APP for *listen*, fricative F for *find*, nasal N for *know*, plosive P for *cook*, and vowel V for *eat*), to account for potential effects of the following word (Klatt, 1976; Umeda, 1977).

SPEAKER, GENDER, AGE, LOCATION and MONOMULTILINGUAL: SPEAKER ID was included to account for general inter-speaker differences in production. GENDER, AGE, and LOCATION, i.e., the place in which the pertinent participant spent the bigger part of their life, were included as well. Additionally, participants who were early bilinguals were categorised as multilingual, while all other participants were categorised as monolingual in MONOMULTILINGUAL.

REAL: Some of the pseudowords in Schmitz et al.'s data set have an orthographically different, but phonologically identical real word counterpart. We introduced the variable REAL to control for this potential confound. This variable is TRUE for pseudowords with such a real word counterpart, and FALSE for those without. We considered the following real words as counterparts as given in Schmitz et al. (2020): *glits* corresponds to *glitz*, *glaiks* corresponds to *Gleicks*, *glifs* corresponds to *glyphs*, and *pleets* corresponds to *pleats*.

All of the following analyses make use of the following non-lexical covariates: BASEDURLOG, SPEAKINGRATE, SLIDENUMBER, and PAUSEBIN as variables concerning speech rate and continuity, PREC and FOLTYPE accounting for coarticulatory effects, LIST taking into consideration potential priming effects, MONOMULTILINGUAL, GENDER, LOCATION, AGE, and SPEAKER to account for speaker-individual differences, and REAL to include potential effects of real word counterparts.

Modelling Strategy

We devised three kinds of model: First, a baseline model with the traditional predictor variables (plus the non-lexical covariates). Second, a model with LDL predictors that also includes AFFIX as a covariate (plus the non-lexical covariates). Third, a model that contains only the LDL predictors (plus the non-lexical covariates).

The three kinds of model will allow us to answer our research questions. Recall that our ultimate goal is to understand how systematic durational differences emerge between words of different, but homophonous morphological categories. Traditional lexical variables are predictive but cannot explain how morphology can make its way into durational differences. But these models can show that such differences exist by looking at the effect of the variable AFFIX. This is our baseline model. As an alternative we implement a model that uses LDL measures. If these measures are predictive, they offer an explanation of the morphologically-induced phonetic differences: they emerge as a by-product of the association of form and meaning in the mental lexicon, and this association is the outcome of discriminative learning. By having a model that also includes AFFIX as an additional predictor, we can see whether the LDL measures completely capture the morphological effect, or whether there is a residue of morphological information that is predictive of duration but is still not captured by the LDL measures.

Model A: Traditional Measures

This model is meant to resemble those in previous studies on word-final S duration (e.g., Plag et al., 2017; Schmitz et al., 2020). Thus, we make use of similar variables: AFFIX, BIPHONEPROBSUMBIN, and BIPHONEPROB, as well as those control variables included in all analyses of this paper. None of these covariates showed high correlation coefficients. Hence, no cautionary measures regarding collinearity were taken before an initial full model was constructed. The model selection process proceeded as explained in section "Model B: LDL Measures and Affix Specification." That is, non-significant variables were excluded in a controlled step-wise fashion.

Then, variance inflation factors were checked. The covariates BIPHONEPROB and PREC showed high VIF values (i.e., 46.53 and 46.88, respectively), indicating potential overfitting of the model (e.g., Zuur et al., 2010; Fox and Weisberg, 2019). Consequently, PREC was removed from the model as it showed the highest VIF value, following the procedure described by Zuur et al. (2010). Re-fitting the model without PREC and re-checking the new variance inflation factor values revealed only non-problematic values.

Finally, the resulting model's residuals were trimmed (e.g., Baayen and Milin, 2010). Data points with residuals larger than 2.5 standard deviations were removed, ensuring a satisfactory distribution of residuals. This procedure led to a loss of 4 data points, i.e., 0.61% of all data points. An overview of all variables used in the initial model is given in **Supplementary Table 2**.

Model B: LDL Measures and Affix Specification

This model makes use of all LDL measures as well as of the AFFIX variable. Additionally, the non-lexical covariates are included. One issue to address when considering a model with such a multitude of variables is collinearity (e.g., Baayen, 2008; Tomaschek et al., 2018). To avoid collinearity related problems later on, all variables were tested for correlation using the languageR package (Baayen and Shafaei-Bajestan, 2019). This correlation check resulted in eight correlation coefficients indicating a high degree of correlation, for which we assume the threshold to be $|rho| \geq 0.5$. The pairs of correlated covariates as well as their correlation coefficients are given in **Table 2**.

Due to the high number of correlated variables, we opted for a principal component analysis (PCA; e.g., Venables and Ripley, 2002; Baayen, 2008; Tomaschek et al., 2018) to address collinearity issues. In a PCA, the dimensionality of the data is reduced by transforming the included variables into principal components. These transformations result in linear combinations of the predictors that are orthogonal to each other. Thus, the resulting principal components are not correlated.

The PCA was carried out using the *PCAmix* function of the PCAmixdata package (Chavent et al., 2017) in R, allowing the simultaneous integration of continuous and discrete variables. All variables given in **Table 2** were included in the computation of the principal component analysis, which yields nine principal components. The next step of the PCA is to determine how many of these principal components are meaningful and thus should be retained for further use. For this decision, we followed several rules of thumb (e.g., O'Rourke et al., 2005; Baayen, 2008). First, any component that displays an Eigenvalue greater than 1 accounts for a greater amount of variance than had been contributed by one variable. Such a component is therefore potentially meaningful. Second, one should retain enough components so that the cumulative percent of variance explained is equal to some minimal value. Following other implementations of principal component analyses, we aim at a value of 80% (e.g., O'Rourke et al., 2005). Third, only interpretable components are to be retained. That is, each component is made up of loadings, i.e., parts of the variables included in the PCA's computation represented by correlation coefficient values. If none of these variables is strongly represented in a component, the interpretability of that component is extremely low, rendering the component of small interest for further analyses. Following these three criteria, we find that the first three of the principal components show an Eigenvalue of one or higher. Also, the first three components account for 84% of variance. Considering the third criterion, all three components are strongly correlated with input variables. We therefore retain components 1 to 3 for further analysis, all of which show an Eigenvalue greater than 1, account for more than eighty percent of variance, and contain strong representations of variables in their loadings.[2] But what do these principal components mean? The highest loadings of the principal components, i.e., the correlation of the original variables to the pertinent component, are given in **Table 3**.

COMPONENT1 is most strongly positively correlated with PATH_COUNTS, PATH_ENTROPIES, and ALDC, while it is most strongly negatively correlated with PATH_SUM and SUPPORT. For PATH_COUNTS, higher values indicate the existence of multiple candidates (and thus paths) in production. It thus functions as an indicator of phonological uncertainty. Values of PATH_ENTROPIES relate to the level of uncertainty concerning the path supports of the predicted candidate form, with higher values indicating a higher level of uncertainty. For ALDC, higher values mean that a word's candidate forms are very different from the intended pronunciation, indicating uncertainty in production. PATH_SUM describes the summed support of paths for a predicted form, with higher values indicating a higher certainty in the candidate form. Higher values for SUPPORT suggest more certainty in the choice of the word-final triphone. COMPONENT1 can thus be described as a dimension that represents phonological or articulatory certainty.

COMPONENT2 is most strongly correlated with L1NORM, L2NORM, NNC, and AFFIX. L1NORM and L2NORM both imply more strong links to many other lexomes with higher values indicating a higher semantic activation diversity. Higher values of NNC suggest a close real word neighbour, which leads to higher levels of co-activation of that real word when confronted with the pseudoword, also leading to higher semantic activation diversity. As for AFFIX, COMPONENT2 is positively correlated with the presence of non-morphemic S data points.

[2] In addition, a cluster analysis was performed. This analysis revealed clusters which align well with the retained components of the principal component analysis. The cluster analysis is also documented in the materials that can be found at https://osf.io/zy7ar/?view_only=ef43a5caf6444270a56074027d7d6482.

TABLE 2 | Correlated variables and their correlation coefficients.

variables		rho	variables		rho
L1NORM	L2NORM	0.98	AFFIX	NNC	−0.89
PATH_COUNTS	PATH_ENTROPIES	0.95	PATH_COUNTS	SUPPORT	−0.65
PATH_COUNTS	ALDC	0.89	PATH_SUM	SUPPORT	0.73
PATH_ENTROPIES	ALDC	0.90	PATH_ENTROPIES	SUPPORT	−0.63

TABLE 3 | Loadings of original predictor variables in the three retained principal components of the first principal component analysis.

	Component1	Component2	Component3
L1NORM		0.397	0.348
L2NORM		0.405	0.363
PATH_COUNTS	0.813		
PATH_ENTROPIES	0.828		
PATH_SUM	−0.430		
ALDC	0.710		
NNC		0.698	
SUPPORT	−0.650		
AFFIX		0.421	0.517

COMPONENT3 is similar to COMPONENT2 as it is also strongly correlated with L1NORM, L2NORM, and AFFIX. Again, for L1NORM and L2NORM higher values indicate higher semantic activation diversity. AFFIX is positively correlated for plural S data points. We will come back to the interpretation of this correlation in Section "Model B: LDL Measures and AFFIX Specification."

In a next step, models were fitted using linear mixed-effects regression in R (R Core Team, 2020) using RStudio (RStudio Team, 2021) and as implemented by lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), and LMERConvenienceFunctions (Tremblay and Ransijn, 2020) to analyse the data on non-morphemic and plural S duration. The dependent variable, duration of S, was log-transformed following standard procedures to reduce the potentially harmful effect of skewed distributions in linear regression models (e.g., Winter, 2019). The name of this variable is SDURLOG.

Following the standard backward step-wise selection process for model selection (e.g., Baayen, 2008), a first model containing all remaining variables is created. That is, COMPONENT1, COMPONENT2, COMPONENT3, DENSITY, ALC, EDNN, BASEDURLOG, SPEAKINGRATE, PAUSEBIN, FOLTYPE, PREC, and REAL were included as fixed effects. The remaining variables, GENDER, LOCATION, MONOMULTILINGUAL, AGE, LIST, and SPEAKER, are included as random intercepts.

This full model was then continuously reduced through step-wise exclusion of non-significant variables. That is, a variable was considered as significant if it passed all of three tests. First, its F-value in the pertinent model had to yield a value below −2 or above 2. Second, the AIC value, i.e., the Akaike information criterion value, of the model including the variable had to be lower than the AIC value of a comparable model without the pertinent variable. Third, the results of log-likelihood tests comparing the model with to a model without the pertinent variable had to yield a p-value below the 0.05 threshold, thus indicating a significant improvement of model fit. This process was verified using the *step* function of R, which resulted in an identical model.

Then, variance inflation factors (VIFs) were computed. Predictors showing variance inflation factor values equal or greater than 3 are to be excluded due to the high risk of introducing multicollinearity and thus overfitting of the model (e.g., Zuur et al., 2010). For the present model, all variance inflation factor values are below 3.

Finally, the resulting model needed trimming of its residuals (e.g., Baayen and Milin, 2010). Data points with residuals larger than 2.5 standard deviations were removed to ensure a more satisfactory residual distribution. This procedure resulted in a loss of six data points (0.92%). An overview of all variables used in the initial model and their distribution is given in **Supplementary Table 2**.

Model C: LDL Measures Only

This model uses all LDL measures but does not incorporate the AFFIX covariate. As in the previous model, there was a high number of highly correlated variables (see **Table 2** with the exception of the correlation of AFFIX and NNC, as AFFIX is not included in this analysis). We therefore again computed a principal component analysis, following the procedure outlined in Section "Model B: LDL Measures and Affix Specification." Following the first two criteria, we find that two principal components are to be retained. However, considering the third criterion, we find that the two components are not readily interpretable as they show relatively high positive or negative correlations with all or almost all variables, without indicating a clearly discernible dimension underlying the patterns of correlations. We therefore turned to another procedure to reduce collinearity issues.

For each set of variables with a correlation of $|rho| > 0.5$, models containing only the pertinent variable and a random intercept for subject are fitted and compared. Using log-likelihood tests for model comparison, the variable contained in a significantly better fit model is retained while those variables highly correlated with it are no longer used. In case of a non-significant difference, the variable of the model with the lower AIC value is retained. This procedure leads to the exclusion of L2NORM, PATH_COUNTS, PATH_ENTROPIES, and PATH_SUM.

Linear mixed-effects regression models were fitted according to the procedure given in Section "Model B: LDL Measures and Affix Specification." That is, an initial full model was fitted with the following variables: L1NORM, ALDC, SUPPORT, DENSITY,

TABLE 4 | *p*-values of fixed effects in the final "traditional" model, fitted to the log-transformed durations of S.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (> F)
AFFIX	0.711	0.711	1	37.90	13.845	0.001
SPEAKINGRATE	0.163	0.163	1	604.07	3.165	0.076
BASEDURLOG	6.278	6.278	1	572.80	122.247	0.000
PAUSEBIN	5.430	5.430	1	635.92	105.722	0.000
BIPHONEPROBSUMBIN	0.646	0.646	1	596.28	12.580	0.000
FOLTYPE	2.199	0.550	4	605.15	10.703	0.000

TABLE 5 | Fixed-effect coefficients and *p*-values as computed by the final "traditional" model (mixed-effects model fitted to the log-transformed duration of S).

	Estimate	Std. Error	*df*	*t*-value	Pre (> | t|)
(Intercept)	−1.202	0.083	407.927	−14.520	0.000
AFFIXPL	−0.087	0.023	37.896	−3.721	0.001
SPEAKINGRATE	−0.022	0.012	604.072	−1.779	0.076
BASEDURLOG	0.635	0.057	572.805	11.057	0.000
PAUSEBINPAUSE	0.234	0.023	635.917	10.282	0.000
BIPHONEPROBSUMBINlow	−0.076	0.021	596.279	−3.547	0.000
FOLTYPEF	−0.001	0.073	610.436	−0.007	0.994
FOLTYPEN	−0.004	0.028	600.528	−0.134	0.893
FOLTYPEP	−0.027	0.025	599.182	−1.107	0.269
FOLTYPEV	−0.145	0.025	610.241	−5.852	0.000

TABLE 6 | *p*-values of fixed effects in the final "LDL measures and Affix" model, fitted to the log-transformed durations of S.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (> F)
COMPONENT1	0.376	0.376	1	618.06	6.970	0.008
COMPONENT3	1.340	1.340	1	627.71	24.819	0.000
BASEDURLOG	6.751	6.751	1	620.55	125.080	0.000
PAUSEBIN	5.805	5.805	1	642.19	107.568	0.000
FOLTYPE	2.093	0.523	4	617.98	9.695	0.000
PREC	0.702	0.234	3	615.33	4.334	0.005
DENSITY	0.219	0.219	1	621.79	4.067	0.044
ALC	0.293	0.293	1	623.25	5.425	0.020

TABLE 7 | Fixed-effect coefficients and *p*-values as computed by the final "LDL measures and Affix" model (mixed-effects model fitted to the log-transformed duration of S).

	Estimate	Std. Error	*df*	*t*-value	Pre (> | t|)
(Intercept)	−1.106	0.124	635.215	−8.952	0.000
COMPONENT1	0.014	0.005	618.057	2.640	0.008
COMPONENT3	−0.041	0.008	627.708	−4.982	0.000
BASEDURLOG	0.652	0.058	620.548	11.184	0.000
PAUSEBINpause	0.237	0.023	642.193	10.371	0.000
FOLTYPEF	−0.014	0.075	621.463	−0.180	0.857
FOLTYPEN	−0.006	0.029	614.760	−0.198	0.843
FOLTYPEP	−0.028	0.025	615.172	−1.126	0.261
FOLTYPEV	−0.141	0.025	620.352	−5.612	0.000
PRECk	−0.023	0.027	614.436	−0.835	0.404
PRECp	−0.040	0.027	614.491	−1.475	0.141
PRECt	−0.095	0.028	615.916	−3.414	0.001
DENSITY	−0.241	0.119	621.790	−2.017	0.044
ALC	−5.302	2.277	623.246	−2.329	0.020

ALC, EDNN, NNC, baseDurLog, speakingRate, pauseBin, folType, preC and real. As for random effects, random intercepts for gender, location, monoMultilingual, age, list, and speaker were included.

This full model was then continuously reduced through step-wise exclusion of non-significant variables, following the aforementioned criteria. Then, variance inflation factors were computed, resulting only in non-problematic values (e.g., Zuur et al., 2010). Finally, the resulting model needed trimming of its residuals (e.g., Baayen and Milin, 2010). That is, data points with residuals larger than 2.5 standard deviations were removed, ensuring a more satisfactory residual distribution. This procedure led to a loss of 8 data points, i.e., 1.2% of all data points. An overview of all variables used in the initial model and their distribution is given in **Supplementary Table 2**.

RESULTS

Model A: Traditional Measures

The final model of traditional measures includes main effects of the following variables: type of S (Affix), speaking rate (speakingRate), log-transformed base duration (baseDurLog), pause (pauseBin), the summed biphone probability (biphoneProbSumBin), and following segmental type (folType). As for random effects, random intercepts for speaker and random slopes for Affix are included. The p-values of the analysis of variance of the final model are given in **Table 4**.

The marginal R-squared value of the model is 0.43, i.e., fixed effects explain 43% of variation in the data. Taking random effects into account as well, the conditional R-squared value is 0.62. That is, the model explains 62% of data variation in total (see Nakagawa et al., 2017, for details on marginal and conditional R-squared computation). Both R-squared values were computed using the MuMIn package (Barton, 2020). The R-squared values are similar to the values found by Schmitz et al. (2020) on their complete data set.

The estimates of the final model and their p-values are given in **Table 5**. The reference levels for the categorical predictors are: for Affix it is NM, for pauseBin it is no-pause, for biphoneProbSumBin it is high, and for folType it is APP.

The predictor strength of individual covariates was checked by taking the final model as template. For each predictor variable, a model was fitted lacking the particular variable. This resulted in seven models, each lacking a different predictor. Then, R-squared values were computed for these models and finally compared. The variable leading to the highest decrease in R-squared value as compared to the final model is thus the variable showing the highest predictor strength. The results of this comparison are reflected in the hierarchy given in (1). The decrease in R-squared is greatest when removing baseDurLog, followed by pauseBin, and so forth. The resulting order is identical to the one found by Schmitz et al. (2020) for the complete data set.

(1) baseDurLog > > pauseBin > > Affix > > folType > > speakingRate > > biphoneProbSumBin

Model B: LDL Measures and affix Specification

In the final model including LDL measures as well as the Affix covariate as parts of the individual components resulting from the principal component analysis, and fitted according to the procedure described in Section "Model B: LDL Measures and Affix Specification," we find main effects of the first principal component (Component1), the third principal component (Component3), density, ALC, base duration (baseDurLog), following pause (pauseBin), following segmental type (folType), and preceding consonant (preC). Regarding random effects, only a speaker-specific random intercept turns out to significantly improve model fit. The p-values of the analysis of variance of the final model are given in **Table 6**.

The marginal R-squared value of the final model is 0.42, thus fixed effects explain 42% of the variation in our data. The conditional R-squared value of the final model is 0.60, that is fixed and random effects taken together explain 60% of variation.

The estimates of the final model and their p-values are given in **Table 7**. The reference levels for the categorical predictors are: for pauseBin it is no-pause, for folType it is APP, and for preC it is f.

Similar to Section "Model B: LDL Measures and affix Specification," the predictor strength of individual covariates was checked by taking the final model as template. For each predictor variable, a model was fitted lacking the pertinent variable. This resulted in seven models, each missing a different covariate. Then, marginal R-squared values were computed and compared. The model showing the lowest of these values in turn missed the covariate with the highest predictor strength. The result of this procedure is reflected in the hierarchy in (2). The decrease in R-squared is greatest when removing baseDurLog, followed by pauseBin, and so forth. In sum, variables containing measures obtained by our LDL analysis appear to be meaningful predictors of S duration.

(2) baseDurLog > > pauseBin > > Component3 > > folType > > ALC > > density > > Component1 > > preC

Figure 3 shows the effect on S duration of the numerical variables included in the model. The estimated values of the dependent variable sDurLog, i.e., S duration, and baseDurLog, i.e., base duration, are back-transformed into seconds. For Component1, higher values lead to longer S durations, while for Component3 (panel A), higher values lead to shorter S durations (panel B). Higher values of density (panel C) and ALC (panel D) come with shorter S durations. Longer bases come with longer S durations (panel E).

The partial effects of the categorical variables included in the final model are illustrated in **Figure 4**. Pauses lead to longer S durations (panel A), which is most likely a case of phrase-final lengthening (e.g., Cooper and Danly, 1981). There is also an effect of the following segment type, with S being shorter when followed by a vowel (panel B). This difference is significant for all consonant types being compared against vowels with the

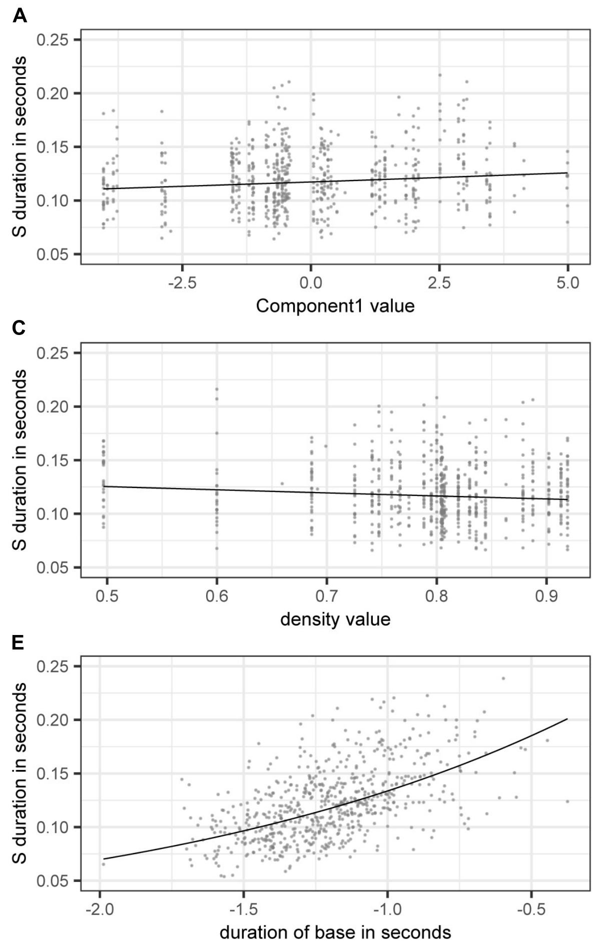

FIGURE 3 | Partial effects of the numerical variables included in the final "LDL measures and AFFIX" model, fitted to the log-transformed values of duration of S. **(A)** COMPONENT1 **(B)** COMPONENT3 **(C)** DENSITY **(D)** ALC **(E)** back-transformed BASEDURLOG.

exception of fricatives. However, as there is only a small number of fricative cases in our data, this non-significant difference is potentially not meaningful. Lastly, there is an effect of preceding consonant on S duration (panel D). S duration is significantly longer if preceded by a voiceless labiodental fricative /f/ or a voiceless velar stop /k/ as compared to cases where S is preceded by a voiceless alveolar stop /t/. All other comparisons are non-significant.

Let us turn to the variables of interest, i.e., those derived from our LDL network. COMPONENT1 acts as a general measure of phonological certainty. High values of COMPONENT1 come with high values of PATH_COUNTS, PATH_ENTROPIES, and ALDC, indicating a high level of phonological uncertainty. At the other end of the COMPONENT1 dimension, high values of PATH_SUM and SUPPORT indicate a high level of phonological certainty. Higher uncertainty appears to lead to longer S durations, while higher certainty appears to lead to shorter S durations.

Recall from Section "Model B: LDL Measures and Affix Specification" that COMPONENT3 relates to semantic activation diversity and to the presence of the plural suffix. Higher values of COMPONENT3 indicate a higher level of semantic activation diversity. Higher levels of activation diversity then lead to shorter S durations (see panel B of **Figure 3**). High values of COMPONENT3 are positively correlated with the presence of plural S. It appears that the presence of plural makes words semantically more similar to each other as they share this meaning component. Hence it is to be expected that plural words live in a space of greater semantic activation diversity. COMPONENT3 is not only a measure of semantic activation diversity, but also indicates that plural pseudowords show a tendency of having a higher degree of semantic activation diversity as compared to monomorphemic pseudowords in general. DENSITY and ALC also tap into the semantics of pseudowords. That is, similar to COMPONENT3, higher values indicate higher levels of semantic activation diversity. These higher levels then lead to shorter S durations.

Model C: LDL Measures Only

The final model of LDL measures only is fitted with main effects of the following variables: L1NORM, ALC, NNC, log-transformed base duration (BASEDURLOG), pause (PAUSEBIN), following segmental type (FOLTYPE), and preceding consonant

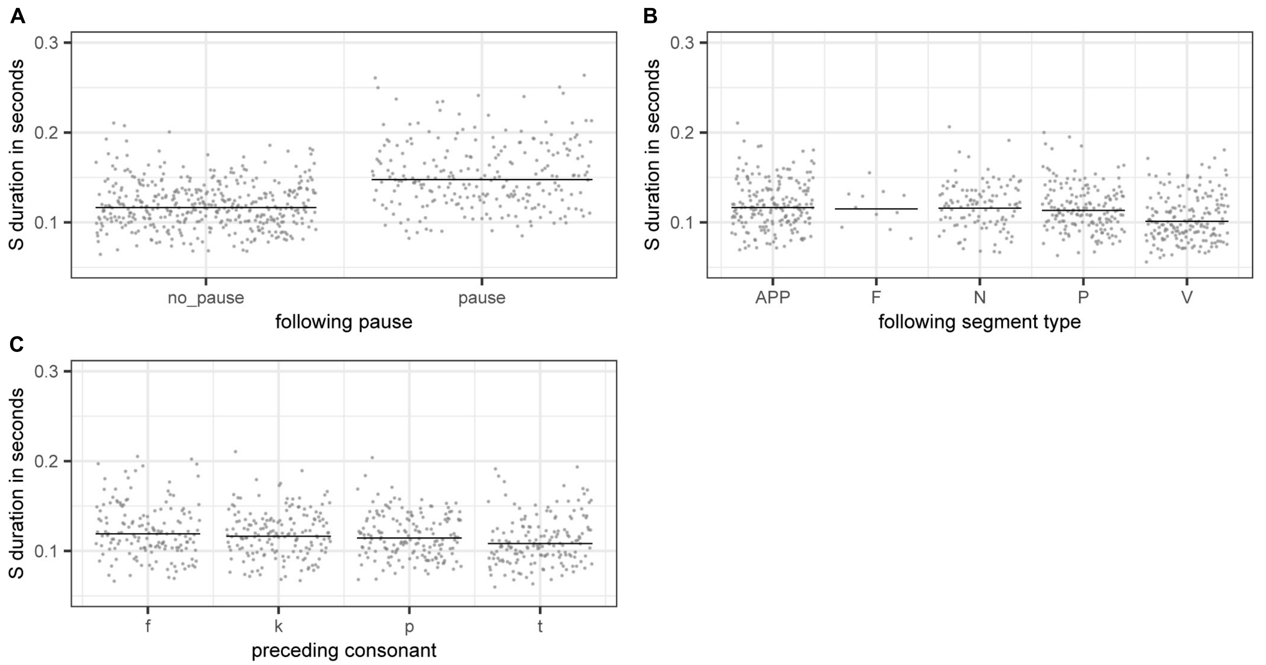

FIGURE 4 | Partial effects of the categorical variables included in the final "LDL measures and AFFIX" model, fitted to the log-transformed values of duration of S. **(A)** PAUSEBIN **(B)** FOLTYPE **(C)** PREC.

(PREC). The SPEAKER variable is included as random intercept. The p-values of the analysis of variance of the final model are given in **Table 8**.

With a marginal R-squared value of 0.41, the fixed effects of this model explain 41% of variation within the data. The conditional R-squared value of the model is 0.61, that is the complete model accounts for 61% of variation.

The coefficients of the final model and their p-values are given in **Table 9**. The reference levels for the categorical covariates are: for PAUSEBIN it is no-pause; for FOLTYPE it is APP, and for PREC it is f.

As for both other final models, the predictor strength of the individual predictors was checked. Models with one of the predictor variables were constructed based on the complete final model. Then, marginal R-squared values were computed for each of these six models. A comparison of R-squared values then revealed the hierarchy of predictor strength given in (3). That is, the decrease in R-squared is greatest when removing BASEDURLOG, followed by PAUSEBIN, and so forth.

(3) BASEDURLOG > > PAUSEBIN >> FOLTYPE >> NNC >> L1NORM >> ALC >> PREC

Base duration and speaking rate show identical effects as compared to the model fitted in Section "Model B: LDL Measures and AFFIX Specification," i.e., longer base durations come with longer S durations, while higher speaking rates lead to shorter S durations. As for categorical variables, pauses again come with longer S durations, and S is shorter if followed by a vowel. There is also an effect of the preceding consonant, with S duration being significantly longer if preceded by a voiceless labiodental fricative /f/ or a voiceless velar stop /k/ as compared to cases where S is preceded by a voiceless alveolar stop /t/. These results are generally in line with those by the analysis in the previous section.

Taking a closer look at the variables of interest, we find that higher values of L1NORM, and ALC, i.e., higher semantic activation diversity, lead to shorter S durations. As in model B, higher levels of semantic activation diversity come with shorter S durations. For NNC, we find that S duration is longer if a pseudoword is semantically similar to a real word. The effects of L1NORM, ALC, and NNC are illustrated in **Figure 5**.

DISCUSSION

The Present Results

Previous studies (Zimmermann, 2016; Seyfarth et al., 2017; Tomaschek et al., 2019; Plag et al., 2020, 2017; Schmitz et al., 2020) reported that there are significant differences in the acoustic duration between different types of word-final S in English. Such durational differences challenge established feed-forward theories of morphology-phonology interaction (e.g., Chomsky and Halle, 1968; Kiparsky, 1982) as well as theories of psycholinguistics (e.g., Levelt et al., 1999; Roelofs and Ferreira, 2019; Turk and Shattuck-Hufnagel, 2020). The present study investigated whether measures derived on the basis of a discriminative learning theory are predictive of S durations in nonce

TABLE 8 | *p*-values of fixed effects in the final "LDL measures only" model, fitted to the log-transformed durations of S.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr (> F)
L1NORM	0.685	0.685	1	611.07	13.473	0.000
BASEDURLOG	6.047	6.047	1	627.51	118.901	0.000
PAUSEBIN	5.440	5.440	1	632.72	106.956	0.000
FOLTYPE	2.056	0.514	4	610.10	10.105	0.000
PREC	0.761	0.254	3	607.96	4.985	0.002
ALC	0.534	0.534	1	615.51	10.504	0.001
NNC	0.778	0.778	1	619.67	15.296	0.000

TABLE 9 | Fixed-effect coefficients and *p*-values as computed by the final "LDL measures" model (mixed-effects model fitted to the log-transformed duration of S).

	Estimate	Std. Error	*df*	*t*-value	Pre (> | t|)
(Intercept)	−2.334	0.320	625.440	−7.301	0.000
L1NORM	−0.044	0.012	611.066	−3.671	0.000
BASEDURLOG	0.624	0.057	627.514	10.904	0.000
PAUSEBINpause	0.233	0.022	632.719	10.342	0.000
FOLTYPEF	−0.019	0.073	613.088	−0.267	0.790
FOLTYPEN	−0.005	0.028	607.324	−0.195	0.845
FOLTYPEP	−0.023	0.024	607.817	−0.950	0.343
FOLTYPEV	−0.140	0.025	611.952	−5.693	0.000
PRECk	−0.029	0.027	607.726	−1.058	0.291
PRECp	−0.053	0.027	607.478	−1.950	0.052
PRECt	−0.101	0.028	608.068	−3.632	0.000
ALC	−6.663	2.056	615.511	−3.241	0.001
NNC	1.221	0.312	619.671	3.911	0.000

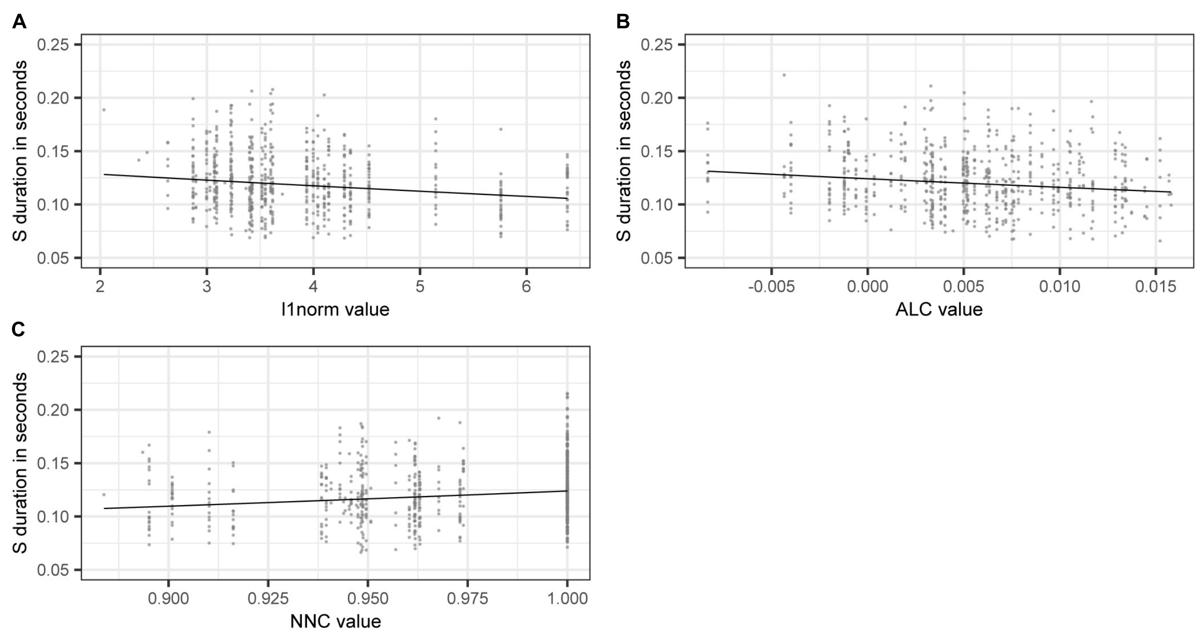

FIGURE 5 | Partial effects of LDL derived variables contained in the final "LDL measures only" model, fitted to the log-transformed values of duration of S. **(A)** L1NORM **(B)** ALC **(C)** NNC.

words. In particular, we implemented LDL networks that model the production of a word based on its relation to the rest of the lexicon.

We explored the predictive possibilities of LDL measures by fitting three different models: a) a model based on the traditional predictors as used in previous studies (Plag et al., 2017; Tomaschek et al., 2019; Schmitz et al., 2020); b) a model with LDL measures and a variable AFFIX specifying the presence or absence of an affix; and c) a model with LDL measures but without a variable specifying the presence or absence of an affix. Both models with LDL measures show that such measures are predictive of S durations. This result is the most important of our study. While traditional variables such as lexical frequencies, bigram frequencies, transitional probabilities or neighbourhood densities measure important lexical properties, it is unclear why they would manifest themselves in a particular morphological effect in speech production. In LDL such effects can emerge through the mapping of form and meaning in a clearly defined process of discriminative learning.

All regression models showed a similar hierarchy of predictor strength for the variables included in the models. For the traditional model A, AFFIX is the third strongest predictor of S duration and for model B this spot is taken by COMPONENT3, while there is no comparable variable included in model C. Comparing the variance explained by the fixed effects of the different models, we find that the traditional model accounts for most variation, i.e., 43%, while the LDL model including the AFFIX variable accounts for 42%, and the LDL model without the AFFIX variable accounts for 41% of variation. Thus, in terms of marginal R-squared values, all three models are close to each other. To check whether these differences in marginal R-squared values are of significance, the three models were refitted to the untrimmed data set and then compared with an analysis of variance. The results suggest that there is no significant difference between the traditional model and the LDL model including the AFFIX variable. However, the LDL model without the AFFIX variable shows a significantly worse fit ($p < 0.01$). This seems to indicate that the LDL measures do not capture the full amount of the variance that is captured by the variable AFFIX. This means that there is still something about the morphological function that translates into duration and that is not properly modelled by the associative measurements of the learning network. The same problem holds, incidentally, for the traditional model (model A), in which the usual lexical measures (such as lexical frequencies, neighbourhood densities, etc.) and phonetic covariates (such as pauses, speech rate, etc.) are also not able to cover all durational variance. The morphological residue in both types of analysis remains a conundrum that calls for more sophisticated approaches in future research.

Comparison of Results to Other Studies

The LDL measures included in our final models are either concerned with semantic activation diversity (COMPONENT3, ALC, and DENSITY in model B; L1NORM, and ALC in model C), semantic similarity (NNC in model C) or with phonological certainty (COMPONENT1 in model B).

Higher degrees of semantic activation diversity come with shorter S durations. This effect is similar to the one which was reported by Tucker et al. (2019b) in a study on stem vowels, and Tomaschek et al. (2019) in their NDL study on S duration. A higher degree of activation diversity makes it "more difficult to discriminate the targeted outcome from its competitors" (Tomaschek et al., 2019:27). As for production, a prolongation of the acoustic signal is dysfunctional if the prolongation maintains or increases the discrimination problem instead of contributing to resolving it (Tomaschek et al., 2019).

In the model without AFFIX as predictor variable, NNC (i.e., a pseudoword's semantic similarity to its closest semantic real word neighbour) emerges as significant (see model C). Why so? As reported in **Table 2**, the AFFIX variable and NNC are strongly negatively correlated (rho = −0.89). Post-hoc analysis shows that plural S has significantly lower NNC values as compared to non-morphemic S (Wilcoxon test, $p < 0.001$). It therefore appears that NNC takes over the role of differentiating between plural and non-morphemic S in model C.

As for phonological certainty, we find that higher phonological certainty leads to shorter S durations, while higher phonological uncertainty leads to longer S durations. Shorter durations in contexts of high phonological certainty may be related to effects of frequency, i.e., highly frequent forms are produced with higher certainty and are thus shorter.

Directions for Future Research and Conclusion

The results of the present study may bring up further questions. First, are the predictive measures found for word-final S duration in pseudowords also predictive for word-final S duration in real words? Tomaschek et al.'s (2019) NDL implementation suggests that it is, but LDL networks still need to be implemented. It would be especially interesting to model those data sets that have yielded seemingly contradictory effects. Second, taking into account that the specification of AFFIX in the modelling process leads to a significantly better model fit, one may ask what the underlying reasons for this significant effect are. This then automatically leads to another question: Is it possible to catch the effect of the AFFIX specification in terms of (new) LDL measures?

To summarize, this paper was the first to investigate durational differences between different types of word-final S (non-morphemic vs. plural S) in pseudowords by means of an LDL implementation, measures, and resulting statistical analyses. The findings yielded important evidence on the question of how such durational difference come to be, i.e., they can be predicted based on their pseudoword's relations to the lexicon. We demonstrated that durational differences emerge from the pseudoword's resonance with the lexicon by way of differing degrees of semantic activation diversity and phonological uncertainty. These manifestations of the relations to other words in the lexicon in turn are the result of discriminative learning.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/zy7ar/?view_only=ef43a5caf6444270a56074027d7d6482.

AUTHOR CONTRIBUTIONS

DS, IP, and DB-H contributed to conception and design of the study, manuscript revisions. DS retrieved the data and performed the computational implementation supported by SS. DS carried out the modelling and statistical analysis, and wrote the first draft of the manuscript. All authors read and approved the submitted version.

FUNDING

This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 'Spoken Morphology', grant PL 151/7-2 'Central project', grants BA 6523/1-1 and PL 151/9-1 'Final S in English: The role of acoustic detail in morphological learning', and grant PL 151/8-2 'Morpho-Phonetic variation in English'), which we gratefully acknowledge.

ACKNOWLEDGMENTS

The authors are grateful to the members of the DFG Research Unit FOR2373, the audience of the Words in the World Conference 2020, Janina Esser, Yu-Ying Chuang, and one reviewer for valuable input. The usual disclaimers apply.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.680889/full#supplementary-material

REFERENCES

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9780511801686

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. *Int. J. Psychol. Res.* 3, 12–28. doi: 10.21500/20112084.807

Baayen, R. H., and Shafaei-Bajestan, E. (2019). *languageR: Analyzing Linguistic Data: A Practical Introduction to Statistics (1.5.0)*. https://cran.r-project.org/package=languageR (accessed June 09, 2021).

Baayen, R. H., Chuang, Y.-Y., and Heitmeier, M. (2019a). *WpmWithLdl: Implementation of Word and Paradigm Morphology with Linear Discriminative Learning (1.3.17.1)*. Available online at: http://www.sfs.uni-tuebingen.de/~hbaayen/software.html (accessed June 09, 2021).

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019b). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (De)composition but in linear discriminative learning. *Complexity* 2019, 1–39. doi: 10.1155/2019/4895891

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–481. doi: 10.1037/a0023851

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: University of Philadelphia.

Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2016). Comprehension without segmentation: a proof of concept with naive discriminative learning. *Lang. Cogn. Neurosci.* 31, 106–128. doi: 10.1080/23273798.2015.1065336

Barton, K. (2020). *MuMIn: Multi-Model Inference (1.43.17)*. Available online at: https://cran.r-project.org/package=MuMIn (accessed June 09, 2021).

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67:51. doi: 10.18637/jss.v067.i01

Ben Hedia, S. (2019). *Gemination and Degemination in English Affixation: Investigating the Interplay Between Morphology, Phonology and Phonetics*. Berlin: Language Science Press, doi: 10.5281/zenodo.3232849

Ben Hedia, S., and Plag, I. (2017). Gemination and degemination in English prefixation: phonetic evidence for morphological organization. *J. Phonetics* 62, 34–49. doi: 10.1016/j.wocn.2017.02.002

Blevins, J. P., Ackerman, F., and Malouf, R. (2016). "Morphology as an adaptive discriminative system," in *Morphological Metatheory*, eds D. Siddiqi and H. Harley (John Benjamins), 271–302. doi: 10.1075/la.229

Boersma, P., and Weenink, D. (2019). *Praat: doing phonetics by computer (6.1.27)*. Available online at: http://www.praat.org/ (accessed October 13, 2020).

Booij, G. E. (1983). Principles and parameters in prosodic phonology. *Linguistics* 21, 249–280. doi: 10.1515/ling.1983.21.1.249

Burnage, G. (1988). *CELEX, A Guide for Users. Centre for Lexical Information*. Nijmegen, Netherlands: Centre for Lexical Information.

Caselli, N. K., Caselli, M. K., and Cohen-Goldberg, A. M. (2016). Inflected words in production: evidence for a morphologically rich lexicon. *Quar. J. Exp. Psychol.* 69, 434–454. doi: 10.1080/17470218.2015.1054847

Chavent, M., Kuentz, V., Labenne, A., Liquet, B., and Saracco, J. (2017). *PCAmixdata: Multivariate Analysis of Mixed Data (3.1)*. Available online at: https://cran.r-project.org/package=PCAmixdata (accessed June 09, 2021).

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. Manhattan, NY: Harper and Row.

Chuang, Y.-Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2020). The processing of pseudoword form and meaning in production and comprehension: a computational modeling approach using linear discriminative learning. *Behav. Res. Methods* 53, 945–976. doi: 10.3758/s13428-020-01356-w

Cooper, W. E., and Danly, M. (1981). Segmental and temporal aspects of utterance-final lengthening. *Phonetica* 38, 106–115.

de Jong, N., and Wempe, T. (2008). *Praat Script Syllable Nuclei [Praat Script]*. Available online at: https://sites.google.com/site/speechrate/Home/praat-script-syllable-nuclei-v2 (accessed August 19, 2020).

Drager, K. K. (2011). Sociophonetic variation and the lemma. *J. Phonetics* 39, 694–707. doi: 10.1016/j.wocn.2011.08.005

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression*. Thousand Oaks, CA: Sage Publishing.

Gahl, S. (2008). Time and thyme are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language* 84, 474–496. doi: 10.1353/lan.0.0035

Goad, H. (1998). Plurals in SLI: prosodic deficit or morphological deficit? *Lang. Acquisition* 7, 247–284. doi: 10.1207/s15327817la0702-4_6

Goad, H. (2002). Markedness in right-edge syllabification: parallels across populations. *Canad. J. Linguistics* 47, 151–186.

Hsieh, L., Leonard, L. B., and Swanson, L. L. (1999). Some differences between english plural noun inflections and third singular verb inflections in the input: the contributions of frequency, sentence position, and duration. *J. Child Lang.* 26, 531–543. doi: 10.1017/S030500099900392X

Ivens, S. H., and Koslin, B. L. (1991). *Demands for Reading Literacy Require New Accountability Methods*. Valley: Touchstone Applied Science Associates.

Jones, M. N., and Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychol. Rev.* 114, 1–37. doi: 10.1037/0033-295X.114.1.1

Kemps, R. J. J. K., Ernestus, M., Schreuder, R., and Harald Baayen, R. (2005a). Prosodic cues for morphological complexity: the case of Dutch plural nouns. *Memory Cogn.* 33, 430–446. doi: 10.3758/BF03193061

Kemps, R. J. J. K., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, R. H. (2005b). Prosodic cues for morphological complexity in Dutch and English. *Lang. Cogn. Proc.* 20, 43–73. doi: 10.1080/01690960444000223

Kiparsky, P. (1982). "Lexical morphology and phonology," in *Linguistics in the Morning Calm: Selected Papers From SICOL1*, ed. I. Yang (Hanshin), 3–91.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acous. Soc. Am.* 59, 1208–1221. doi: 10.1121/1.380986

Krivokapić, J. (2007). Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phonetics* 35, 162–179. doi: 10.1016/j.wocn.2006.04.001

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). Lmertest package: tests in linear mixed effects models. *J. Statist. Softw.* 82:35902. doi: 10.18637/jss.v082.i13

Landauer, T. K., and Dumais, S. T. (1997). A Solution to plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240. doi: 10.1037/0033-295X.104.2.211

Lee, S., and Oh, Y. H. (1999). Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems. *Speech Commun.* 28, 283–300. doi: 10.1016/S0167-6393(99)00014-X

Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sc.* 22, 1–75. doi: 10.1017/S0140525X99001776

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. advances in neural information processing systems. *Arxiv* [preprint]. arxiv. 1310.4546,

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., and Baayen, R. H. (2017). Discrimination in lexical decision. *PLoS One* 12:e0171935. doi: 10.1371/journal.pone.0171935

Moore, H. E. (1920). On the reciprocal of the general algebraic matrix. *Bull. Am. Mathemat. Soc.* 26, 394–395.

Nakagawa, S., Johnson, P. C. D., and Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. Royal Soc. Int.* 14. doi: 10.1098/rsif.2017.0213

O'Rourke, N., Hatcher, L., and Stepanski, E. J. (2005). *A Step-by-Step Approach to Using SAS for Univariate & Multivariate Statistics*. Cary: SAS Publishing.

Penrose, R. (1955). A generalized inverse for matrices. *Mathemat. Proc. Cambridge Philos. Soc.* 51, 406–413. doi: 10.1017/S0305004100030401

Plag, I. (2018). *Word-Formation in English (Second Edition)*. Cambridge: University Press, doi: 10.1017/CBO9780511841323

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: the acoustics of word-final S in English. *J. Linguistics* 53, 181–216. doi: 10.1017/S0022226715000183

Plag, I., Lohmann, A., Ben Hedia, S., and Zimmermann, J. (2020). "An <s> is an <s'>, or is it? Plural and genitive-plural are not homophonous," in *Complex Words*, eds L. Körtvélyessy and P. Štekauer (Cambridge University Press).

Ramscar, M., and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: a new approach to the logical problem of language acquisition. *Cogn. Sci.* 31, 927–960. doi: 10.1080/03640210701703576

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957. doi: 10.1111/j.1551-6709.2009.01092.x

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available online at: https://www.r-project.org/ (accessed February 15, 2021).

Rescorla, R. A. (1988). Pavlovian conditioning: it's not what you think it is. *Am. Psychol.* 43, 151–160. doi: 10.1037/0003-066X.43.3.151

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 Research and Theory*, eds A. H. Black and W. F. Prokasy (Newyork, NY: Appleton Century Crofts), 64–99.

Roelofs, A., and Ferreira, V. S. (2019). "The architecture of speaking," in *Human Language: From Genes and Brains to Behavior*, ed. P. Hagoort (MIT Press), 35–50.

RStudio Team (2021). *R, RStudio: Integrated Development for R (1.4.1103)*. Available online at: http://www.rstudio.com/ (accessed January 20, 2021).

Schmid, H. (1999). "Improvements in part-of-speech tagging with an application to german," in *Natural Language Processing Using Very Large Corpora. Text, Speech and Language Technology*, eds S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann, and D. Yarowsky (Springer), 11. doi: 10.1007/978-94-017-2390-9_2

Schmitz, D., Baer-Henney, D., and Plag, I. (2020). *The Duration of Word-Final /s/ Differs Across Morphological Categories in English: Evidence From Pseudowords [Manuscript submitted for publication]*. Germany: English Language and Linguistics, Heinrich Heine University Düsseldorf.

Selkirk, E. (1996). "The prosodic structure of function words," in *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition*, eds K. Demuth and J. Morgan (Routledge), 187–213.

Sering, T., Milin, P., and Baayen, R. H. (2019). Language comprehension as a multiple label classification problem. Statistica Neerlandica. *Statistica Neerlandica* 1–15.

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2017). Acoustic differences in morphologically-distinct homophones. *Lang. Cogn. Neurosci.* 33, 32–49. doi: 10.1080/23273798.2017.1359634

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413. doi: 10.3758/BRM.42.2.393

Swanson, L. A., and Leonard, L. B. (1994). Duration of function-word vowels in mothers' speech to young children. *J. Speech Hearing Res.* 37, 1394–1405. doi: 10.1044/jshr.3706.1394

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *J. Phonetics* 71, 249–267. doi: 10.1016/j.wocn.2018.09.004

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Phonetic effects of morphology and context: modeling the duration of word-final S in English with naïve discriminative learning. *J. Linguistics* 2019, 1–39. doi: 10.1017/S0022226719000203

Tremblay, A., and Ransijn, J. (2020). *LMERConvenienceFunctions: Model Selection and Post-Hoc Analysis for (G)LMER Models (3.0)*. Available online at: https://cran.r-project.org/package=LMERConvenienceFunctions (accessed June 09, 2021).

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., and Sims, M. (2019a). The massive auditory lexical decision (mald) database. *Behav. Res. Methods* 51, 1187–1204. doi: 10.3758/s13428-018-1056-1

Tucker, B., Sims, M., and Baayen, R. H. (2019b). Opposing forces on acoustic duration. *PsyArXiv* [preprint] 1–38. doi: 10.31234/osf.io/jc97w

Turk, A., and Shattuck-Hufnagel, S. (2020). *Speech Timing*. Oxford: Oxford University Press, doi: 10.1093/oso/9780198795421.001.0001

Umeda, N. (1977). Consonant duration in American English. *J. Acoust. Soc. Am.* 61, 846–858. doi: 10.1121/1.381374

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics With S*. New York: Springer, doi: 10.1007/978-0-387-21706-2

Vitevitch, M. S., and Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behav. Res. Methods Instru. Comput.* 36, 481–487. doi: 10.3758/BF03195594

Wagner, A. R., and Rescorla, R. A. (1972). "Inhibition in Pavlovian conditioning: application of a theory," in *Inhibition and Learning*, eds R. A. Boakes and M. S. Halliday (London: Academic Press), 301–334.

Walsh, T., and Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *J. Phonetics* 11, 201–206. doi: 10.1016/s0095-4470(19)30816-2

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717. doi: 10.1121/1.402450

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Milton Park: Routledge, doi: 10.4324/9781315165547

Yao, Y. (2007). Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech. *UC Berkeley Phonol. Lab Ann. Rep.* 8, 183–225.

Zimmermann, J. (2016). "Morphological status and acoustic realization," in *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology (SST-2016)*, eds C. Carignan and M. D. Tyler (ASSTA), 201–204.

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evolu.* 1, 3–14. doi: 10.1111/j.2041-210X.2009.00001.x

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

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fpsyg.2021.720017

Paradigmatic Relations Interact During the Production of Complex Words: Evidence From Variable Plurals in Dutch

Tim Zee[1]*, *Louis ten Bosch*[1], *Ingo Plag*[2] *and Mirjam Ernestus*[1]

[1] *Centre for Language Studies, Radboud University, Nijmegen, Netherlands,* [2] *Department of English Language and Linguistics, Heinrich Heine University, Düsseldorf, Germany*

OPEN ACCESS

Edited by:
Walter Daelemans, University of Antwerp, Belgium

Reviewed by:
Marc Brysbaert, Ghent University, Belgium
Clara Cohen, University of Glasgow, United Kingdom
Marcel Pikhart, University of Hradec Králové, Czechia
Claudia Marzi, Istituto di linguistica computazionale "Antonio Zampolli" (ILC), Italy

***Correspondence:**
Tim Zee
tim.zee@ru.nl

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

Received: *03 June 2021*
Accepted: *29 July 2021*
Published: *01 September 2021*

Citation:
Zee T, ten Bosch L, Plag I and Ernestus M (2021) Paradigmatic Relations Interact During the Production of Complex Words: Evidence From Variable Plurals in Dutch. Front. Psychol. 12:720017. doi: 10.3389/fpsyg.2021.720017

A growing body of work in psycholinguistics suggests that morphological relations between word forms affect the processing of complex words. Previous studies have usually focused on a particular type of paradigmatic relation, for example the relation between paradigm members, or the relation between alternative forms filling a particular paradigm cell. However, potential interactions between different types of paradigmatic relations have remained relatively unexplored. This paper presents two corpus studies of variable plurals in Dutch to test hypotheses about potentially interacting paradigmatic effects. The first study shows that generalization across noun paradigms predicts the distribution of plural variants, and that this effect is diminished for paradigms in which the plural variants are more likely to have a strong representation in the mental lexicon. The second study demonstrates that the pronunciation of a target plural variant is affected by coactivation of the alternative variant, resulting in shorter segmental durations. This effect is dependent on the representational strength of the alternative plural variant. In sum, by exploring interactions between different types of paradigmatic relations, this paper provides evidence that storage of morphologically complex words may affect the role of generalization and coactivation during production.

Keywords: morphology, phonetics, paradigms, reduction, inflection, Dutch, plural, variation

1. INTRODUCTION

Most psycholinguistic accounts of lexical processing agree that the comprehension and production of a word form can be affected by its morphological relations with other word forms (see, for example, the recent overview in Arndt-Lappe and Ernestus, 2020). In very general terms, two words can be seen as morphologically related if they share phonological features that also reflect a similarity in meaning. Broadly, two types of morphological relations can be distinguished: relations between words that share a base (e.g., *burn* and *burned*) and relations between words with shared inflectional or derivational exponence (e.g., *burned* and *cared*). In this paper, we will refer to the former as relations within paradigms and to the latter as relations between paradigms. We will make a further distinction between two types of within-paradigm relations: those between the base and a complex form (e.g., *burn* and *burned*), and those between two alternative forms (e.g., *burned* and *burnt*). Previous psycholinguistic studies on morphological relations have mostly focused on how the different relation types individually affect word processing (e.g., Ernestus and Baayen, 2003 for between-paradigm relations; Hay, 2001 for base-complex relations; Cohen, 2015 for relations between alternatives). Potential interactions between these different types of paradigmatic relations have remained relatively unexplored (but see Milin et al., 2009). In the current research, we will use Dutch plurals to investigate the potentially interacting effects of between- and within-paradigm relations. In doing so, we aim to contribute to a more complete understanding of the mechanisms of generalization, storage and coactivation that are involved in the processing of complex words.

Most Dutch plural nouns are inflected for number by suffixing the singular base form with either of the two regular suffixes -*en* and -*s*. In addition, a few plurals are formed with irregular suffixes such as -*eren* or -*a*. As noted by dictionaries (e.g., Van Dale, 2020) and textbooks on Dutch morphology (e.g., de Haas and Trommelen, 1993), for certain nouns more than one suffix is acceptable: *artikel* "article" can be inflected as both *artikels* and *artikelen*, and both *keuzes* and *keuzen* are acceptable plurals of *keuze* "choice." Although some of this variability can be attributed to differences in modality (Kürschner, 2009), register (Baayen et al., 2002) and dialect (Goeman et al., 2005), different plural forms of the same noun can be found in a single utterance, see (1) which was taken from Wilde Haren De Podcast (2019).

(1) de **piramides** van Gizeh, hè, de drie bekende
'the pyramids of Giza, right, the three famous
piramiden
pyramids'

Baayen et al. (2002) argue that this type of variation occurs when the factors that govern the allomorphy within the Dutch plural system are inconclusive. For instance, most accounts of the Dutch plural agree that the distribution of -*en*, pronounced /ə(n)/, vs. -*s*, pronounced /s/, seems to reflect a prosodic preference for a word-final disyllabic trochee. As a result, most nouns with an unstressed final syllable, e.g., *bakker* /ˈbɑkər/, are pluralized with -*s*, whereas most nouns ending in a stressed syllable are pluralized with -*en*, e.g., *dier* /ˈdir/. However, if a singular noun already ends in schwa, e.g., *piramide* /ˌpiraˈmidə/, the -*en* suffix is simplified to -*n*, such that adding either suffix would result in a word-final trochee and, as a result, an acceptable plural (Kürschner, 2009). Variation may also occur when two factors are in conflict. For instance, the phonological generalization that nouns ending in stressed vowels have the plural suffix -*s* sometimes conflicts with the preference for a trochee. This may explain the variation in the plural of the noun *individu* /ˌɪndiviˈdy/: *individu's* and *individuen*. In sum, previous discussions of Dutch variable plurals suggest that two alternative forms may exist as a consequence of the application of non-deterministic phonological generalizations. However, we will argue that storage and coactivation mechanisms might also be expected to affect the production of variable plurals, given the different paradigmatic relations that apply to variable plurals. As such, Dutch variable plurals provide an excellent opportunity to investigate how different types of paradigmatic relations interact.

1.1. Paradigmatic Relations

Dutch variable plurals are a suitable phenomenon to illustrate how between-paradigm relations may affect morphological processing. The morpho-phonological patterns that, according to Baayen et al. (2002), govern both the distribution of invariable and variable Dutch plurals can be seen as generalizations among noun paradigms. In fact, these between-paradigm generalizations can be explicitly modeled using the mechanism of analogy. For instance, in order to produce the plural form of *vampier* "vampire," generalization by analogy relies on morpho-phonological similarities to singular base forms from other paradigms such as *pionier* "pioneer" and generalizes their plural forms, i.e., *pioniers*, to the original base form, resulting in *vampiers*. An advantage of such an analogical approach is that the production of variation is built-in: the plural of *vampier* can also be generalized from the *papier-papieren* "paper(s)" paradigm, resulting in *vampieren*, which is also an acceptable form. Previous work has shown that computational analogical models accurately predict the variation observed for various phonological and morphological phenomena, and affix choice in particular (e.g., Krott et al., 2001; Wulf, 2002; Ernestus and Baayen, 2003; Keuleers et al., 2007; Arndt-Lappe, 2014). Although analogical models elegantly predict the occurrence of many affixed forms that would be classified as exceptions in categorical rule-based models, analogical mechanisms are not completely successful in their predictions either. The model implemented by Keuleers et al. (2007) shows that inaccurate predictions also exist for Dutch plurals. Although this model improved on the performance of a deterministic rule-based model, it still attributed the wrong allomorph to around 9% of the plural forms they considered. This suggests that not every Dutch plural form can be predicted from between-paradigm relations.

It has been argued that the influence of between-paradigm relations on lexical processing is limited for word forms with high token frequencies (e.g., Bybee, 1995). The reasoning behind this claim is that repeated exposure to a word form results in a strong representation which is easier to access directly, compared to weaker representations of infrequent word forms, which may be easier to process by generalization from related word forms (e.g., Divjak and Caldwell-Harris, 2019). Such storage effects might affect the distribution of morphological structure in a language. For example, Bybee (1995) argues that the irregular past tense in English tends to occur in frequent verbs because their strong representations have resisted the generalization from phonologically similar regular past tense forms (see also Cuskley et al., 2014). This suggests that absolute token frequency is a measure of representational strength. However, some studies (Hay, 2001, 2007; Blumenthal-Dramé, 2012) have claimed that representational strength of complex forms is best measured as the token frequency of the complex word relative to its base word. Hay (2001) observes that models of lexical processing which incorporate both computation and whole-word access involve some type of competition between whole-word representations and representations of the base (e.g., Baayen et al., 1997b). It follows, according to Hay (2001), that relative frequency between these forms, rather than absolute frequency of the complex form, is a better predictor of the degree to which complex representations are accessed directly in lexical processing. Psycholinguistic evidence for this base-complex frequency relation has come from studies on derived words (e.g., Hay, 2001, 2007; Blumenthal-Dramé, 2012) in addition to findings from plural inflection (e.g., Baayen et al., 1997a,b, 2003; New et al., 2004; Biedermann et al., 2013; Beyersmann et al., 2015). For instance, Baayen et al. (1997b) showed that Dutch singular nouns are processed faster than their plural inflections but only if they are singular-dominant, i.e., if the singular forms are more frequent than the corresponding plurals. These findings have led researchers to posit that processing of singular-dominant plurals often requires computation based on the singular, resulting in slower and less accurate processing (Beyersmann et al., 2015). Conversely, in a picture naming study, Baayen et al. (2008) concluded that shorter production latencies for Dutch plural-dominant plurals may reflect that their production is less dependent on analogical generalization. In sum, the base-complex frequency relation has been argued to mediate between distinct processing mechanisms: direct activation of a representation vs. some form of generalization, be it through rules or analogy.

Within-paradigm frequency relations have also been found to affect the phonetic realization of morphologically complex words. For instance, Cohen (2014) found that when speakers read aloud sentences like *The choir for the church services seems nervous*, the verb agreement suffix *-s* was longer if the 3rd person singular form (e.g., *seems*) was frequent compared to the uninflected form (e.g., *seem*). Various studies have found similar phonetic enhancement of complex words with a higher frequency relative to one or more members of their paradigm (Kuperman et al., 2007; Schuppler et al., 2012; Bell et al., 2020; Tomaschek et al., 2021b, but see Hanique and Ernestus, 2012). This so-called *paradigmatic enhancement* effect has been argued to occur when the choice between multiple paradigm members is probabilistic (Kuperman et al., 2007). However, studies vary considerably with regard to the paradigm members they deem to contribute to this effect. In the current research, we will follow Cohen (2014) and Cohen (2015) by only considering the *paradigmatic enhancement* effect associated with the frequency relation between paradigm members that are allowed by the syntactic context and that result in a very similar meaning. We can illustrate such *paradigmatic alternatives* using Dutch variable plurals: in *De drie bekende piramides/piramiden* "the three famous pyramids," both plurals are allowed by the syntactic context and the resulting semantics are very similar (if they differ at all). If paradigmatic enhancement applies to Dutch variable plurals we would expect the frequency ratio between plural variants to affect their pronunciation.

Paradigmatic enhancement can be formulated in terms of probability: words with a higher paradigmatic probability have more enhanced pronunciations. In that light, paradigmatic enhancement is a surprising effect, given many previous studies which show that increased probability of a linguistic structure generally results in reduced pronunciations. For instance, it has been shown that contextually probable segments (e.g., van Son and Pols, 2003), syllables (e.g., Aylett and Turk, 2006), and words (e.g., Bell et al., 2009) are reduced in terms of duration and/or spectral qualities. Moreover, there is even some evidence that increased probability of a complex word relative to its base results in reduced pronunciation (Hay, 2001). This tendency to reduce predictable units can be explained from a communicative perspective if we assume that speakers reduce elements that contribute less to listener comprehension (e.g., Aylett and Turk, 2004). In addition to this listener-oriented account, an alternative, potentially better supported (Bell et al., 2009; Ernestus, 2014), speaker-driven account of reduction has been proposed. In such an account, the reduction of predictable words can be explained using two mechanisms that are relevant to the current study. Firstly, it has been proposed that representations of more predictable words are easier to access, which allows for faster articulation (e.g., Bell et al., 2009). Secondly, the reduction of high probability words can been explained as a direct result of practicing the same articulations over and over (e.g., Bybee and Hopper, 2001). Neither of these mechanisms, however, predicts paradigmatic enhancement, which seems to require a different explanation.

The first detailed theoretical account of paradigmatic enhancement is given by Cohen (2015), who adopts an exemplar theoretic approach (e.g., Goldinger, 1998) in which the pronunciation of a word is codetermined by all exemplars that are activated during production (e.g., Walsh et al., 2010). According to Cohen (2015), during lexical access, multiple representations of paradigmatically related words may be activated. This coactivation is mediated by the linguistic context, which means that paradigm members that are contextually plausible are activated more strongly. For example, in the Dutch sentence *de antilopen/antilopes rennen* "the antelopes are running," both the *-en* and the *-s* form are allowed, and, as a result, activation of the *-s* form may lead to coactivation of the *-en* form. Importantly, the degree to which the exemplars of the coactivated form contribute to the pronunciation of the word depends on the number of exemplars of each activated form, i.e., how often the speaker has encountered the respective forms. For instance, the pronunciation of the *-s* suffix in Dutch *antilopes* might be strongly influenced by *antilopen* exemplars because the *-en* form is much more frequent for this noun. Cohen (2015) argues that the nature of this influence can be predicted by comparing the target pronunciation and the coactivated pronunciation. In our example, final [s] in the target pronunciation [ɑntilopəs] would be reduced because the coactivated pronunciation [ɑntilopə] does not have a final [s] (the /n/ in the *-en* suffix is usually omitted). However, if the target form, e.g., *piramides*, is more frequent than the coactivated form, *piramiden*, we would expect the [s] in the target pronunciation to be less reduced. According to this account, then, paradigmatic enhancement reflects a relative lack of reduction due to the relative infrequency of coactivated word forms. While direct phonetic influence of the coactivated variants on pronunciation works for this example and the phenomena described by Cohen (2014) and Cohen (2015), it does not explain other manifestations of paradigmatic enhancement (e.g., Tomaschek et al., 2021b). It may also be that coactivation of paradigmatic alternatives indirectly disrupts articulation of the target form. Bell et al. (2020) propose that enhancement of a particular segment depends on the amount of activation available for its articulation, which in turn is decreased by paradigmatic alternatives with a different (or no) segment in the same position. In such an account, a strong representation of an alternative plural variant would take away activation from the articulation of final [s], resulting in reduced pronunciation. Regardless of the precise implementation of reduction, an account in which articulation is affected by coactivated representations of paradigmatic alternatives may explain why produced forms with higher frequencies relative to paradigmatic alternatives have less reduced pronunciations.

1.2. Interactions Between Paradigmatic Relations

While it has been shown that base-complex relations and relations between paradigmatic alternatives affect production, it is unknown whether these different within-paradigm relations interact with each other. Such an interaction might be expected given previous theoretical assumptions about the respective relations. The first assumption is that the base-complex frequency relation (e.g., *piramide-piramides*) reflects the representational strength of complex words (e.g., Hay, 2001). Whether this assumption applies to Dutch variable plurals is tested separately in our Study 1. The second assumption (as proposed by Cohen, 2015) is that the degree of paradigmatic enhancement depends on the representational strengths of the produced form (e.g., *piramides*) and the co-activated alternative form (e.g., *piramiden*). If we apply the definition of representational strength in the first assumption to the second assumption, a hypothesis can be constructed about how paradigmatic enhancement should be affected by an interaction between base-complex relations and relations among paradigmatic alternatives. The first assumption implies that the greatest disparity in representational strength between paradigmatic alternatives can be found if one alternative (A1) is much more frequent than the base form (B) whereas the other alternative (A2) is much less frequent compared to the base (i.e., A2 < B < A1). According to the second assumption, we would expect to see a strong paradigmatic enhancement effect in this case. Conversely, if both alternatives are much less frequent than the base (i.e., A1, A2 < B), we would not expect to see a strong paradigmatic enhancement effect. In terms of processing mechanisms, this means that a paradigmatic enhancement effect would not be expected to surface if production of paradigmatic alternatives might be mostly computational, i.e., if production does not involve strong representations of complex words. Applied to Dutch variable plurals, this interaction hypothesis would predict that the relative frequency of plural variants has a greater effect on pronunciation if the noun paradigm is plural-dominant. After all, in plural-dominant paradigms the differences in representational strength between plural variants are potentially greatest (i.e., A2 < B < A1; or A1 < B < A2). This interaction hypothesis is tested in our Study 2.

Dutch plural variation has a number of features that makes it a suitable phenomenon to test the interaction between base-complex relations and relations among paradigmatic alternatives. Firstly, as the plural variants have the same morphological function (see also *morphological overabundance*; Thornton, 2019), they form paradigmatic alternatives in every context. Consequently, the context of the plural variants does not need to be controlled in an experiment to collect enough data points, which means that the relations among plural variants can be studied in natural communicative settings. Secondly, for Dutch variable plurals the base-complex relation and the relation between paradigmatic alternatives are not conflated. Such a conflation of relations can be found in English verb agreement to collective nouns: in the *the family seem/seems* example, *seem* is both the base form and the alternative of *seems* (see also Cohen, 2014). Finally, the range of relative frequencies between the members of the noun paradigms that contain variable plurals is large enough to measure their effect on pronunciation. Importantly, given the assumption that the frequency of a complex word relative to its base reflects how it is processed, paradigms should be included in which the singular base is more frequent than the complex plurals as well as paradigms with relative frequencies in favor of the plural forms. Conveniently, a fair number of Dutch nouns are plural-dominant, providing the necessary spread in the relative frequency between complex and base forms. In sum, Dutch variable plurals provide an excellent opportunity to investigate how the different relations within paradigms interact during production.

1.3. The Present Studies

The current research approaches the interaction between singular-plural relations and relations among plural variants in two studies. The first study tests whether previous assumptions about the singular-plural relation for invariable plurals and base-complex relations in general also apply to variable plurals. The second study of this research tests whether the singular-plural relation interacts with the relation between plural variants in affecting the processing of variable plurals. In both studies, we will focus on how production of a single variant is affected by paradigmatic effects. Specifically, we will focus on the *-s* variant because affixes realized as [s] have reliably shown morphological effects on duration in previous research (e.g., Walsh and Parker, 1983; Cohen, 2014; Plag et al., 2017, 2020; Tomaschek et al., 2021a for *-s* suffix in English; Kuperman et al., 2007 for *-s-* interfix in Dutch).

Our first study tests the association between the base-complex frequency relation and the representational strength of complex words. As strong representations have been argued to limit the influence of generalization (e.g., Divjak and Caldwell-Harris, 2019), this association can be evidenced by showing that relatively frequent complex words are less affected by generalization. Specifically, our first study investigates whether Plural Dominance, measured as the combined frequency of the plural variants divided by the frequency of the singular form, moderates the influence of phonological generalizations on the choice between plural variants. It has been shown that phonological generalizations can be used to accurately predict the plural suffix of many Dutch nouns (Baayen et al., 2002; Keuleers et al., 2007). Given psycho-linguistic studies on Plural Dominance (Baayen et al., 2008; Beyersmann et al., 2015), we would expect that the plural variant of plural-dominant plurals is harder to predict using phonological patterns. In order to test the predictability of a plural variant, we needed a measure of the distribution of plural variants that would be predicted by phonological generalizations and a measure of the actual distribution. The actual distribution can be extracted from a corpus of written Dutch. By counting the number of -*s* (e.g., *piramides*) tokens and the number of competing (e.g., *piramiden*) tokens, the ratio of -*s* tokens, henceforth -S BIAS, can be computed for each noun. The predicted distribution can be obtained using a computational model that predicts the plural variant based on phonological features of the noun. We adopted the analogical model of Dutch plural formation described by Keuleers et al. (2007) to predict the plural allomorph of variable plurals. In this model, which is implemented using the TiMBL software (Tilburg Memory Based Learner; Daelemans et al., 2018), conflicts between analogies with different nouns are possible, resulting in uncertainty about the plural allomorph that should be chosen. By expressing this uncertainty as the probability of obtaining the -*s* allomorph and entering it as the -S PREDICTION variable into a regression model of the -S BIAS, we can assess the extent to which phonological generalization predicts the variation. We expect that the positive effect of -S PREDICTION on the observed -S BIAS will be smaller for more frequently pluralized nouns, that is, for nouns with higher PLURAL DOMINANCE. This outcome would support the hypothesis that the frequency relation between variable plurals and their singular forms reflects the influence of different processing mechanisms (generalization vs. whole-word access) on the production of variable plurals. More generally, such an outcome supports the assumption that the base-complex frequency relation reflects the representational strength of complex words.

In our second study, we test the hypothesis that base-complex relations interact with relations among paradigmatic alternatives. Specifically, we used the paradigmatic enhancement phenomenon to investigate the interaction between the singular-plural dominance relation and the coactivation among plural variants. On the basis of Cohen's (2015) theoretical account of paradigmatic enhancement, we can predict that a plural variant that is infrequent relative to its alternative should be pronounced with a more reduced plural suffix. As such, we expect that final -*s* is shorter for plurals with a more frequent -*en* or irregular variant. Crucially, we expect that this effect of -S BIAS is mediated by the PLURAL DOMINANCE measure. For noun paradigms with high PLURAL DOMINANCE, a low -S BIAS means that the competing plural variant is frequent relative to both the -*s* variant and the singular. As such, the final [s] of these nouns is expected to be shorter due to interference of the much stronger representation of the alternative variant. Conversely, a high -S BIAS for plural-dominant nouns suggests that the -*s* variant has a much stronger representation than the alternative variant, which is therefore not expected to reduce the duration of final [s]. For infrequently pluralized nouns, i.e., nouns with low PLURAL DOMINANCE, we do not expect a strong paradigmatic enhancement effect as neither plural variant is assumed to have a strong representation. These outcomes would provide evidence for an account of plural production in which the representational strength of the plural variants negotiates between the influence of generalizations across different noun paradigms and the influence of alternatives within its own paradigm. In such an account, plural variants that have strong representations are mostly produced by accessing whole word representations, whereas plural variants with weak representations are mostly produced by a generalization mechanism. The influence of the competing plural variant on production is dependent on its representational strength relative to that of the produced variant. More generally, such an outcome would be in line with the hypothesis that base-complex relations interact with relations among paradigmatic alternatives.

TABLE 1 | Mean, minimum and maximum values of the variables in the distributional study.

Dependent variable	**Mean**	**Min.**	**Max.**
-s Bias	−0.222	−7.749	6.564
Predictors of interest			
-s Prediction	0.482	0.000	1.000
Plural Dominance	−1.079	−7.726	7.953
Covariate			
Plural Frequency	3.329	1.099	8.033

2. DISTRIBUTIONAL STUDY

2.1. Materials and Methods

2.1.1. Frequency Data

Most of the variables used in this study (see **Table 1**) were based on word frequency data. The corpus used to compute these word frequencies had to meet a number of criteria. Most importantly, it needed to be sufficiently large. Numerous examples of variable plurals are discussed in the literature (e.g., de Haas and Trommelen, 1993), but many of these are low frequency words and are therefore not likely to occur frequently in small text corpora, which would hamper the computation of reliable ratios of the occurrence of -*s* vs. other plural affix variants. The second criterion related to the level of annotation. Word tokens needed to be morphologically annotated for the data processing step, which consisted of automatically selecting nouns, identifying which word forms were part of the same inflectional paradigm, and distinguishing between invariable and variable plurals. Finally, we preferred a corpus that was not solely based on formal written language. This was important as formal texts are more sensitive to prescriptive rules and conscious linguistic processing, which might have limited the amount of variation in plural suffixes.

The SUBTLEX-NL corpus was found to best match these criteria. With more than 400,000 unique, morphologically annotated word forms, it met two of our requirements. Furthermore, it is based on subtitles, which have word frequency distributions that have been shown to predict word processing measures more accurately than frequencies from alternative sources (Keuleers et al., 2010), presumeably because subtitle frequencies approximate those in natural speech. Using the morphological annotations of the SUBTLEX-NL corpus, we

TABLE 2 | An example of a TiMBL feature vector and class label for the plural *vaders* "fathers".

Penultimate syllable				Final syllable					Plural type
Onset	Nucleus	Coda	Stress	Onset	Nucleus	Coda	Stress	Final letter	
v	a	–	+	d	@	r	–	r	-s

automatically separated the nouns that had a single plural form from those that had multiple. As we focused on *-s* plural variants, we only considered nouns with multiple plurals if one of those was an *-s* variant. From this set of variable plurals, we manually excluded nouns that were incorrectly identified as having a variable plural. For instance, certain orthographically identical but phonologically and semantically separate words with different plurals, e.g., *sportster+s* "female athletes" and *sportster+en* "sports stars", were incorrectly conflated under a single lexical entry. Similarly, we excluded nouns if their different plural forms had separate (though sometimes related) meanings, such as *wortelen* "carots" and *wortels* "roots" (see Haeseryn et al., 1997). Other cases we excluded involved incomplete interfixed compounds, such as *functionerings(gesprek)* "appraisal [meeting]", which were sometimes analyzed as *-s* plurals by the morphological tagger used for SUBTLEX. Apart from removing obvious mistakes, we also excluded plural forms that occur in very few paradigms such as *brandweerman-brandweerlieden* "firefighter(s)", and *-en* plurals that could also be analyzed as infinitive verb forms such as *testen* in *De onderzoeker houdt van testen* "The researcher loves tests/to test". Finally, we removed forms that occurred more frequently in foreign utterances than in Dutch utterances, e.g., *rings*.

After excluding mistakes and potentially unreliable data, the selection of variable plurals consisted of 384 noun types. For each of these nouns the dependent variable -S BIAS was computed by dividing the number of *-s* tokens by the number of tokens with the alternative plural variant and taking the natural logarithm of the resulting ratio. Additionally, the predictor PLURAL DOMINANCE was calculated for each noun type by dividing the total number of plural tokens by the total number of singular tokens and taking the natural logarithm of the resulting ratio. Following Cohen (2015), we expressed these within-paradigm frequency relations using log-transformed ratios to compensate for the enormous range in token frequencies. A positive log-ratio indicates that the numerator (e.g., plural frequency for PLURAL DOMINANCE) is greater than the denominater (e.g., singular frequency for PLURAL DOMINANCE). The reverse frequency relation is true for a negative log-ratio, and a log-ratio of zero indicates that numerator and denominator are equally frequent. In other words, -S BIAS and PLURAL DOMINANCE are centered around the point of equal proportion. In addition to the paradigmatic predictors, the PLURAL FREQUENCY variable was computed by taking the natural logarithm of the total number of plural tokens for each noun. For lower values of PLURAL FREQUENCY, the -S BIAS measure is biased toward 0. In fact, -S BIAS is exactly 0 for all variable plurals that occur only twice in the corpus. These plurals were excluded, as they would lead to less reliable estimates of the regression model. The final set consists of 361 noun types. Section 2.1.3 describes how we used PLURAL FREQUENCY to account for the tendency of *-s Bias* toward 0 in the remaining data when estimating the effects of the predictors of interest.

2.1.2. Generating -s Predictions With TiMBL

In order to model the influence of between-paradigm relations on the choice of plural variant, we needed detailed phonological transcriptions for the nouns that were identified in the SUBTLEX corpus. As such, we used the CELEX corpus (Baayen et al., 1996) to collect phoneme and word stress features for the singulars forms of both the variable and invariable plurals that were selected from SUBTLEX. In addition to these features, we also needed a computational model that could use them to predict the plural variant. We adopted the approach by Keuleers et al. (2007), who used the TiMBL classifier (Daelemans et al., 2018) to implement a probabilistic model based on phonological and orthographic analogy that predicts the suffix of Dutch plurals. In this approach, each plural was represented as a vector of phonological and orthographical features and a class label indicating the correct plural type; see **Table 2** for the example *vaders*.

In the present study, we recognized 3 plural suffix types: *-s*, *-en*, and *other*. TiMBL uses the k-nearest neighbors algorithm (kNN) to predict the plural suffix of noun types that are unseen by TiMBL. This algorithm compares the feature vector of an unseen noun to the feature vectors of nouns for which the plurals are known. The noun with the feature vector most similar to that of the unseen noun is the closest neighbor at $k = 1$. Similarly, the second-most similar noun is at distance $k = 2$, *et cetera*. Consequently, if the parameter k is set to larger numbers, more dissimilar nouns are considered in the comparison. In the standard configuration of the kNN algorithm, the unseen noun is assigned the plural type that was associated with the majority of the neighbors. If distance weighting is enabled, closer neighbors count for more than distant neighbors.

Although this standard implementation of TiMBL has been shown to model phonological factors on invariable Dutch plurals quite well (Keuleers et al., 2007), its categorical output is not a very useful predictor for variable plurals. Therefore, we had our TiMBL model produce two types of output: categorical classifications for training and validation based on the invariable plurals, and continuous probabilities for prediction of the variable plurals. Accordingly, we separated our plural data into a *training set*, which consisted of 9908 invariable plural types, a *validation set*, which contained another 1532 invariable plurals, and a *test set*, which contained 361 variable plurals. The model

TABLE 3 | Beta-binomial model of -s Bias.

μ **Coefficients**	**Estimate**	**Std. Error**	***z*-value**	***p***
Intercept	0.588	0.277	2.123	0.034
-s Prediction	−1.238	0.497	−2.493	0.013
Plural frequency	−0.379	0.063	−5.974	0.000
Plural dominance	−0.058	0.069	−0.833	0.405
-s Prediction : *Plural frequency*	0.614	0.114	5.392	0.000
-s Prediction : *Plural dominance*	−0.248	0.120	−2.069	0.039
ϕ **Coefficients**	**Estimate**	**Std. Error**		
Intercept	0.350	0.016		

p-values were estimated using Wald tests.

was subsequently trained and optimized on the training and validation sets using categorical labels. This process involved comparing the validation accuracies for every combination of the hyperparameters listed in **Supplementary Table 1**.

The best validation accuracy of 0.949 was achieved by a model that used inverse distance decay with k=5, trained on type merged data with feature vectors of 2 syllables (see **Supplementary Table 1** for descriptions of features). Subsequently, this model was used to provide probabilities of the respective plural classes for the variable nouns in the test set. The predictor of interest -s Prediction (see **Table 1**) was extracted from the resulting probability distributions.

2.1.3. Modeling -s Bias

To assess the potential for collinearity in our data, we calculated correlations between all the variables in this study. None of the pairwise Pearson correlations between predictor variables exceeded $r = 0.20$ (see **Supplementary Figure 1** for full documentation of all correlations).

In choosing an appropriate statistical model of the interaction effect between Plural Dominance and -s Prediction on -s Bias, we considered the nature of the dependent variable. As -s Bias can be described as a log odds ratio, a binomial model seemed the obvious choice. Binomial models are suitable for our data as they can take into account differences in sample size, i.e., plural frequencies, when calculating the standard errors of the estimated log odds. However, when we considered that the dependent variable is based on characteristics of specific words (see *language-as-fixed-effect fallacy*, Clark, 1973), it became clear that regular logistic regression would lead to a poorly estimated model. We know from research on invariable Dutch plurals (Keuleers et al., 2007) that the choice of allomorph does not always follow a predictable pattern. A calculation based on the data from Keuleers et al. (2007) shows that around 9% of invariable plurals does not have the allomorph predicted by TiMBL. In other words, for some nouns the choice of plural allomorph is noun-specific. Likewise, we might expect that the distribution of plural variants for certain variable plurals is at least partly specific to the noun. It is therefore likely that modeling -s Bias using logistic regression would lead to overdispersion, i.e., a case in which the data show more variability than expected on the basis of a regular binomial model. After all, simple logistic regression assumes that the -s Bias of each noun can be predicted exclusively from fixed effects (e.g., phonological patterns). Instead, an approach was needed which treated the underlying probability of an -*s* variant as a random variable. Although random structure in binomial data can be modeled using generalized mixed effects models, previous research has shown that beta-binomial regression more reliably results in robust parameter estimates (Harrison, 2015). Beta-binomial regression assumes that the probability parameter of the binomial model is randomly chosen from a beta-distribution for each noun. The additional free parameter of this beta-distribution is estimated when the beta-binomial model is fitted. This allowed us to model both fixed and noun-specific effects on -s Bias. As such, we used beta-binomial regression, as implemented in the R package `aods3` (Lesnoff and Lancelot, 2018), to model -s Bias. Model diagnostics did indeed reveal that a beta-binomial model fitted the data significantly better than a binomial model, see **Supplementary Figure 2**.

The -s Prediction × Plural Dominance interaction was included to test our hypothesis that the representational strength of a plural limits the degree to which the choice between plural variants is governed by analogical generalization. We expected that higher values of Plural Dominance, which are assumed to reflect stronger plural representations, would be associated with a weaker relation between -s Prediction and -s Bias. Additionally, the -s Prediction × Plural Frequency interaction was included to account for the tendency of -s Bias toward 0 for infrequent plurals. Biased values of -s Bias for low frequency plurals limit the amount of variance that can be explained by -s Prediction. As such, we expected that the positive relation between -s Prediction and -s Bias would diminish for lower values of Plural Frequency. By accounting for this effect, the estimation of the -s Prediction × Plural Dominance interaction should be less influenced by the limited effect of -s Prediction at lower Plural Frequency.

2.2. Results

Table 3 summarizes the outcome of the fitted beta-binomial model of -s Bias. The μ coefficients describe

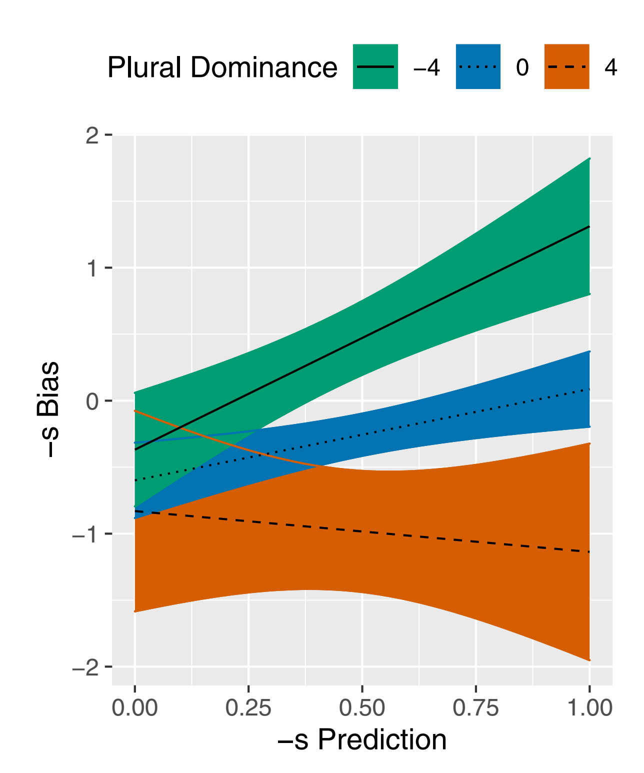

FIGURE 1 | Partial effect plot for the -S PREDICTION × PLURAL DOMINANCE interaction in the model of -S BIAS. PLURAL FREQUENCY is held constant at the median value. Coloured bands reflect 95% confidence intervals.

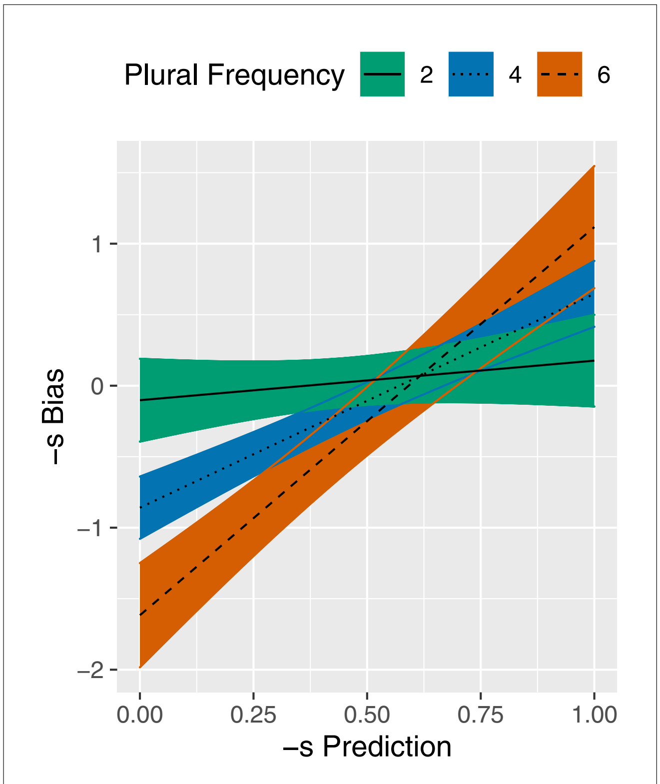

FIGURE 2 | Partial effect plot for the -S PREDICTION × PLURAL FREQUENCY interaction in the model of -S BIAS. PLURAL DOMINANCE is held constant at the median value. Coloured bands reflect 95% confidence intervals.

the average relations between the predictors and -S BIAS. The ϕ coefficient, a dispersion parameter, describes the estimated shape of the underlying probability distribution of -S BIAS.

Table 3 reveals a significant interaction between the predictors of interest, -S PREDICTION and PLURAL DOMINANCE. The fitted lines in **Figure 1** illustrate the estimated effect of -S PREDICTION on -S BIAS at different values of PLURAL DOMINANCE. A PLURAL DOMINANCE of 4 amounts to a plural/singular ratio of more than 50/1 and it is indicated by the dashed line with an orange confidence band; a value of 0 corresponds to a plural/singular ratio of exactly 1/1 which is represented by the dotted line with a blue confidence band; and a value of −4 reflects a plural/singular ratio of less than 1/50 and it is visualized by the solid line with a teal confidence band. As PLURAL DOMINANCE decreases, the slopes of these lines increase. This result is in line with our expectations, which suggested that generalization, represented by -S PREDICTION, mainly affects the plural variation of plurals with less representational strength (PLURAL DOMINANCE).

Additionally, **Table 3** indicates a significant interaction between -S PREDICTION and PLURAL FREQUENCY. The fitted lines in **Figure 2** visualize the effect of -S PREDICTION on -S BIAS at different values of PLURAL FREQUENCY. The log-transformed values of 2, 4, and 6 correspond to approximate untransformed frequencies of 7, 55, and 403, respectively. As illustrated by the nearly horizontal line, -S PREDICTION does not have a clear effect on -S BIAS for nouns with low PLURAL FREQUENCY. Conversely, for nouns with high PLURAL FREQUENCY, the rising line indicates a positive effect of -S PREDICTION on -S BIAS. This interaction was expected because -S BIAS has a tendency toward 0 for low frequency nouns.

3. DURATIONAL STUDY

3.1. Materials and Methods

3.1.1. Acoustic Data

The speech material analyzed in this study was extracted from the Dutch speech corpora listed in **Table 4**. We limited our dataset to Netherlandic Dutch, as the Dutch-Belgian border coincides with a different distribution of plural allomorphs for a number of nouns Goeman et al. (2005). Variable plural tokens were automatically identified using the orthographic transcriptions of the speech corpora and the selection of 361 noun types that occurred with multiple plural forms in SUBTLEX. We arived at a final dataset after discarding observations that would have resulted in unreliable duration measurements. This included tokens in which the final /s/ was preceded or followed by

/s/, /z/, /ʃ/, /ʒ/, /t/, /d/ or /j/, as it is very difficult to segment the speech signal into two distinctive sounds in such cases. Furthermore, in certain recordings that involved multiple speakers, the respective speakers' voices were not recorded on separate audio channels. As a result, overlapping speech in those recordings is more difficult to segment, and durational measurements of such data may not be reliable. Therefore, final /s/ tokens from these recordings were excluded if they occurred in overlapped speech. The final data set consisted of 594 *-s* plural tokens.

The final /s/ duration of the variable plural tokens was measured by the Kaldi-based (Povey et al., 2011) CLST forced-aligner (Kuijpers et al., 2018) to limit the influence of human biases and inconsistencies. The pronunciation dictionary of the forced-aligner was enriched to allow for reduced pronunciation variants according to the rules laid out by Schuppler et al. (2011). The parameters of the forced-aligner were validated on a separate set of manually annotated utterances in the Spoken Dutch Corpus (Oostdijk, 2000). Using this procedure we selected the settings that resulted in the smallest number of phonetic feature changes, insertions or deletions (as measured by *weighted feature edit distance*; Mortensen et al., 2016) between the automatic and manual transcriptions. The extracted segment durations from the automatically aligned speech were log-transformed to arrive at our dependent variable -s Duration.

3.1.2. Predictors

Our paradigmatic predictors of interest -s Bias and Plural Dominance were extracted from the data set used in the distributional study. Additionally, we used SUBTLEX to calculate two alternative measures of lexical representation, -s Frequency and Relative -s Frequency, which have been used in previous research. -s Frequency was computed to represent an account of the lexical representation of the *-s* plural based on its log-transformed absolute frequency instead of paradigmatic relations (e.g., Schuppler et al., 2012). We also included the Relative -s Frequency to account for the proposal that paradigmatic effects should be measured by dividing the frequency of the *-s* plural by the lexeme frequency and log-transforming the resulting proportion (e.g., Cohen, 2015).

In order to account for the variance in -s Duration that is unrelated to our paradigmatic predictors, we included a number of covariates. Specifically, we used covariates that have been used in previous studies that looked at segmental durations in corpus data (e.g., Plag et al., 2017).

One of the more obvious influences on segmental duration comes from the relative speed with which the surrounding speech is uttered. We measured this influence using two different variables. Firstly, Speech rate was calculated in syllables per second by counting the number of syllables in the current utterance and dividing it by the duration of the utterance. Utterances were defined as uninterrupted chunks of speech. The number of syllables was determined by counting the number of vowels that were recognized by the forced aligner. Secondly, Base duration was defined as the natural logarithm of the duration of the word excluding the final /s/. This measure was included to account for the variation in local speaking rate that was not captured by the speech rate variable.

The duration of final /s/ might also be influenced by the phonological characteristics of the word containing and the word following it. As such, Number of syllables was included as a variable to account for the segmental reduction that increases with the number of syllables in a word (e.g., Nooteboom, 1972). Additionally, the phonetic class of the Previous segment was taken into account, as it might influence the duration of the final /s/. For instance, final /s/ might be shorter if it forms a consonant cluster with the preceding segment (e.g., Klatt, 1976). The phonetic context following final consonants has also been shown to influence segmental duration (e.g., Luce and Charles-Luce, 1985). Therefore, the broad phonetic class of the Next segment was also included as a variable. We considered the following classes for Previous segment and Next segment: *vowels*, *liquids*, *approximants*, *nasals*, *fricatives*, *plosives* and *silence*.

A number of prosodic variables have been shown to affect the pronunciation of consonants (e.g., Cho and McQueen, 2005). On a word level, stressed syllables result in longer segments. Therefore, we used CELEX to implement Word stress as a categorical variable which indicated whether the stressed syllable contained the final /s/. The larger prosodic context also influences segmental duration (Cho and McQueen, 2005). Particularly relevant for the current study is the phenomenon known as final lengthening, in which segments that occur before a prosodic boundary are lengthened (e.g., Hofhuis et al., 1995). Unfortunately, the corpora used in this study were not prosodically annotated. To get around this problem some corpus studies (e.g., Plag et al., 2017) use syntactic boundaries instead, as these sometimes co-occur with prosodic boundaries. We took a similar approach by generating syntactic annotations using the dependency parser (Canisius et al., 2006) included in the FROG natural language processing tool (Hendrickx et al., 2016). We then derived features from these annotations that have been shown to predict prosodic boundaries, such as intermediate or intonational phrase breaks (see features F2–F8 in Ingulfsen, 2004, pp. 36–38). In order to limit the number of prosodic boundary variables, we used a principle component analysis to identify 5 principle components, Prosody$_{PC1-5}$, that accounted for more than 94% of the variance described by the 7 original features.

We also considered the distributional characteristics of the words containing and surrounding the /s/. It has been shown, for instance, that words which are predictable given the surrounding words have more reduced realizations (e.g., Pluymaekers et al., 2005; Bell et al., 2009). As such, we used the NLCOW14 corpus (Schäfer, 2015) to measure the bigram frequency of the plural and the word preceding it in addition to the bigram frequency of the plural and the word following it. By dividing these respective bigram frequencies by the frequency of the plural form in the NLCOW14 corpus, we calculated conditional probabilities of the plural form given the preceding and subsequent word. These were log-transformed, resulting in Probability from previous word and Probability from next word, respectively. Similarly, whether or not a word has been recently mentioned may also affect its pronunciation

TABLE 4 | Overview of the corpora used in the durational study, including the number of -*s* plural tokens that were selected.

Register	Corpus	Reference	Tokens
Spontaneous conversation	Spoken Dutch corpus part a	Oostdijk, 2000	196
	Spoken Dutch corpus part c		47
	Spoken Dutch corpus part d		45
	Ernestus corpus of spoken Dutch	Ernestus, 2000	9
	IFA dialog video corpus	van Son et al., 2008	8
News broadcasts	Spoken Dutch corpus part k	Oostdijk, 2000	74
Read stories	Spoken Dutch corpus part o		215
		Total:	594

(e.g., Pluymaekers et al., 2005). This was encoded as the binary RECENTLY MENTIONED variable by checking whether the same plural had been uttered in the 30 seconds prior.

Another feature that may influence phonetic reduction concerns a word's phonological similarity to other words. This similarity has been implemented by counting the number of PHONOLOGICAL NEIGHBORS, which are the words that differ from the target word by one sound. Higher neighborhood density has been associated with both more and less reduced segments (see discussion in Gahl et al., 2012). For each plural, we used the pronunciation lexicon that came with the CLST forced aligner (Kuijpers et al., 2018) to find the number of lexical neighbors.

Finally, previous research has shown that more careful speech is associated with longer durations (e.g., van Son and Pols, 1999). We expected that some of the speech used in this study, such as the read-aloud stories, would be more careful compared to speech from spontaneous conversations. Consequently, speech REGISTER was the final influence on the duration of final /s/ that we considered. This variable had three levels: *Conversation*, *Stories* and *News*.

3.1.3. Modeling -s Duration

We used linear mixed effects regression, as implemented in the R package `lme4` (Bates et al., 2015), to model -S DURATION. By analyzing the effect of the interaction between -S BIAS and PLURAL DOMINANCE on -S DURATION we hoped to test our hypothesis that the effect of competition between plural variants on pronunciation is more noticeable if the plural variants are representationally strong relative to the singular. Additionally, we wanted to know how well our paradigmatic predictors explained differences in -S DURATION compared to alternative measures like the absolute -S FREQUENCY. As such, we created multiple models.

First, we fitted a *Paradigmatic model* containing -S BIAS, PLURAL DOMINANCE and their interaction term, all covariates, and random intercepts for SPEAKER and NOUN, which was the maximal random structure that was supported by the data. Additionally, we fitted two alternative models in which the -S BIAS and PLURAL DOMINANCE variables were replaced by alternative measures of representational strength. In the *Absolute frequency model* we replaced the paradigmatic measures with a single -S FREQUENCY predictor. We also fitted a *Relative frequency model*, in which we used the RELATIVE -S FREQUENCY measure. Using the AIC scores of the resulting three models, we calculated their relative likelihood to determine whether our paradigmatic predictors provided the best fit to the data.

Subsequently, we wanted to interpret the predictors of interest in our paradigmatic model. As such, we needed to avoid collinearity between our predictors of interest and any covariates. To assess the potential for collinearity in our data, we calculated correlations between all covariates and our predictors of interest; see **Supplementary Figure 3**. This showed us that -S BIAS was correlated (Pearson's $r \geq 0.4$) with the covariates WORD STRESS and NUMBER OF SYLLABLES. This was not very surprising, as both of these covariates can be related to the stress pattern of a noun, which has been shown to affect the choice of plural suffix (Baayen et al., 2002). Removing these covariates would make sure that they could not lead to collinearity issues. However, we wanted to make sure that any potential effect of -S BIAS and its interaction with PLURAL DOMINANCE would not actually be better modeled by the correlated covariates. Therefore, we fitted three linear regression models of -S DURATION: for -S BIAS, WORD STRESS and NUMBER OF SYLLABLES, respectively. Each model contained one of the three correlated variables, the PLURAL DOMINANCE variable and their interaction. An AIC comparison showed that the model containing -S BIAS performed best. As such we excluded the correlated covariates from further analysis. Starting from the resulting *Paradigmatic model*, we used backward elimination (as implemented in Kuznetsova et al., 2017) on to arrive at a model in which only the significant predictors remained. After fitting the model with the remaining variables, we trimmed the data with residuals that exceeded 2.5 standard deviations and refitted the model on the trimmed data set, following Baayen (2008). The residuals of this final model were approximately normally distributed, see **Supplementary Figure 4**.

3.2. Results

The full paradigmatic model of -S DURATION containing the -S BIAS × PLURAL DOMINANCE interaction had an AIC of 690.90. By comparison, the best performing alternative model, which contained the -S FREQUENCY predictor, had an AIC of 697.86; see **Supplementary Table 2** for full models. This means that the *Absolute frequency model* was $exp(\frac{690.90-697.86}{2}) = 0.031$

TABLE 5 | Mixed effects model of -s Duration.

Fixed effects	**Estimates**	**Std. Error**	***t*-value**	***p***
Intercept	−2.610	0.038	−68.979	0.000
Speech rate	−0.142	0.018	−7.891	0.000
Prosody$_{PC2}$	0.034	0.016	2.124	0.034
Next segment: Approximant	−0.334	0.071	−4.713	0.000
Next segment: Fricative	−0.187	0.047	−3.982	0.000
Next segment: Liquid	−0.219	0.191	−1.145	0.253
Next segment: Nasal	−0.077	0.078	−0.995	0.320
Next segment: Plosive	−0.096	0.070	−1.369	0.172
Next segment: Silence	0.562	0.042	13.523	0.000
Register: Stories	0.170	0.039	4.374	0.000
Register: News	0.016	0.069	0.233	0.817
-s Bias	−0.000	0.007	−0.058	0.954
Plural Dominance	−0.023	0.010	−2.432	0.015
-s Bias : *Plural Dominance*	0.017	0.004	4.271	0.000
Random effects	**Variance**			**Std. Deviation**
Speaker (Intercept)	0.019			0.137
Residual	0.125			0.353

p-values were calculated using Satterthwaite's method. Reference levels are Next segment: *Vowel and* Register: *Conversation.*

times as likely to minimize the information loss compared to the *Paradigmatic model* (Burnham and Anderson, 2004). In other words, the *Paradigmatic model* performed much better than the models with alternative measures of representational strength.

Table 5 summarizes the parameters of the final model, that is, the *Paradigmatic model* after removal of correlated covariates and insignificant predictors. In addition to the -S BIAS × PLURAL DOMINANCE interaction, this model contains the covariates SPEECH RATE, PROSODY$_{PC2}$, NEXT SEGMENT, and REGISTER and the random variable SPEAKER. As indicated by the estimates in **Table 5**, the covariates show the expected effects, e.g., a higher SPEECH RATE reduces -S DURATION and a subsequent *Silence* is associated with a longer -S DURATION. Importantly, **Table 5** also reveals a significant interaction between the predictors of interest, -S BIAS and PLURAL DOMINANCE. The fitted lines in **Figure 3** illustrate the estimated effect of -S BIAS on -S DURATION at three different values of PLURAL DOMINANCE (see section 2.2 for interpretation of these values). At high PLURAL DOMINANCE, the slope of the line is positive, which means that final *-s* becomes longer if -S BIAS becomes larger. This result supports the expected *paradigmatic enhancement* effect. Unexpectedly, we find the opposite effect at low PLURAL DOMINANCE: for these nouns, final *-s* becomes shorter as -S BIAS becomes larger. We expected that -S BIAS would have very little effect on -S DURATION at negative PLURAL DOMINANCE, resulting in a horizontal line. However, the model predicts that the *paradigmatic enhancement* effect is already nullified at a PLURAL DOMINANCE of zero. In noun paradigms with negative PLURAL DOMINANCE, a reduction effect is predicted.

4. DISCUSSION

The current research explored how paradigmatic structure relates to the mechanisms that are involved in the processing of complex words. Dutch variable plurals were chosen as the subject of inquiry, as they are involved in paradigmatic relations that have been associated with generalization, storage and coactivation mechanisms.

In our first study we investigated whether the singular-plural frequency relation of a noun influences the distribution of its plural variants in a Dutch subtitles corpus. We hypothesized that the distribution of variants for nouns with higher PLURAL DOMINANCE would be less predictable by a measure of phonological generalization. The results supported this account by showing that the positive effect of the generalization measure -S PREDICTION on the distributional measure -S BIAS decreases with higher values of PLURAL DOMINANCE. These findings are in line with previous accounts of *in*variable plurals (Baayen et al., 2008; Beyersmann et al., 2015) which suggest that higher plural dominance limits the influence of generalization on plural processing. Presumably, plural-dominant variable plurals are less affected by generalization because they have representations that are more stable or are easier to retrieve during the speech production process. The distributional results also contribute to the wider discussion about the role of token frequency in the generalization of morphological exponents. Whereas, previous distributional research has generally focused on absolute token frequency as an inhibitor of generalization (e.g., Cuskley et al., 2014), this study showed that frequency relative to the base form

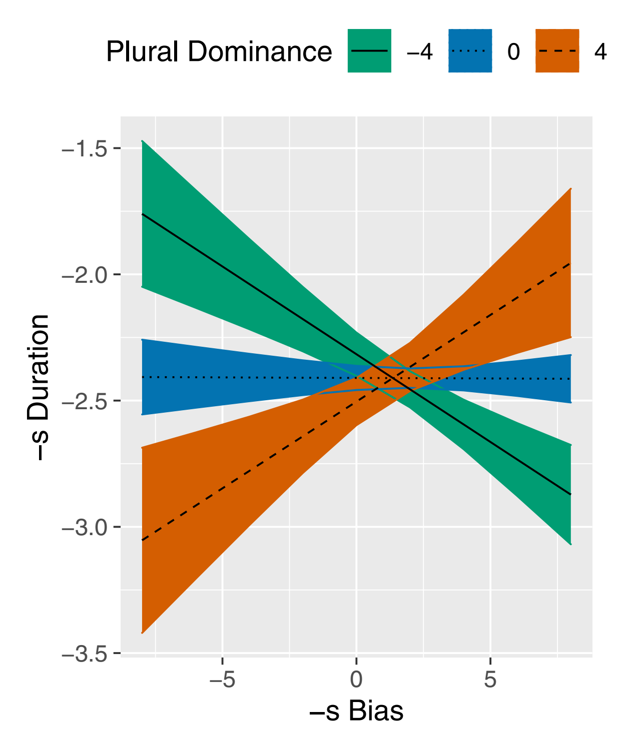

FIGURE 3 | Partial effect plot for the -S BIAS × PLURAL DOMINANCE interaction in the model of -S DURATION. Coloured bands reflect 95% confidence intervals.

may also affect the scope of general phonological patterns[1] (see also Tiersma, 1982; Collie, 2008).

The goal of the second study was two-fold. Firstly, we wanted to investigate whether the *paradigmatic enhancement hypothesis* applies to Dutch variable plurals. That is, an *-s* plural variant that is more frequent than its alternative should be phonetically enhanced compared to an *-s* variant that is less frequent than its alternative. This hypothesis was based on the assumption that the more frequent *-s* variant has a stronger representation and therefore its pronunciation is less affected by the coactivated representation of the alternative variant. Additionally, we hypothesized that such a paradigmatic enhancement effect should primarily occur in noun paradigms with relatively high PLURAL DOMINANCE. This qualification was based on the assumption that the representational strengths of plural variants primarily depend on their frequencies relative to the singular. As such, we expected that the differences in representational strengths measured by -S BIAS would be greatest at high PLURAL DOMINANCE. The results revealed an interaction effect of -S BIAS and PLURAL DOMINANCE on -S DURATION. For plural-dominant plurals, a higher -S BIAS was associated with a longer -S DURATION, which suggests that paradigmatic enhancement is reflected in our data. This finding supports previous accounts of paradigmatic enhancement that interpret the frequency relation between paradigmatic alternatives as a measure of their relative representational strengths Cohen (2014, 2015). Furthermore, the results showed that an increased -S BIAS was associated with a shorter -S DURATION for singular-dominant plurals, which was surprising as we expected that the pronunciation of infrequently pluralized plurals would not be affected by the frequency relation between variants. Nonetheless, this interaction effect was in line with our hypothesis that paradigmatic enhancement would primarily affect plural-dominant plurals. As such, this study is the first to provide evidence that, in certain paradigms, paradigmatic enhancement is mediated by the base-complex relation.

By combining the findings from both studies, we might better understand the unexpected reduction effect of -S BIAS on -S DURATION for singular-dominant plurals that was observed in our second study. The interpretation of the -S BIAS predictor is crucial to this understanding. The combined results suggest that what -S BIAS represents depends on the value of PLURAL DOMINANCE. At high PLURAL DOMINANCE, the paradigmatic enhancement effect in the durational study suggests that -S BIAS is a measure of the representational strength of the *-s* variant relative to its competitor. However, at low PLURAL DOMINANCE, the distributional study suggests that -S BIAS represents the amount of phonological support from similar paradigms, i.e., the -S PREDICTION. In formulating the hypotheses for the duration study, we did not consider that increased analogical support could result in the reduction of final *-s*, given the lack of precedents for such an effect (but see *gang size* effect in Tucker et al., 2019). However, the association of reduced final *-s* with increased -S PREDICTION would fit the more general theory that predictable linguistic elements are reduced (e.g., Bell et al., 2009). Importantly, as this explanation assumes that -S BIAS primarily reflects -S PREDICTION for singular-dominant nouns, it does not conflict with our account of paradigmatic enhancement, which mostly affects frequently pluralized nouns.

The combined results have implications for psycholinguistic models of morphological processing. These models can be categorized according to the relative importance they attribute to abstract rules and lexical storage (see the overviews in Arndt-Lappe and Ernestus, 2020; Fábregas and Penke, 2020). At one end of the spectrum are models that emphasize the role of rules in explaining the paradigmatic structure that arises from commonalities in form and function among the words of a language. In these models, complex words are only stored if they do not submit to morphological rules (e.g., Wunderlich, 1996). Such models often assume that stored exceptions to the rule do not influence regular application of the rule. Our results suggest that the base-complex frequency relation indicates the extent to which variable plurals follow the morpho-phonological rules. As this frequency relation must be stored somehow, either in representations of individual nouns or in

[1] A reviewer pointed out that the interaction effect of our absolute token frequency measure, PLURAL FREQUENCY, and our generalization measure, -S PREDICTION, seems to suggest that higher token frequency facilitates generalization, which would be completely contrary to previous findings. However, this interaction is confounded by sampling bias: a frequent plural is more likely to show variation compared to an infrequent plural, especially when the general patterns strongly favour a particular variant. As we only included plurals that showed variation in our data set, this resulted in a better match between -S PREDICTION and observed -S BIAS for high frequency plurals. The interaction between the absolute token frequency and the generalization measure was included to account for the effect of this bias in the estimation of the other predictors, see section 2.1.3.

weighted connections between morphological exponents and specific semantic representations, the current research questions the complete separation of generalization and storage in rule-based models.

In a second category of models, both abstract rules and lexical storage may affect the production of complex words (see Arndt-Lappe and Ernestus, 2020). In one such a model, the *Parallel Dual Route Model* (e.g., Baayen et al., 1997b), production of a complex word involves simultaneous retrieval of the complex representation and composition involving the base representation. The relative speed of the composition and retrieval routes determines which route affects production the most. In some dual route models, complex words that are less frequent than their bases are assumed to be more easily (de)composable Hay (2001), which would speed up the (de)composition route. This conceptualization of the base-complex relation can be applied to the PLURAL DOMINANCE variable in our studies. It would predict that singular-dominant plurals are produced using the composition route, whereas plural-dominant plurals are primarily produced using the retrieval route. Such an account would explain why the distribution of plural variants in singular-dominant noun paradigms follows phonological patterns. It would also explain why plural variants in plural-dominant noun paradigms are subject to paradigmatic enhancement. At high PLURAL DOMINANCE, an *-s* variant with high -S BIAS would be more frequent than the singular. As such it would be produced using the retrieval route, and its pronunciation would not be affected by the alternative variant. However, an *-s* variant with low -S BIAS would likely be less frequent than the singular while its alternative would be more frequent compared to the singular. The *-s* variant would then be produced using the composition route, and its pronunciation would be affected by the alternative variant, which was simultaneously activated through the retrieval route.

In a third category of models, processing of complex words involves no abstract computation. In those word-based models (e.g., Bybee, 1995), paradigmatic enhancement findings are easily accounted for, as all word forms including their frequencies of occurrence can be stored. Analogy between stored word forms can be used to explain morpho-phonological patterns across paradigms. Our first study showed that such an analogical mechanism can also account for variation observed for Dutch variable plurals. The reduced influence of analogy on the production of plural-dominant nouns can be explained through a weaker activation level of the singular representation relative to the plural representation: a relatively infrequent singular form results in decreased activation of a noun's singular representation, which, in turn, leads to decreased analogical influence of other noun paradigms with phonologically similar singular forms. As such, models without a separate rule-based processing route can account for the Dutch variable plural data as well.

The current findings shed light on how paradigmatic relations may be related to the mechanisms that are involved in the processing of complex words. While the results cannot be explained by the mechanisms of a primarily rule-based model, both a dual-route model and a word-based model are compatible with the results. Regardless of theoretical framework, the novel implication of this research is that the role of the base-complex relation, whether it is conceptualized using activation levels or (de)composability, should be considered when the effect of additional within-paradigm relations, such as those between plural variants, are investigated. It follows that measures which conflate base-complex relations and relations among paradigmatic alternatives, such as form frequency relative to lexeme frequency, might not adequately capture how processing mechanisms interact. This was evidenced in our durational study by the fact that the model which distinguished between -S BIAS and PLURAL DOMINANCE predictors performed much better than the model that combined them into a single RELATIVE -S FREQUENCY predictor. More generally, these findings show that the nature of the individual morphological relations within a paradigm should be considered when their effect on processing is investigated.

In addition to providing answers about paradigmatic relations, our findings also raise questions. This research was concerned with paradigmatic relations and their psycholinguistic relevance during speech production. It would therefore be interesting to know whether our interpretations of the -S BIAS, PLURAL DOMINANCE and -S PREDICTION relations are representative for the processing mechanisms of individual speakers. However, these relations were measured using type and token frequencies from corpus data. As Blumenthal-Dramé (2012) points out, corpus frequencies do not necessarily reflect the input frequencies of individual language users, but rather a simplified and likely biased approximation of the input of multiple language users. With regard to Dutch plurals in particular, it seems unlikely that all speakers encounter and/or produce the different variants of a plural with the same -S BIAS. Presumably, this also leads to differences among speakers in the processing of variable plurals. It is therefore likely that the paradigmatic effects found in this research do not affect the speech of all language users equally. This is particularly true for the distributional study, as it does not relate the paradigmatic measures to the production of individual speakers. Unfortunately, the small size of our data set meant that we could not investigate inter-speaker differences in the paradigmatic enhancement effect. Additionally, due to the nature of the data, we could not take other potentially relevant factors, such as register, into account in our distributional study. These issues may be addressed by studies with better control over the relevant variables.

Furthermore, the findings from the durational study are primarily relevant for a narrow definition of paradigmatic enhancement. In this account, the coactivation resulting in paradigmatic enhancement only involves paradigm members that occur in the same linguistic context. In other words, the context works as a filter that determines which representations are coactivated: in utterances like *the boy runs/run/running* only one paradigm member (*runs*) is likely and therefore no paradigmatic enhancement effect would be expected. As such, this account does not provide clear explanations of paradigmatic enhancement effects on forms that can be predicted from the communicative context (e.g., Kuperman et al., 2007; Schuppler et al., 2012). Our research does provide naturalistic support for previous experimental findings of paradigmatic enhancement in which the linguistic context was controlled to allow for multiple paradigm members (e.g., Cohen, 2014, 2015; Bell et al., 2020; Tomaschek et al., 2021b).

Regardless of the limitations of the current research, its relevance is not limited to obscure morphological alternations. As documented by work on morphological overabundance (e.g., Thornton, 2019), the existence of paradigmatic alternatives is far from exceptional. As such, this research paves the way for similar investigations of paradigmatic relations using other overabundance phenomena. Apart from highlighting the underexplored variation in the realization of complex words, such research would contribute to morphological theory by identifying paradigmatic effects on processing that must be accounted for by psycho-linguistic models. As this research has emphasized, those paradigmatic effects can only be understood if paradigmatic relations are considered both individually and taken together.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.17026/dans-xvr-qscf.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Assessment Committee Humanities Radboud University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TZ conducted the research and wrote the initial draft of the manuscript. LB, ME, and IP provided feedback and suggestions for the research, and contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Deutsche Forschungsgemeinschaft (Research Unit FOR2373 'Spoken Morphology,' Project ER 547/1-1 'Dutch morphologically complex words: The role of morphology in speech production and comprehension'), which we gratefully acknowledge.

ACKNOWLEDGMENTS

We acknowledge the authors of the speech corpora for providing access to their data, the Centre for Language and Speech Technology at the Radboud University for providing technical support, the reviewers for providing helpful feedback, and Taki Bremmers for the example in (1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.720017/full#supplementary-material

REFERENCES

Arndt-Lappe, S. (2014). Analogy in suffix rivalry: the case of English -ity and -ness. *Engl. Lang. Linguist*. 18, 497–548. doi: 10.1017/S136067431400015X

Arndt-Lappe, S., and Ernestus, M. (2020). "Morpho-phonological alternations: the role of lexical storage," in *Word Knowledge and Word Usage*, eds V. Pirrelli, I. Plag and W. U. Dressler (Berlin: De Gruyter Mouton), 191–227.

Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech*. 47, 31–56. doi: 10.1177/00238309040470010201

Aylett, M., and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc Am*. 119, 3048–3058. doi: 10.1121/1.2188331

Baayen, H., Burani, C., and Schreuder, R. (1997a). "Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian," in *Yearbook of Morphology 1996* (Dordrecht: Kluwer), 13–34.

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997b). 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., Levelt, W., Schreuder, R., and Ernestus, M. (2008). "Paradigmatic structure in speech production," in *Proceedings From the Annual Meeting of the Chicago Linguistic Society, Vol. 43*, eds M. Elliott, J. Kirby, O. Sawada, E. Staraki and S. Yoon (Chicago: Chicago Linguistic Society), 1–29.

Baayen, R. H., McQueen, J. M., Dijkstra, T., and Schreuder, R. (2003). "Frequency effects in regular inflectional morphology: revisiting Dutch plurals," in *Morphological Structure in Language Processing*, eds R. H. Baayen and R. Schreuder (Berlin: De Gruyter Mouton), 355–390.

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1996). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium. doi: 10.35111/gs6s-gm48

Baayen, R. H., Schreuder, R., De Jong, N., and Krott, A. (2002). "Dutch inflection: the rules that prove the exception," in *Storage and Computation in the Language Faculty*, eds S. G. Nooteboom, F. Weerman and F. N. K. Wijnen (Dordrecht: Kluwer), 61–92.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw*. 67, 1–48. doi: 10.18637/jss.v067.i01

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang*. 60, 92–111. doi: 10.1016/j.jml.2008.06.003

Bell, M. J., Ben Hedia, S., and Plag, I. (2020). How morphological structure affects phonetic realisation in English compound nouns. *Morphology* 31, 87–120. doi: 10.1007/s11525-020-09346-6

Beyersmann, E., Dutton, E. M., Amer, S., Schiller, N. O., and Biedermann, B. (2015). The production of singular- and plural-dominant nouns in Dutch. *Lang. Cogn. Neurosci*. 30, 867–876. doi: 10.1080/23273798.2015.1027236

Biedermann, B., Beyersmann, E., Mason, C., and Nickels, L. (2013). Does plural dominance play a role in spoken picture naming? a comparison of unimpaired and impaired speakers. *J. Neurolinguist*. 26, 712–736. doi: 10.1016/j.jneuroling.2013.05.001

Blumenthal-Dramé, A. (2012). *Entrenchment in Usage-Based Theories: What Corpus Data do and do Not Reveal About the Mind*. Berlin: Walter De Gruyter.

Burnham, K. P., and Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res*. 33, 261–304. doi: 10.1177/0049124104268644

Bybee, J. (1995). Regular morphology and the lexicon. *Lang. Cogn. Process* 10, 425–455. doi: 10.1080/01690969508407111

Bybee, J., and Hopper, P. (2001). "Introduction to frequency and emergence of linguistic structure," in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Amsterdam: John Benjamins Publishing), 1–24.

Canisius, S., Bogers, T., Van Den Bosch, A., Geertzen, J., and Sang, E. T. K. (2006). "Dependency parsing by inference over high-recall dependency predictions," in *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)* (New York, NY), 176–180.

Cho, T., and McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. *J. Phon*. 33, 121–157. doi: 10.1016/j.wocn.2005.01.001

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verbal Learn. Verbal Behav*. 12, 335–359. doi: 10.1016/S0022-5371(73)80014-3

Cohen, C. (2014). Probabilistic reduction and probabilistic enhancement. *Morphology* 24, 291–323. doi: 10.1007/s11525-014-9243-y

Cohen, C. (2015). Context and paradigms: Two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *Ment Lex*. 10, 313–338. doi: 10.1075/ml.10.3.01coh

Collie, S. (2008). English stress preservation: the case for 'fake cyclicity'. *Engl. Lang. Linguist*. 12, 505–532. doi: 10.1017/S1360674308002736

Cuskley, C. F., Pugliese, M., Castellano, C., Colaiori, F., Loreto, V., and Tria, F. (2014). Internal and external dynamics in language: evidence from verb regularity in a historical corpus of English. *PLoS ONE* 9:e102882. doi: 10.1371/journal.pone.0102882

Daelemans, W., Zavrel, J., van der Sloot, K., and van den Bosch, A. (2018). *TiMBL: Tilburg Memory-Based Learner, Version 6.4*. Technical Report 11-01, Induction of Linguistic Knowledge Research Group, Tilburg University. Reference Guide.

de Haas, W., and Trommelen, M. (1993). *Morfologisch Handboek van het Nederlands: Een Overzicht van de Woordvorming*. The Hague: SDU Uitgeverij.

Divjak, D., and Caldwell-Harris, C. L. (2019). "Frequency and entrenchment," in *Cognitive Linguistics*, eds E. Dąbrowska and D. Divjak (Berlin: De Gruyter), 61–86.

Ernestus, M. (2000). *Voice Assimilation and Segment Reduction in Casual Dutch*. Utrecht: Holland Institute of Generative Linguistics.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142, 27–41. doi: 10.1016/j.lingua.2012.12.006

Ernestus, M., and Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79, 5–38. doi: 10.1353/lan.2003.0076

Fábregas, A., and Penke, M. (2020). "Word storage and computation," in *Word Knowledge and Word Usage*, eds V. Pirrelli, I. Plag, and W. U. Dressler (Berlin: De Gruyter Mouton), 455–505.

Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *J. Mem. Lang*. 66, 789–806. doi: 10.1016/j.jml.2011.11.006

Goeman, T., de Schutter, G., van den Berg, B., and de Jong, T. (2005). *Morfologische Atlas van de Nederlandse Dialecten, Vol. 1*. Amsterdam: Meertens Instituut; KANTL; Amsterdam UP.

Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychol. Rev*. 105, 251. doi: 10.1037/0033-295X.105.2.251

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., and Van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst, 2nd Edn*. Groningen; Deurne: Martinus Nijhoff uitgevers; Wolters Plantyn.

Hanique, I., and Ernestus, M. (2012). The role of morphology in acoustic reduction. *Lingue e Linguaggio* 11, 147–164. doi: 10.1418/38783

Harrison, X. A. (2015). A comparison of observation-level random effect and beta-binomial models for modelling overdispersion in binomial data in ecology evolution. *PeerJ*. 3:e1114. doi: 10.7717/peerj.1114

Hay, J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics* 39, 1041–1070. doi: 10.1515/ling.2001.041

Hay, J. (2007). "The phonetics of 'un'," in *Lexical Creativity, Texts and Contexts*, ed J. Munat (Amsterdam: John Benjamins), 39–57.

Hendrickx, I., van den Bosch, A., van Gompel, M., and van der Sloot, K. (2016). *Frog, a Natural Language Processing Suite for Dutch*. Nijmegen: Reference guide, Radboud University Nijmegen.

Hofhuis, E., Gussenhoven, C., and Rietveld, T. (1995). "Final lengthening at prosodic boundaries in Dutch," in *Proceedings of the XIIIth International Congres of Phonetic Sciences, Vol. 1*, eds E. Elenius and P. Branderud (Stockholm: Stockholm University), 154–157.

Ingulfsen, T. (2004). *Influence of syntax on prosodic boundary prediction*. Technical Report 610, University of Cambridge Computer Laboratory.

Keuleers, E., Brysbaert, M., and New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behav. Res. Methods* 42, 643–650. doi: 10.3758/BRM.42.3.643

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., and Martens, E. (2007). Dutch plural inflection: the exception that proves the analogy. *Cogn. Psychol*. 54, 283–318. doi: 10.1016/j.cogpsych.2006.07.002

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J. Acous.t Soc. Am*. 59, 1208–1221. doi: 10.1121/1.380986

Krott, A., Baayen, R. H., and Schreuder, R. (2001). Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* 39, 51–93. doi: 10.1515/ling.2001.008

Kuijpers, L., Ganzeboom, M., and Wei, X. (2018). *CLST ASR Forced Aligner*. Available online at: https://www.cls.ru.nl/clst-asr/doku.php?id=forced-aligner (accessed April 22, 2021).

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *J. Acoust. Soc. Am*. 121, 2261–2271. doi: 10.1121/1.2537393

Kürschner, S. (2009). Morphological non-blocking in Dutch plural allomorphy: a contrastive approach. *Lang. Typol. Univer*. 62, 285–306. doi: 10.1524/stuf.2009.0022

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw*. 82, 1–26. doi: 10.18637/jss.v082.i13

Lesnoff, M., and Lancelot, R. (2018). *aods3: Analysis of Overdispersed Data Using S3 Methods*. aods3 package version 0.4-1.1.

Luce, P. A., and Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *J. Acous. Soc. Am*. 78, 1949–1957. doi: 10.1121/1.392651

Milin, P., Đurđević, D. F., and del Prado Martín, F. M. (2009). The simultaneous effects of inflectional paradigms and classes on lexical recognition: evidence from Serbian. *J. Mem. Lang*. 60, 50–64. doi: 10.1016/j.jml.2008.08.007

Mortensen, D. R., Littell, P., Bharadwaj, A., Goyal, K., Dyer, C., and Levin, L. (2016). "PanPhon: a resource for mapping IPA segments to articulatory feature vectors," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (Osaka: The COLING 2016 Organizing Committee), 3475–3484.

New, B., Brysbaert, M., Segui, J., Ferrand, L., and Rastle, K. (2004). The processing of singular and plural nouns in French and English. *J. Mem. Lang*. 51, 568–585. doi: 10.1016/j.jml.2004.06.010

Nooteboom, S. G. (1972). *Production and Perception of Vowel Duration*. Utrecht: Utrecht University.

Oostdijk, N. (2000). "The Spoken Dutch Corpus: Overview and first evaluation," in *LREC-2000 Proceedings* (Athens), 887–894.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: the acoustics of word-final S in English. *J. Linguist*. 53, 181–216. doi: 10.1017/S0022226715000183

Plag, I., Lohmann, A., Ben Hedia, S., and Julia, Z. (2020). "An S is an 'S, or is it? plural and genitive plural are not homophonous," in *Complex Words: Advances in Morphology*, eds L. Körtvélyessy and P. Štekauer (Cambridge: Cambridge University Press), 260–292.

Pluymaekers, M., Ernestus, M., and Baayen, R. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146–159. doi: 10.1159/000090095

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., et al. (2011). "The Kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* (Hawaii: IEEE Signal Processing Society; IEEE Catalog No.: CFP11SRW-USB).

Schäfer, R. (2015). "Processing and querying large web corpora with the COW14 architecture," in *Challenges in the Management of Large Corpora (CMLC-3)* (Lancaster), 28.

Schuppler, B., Ernestus, M., Scharenborg, O., and Boves, L. (2011). Acoustic reduction in conversational Dutch: a quantitative analysis based on automatically generated segmental transcriptions. *J. Phon*. 39, 96–109. doi: 10.1016/j.wocn.2010.11.006

Schuppler, B., van Dommelen, W. A., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: studies at the phonemic and sub-phonemic level. *J. Phon*. 40, 595–607. doi: 10.1016/j.wocn.2012.05.004

Thornton, A. M. (2019). "Overabundance in morphology," in *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press.

Tiersma, P. M. (1982). Local and general markedness. *Language* 58, 832–849. doi: 10.2307/413959

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2021a). Phonetic effects of morphology and context: modeling the duration of word-final s in English with naïve discriminative learning. *J. Linguist*. 57, 123–161. doi: 10.1017/S0022226719000203

Tomaschek, F., Tucker, B. V., Ramscar, M., and Baayen, R. H. (2021b). Paradigmatic enhancement of stem vowels in regular English inflected verb forms. *Morphology* 31, 171–199. doi: 10.1007/s11525-021-09374-w

Tucker, B. V., Sims, M., and Baayen, R. H. (2019). Opposing forces on acoustic duration. *PsyArXiv [Preprint]*. doi: 10.31234/osf.io/jc97w

Van Dale (2020). *Gratis woordenboek*. Available online at: https://www.vandale.nl/opzoeken (accessed April 22, 2021).

van Son, R., and Pols, L. C. (1999). An acoustic description of consonant reduction. *Speech Commun*. 28, 125–140. doi: 10.1016/S0167-6393(99)00009-6

van Son, R., and Pols, L. C. (2003). How efficient is speech. *Proc. Inst. Phon. Sci*. 25, 171–184.

van Son, R., Wesseling, W., Sanders, E., and van den Heuvel, H. (2008). "The IFADV corpus: a free dialog video corpus," in *LREC-2008 Proceedings*, (Marrakech) 501–508.

Walsh, M., Möbius, B., Wade, T., and Schütze, H. (2010). Multilevel exemplar theory. *Cogn. Sci*. 34, 537–582. doi: 10.1111/j.1551-6709.2010.01099.x

Walsh, T., and Parker, F. (1983). The duration of morphemic and non-morphemic /s/ in English. *J. Phon*. 11, 201–206. doi: 10.1016/S0095-4470(19)30816-2

Wilde Haren De Podcast (2019). *Huub Pragt: WHDP S06E05*. Available online at: https://www.youtube.com/watch?v=fBoIBI_hlUk (accessed April 22, 2021).

Wulf, D. J. (2002). "Applying analogical modeling to the German plural," in *Analogical Modeling*, eds R. Skousen, D. Lonsdale and D. B. Parkinson (Amsterdam: John Benjamins), 109–122.

Wunderlich, D. (1996). "Minimalist morphology: the role of paradigms," in *Yearbook of Morphology 1995*, eds G. Booij and J. van Marle (Dordrecht: Springer), 93–114.

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

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fpsyg.2021.720228

Form and Function: A Study on the Distribution of the Inflectional Endings in Italian Nouns and Adjectives

Valentina Nicole Pescuma[1†], Chiara Zanini[2†], Davide Crepaldi[1] and Francesca Franzon[2]*

[1] Neuroscience Area, International School for Advanced Studies, Trieste, Italy, [2] Romanisches Seminar, University of Zurich, Zurich, Switzerland

Inflectional values, such as singular and plural, sustain agreement relations between constituents in sentences, allowing sentence parsing and prediction in online processing. Ideally, these processes would be facilitated by a consistent and transparent correspondence between the inflectional values and their form: for example, the value of plural should always be expressed by the same ending, and that ending should only express plural. Experimental research reports higher processing costs in the presence of a non-transparent relation between forms and values. While this effect was found in several languages, and typological research shows that consistency is far from common in morphological paradigms, it is still somewhat difficult to precisely quantify the transparency degree of the inflected forms. Furthermore, to date, no accounts have quantified the transparency in inflection with regard to the declensional classes and the extent to which it is expressed across different parts of speech, depending on whether these act as controllers of the agreement (e.g., nouns) or as targets (e.g., adjectives). We present a case study on Italian, a language that marks gender and number features in nouns and adjectives. This work provides measures of the distribution of forms in the noun and adjective inflection in Italian, and quantifies the degree of form-value transparency with respect to inflectional endings and declensional classes. In order to obtain these measures, we built Flex It, a dedicated large-scale database of inflectional morphology of Italian, and made it available, in order to sustain further theoretical and empirical research.

Keywords: grammatical gender, grammatical number, adjective inflection, noun inflection, declensional classes, inflectional morphology, language resource, contextual and inherent inflection

OPEN ACCESS

Edited by:
Robert Malouf,
San Diego State University,
United States

Reviewed by:
Andreas Opitz,
Leipzig University, Germany
Sendy Caffarra,
Stanford University, United States
Noémie Auclair-Ouellet,
McGill University, Canada

***Correspondence:**
Francesca Franzon
ffranzon@sissa.it

[†]These authors share first authorship

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

Received: *03 June 2021*
Accepted: *09 September 2021*
Published: *07 October 2021*

Citation:
Pescuma VN, Zanini C, Crepaldi D and Franzon F (2021) Form and Function: A Study on the Distribution of the Inflectional Endings in Italian Nouns and Adjectives. Front. Psychol. 12:720228. doi: 10.3389/fpsyg.2021.720228

1. INTRODUCTION

Languages can express grammatical features through inflectional morphology. For instance, in English the singular and plural values of the grammatical feature of number can be expressed through the forms *apple* (SG, singular) and *apples* (PL, plural), whereby the plural form is realized through the ending *-s*. On the language processing side, the relevance of the role of inflectional features for comprehension is attested, for instance, by the ability to pick up inflectional regularities from the first stages of language development shown by children as young as 12 months (Ferry et al., 2020). One might expect these processes to be enabled and facilitated by consistency in the correspondence between an inflectional feature's value and a word form. For example, the value of plural should ideally always be expressed by the same affix in a certain language (for a review: Huettig et al., 2011). In line with this account, transparency does appear to facilitate the acquisition of inflectional features, as shown in a recent study with Bulgarian- and Russian-speaking children on the acquisition of grammatical gender (Ivanova-Sullivan and Sekerina, 2019). Similarly, it has been noted that, in sentence comprehension, speakers of morphologically rich languages (like Italian or German) are more likely to use inflectional cues than speakers of languages having highly constrained word order (like English; Bates et al., 1982; MacWhinney et al., 1984; MacWhinney and Bates, 1989) and that in second language acquisition less proficient speakers are more likely to rely on ending cues than more proficient speakers and, as a result, are faster and more accurate in retrieving the gender of nouns whose endings transparently convey the corresponding morphological value (e.g., for German-English bilinguals: Bordag et al., 2006; for Basque-Spanish bilinguals: Caffarra et al., 2017). Unsurprisingly, a facilitation in the processing of grammatical gender information when the relation between ending and value is transparent or regular has been observed in a wealth of studies, comprising behavioral paradigms (e.g., Bates et al., 1995, 1996; Taft and Meunier, 1998; Gollan and Frost, 2001; De Martino et al., 2011), electrophysiological (e.g., Caffarra et al., 2015) and neural evidence (e.g., Miceli et al., 2002; Russo et al., 2021), including studies on aphasia and semantic dementia (Luzzatti and De Bleser, 1996; Lambon Ralph et al., 2011; Franzon et al., 2013).

1.1. The Form-Function Inconsistency Issue

However, consistency is not always observed in the inflectional paradigms of natural languages (Corbett, 2006). In fact, a lack of transparency between forms and feature values is more the rule than the exception (e.g., one value expressed through several different endings, different values expressed through the same ending, or values apparently expressed by no ending). For example, in English the plural value is not always conveyed by the final *-s*. Cases of allomorphy like *ox/oxen*, suppletivism like *child/children*, and apophony like *foot/feet* are not infrequent, such that the very Bloomfieldian notion of morphemes as the smallest linguistic units bearing meaning (Bloomfield, 1933) has been questioned (Matthews, 1974; Anderson, 1992; Aronoff, 1994; Baayen et al., 2011). Although these forms can be seen as sub-regularities (due to the fact that they are fossils of grammatical rules that are no longer active in a synchronic perspective, and therefore no longer productive; e.g., Anderson, 1992), they are nonetheless "irregular" since to say that a form has a regular inflection is to say that it has the inflection one would expect unless one knew that it was different (Matthews, 1991, p. 130).

Some accounts suggest that the presence of irregular inflectional paradigms, which may initially yield errors related to an over-generalization of regular patterns, ultimately supports learning processes (Ramscar et al., 2018). Furthermore, as noted in relation to verb inflection by Marzi et al. (2019), while one would expect maximal contrast between forms to yield immediate discrimination and recognition of inflectional values, this has a cost in terms of the storage space required for too many different forms. The coexistence of regular and irregular forms within the language has indeed been ascribed to an inevitable trade-off between maximal discriminability, on the one hand, and a degree of regularity sufficient to allow successful generalization, on the other (Blevins et al., 2017). A way in which this relation between ambiguity and informativeness of inflectional systems has been operationalized is the implementation of entropy metrics (Dye et al., 2017; Mickus et al., 2019; Williams et al., 2020; Franzon and Zanini, 2021) as defined by Shannon, 1948[1]. In this sense, entropy allows to quantify the probability for a feature to be associated to one or more given forms, and vice versa, assessing the consistency of this association.

1.2. Noun and Adjective Forms in Italian

The inflectional system of Italian nouns and adjectives comprises four combinations of inflectional values (i.e., masculine singular, masculine plural, feminine singular, and feminine plural). However, noun inflection in Italian has hardly been investigated (Franzon and Zanini, 2021), and an account of adjective inflection in Italian is completely missing, up to date. Furthermore, although the reason why form-value inconsistency occurs for inflectional features is still debated, form-value inconsistency in Italian inflection has hardly ever been quantified in these terms. Given that the notion of transparency is pivotal for psycholinguistic accounts describing the architecture of the mental lexicon (Crepaldi et al., 2010; Davis and Rastle, 2010; Amenta and Crepaldi, 2012; Marelli et al., 2015; Milin et al., 2017; Marelli and Amenta, 2018), measures of the transparency of the inflectional systems can significantly contribute to the understanding of how words are processed both in isolation and in sentence contexts. In the present study, a first step is taken to assess the extent to which inflectional forms consistently represent a given value in Italian. We will assess how the inflected forms of nouns and adjectives are distributed within the finite set of the combinations of feature values of gender and number.

Indeed, Italian nouns and adjectives are necessarily inflected for number (singular vs. plural) and gender (masculine vs. feminine), whose values are both expressed in a single fusive ending [e.g., *gatto* "cat(M).SG," *gatta* "cat(F).SG," *gatti* "cat(M).PL," *gatte* "cat(F).PL"]. Crucially, the endings of Italian nouns and adjectives cannot be considered unambiguous formal cues for gender and number values; nonetheless, the correspondence between forms and functions displays some recurrent patterns. Nouns have traditionally been divided into declensional classes according to the inflectional endings of their singular and plural forms. Considering a declensional class as a set of lexemes whose members each select the same set of inflectional realizations (Aronoff, 1994, p. 64), six declensional classes have been described for nouns in Italian (Iacobini and Thornton, 2016, p. 195): Class I (SG: *-o*; PL: *-i*, *libro - libri* "book

[1] $H = -\sum_x p(x) \log_2 p(x)$ where $p(x)$ is the probability of occurrence of a given word form.

- books"); Class II (SG: *-a*; PL: *-e, rosa - rose* "rose - roses"); Class III (SG: *-e*; PL: *-i, fiore - fiori* "flower - flowers"); Class IV (SG: *-a*; PL: *-i, problema - problemi* "problem - problems"); Class V (SG: *-o*; PL: *-a, uovo - uova* "egg - eggs"); Class VI (invariable nouns, various endings: e.g., *re* "king / kings"). For adjectives, five declensional classes have been identified (Iacobini and Thornton, 2016, p. 204): Class I (M.SG: *-o*; M.PL: *-i*; F.SG: *-a*; F.PL: *-e, bello - belli - bella - belle* "beautiful"); Class II (M.SG and F.SG: *-e*; M.PL and F.PL: *-i, grande - grandi* "big"); Class III (M.SG and F.SG: *-a*; M.PL: *-i*; F.PL: *-e, belga - belgi - belghe* "Belgian"); Class IV (M.SG and F.PL: *-e*; M.PL: *-i*; F.SG: *-a, sornione - sornioni - sorniona* "seemingly friendly"); Class V (invariable adjectives, various endings: e.g., *blu* "blue").

Noun Classes I and II are quite transparent with respect to gender features (comprising, respectively, mostly feminine and mostly masculine nouns), and so is adjective Class I. However, there is no straightforward correspondence between declensional classes and gender features. This entails that, considering the whole declensional system, no ending is unambiguously related to one value, and likewise no value is unambiguously related to one ending. This is possibly due to the fact that Italian, unlike languages such as English or Spanish, has a non-additive, non-sigmatic plural and, in general, its words must end with a vowel. As such, Italian noun and adjective forms are distributed in a narrow space subtended by just four vowels: *-o*, *-a*, *-e*, *-i*. In principle, a speaker exposed to a novel noun ending in *-e*, in the absence of other cues (such as an article or any other determiner), would not be able to disentangle whether the noun is a feminine plural of the first class like *sedie* "chairs," a feminine singular of the third class like *tigre* "tiger," or a masculine singular of the third class like *elefante* "elephant." Similarly, the masculine singular value is realized with different endings, such as *-o* (*divano* "couch") and *-e* (*elefante* "elephant"), *-a* (*problema* "problem").

We are aware that in our experience as speakers and readers we are hardly exposed to nouns in isolation. Therefore, a transparent form-value relation may not be a necessary nor a sufficient cue to sustain learning processes. Indeed, inflection plays a functional role in establishing morpho-syntactic agreement (e.g., *the apple*.SG *is*.SG *red* vs. *the apples*.PL *are*.PL *red*). Agreement can be described as the systematic covariance between a semantic or formal property of one element and a formal property of another (Steele, 1978, p. 610). It has been noted that agreement involving inflectional features, such as the feature of number with its singular and plural values, allows to disambiguate the relations between words in sentence parsing, reducing processing effort by favoring word predictions (Wicha et al., 2004; Huettig et al., 2011; Dye et al., 2017). More precisely, nouns are generally the "controllers," i.e., the elements that determine the agreement and whose expression of agreement features is usually covert. On the other hand, adjectives (as well as other functional elements such as articles) are "targets," i.e., the elements whose form is determined by the controllers (Corbett, 2006). In turn, this relates to another related aspect, that is, the difference between inherent and contextual inflection proposed in theoretical linguistics accounts (Booij, 1993, 1996; Di Domenico, 1997), and seldom explored experimentally (De Vincenzi and Di Domenico, 1999; Franzon et al., 2014). Inherent and contextual inflection are here exemplified, respectively, by nouns, which have an inherent, context-autonomous gender, and determine the form of other parts of speech, and adjectives whose gender and number will be determined by those of the noun they are related to. As we will discuss in section 4.1.3, this entails interesting differences in the distribution of inflectional features of Italian nouns and adjectives. It follows that less variability is expected in the target, i.e., the adjective forms having contextual inflection, since gender and number play here a merely functional, context-driven role and, as such, on the computational side, can serve more for prediction purposes allowing a maximal discriminability between gender and number values. A new metric therefore appears more suitable to quantify form-value consistency, while moving away from binary, categorical and non-quantifiable distinctions such as "transparent vs. opaque" or "regular vs. irregular."

1.3. Objectives of the Study

In Italian, studies concerning nominal inflection or nominal agreement have often relied on the morphological competence of the experimenters in controlling the transparency of the stimuli selection, even when the processing of inflected word forms was a central part of the study (Luzzatti and De Bleser, 1996; Caffarra et al., 2015; Franzon et al., 2016; Arcara et al., 2019; Zanini et al., 2020). This shortcoming has been likely due to the long-standing unavailability of suitable linguistic resources to measure noun transparency. To our knowledge, a resource for nominal inflection in Italian was released only recently: the database DeGNI (De Martino et al., 2019), which is based on the Colfis corpus (Bertinetto et al., 2005), containing type frequency information for mostly singular forms. Token frequency information, which is considered a better estimate of actual language use, is not provided.

The present work aims at providing an account of the distributional properties of noun and adjective inflection in Italian, to quantify the degree of form-value transparency and to investigate the distribution of forms across inflectional values and declensional classes. In order to compute such metrics, we built a dedicate large scale resource: Flex It, a database of inflectional morphology of Italian, which will be described in section 2. Flex It is set available as a freely usable resource, with the aim to enable further empirical and theoretical research.

1.4. Definition of the Terms Used in the Study

Before moving to a more thorough description of the Flex It database, it is worth summarizing and defining a few terms used in this paper (especially in the light of inconsistent terminology in the literature): **word form**, any inflected word (e.g., *gatti* "cats," is the Italian plural form of the noun *gatto* "cat"); **ending**, the inflectional termination of a word (e.g., *-i* in the Italian noun *gatti* "cats"); **declensional class**, set of lexemes whose members each select the same set of inflectional realizations (e.g., the Italian nouns *gatto* "cat," and *cane* "dog," belong to two different declensional classes since they do not share the same endings: *gatt-o/-i* "cat/cats," vs. *can-e/-i* "dog/dogs"); **feature**, any grammatical characteristic/property for which a word can be specified (e.g., Italian nouns can be specified for number: *gatto* "cat," vs. *gatti* "cats"); **value**, any possible specification of a given feature (e.g., in Italian, the feature of number has two values: singular and plural); **token**, the total number of occurrences of a word form in the database (e.g., the plural word form *gatti* "cats," occurs N times); **type**, every different type of word form in the database, regardless of its total number of occurrences (e.g., even if the plural word form *gatti* "cats," occurs N times, it is counted only once).

2. METHODS

2.1. The Flex It Database

In building Flex It, our goal was to gather data for the present study, as well as to provide a large-scale morphologically annotated database and set it available for further research. The database and its descriptive analyses were developed using R (R Core Team, 2021) and can be downloaded from: https://github.com/franfranz/Flex_it.

The database contains the token frequencies of 71,954 Italian word forms (33,637 noun types and 38,317 adjective types), annotated for inflectional ending, gender, number, declensional class, lemma, grade of adjectives, raw and standardized measures of frequency. We obtained token frequency measures from ItWaC, the largest freely available corpus of Italian, consisting of 1.9 billion tokens from web-collected texts (Baroni et al., 2009). While the size and text variety of this corpus suffice in providing an excellently representative sample of language use, its morphological tagging is at the part of speech (POS) level. A finer-grained morphological annotation, comprising also the indication of gender (feminine - masculine) and number (singular - plural) feature values for adjective and noun types, was retrieved from Morph-it!, a list containing approximately 500,000 word forms, tagged for lemma (Zanchetta and Baroni, 2005).

The Flex It database provides morphological information on a wide scale: besides tags for gender, number and for inflectional endings, we reported a tag for inflectional class. As stated in section 1, in Italian, the inflectional ending corresponds to the last phoneme of a word form, in the noun as well as in adjective declension. In written text, it will in turn correspond to the last letter, due to the orthographically transparent writing system. In order to obtain the inflectional ending, the last character of each word form was stripped. Inflectional paradigms were reconstructed by coupling the endings occurring for the same lemma. Embracing the inherent *vs.* contextual theoretical distinction (as discussed in section 1), inflectional paradigms for nouns include the number values as a two-cell paradigm, whereas inflectional paradigms of adjectives include gender and number values as a four-cell paradigm (In other words, the lemma of a noun is lexically specified for gender, can be inflected in the singular or in the plural, and thus can assume two combinations of values. Instead, the form of an adjective is determined by the values of the noun it modifies and, thus, each lemma can assume four combinations of values). In some cases, only one form was attested for a lemma; in this case, a "NA" tag was assigned in place of the ending not attested in our database even if supposed from a theoretical point of view. In the case of identical word forms for the singular and the plural, an "Inv" tag signals the invariance. In order to avoid some possible confounds derived from the tagging of the original resources, invariance and other phenomena that lead to the presence of ambiguous forms had to be tackled before quantifying the morphological transparency of inflectional classes and exponents.

2.2. Ambiguous Forms

Some noun types are homograph to other POS, such as *apparecchio* noun(M).SG "device," or verb-I.SG.PRES "I prepare." These cases were not problematic for the database, as we collected the token frequency for the occurrences of words tagged as nouns in the ItWaC corpus. The same method was applied to the collection of adjective types homograph to types tagged as other POS. Similarly, in word forms occurring as nouns as well as adjectives, such as *manifesto* noun(M).SG "poster," or adj(M).SG "evident," or *sole* noun(M).SG "sun," or adj(F).PL "alone," the token frequency measures reported in the noun and in the adjective lists refer to the occurrences, respectively, tagged as nouns and as adjectives in the corpus. Since ItWaC is tagged at the POS level, no confounds should occur in measures taken on homograph forms belonging to different POS.

Nevertheless, some types sharing the same POS inflected in different feature values do surface with an identical word form. This can be due to several factors. In cases like *latte* noun(M).SG "milk"/noun(F).PL "tin cans," the difference in meaning undoubtedly points to two different lemmas incidentally surfacing in a homograph form. In other cases, homography is observed in semantically related words and has a more systematic aspect due to the intersection of inflectional classes, as in *cameriere* noun(M).SG "waiter," and noun(F).PL "waitresses"; here, the singular masculine in the *e_i* class is confounded with the plural feminine in the *a_e* class. Similarly, other types surface in the same word form in the singular, like *musicista* noun(M).SG, noun(F).SG "(male/female) musician," showing different forms in the plural, respectively, feminine, *musiciste*(F).PL "female musicians," and masculine, *musicisti*(M).PL "male musicians." Other types are identical in the singular and in the plural: this lack of change of form will be the hallmark of an "invariant" inflectional class. Finally, some nouns, mostly denoting humans (*portavoce*) "spokesperson," show the same form for all four features. We collected all the ambiguous forms, independently of the factors that determine their ambiguity. Ambiguous forms make up the 0.056 (in proportion) of the total noun list, and adjectives make up the 0.169 (in proportion) of the adjective list.

In **Tables 1**, **2**, we report the number of ambiguous forms for nouns and adjectives, respectively, and their occurrence across the inflectional features. A form is reported as an example for each kind of ambiguity.

For each of the forms ambiguously surfacing in more than one combination of feature values, it is possible to retrieve its type frequency, due to the tag provided by the Morph-it! list. However, the token frequency for each of these types cannot be disambiguated into the different values. For example, it is not possible to state how many of the 5,232 tokens of the word form *cameriere* are occurrences of the type noun(M).SG "waiter" and how many of the type noun(F).PL "waitresses." In order to avoid this potential confound, we considered the type frequencies of ambiguous nouns in our analysis, but we limited our counts on the token frequency of non-ambiguous forms.

3. RESULTS

We measured the distribution of Italian nouns and adjectives in the Flex It database to assess the entropy of the morphological systems with respect to the features of gender and number. To this end, we considered the distribution of the word forms from two different points of view: (i) first, the arrangement of the word forms according to each declensional class (e.g., the amount of word forms that belong to Class I, sharing the same endings *o_i* for the singular and the plural, and convey the value of masculine vs. the amount of word forms that belong to Class I and instead convey the value of feminine; section 3.1.1 for nouns and section 3.2.1 for adjectives); (ii) second, the distribution of the word forms across all possible combinations of values (F.SG, F.PL, M.SG, M.PL) with respect to each inflectional ending (e.g., the amount of word forms in *-o* that convey the value combination of masculine singular vs. the amount of word forms in *-o* that instead convey the value combinations of masculine plural or feminine singular or feminine plural; section 3.1.2 for nouns and section 3.2.2 for adjectives).

TABLE 1 | Number of ambiguous noun forms.

Feature values		N.forms
F. SG. - F. PL. - M. SG. - M. PL.	*portavoce* "(fe)male spokesperson(s)"	43
F. SG. - F. PL. - M. SG.	*radio* "radio - radios - radius bone"	11
F. PL. - M. SG. - M. PL.	*marine* "marinas - mariner - mariners"	6
F. SG. - M. SG. - M. PL.	*boa* "buoy - boa - boas"	9
F. SG. - F. PL.	*analisi* "analysis - analyses"	246
F. PL. - M. PL.	*abitanti* "female residents - male residents"	190
F. PL. - M. SG.	*cameriere* "waitresses - waiter"	34
F. SG. - M. PL.	*sequestri* "kidnapping unit - requisitions"	1
F. SG. - M. SG.	*abitante* "female resident - male resident"	240
M. SG. - M. PL.	*quiz* "quiz - quiz"	969
Total		1,749

TABLE 2 | Number of ambiguous adjective forms.

Feature values		N.forms
F. SG. - F. PL. - M. SG. - M. PL.	*antidroga* "antidrug"	268
F. SG. - F. PL. - M. SG.	*molle* "soft"	9
F. PL. - M. PL.	*abili* "skilled"	2,451
F. SG. - M. SG.	*abile* "skilled"	2,663
Total		5,391

3.1. Noun Inflection

3.1.1. Declensional Classes

The number of type and tokens for each declensional class are reported in **Table 3**. The invariant nouns are grouped together as a single class "Inv." The "Other" tag in the table collects the nouns that would be expected to be invariant but are attested as inflected in some cases, as *sport*(M).SG/*sports*(M).PL - *corpus*(M).SG/*corpora*(M).PL. For each of the declensional classes, we report an entropy value *H*, calculated in the way indicated by Shannon (1948), based on the probability for each set of endings to realize the feminine or the masculine forms. In this sense, entropy is a measure of consistency in the association of a declensional class with a gender value. Low entropy values correspond to a more stable association between a declensional class and a gender value.

3.1.2. Inflectional Endings

The distribution of noun types across the four most frequent inflectional endings *-a, -e, -i, -o* is reported in **Table 4** and plotted in **Figure 1A**. The distribution of noun tokens across the four

TABLE 3 | Distribution of noun lemmas across the declensional classes (types - token).

	Noun types				Noun tokens			
Class	F.	M.	Total	*H*	F.	M.	Total	*H*
o_i	2	11,957	11,959	0.0023	442,454	128,733,286	129,175,740	0.033
a_e	8,318	0	8,318	0	81,295,823	0	81,295,823	0
e_i	3,268	3,907	7,175	0.9943	47,533,225	26,577,783	74,111,008	0.9415
a_i	4	932	936	0.0398	207,807	5,900,697	6,108,504	0.2142
o_a	23	23	46	1	405,856	472,125	877,981	0.9959
o_a_i[*]	14	42	56	0.8113	136,379	684,209	820,588	-
Inv	2	40	42	0.2761	893	9,004	9,897	0.4372
Other	55	89	144	0.9594	233,392	979,046	1,212,438	0.7066

[*] In Italian, a handful of nouns can have one form for the singular, but two forms for the plural (e.g., muro(M).SG "wall," muri(M).PL "walls," mura(F).PL "city wall"). These forms are listed as the same lemma in the resources used for the compilation of Flex It. The count for these forms are reported here for informative purposes only. It is not among the scopes of the present paper to describe these plural forms (e.g., muri(M).PL "walls" and mura(F).PL "city wall") as resulting from the inflection of the same lexeme or two distinct, although homophonous, lexemes (each linked to a diverse plural form). For theoretical accounts concerning these forms, see (Acquaviva, 2002, 2008; Thornton, 2013).

TABLE 4 | Number of noun types for the most frequent inflectional endings.

Nouns - types						
Ending	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
-a	23 (0.0047)	4,331 (0.8943)	12 (0.0025)	477 (0.0985)	4,843	0.5316
-e	4,286 (0.5353)	1,715 (0.2142)	3 (0.0004)	2,002 (0.2501)	8,006	1.4631
-i	1,667 (0.1613)	0 (0)	8,661 (0.8383)	4 (0.0004)	10,332	0.6424
-o	0 (0)	5 (0.0008)	7 (0.0011)	6,205 (0.9981)	6,217	0.0221

The first column lists the endings, the second to fifth report the counts and probabilities (the proportional values in brackets) for each combination of features. In the last, we report the entropy for the ending indicated in the row.

TABLE 5 | Number of noun tokens for the most frequent inflectional endings.

Nouns - tokens						
Ending	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
-a	405,856 (0.0063)	59,533,689 (0.9270)	8,213 (0.0001)	4,273,650 (0.0665)	64,221,408	0.4094
-e	24,251,151 (0.3002)	39,817,093 (0.4929)	134 (0.000)	16,710,610 (0.2069)	80,778,988	1.4945
-i	11,290,598 (0.1739)	0 (0)	53,488,244 (0.8241)	129,009 (0.0020)	64,907,851	0.6868
-o	0 (0)	289,291 (0.0032)	5,963 (0.0001)	90,593,727 (0.9968)	90,888,981	0.0320

The first column lists the endings, the second to fifth report the counts and probabilities (the proportional values in brackets) for each combination of features. In the last, we report the entropy for the ending indicated in the row.

most frequent inflectional endings is reported in **Table 5** and plotted in **Figure 1B**.

We counted how many forms occur for each combination of inflectional values. Based on the probability of each ending to realize one of the possible forms, we calculated the entropy as a proxy to transparency of each of the endings. Transparency is related to a low entropy, corresponding to the fact that an ending is mostly likely to realize a specific combination of values. Such an ending will be informative of the presence of an inflectional value or combination of values. The entropy for each ending is reported in the *H* columns in the tables, respectively, calculated on the types and on the tokens.

3.2. Adjective Inflection

3.2.1. Declensional Classes

The number of type and tokens for each declensional class are reported in **Tables 6**, **7**. We find a consistent representation of the first and second declensional classes predicted by theoretical descriptions. Due to the less precise representation of adjectives in the corpus, possibly related to their lower frequency of occurrence (as shown in **Figures 1**, **2**), we reported several defective types for which some inflected forms are not present in the corpus. In this regard, it is worth noticing that not all the possible forms of an adjective lemma predicted on a theoretical basis occur in our database (for example, an adjective lemma that can be inflected in all combinations of values, i.e., F.SG, F.PL, M.SG, and M.PL, is attested only in the F.SG). Moreover, only the first two declensional classes (which are also the most represented) include adjective forms per all possible combinations of values. Hence, the lack of occurrences of some forms in the database, instead assumed at a theoretical level, explains the apparent discrepancy between the number of declensional classes identified in the literature (i.e., five) and the number of rows in **Tables 6**, **7** (i.e., 12). For each of the declensional classes, we report an entropy value, which refers to the probability for each set of endings to realize the combination of feminine plural, feminine singular, masculine plural or masculine singular values. These declensional classes stem from the realization of an inflected adjective lemma. For example, the first class collected the lemmas whose occurrences end in *-a* in the feminine singular, in *-e* in the feminine plural, in *-o* in the masculine singular and in *-i* in the masculine plural. In this case, the transparency of the forms is evident in the column "Class" of **Tables 6**, **7**, which lists four different forms. The columns *H* represents the probability for which each of the lemmas occurs as inflected in each of the value combination.

3.2.2. Inflectional Endings

The distribution of adjective types across the four most frequent inflectional endings *-a, -e, -i, -o* is reported in **Table 8** and plotted in **Figure 2A**. The distribution of adjective tokens across the four most frequent inflectional endings is reported in **Table 9** and plotted in **Figure 2B**. We counted how many forms occur for each combination of inflectional values. Based on the probability of each ending to realize one of the possible forms, we calculated the transparency of each of the endings. Transparency is related to a low entropy, corresponding to the fact that an ending is mostly likely to realize a specific combination of values; such ending will be informative of the presence of an inflectional value or combination of values.

4. DISCUSSION

For the first time, we measured the distribution of Italian nouns and adjectives across the feature values for which they can be specified to assess the entropy of the morphological inflectional system. For this purpose, we created a large database, Flex It, combining the corpus ItWaC, the largest freely available corpus of Italian which is tagged at the part of speech level, and Morph-it!, a list of word forms comprising a finer-grained morphological annotation. Based on the probability of the inflectional endings to convey the possible feature values, we calculated the entropy as a proxy to transparency of each of the endings. More precisely, we considered the distribution of the word forms from two different points of view: (i) the distribution of the word forms across all possible combinations of values (F.SG, F.PL, M.SG, M.PL) with respect to each inflectional ending (e.g., the amount of word forms in *-o* that convey the value combination of masculine

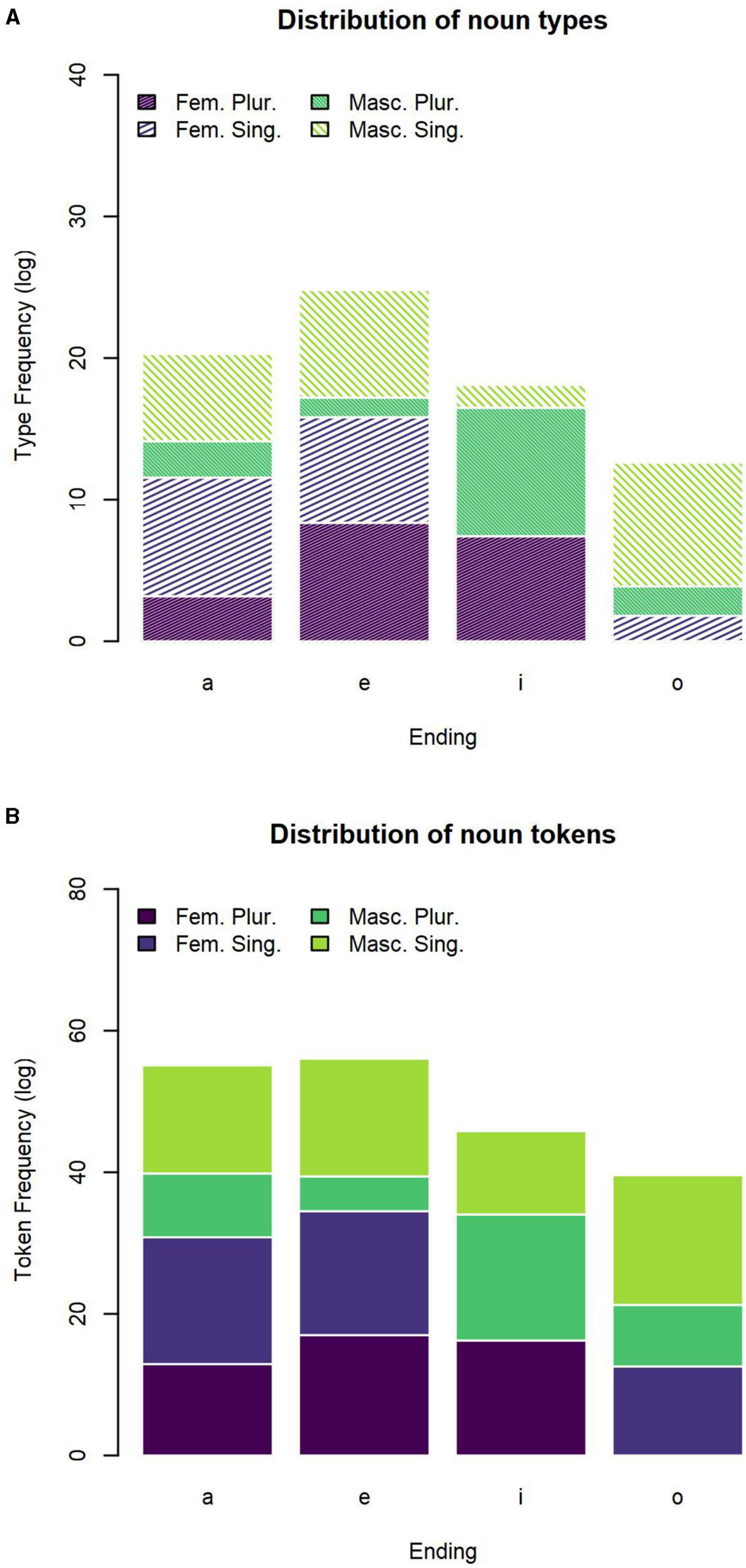

FIGURE 1 | Distribution of nouns across the most frequent inflectional endings. **(A)** Number of noun types for the most frequent inflectional endings. **(B)** Number of noun tokens for the most frequent inflectional endings.

TABLE 6 | Distribution of adjective lemmas across the declensional classes (types).

Adjectives - types per declensional class						
Class	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
a_e o_i	5,727	6,068	5,831	6,137	23,763	1.9994
e_i e_i	148	148	148	148	592	2
NA o_i	0	0	285	287	572	1
a_NA NA	0	458	0	0	458	0
NA o_NA	0	0	0	395	395	0
a_e NA	128	131	0	0	259	0.9999
NA_e NA	242	0	0	0	242	0
NA NA_i	0	0	240	0	240	0
NA e_i	0	0	19	19	38	1
e_NA NA	0	11	0	0	11	0
e_i NA	2	2	0	0	4	1
NA e_NA	0	0	0	2	2	0

TABLE 7 | Distribution of adjective lemmas across declensional classes (tokens).

Adjectives - tokens per declensional class						
Class	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
a_e o_i	12,516,462	24,500,447	15,542,053	27,661,762	80,220,724	1.9292
e_i e_i	49,146	98,955	130,777	157,807	436,685	1.8916
NA o_i	0	0	52,466	165,194	217,660	0.7968
NA_e NA	154,248	0	0	0	154,248	0
NA NA_i	0	0	117,487	0	117,487	0
NA o_NA	0	0	0	82,563	82,563	0
a_e NA	35,220	47,231	0	0	82,451	0.9846
a_NA NA	0	27,666	0	0	27,666	0
NA e_i	0	0	1,891	2,530	4,421	0.9849
NA e_NA	0	0	0	2,167	2,167	0
e_i NA	122	483	0	0	605	0.7252
e_NA NA	0	132	0	0	132	0

TABLE 8 | Number of adjective types for the most frequent inflectional endings.

Adjectives - types						
Ending	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
-a	0 (0)	6,657 (1)	0 (0)	0 (0)	6,657	0
-e	6,097 (0.9487)	161 (0.0251)	0 (0)	169 (0.0263)	6,427	0.3434
-i	150 (0.0225)	0 (0)	6,523 (0.9775)	0 (0)	6,673	0.1551
-o	0 (0)	0 (0)	0 (0)	6,819 (1)	6,819	0

The first column lists the endings, the second to fifth report the counts and probabilities (the proportional values in brackets) for each combination of features. In the last, we report the entropy for the ending indicated in the row.

TABLE 9 | Number of adjective tokens for the most frequent inflectional endings.

Adjectives - tokens						
Ending	**F.PL**	**F.SG**	**M.PL**	**M.SG**	**Total**	***H***
-a	0 (0)	24,575,344 (1)	0 (0)	0 (0)	24,575,344	0
-e	12,705,930 (0.9798)	99,570 (0.0077)	0 (0)	162,504 (0.0125)	12,968,004	0.1620
-i	49,268 (0.0031)	0 (0)	15,844,674 (0.9969)	0 (0)	15,893,942	0.0303
-o	0 (0)	0 (0)	0 (0)	27,909,519 (1)	27,909,519	0

The first column lists the endings, the second to fifth report the counts and probabilities (the proportional values in brackets) for each combination of features. In the last, we report the entropy for the ending indicated in the row.

singular vs. the amount of word forms in *-o* that instead convey the other combinations of values), and (ii) the arrangement of the word forms according to each declensional class (e.g., the amount of word forms that belong to Class I, sharing the same endings *o_i* for singular and plural, and convey the value of masculine vs. the amount of word forms that belong to Class I and instead convey the value of feminine).

4.1. Form-Function Consistency

4.1.1. Transparency of Inflectional Endings

First, we found that masculine singular nouns mostly end in -*o*, which is indeed associated to the lower close-to-min entropy of the distribution (of both types and tokens). A higher entropy is instead detected for *-a* (which mostly realizes -but it is not restricted to- feminine singular forms) and for *-i* (which mostly realizes -but it is not restricted to- masculine plural forms). The highest entropy of the distribution was spotted for *-e* which is almost equally likely to form feminine singular, feminine plural, and masculine singular nouns. Thus, the overall system seems to reflect the trade-off between maximal discriminability and maximal regularity that has been argued in the literature for other languages and other grammatical systems (as mentioned in section 1; e.g., Blevins et al. 2017).

The distribution of nouns mirrors, in broad terms, that of adjectives, albeit with some non-negligible differences. Indeed, we observed much less variability in the distribution of adjective forms across feature values in comparison to nouns. In this case, the endings *-a* and *-o* are both associated to the minimum entropy being the unambiguous marks of feminine singular and masculine singular, respectively; and the endings *-e* and -*i*, even if associated to a slightly higher entropy, are mostly used for feminine plural and masculine plural, respectively. Put in different terms, the overall association between inflectional endings and feature values tends to be more transparent and clear-cut in the adjective forms than in the noun forms. If, from a theoretical perspective (see sections 1 and 4.1.3), this may be anything but an unexpected result, nevertheless, it is the first time that these different distributions are quantified and caught in terms of entropy metrics as for the Italian inflectional system.

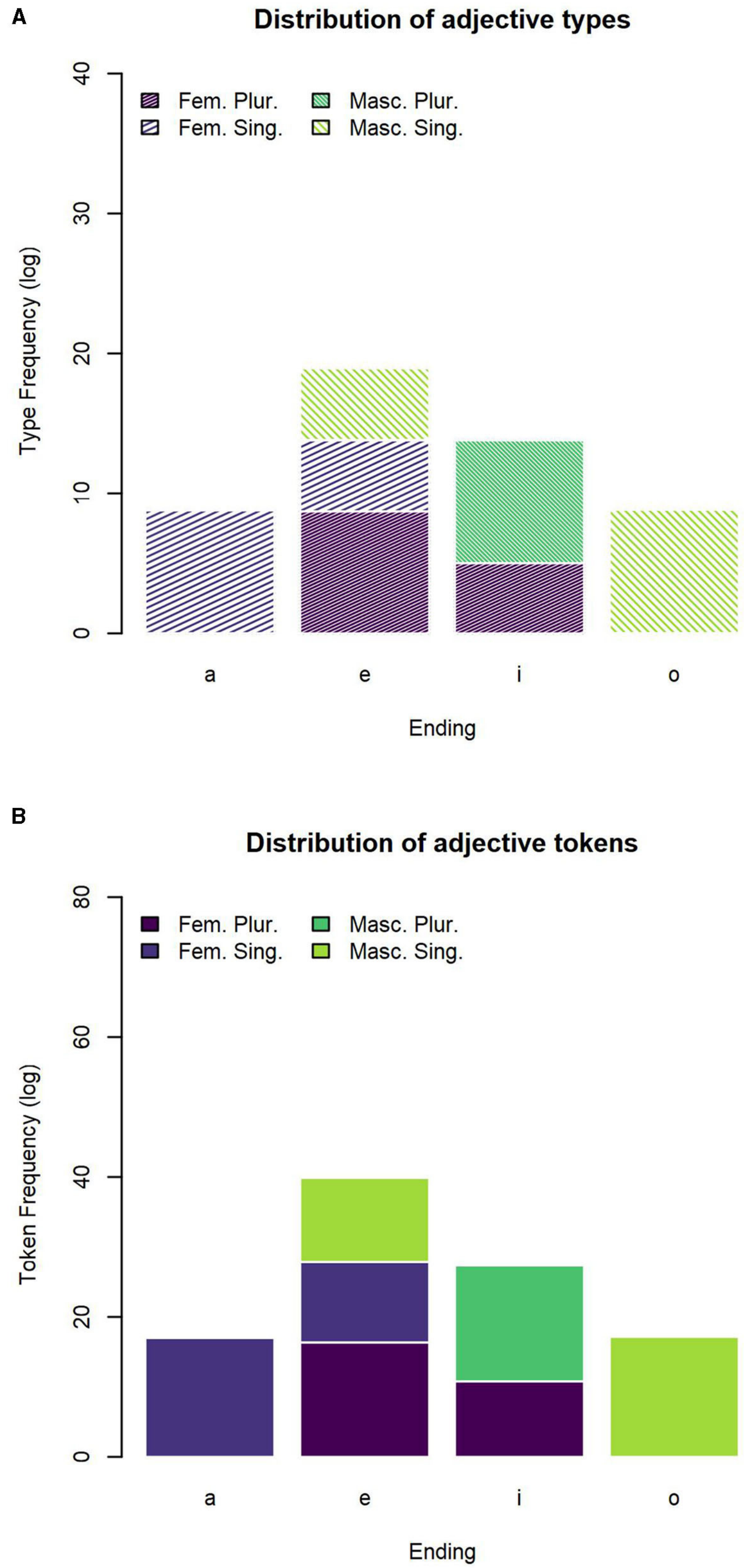

FIGURE 2 | Distribution of adjectives across the most frequent inflectional endings. **(A)** Number of adjective types for the most frequent inflectional endings. **(B)** Number of adjective tokens for the most frequent inflectional endings.

4.1.2. Transparency of Declensional Classes

When considering declensional classes, the distribution of forms is arranged a little differently across nouns and adjectives. By definition, Italian nouns have a two-cell paradigm, whereas adjectives have a four-cell paradigm (see section 2.1; Iacobini and Thornton, 2016). It follows that the maximum entropy for noun paradigms will be 1 bit, while for adjective paradigms it will be 2 bits. At the same time, entropy allows to quantify information content across paradigms with different numbers of cells. Nonetheless, it is noteworthy that in Italian adjectives information tends to be higher when considering form distribution across feature values, whereas in Italian nouns information grows when considering form distribution across declensional classes.

As for nouns, the most represented classes are *a_e, o_i*, and *e_i*. While the first two classes show the minimum entropy as they almost always host feminine and masculine nouns, respectively, the third class *e_i* shows the close-to-max entropy as masculine and feminine nouns share almost the same probability of being comprised. This is consistent with what is usually stated in the literature, namely that the first two inflectional classes tend to be the most productive as they are more transparent with respect to gender and number features. In other words, newly formed lexical entries are more likely to be assigned to one of the first two Italian declensional classes because these are the most informative ones (Thornton, 2004; D'Achille and Thornton, 2008; Acquaviva, 2009). Even in this case, the overall declensional system seems to reflect a trade-off between maximal discriminability and maximal regularity.

When it comes to adjectives, once again, much less variability is found than for nouns. By far the most represented class, *a_e o_i* is associated to the close-to-max entropy of the distribution since it equally comprises masculine singular, feminine singular, masculine plural, and feminine plural forms.

4.1.3. Form-Value Transparency in Nouns and Adjectives

We suggest that the different distributions of noun and adjective forms we observed so far are related to the distinct functions played by these two parts of speech in agreement relations, in which they usually act, respectively, as controllers and targets. In Italian, the gender and number of nouns are inherent to the lexeme because their encoding is context-autonomous, while the gender and number of adjectives are contextual because their encoding is obligatorily driven by morpho-syntactic agreement (see section 1; for more on inherent vs. contextual inflection: Booij, 1993, 1996). Therefore, since it is the target which is the locus of agreement (Corbett, 2006, p. 12), in the sense that the signpost of agreement surfaces in the form of the target, we expect a more transparent form-value relation in targets than in controllers. Consequently, we expect this to be reflected in their distribution in the language. Word-formation processes in the adjective domain confirm this aspect, with superlative forms in *-issima, -issime, -issimo, -issimi* being assigned to the maximally discriminative class.

To a certain extent, this also applies to adjectives such as *grande-grandi*, "big.SG – big.PL," in which *-e* is the ending for both masculine and feminine singulars and *-i* is the ending for both masculine and feminine. Although syncretism blurs the gender distinction, the number opposition is still clear-cut. This resonates with general typological trends whereby the feature of number is prioritized over the feature of gender. Indeed, grammatical gender is less widespread across languages (Corbett, 1991) and, as stated in Greenberg's Universal 34, in a language, the presence of number is a necessary condition for gender to surface (Greenberg, 1963) possibly, due to a preminence of the semantic information conveyed by number (Franzon et al., 2019, 2020). Thus, noun forms are less informative with respect to gender (and, to a lesser extent, number) since their main role is to distinguish classes of words mainly favoring discriminability between diverse forms as a whole rather than between gender and number values. Conversely, almost all Italian adjective forms manage to maintain close-to-max discriminability, at least between number values. This result can be interpreted in light of a language processing mechanism; the transparency of targets disambiguates the features of theirs controllers, making their agreement relation explicit. Since targets favor prediction in language processing, it seems reasonable that their form-to-value consistency tends to be more transparent when compared to controllers.

4.2. Pending Issues and Conclusions

Our results on Italian nouns and adjectives are compatible with current Word and Paradigm-based approaches from both a theoretical and computational perspective (for an overview see Marzi et al., 2020). However, the differences found in the distributions of nouns (that are generally controllers and have an inherent inflection) and adjectives (that are generally targets and have a contextual inflection) needs to be deepened on. In this respect, while some recent accounts have explored the differences in the effect of syntagmatic and paradigmatic cues in comprehension (Đurđević and Milin, 2019), only few studies have been dedicated on how contextual and inherent inflection are parsed during language processing (Franzon et al., 2013, 2014). Do the distributional properties measured in this study reflect only mechanisms internal to the morphological organization of (Italian) forms, or are they also a reflection of more general cognitive mechanisms? Literature is scarce in this regard. Despite the fact that consistency between formal cues and gender values has been shown to impact gender retrieval in both isolated word presentation and sentence processing (see, for example, Caffarra et al., 2015), to date no psycholinguistic study has tested whether the observed differences in the distribution of noun forms vs. adjective forms with respect to gender and number values also correspond to differences in processing. Yet, it is possible that inherent inflection and contextual inflection are not merely theoretical constructs. For example, it has been found that, in contact language situations (when a recipient language changes as an effect of contact with a source language), inherent inflection is more likely to be borrowed than contextual inflection since this latter is more entrenched in the grammar and altering it in a resource language causes huge changes in agreement mechanisms. By contrast, the introduction of endings realizing inherent inflection impacts less on the overall morpho-syntactic structure of the recipient language (Gardani, 2012, 2020).

Our results are also consistent with psycholinguistic studies that have related effects on processing to the distributional properties of Italian nouns, reporting slower and less accurate responses to noun forms opaque with respect to gender (De Martino et al., 2011, 2017; Caffarra et al., 2015). Mismatches between declensional class and gender value have been proven costly in processing terms and, in particular, fMRI data showed increased cortical activity for an extensive network (involving frontal and temporal areas, cingulate cortex and cerebellum) linked to inflectional operations for Italian non-transparent declensional classes (Russo et al., 2021). We expect our results to provide a better estimate of Italian nouns' transparency for future neuro- and psycholinguistic studies on inflection.

In this respect, we are well aware that our approach is only one possible way to quantify the regularity of morphological cues. For example, under the umbrella of the competition model, MacWhinney et al. (1984) argued that each mapping between a form and a function can be assigned a weight or strength. The weight of a cue would depend on its validity, i.e., the combination of cue reliability (how many times the cue relates to a specific function) and cue availability (how many times a specific cue is present in the lexicon). We do not comment on the substance of this model. Yet, it is worth noticing that, in the present study, we propose entropy as a measure based on the properties of the signal, as observed in linguistic corpora. To use Mandelbrot's words, three elements are to be considered [for a theory of communication]: (1) The structure of language, or shortly, message; (2) The way in which information is coded by the brain; (3) The economical "criterion of matching" which links 1 and 2 (Mandelbrot, 1953, p. 486). The present work aims at contributing to the knowledge regarding the first element, which is necessary to inform the other two. With this purpose in mind, and with the currently available material, it is not possible to completely disentangle entropy from other linguistic and psycholinguistic variables. However, we believe that this work will nonetheless provide researchers with a useful metric of form-value, that has thus far scarcely been considered (especially with regard to Italian noun and adjective forms), and that this will provide them with a solid ground for the experimental assessment of inflectional morphology-related hypotheses. Moreover, entropy metrics seem to be a suitable and well-grounded tool when comparing typologically diverse languages.

Eventually, although we have measured the entropy of purely morphological systems, the distribution of word forms across inflectional feature values, overall, seems to reflect factors which relate to the morpho-syntactic level and the functions that parts of speech such as nouns and adjectives play at this level. Hence, these entropy metrics are valuable both when testing words in isolation and in sentence context. For all these reasons, we believe that the set of observations we have provided in the present work are potentially relevant for any future study focusing on inflection, in light of the implications that form-value (in)consistency can have for sentence processing, especially with respect to nouns and adjectives. We encourage further research on this topic.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/franfranz/Flex_it.

AUTHOR CONTRIBUTIONS

VP: original idea of the study, manuscript drafting, previous versions of the resource development, and final draft revision. CZ: original idea of the study, manuscript drafting, and final draft revision. DC: scientific supervision to the study and final draft revision. FF: original idea of the study, manuscript drafting, data analysis, development of the current version of the resource, and final draft revision. All authors contributed to the discussion that laid the theoretical foundations of the study and approved the submitted version of the manuscript.

FUNDING

This study was supported by H2020 European Research Council [grant number: 679010 (ERC STATLEARN to DC)].

REFERENCES

Đurđević, D. F., and Milin, P. (2019). Information and learning in processing adjective inflection. *Cortex* 116, 209–227. doi: 10.1016/j.cortex.2018.07.020

Acquaviva, P. (2002). Il plurale in-a come derivazione lessicale. *Lingue e linguaggio* 1, 295–326. doi: 10.1418/7862

Acquaviva, P. (2008). *Lexical Plurals: A Morphosemantic Approach*. New York, NY: Oxford University Press.

Acquaviva, P. (2009). "Roots and lexicality in distributed morphology," in *Fifth York-Essex Morphology Meeting (YEMM)* (Department of Language and Linguistic Science, University of York).

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

Anderson, S. R. (1992). *A-Morphous Morphology*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511586262

Arcara, G., Franzon, F., Gastaldon, S., Brotto, S., Semenza, C., Peressotti, F., et al. (2019). One can be some but some cannot be one: ERP correlates of numerosity incongruence are different for singular and plural. *Cortex* 116, 104–121. doi: 10.1016/j.cortex.2018.10.022

Aronoff, M. (1994). *Morphology by Itself: Stems and Inflectional Classes*. Cambridge, MA: MIT Press.

Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118:438. doi: 10.1037/a0023851

Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Lang. Resour. Eval.* 43, 209–226. doi: 10.1007/s10579-009-9081-4

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

Bates, E., Devescovi, A., Pizzamiglio, L., D'amico, S., and Hernandez, A. (1995). Gender and lexical access in Italian. *Percept. Psychophys*. 57, 847–862. doi: 10.3758/BF03206800

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., and Smith, S. (1982). Functional constraints on sentence processing: a cross-linguistic study. *Cognition* 11, 245–299. doi: 10.1016/0010-0277(82)90017-8

Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., et al. (2005). *Colfis (Corpus e lessico di Frequenza Dell'italiano Scritto)*. Available online at: http://www. istc.cnr.it/material/database

Blevins, J. P., Milin, P., and Ramscar, M. (2017). "The zipfian paradigm cell filling problem," in *Perspectives on Morphological Structure: Data and Analyses*, eds F. Kiefer, J. P. Blevins, and H. Bartos (Leiden: Brill), 139–158. doi: 10.1163/9789004342934_008

Bloomfield, L. (1933). *Language*. New York, NY: Holt.

Booij, G. (1993). "Against split morphology," in *Yearbook of Morphology 1993*, eds G. Booij and J. van Marle (Dordrecht: Springer), 27–49. doi: 10.1007/978-94-017-3712-8_2

Booij, G. (1996). "Inherent versus contextual inflection and the split morphology hypothesis," in *Yearbook of Morphology 1995*, eds G. Booij and J. van Marle (Dordrecht: Springer), 1–16. doi: 10.1007/978-94-017-3716-6_1

Bordag, D., Opitz, A., and Pechmann, T. (2006). Gender processing in first and second languages: the role of noun termination. *J. Exp. Psychol. Learn. Mem. Cogn*. 32:1090. doi: 10.1037/0278-7393.32.5.1090

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

Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., and Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology* 52, 1019–1030. doi: 10.1111/psyp.12429

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

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

Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *J. Mem. Lang*. 63, 83–99. doi: 10.1016/j.jml.2010.03.002

D'Achille, P., and Thornton, A. M. (2008). "I nomi femminili in-o," in *Prospettive nello studio del lessico Italiano. Atti del IX Congresso Internazionale della Societá di Linguistica e Filologia Italiana* (Firenze: Firenze University Press), 473–481.

Davis, M. H., and Rastle, K. (2010). Form and meaning in early morphological processing: comment on Feldman, O'connor, and Moscoso del prado Martin (2009). *Psychon. Bull. Rev*. 17, 749–755. doi: 10.3758/PBR.17.5.749

De Martino, M., Bracco, G., and Laudanna, A. (2011). The activation of grammatical gender information in processing Italian nouns. *Lang. Cogn. Process*. 26, 745–776. doi: 10.1080/01690965.2010.491977

De Martino, M., Bracco, G., Postiglione, F., and Laudanna, A. (2017). The influence of grammatical gender and suffix transparency in processing Italian written nouns. *Mental Lexicon* 12, 107–128. doi: 10.1075/ml.12.1.05dem

De Martino, M., Postiglione, F., Bracco, G., and Laudanna, A. (2019). Declinazione e genere dei nomi Italiani: degni, una banca dati lessicale. *Giornale Italiano di Psicologia* 46, 247–268. doi: 10.1421/93789

De Vincenzi, M., and Di Domenico, E. (1999). A distinction among features: The role of gener and number in the retrieval of pronoun antecendents. *Ital. J. Linguist*. 11, 41–74.

Di Domenico, E. (1997). *Per una teoria del Genere Grammaticale*. Padova: Unipress.

Dye, M., Milin, P., Futrell, R., and Ramscar, M. (2017). "A functional theory of gender paradigms," in *Perspectives on Morphological Structure: Data and Analyses*, eds F. Kiefer, J. P. Blevins, and H. Bartos (Leiden: Brill), 212–239. doi: 10.1163/9789004342934_011

Ferry, A., Nespor, M., and Mehler, J. (2020). Twelve to 24-month-olds can understand the meaning of morphological regularities in their language. *Dev. Psychol*. 56, 40–52. doi: 10.1037/dev0000845

Franzon, F., Arcara, G., and Zanini, C. (2016). "Lexical categories or frequency effects? A feedback from quantitative methods applied to psycholinguistic models in two studies on Italian," in *Proceedings of the Third Italian Conference on Computational Linguistics-CLiC-it* (Torino), 152–156. doi: 10.4000/books.aaccademia.1778

Franzon, F., Bertocci, D., and Semenza, C. (2013). Exploring gender inflection: an insight from errors in aphasia. *Stem Spraak Taalpathologie* 18(Suppl. 1), 171–173.

Franzon, F., Peressotti, F., Arcara, G., and Semenza, C. (2014). Semantic interpretability speeds up the processing of morphological features. A psycholinguistic experiment on gender agreement. *Stem Spraak Taalpathologie* 19, 189–191.

Franzon, F., and Zanini, C. (2021). The entropy of morphological systems in natural languages is modulated by functional and semantic properties. *[Preprint]*. doi: 10.31234/osf.io/qyd4g

Franzon, F., Zanini, C., and Rugani, R. (2019). Do non-verbal number systems shape grammar? Numerical cognition and number morphology compared. *Mind Lang*. 34, 37–58. doi: 10.1111/mila.12183

Franzon, F., Zanini, C., and Rugani, R. (2020). Cognitive and communicative pressures in the emergence of grammatical structure: a closer look at whether number sense is encoded in privileged ways. *Cogn. Neuropsychol*. 37, 355–358. doi: 10.1080/02643294.2020.1802241

Gardani, F. (2012). "4. Plural across inflection and derivation, fusion and agglutination," in *Copies versus Cognates in Bound Morphology* (Leiden: Brill).

Gardani, F. (2020). Borrowing matter and pattern in morphology. An overview. *Morphology* 30, 263–282. doi: 10.1007/s11525-020-09371-5

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

Greenberg, J. (ed.). (1963). *Some Universals of Grammar With Particular Reference to the Order of Meaningful Elements*. Cambridge, MA: Universals of Language.

Huettig, F., Rommers, J., and Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychol*. 137, 151–171. doi: 10.1016/j.actpsy.2010.11.003

Iacobini, C., and Thornton, A. M. (2016). "8. Morfologia e formazione delle parole," in *Manuale di Linguistica Italiana*, ed S. Lubello (Berlin; Boston, MA: De Gruyter), 190–221. doi: 10.1515/9783110360851-010

Ivanova-Sullivan, T., and Sekerina, I. A. (2019). "Distributional regularity of cues facilitates gender acquisition: a contrastive study of two closely related languages," in *Proceedings of the 43rd Boston University Conference on Language Development*, eds M. M. Brown and B. Dailey (Somerville, MA: Cascadilla Press), 311–323.

Lambon Ralph, M., Sage, K., Heredia, C. G., Berthier, M. L., Martinez-Cuitino, M., Torralva, T., et al. (2011). El-la: the impact of degraded semantic representations on knowledge of grammatical gender in semantic dementia. *Acta Neuropsychol*. 9, 115–131. Available online at: https://actaneuropsychologica.com/resources/html/article/details?id=18901

Luzzatti, C., and De Bleser, R. (1996). Morphological processing in Italian agrammatic speakers: eight experiments in lexical morphology. *Brain Lang*. 54, 26–74. doi: 10.1006/brln.1996.0060

MacWhinney, B., and Bates, E. (1989). "Functionalism and the competition model," in *The Crosslinguistic Study of Sentence Processing*, eds B. MacWhinney and E. Bates (New York, NY: Cambridge University Press), 3–73.

MacWhinney, B., Bates, E., and Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *J. Verb. Learn. Verb. Behav*. 23, 127–150. doi: 10.1016/S0022-5371(84)90093-8

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Commun. Theory* 84, 486–502.

Marelli, M., and Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behav. Res. Methods* 50, 1482–1495. doi: 10.3758/s13428-018-1017-8

Marelli, M., Amenta, S., and Crepaldi, D. (2015). Semantic transparency in free stems: the effect of orthography-semantics consistency on word recognition. *Q. J. Exp. Psychol*. 68, 1571–1583. doi: 10.1080/17470218.2014.959709

Marzi, C., Blevins, J., Booij, G., and Pirrelli, V. (2020). "Inflection at the morphology-syntax interface," in *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*, eds V. Pirrelli, I. Plag, and W. Dressler (Berlin; Boston, MA: De Gruyter Mouton), 228–294. doi: 10.1515/9783110440577-007

Marzi, C., Ferro, M., and Pirrelli, V. (2019). A processing-oriented investigation of inflectional complexity. *Front. Commun*. 4:48. doi: 10.3389/fcomm.2019.00048

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. London; New York, NY: Cambridge University Press.

Matthews, P. H. (1991). *Morphology. Cambridge Textbooks in Linguistics, 2nd Edn*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139166485

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

Mickus, T., Bonami, O., and Paperno, D. (2019). Distributional effects of gender contrasts across categories. *Proc. Soc. Comput. Linguist*. 2, 174–184. doi: 10.7275/g11b-3s25

Milin, P., Feldman, L. B., Ramscar, M., Hendrix, P., and Baayen, R. H. (2017). Discrimination in lexical decision. *PLoS ONE* 12:e0171935. doi: 10.1371/journal.pone.0171935

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramscar, M., Dye, M., Blevins, J., and Baayen, H. (2018). 'Morphological development," in *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*, eds E. Dattner and D. Ravid (Berlin; Boston, MA: De Gruyter Mouton), 181–202. doi: 10.1515/9781614514909-010

Russo, A. G., Esposito, F., Laudanna, A., Mancuso, A., Di Salle, F., Elia, A., et al. (2021). The neural substrate of noun morphological inflection: a rapid event-related fMRI study in Italian. *Neuropsychologia* 151:107699. doi: 10.1016/j.neuropsychologia.2020.107699

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J*. 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Steele, S. (1978). "Word order variation. A typological study," in *Universals of Human Language, Volume 4: Syntax*, eds J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford, CA: Stanford University Press), 585–623.

Taft, M., and Meunier, F. (1998). Lexical representation of gender: a quasiregular domain. *J. Psycholinguist. Res*. 27, 23–45. doi: 10.1023/A:1023270723066

Thornton, A. M. (2004). "Conversione," in *La formazione delle parole in italiano*, eds M. Grossmann and F. Rainer (Tübingen: Niemeyer), 419–477.

Thornton, A. M. (2013). La non canonicitá del tipo it. bracciobraccia/bracci: sovrabbondanza, difettivitá o iperdifferenziazione. *Studi di Grammatica Italiana*, Vol. 29–30, 419–477.

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

Williams, A., Pimentel, T., McCarthy, A. D., Blix, H., Chodroff, E., and Cotterell, R. (2020). Predicting declension class from form and meaning. *arXiv preprint arXiv:2005.00626*. doi: 10.18653/v1/2020.acl-main.597

Zanchetta, E., and Baroni, M. (2005). Morph-it. A free corpus-based morphological resource for the Italian language. *Corpus Linguist*. 1:2005. Available online at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.211.8926

Zanini, C., Rugani, R., Giomo, D., Peressotti, F., and Franzon, F. (2020). Effects of animacy on the processing of morphological number: a cognitive inheritance? *Word Struct*. 13, 22–44. doi: 10.3366/word.2020.0158

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

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fpsyg.2021.719657

Patterns of Adaptation in Child-Directed and Child Speech in the Emergence of Hebrew Verbs

Elitzur Dattner[1]*, *Ronit Levie*[1], *Dorit Ravid*[1,2] *and Orit Ashkenazi*[3]

[1] Department of Communication Disorders, Tel Aviv University, Tel Aviv, Israel, [2] School of Education, Tel Aviv University, Tel Aviv, Israel, [3] Department of Communication Disorders, Haddassa College, Jerusalem, Israel

OPEN ACCESS

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
Marilyn Vihman,
University of York, United Kingdom
Sharon Armon-Lotem,
Bar-Ilan University, Israel
Elisabet Serrat,
University of Girona, Spain

***Correspondence:**
Elitzur Dattner
elitzurd@gmail.com

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

Received: *02 June 2021*
Accepted: *31 August 2021*
Published: *13 October 2021*

Citation:
Dattner E, Levie R, Ravid D and Ashkenazi O (2021) Patterns of Adaptation in Child-Directed and Child Speech in the Emergence of Hebrew Verbs. Front. Psychol. 12:719657. doi: 10.3389/fpsyg.2021.719657

Children approach verb learning in ways that are specific to their native language, given the differential typological organization of verb morphology and lexical semantics. Parent-child interaction is the arena where children's socio-cognitive abilities enable them to track predictive relationships between tokens and extract linguistic generalizations from patterns and regularities in the ambient language. The current study examines how the system of Hebrew verbs develops as a network over time in early childhood, and the dynamic role of input-output adaptation in the network's increasing complexity. Focus is on the morphological components of Hebrew verbs in a dense corpus of two parent-child dyads in natural interaction between the ages 1;8-2;2. The 91-hour corpus contained 371,547 word tokens, 62,824 verb tokens, and 1,410 verb types (lemmas) in CDS and CS together. Network analysis was employed to explore the changing distributions and emergent systematicity of the relations between verb roots and verb patterns. Taking the Semitic root and pattern morphological constructs to represent linked nodes in a network, findings show that children's networks change with age in terms of node degree and node centrality, representing linkage level and construct importance respectively; and in terms of network density, as representing network growth potential. We put forward three main hypotheses followed by findings concerning (i) changes in verb usage through development, (ii) CS adaptation, and (iii) CDS adaptation: First, we show that children go through punctuated development, expressed by their using individual constructs for short periods of time, whereas parents' patterns of usage are more coherent. Second, regarding CS adaptation within a dynamic network system relative to time and CDS, we conclude that children are attuned to their immediate experience consisting of current CDS usage as well as previous usage in the immediate past. Finally, we show that parents (unintentionally) adapt to their children's language knowledge in three ways: First, by relating to their children's current usage. Second, by expanding on previous experience, building upon the usage their children have already been exposed to. And third, we show that when parents experience a limited network in the speech of their children, they provide them with more opportunities to expand their system in future interactions.

Keywords: CS-CDS adaptation, network analysis, Hebrew, roots and patterns, dynamic network analysis

INTRODUCTION

Network analysis is increasingly common in various areas of science, from social studies to the spread of epidemics (Kolaczyk, 2009), as it captures relations within the data and allows the statistical assessment of the structure of links between data components (Chen et al., 2018). In linguistics, network analysis has mostly been used to explain the structure and development of semantic networks (Beckage et al., 2011). The present study aims to model the development of Hebrew verb morphology—that is, the system of relations between roots and inflected patterns. We look at patterns of adaptation between Child Speech and Child Directed Speech (van Geert, 1991; van Dijk et al., 2013), expressed in changes within their respective morphological systems. The development of the system is shown to be complex and dynamic, such that attributes of the child's system are affected by other attributes within the system, as well as by the parent's system, and vice versa. In order to account for the verb lexicon morphology as a system, we adopt a network-based framework that allows for measuring complex relations between morphological constructs and their dynamic changes as a function of development and adaptation.

In light of these objectives, the current paper extends linguistic network analysis in two important directions. One is developmental: while language learning makes use of low-level generalizations, taking into account frequency and similarity of exemplars (Ambridge, 2020), the adaptive nature of language development entails the growing complexity of networks (Beckner et al., 2009). A second direction is morphological: Network analysis makes it possible to underscore the role of links between morphemes and the structure that emerges from these connections. The present study utilizes measures of network structure to explain early morphological development of the Hebrew verb system in the context of parent-child interaction and adaptation.

Input–Output Relations in Language Development

Parent-child interactions constitute the arena in which children use their cognitive and social abilities to extract patterns and regularities from the ambient language. Interactional support, linguistic adaptation and conceptual challenge promote language learning during these interactions (Rowe and Snow, 2020). In the realm of usage-based language acquisition, this type of linguistic input, also termed Child Directed Speech (CDS), is fine-tuned to the child's age and linguistic abilities (Snow, 1995; Ko, 2012). For the child, CDS is the major source of information about the morphology, syntax and semantics of the language being acquired (Hoff-Ginsberg, 1985; Maslen et al., 2004; Behrens, 2006). Usage-based analyses have shown that children detect patterns in the speech they hear and form generalizations by using the socio-cognitive abilities of intention reading, coupled with statistical learning and consequent schematization (Saffran, 2003; Tomasello, 2003, 2006, 2009). Abstract categories gradually emerge out of the items children have learned, based on the distributional and frequency properties of the input (Lieven et al., 2003; Tomasello, 2004; Lieven, 2008).

Focusing on the acquisition and development of verbs, studies on input-output relations have revealed clear correlations between features of verbs in CDS and their realization in Child Speech (CS). These include morphological characteristics of verbs (Acsu-Koc, 1998; Xantos et al., 2011) and their lexical semantics (Montag et al., 2015); syntactic properties of verbs in their environments (Naigles and Hoff-Ginsberg, 1998; Goldberg, 2006; Arunachalam et al., 2011); and their pragmatic features (Cameron-Faulkner, 2012; Clark and de Marneffe, 2012; Ninio, 2014). Of particular interest to the present study is the development of Hebrew verb morphology as a system that develops over time in early childhood, and the role of input-output relations and adaptation in the network's increasing complexity.

Recent studies have shown that Hebrew acquiring toddlers rely on stable, frequently occurring inflectional verb affixes in maternal input to gain salient information on the opaque, irregular verbs they frequently encounter (Ashkenazi et al., 2016). Furthermore, correlations were found between Child Directed Speech and Child Speech in terms of verb lemmas and their morphological components—structural root categories, *binyan* conjugations, and derivational verb families. Clear CDS-CS relations were also found between lexical-derivational development and inflectional growth as measured by Mean Size of Paradigm (MSP; Ashkenazi et al., 2020). The current study delves deeper into the development of morphological complexity in the verb domain by computing developmental changes in root, pattern and inflectional morphology in the dyadic interactions of two toddlers and their respective parents.

Morphological Constructs in Hebrew Verbs

Three morphological constructs are relevant to the current study: Semitic roots, *binyan* patterns, and subject-verb agreement markers.

The Semitic Root Network

The morphological construct termed the Semitic root (e.g., *m-s-r* "deliver," *g-d-l* "grow") is a central feature of Semitic languages. This is a (usually) tri-literal consonantal string that constitutes the formal and semantic core of many Hebrew words, and most especially, of all Hebrew verbs (Laks, 2013; Kastner, 2019; Ravid, 2019). Many studies point to the Semitic root as the most accessible Hebrew morpheme in spoken and written language development and usage (Ravid and Bar-On, 2005; Gillis and Ravid, 2006; Schiff et al., 2012; Ben-Zvi and Levie, 2016; Deutsch and Kuperman, 2019), including contexts of language disability or environmental deprivation (Ravid and Schiff, 2006; Schiff and Ravid, 2007; Levie et al., 2017, 2019). Young Hebrew-speaking children demonstrate an early ability to extract roots from familiar words and use them in novel forms (Berman, 1985, 2000, 2012; Ravid, 2003). While a root is not a verb, it functions as a consonantal skeleton shared by different verbs—e.g., *r-d-m* in *nirdam* "fall asleep," *hirdim* "make sleep;" or *r-g-l* in *hirgil* "make familiar" and *hitragel* "get used"—often carrying a shared basic lexical semantics, creating derivational verb families (Levie et al., 2020). Therefore, roots are key in Hebrew morpho-lexical development as the organizers of root-based networks in the verb lexicon. Within the network-based framework of the current paper, the root is a morphological construct which is conceived as a node in a morphological network.

The Semitic *Binyan* Network

As a consonantal, discontinuous entity, the Semitic root is not pronounceable, and as a sub-lexical bound morpheme, it has no lexical category. It is thus always complemented by the Semitic *binyan* (lit. "building"), a prosodic template interspersing root radicals with vowels, often preceded or followed by a small set of pattern affixes, as in *maskim* "agrees," pattern *maCCiC*. There are seven *binyan* conjugations respectively termed *Qal, Nif'al, Hif'il, Huf'al, Pi'el, Pu'al*, and *Hitpa'el*, which are affixed to roots to create verb lemmas (Schwarzwald, 1981; Berman, 1993a,b; Berman, 2012). For example, *siper* "tell" is expressed by the combination of root *s-p-r* and *binyan Pi'el*; *yarad* "go down" as the combination of root *y-r-d* with *Qal*, and *horid* "take down" as the combination of root *y-r-d* with *Hif'il* (the last two sharing a root, but being two discrete verb lemmas).

In tandem with roots, *binyan*-based conjugations thus constitute networks organizing the Hebrew verb lexicon in morpho-phonological patterns associated with a set of transitivity and Aktionsart functions (Berman, 1993a,b; Kastner, 2016; Ravid, 2019). On the one hand, root-*binyan* verb lemmas form derivational verb families, where verbs with different *binyan* patterns are based on a single shared root (Bolozky, 1999; Ravid, 2019). Consider, for example, the network of verbs sharing root *l-m-d*: *lamad* "learn" (in *Qal*), its passive counterpart *nilmad* "be learned" (*Nif'al*), the causative verb *limed* "teach" (*Pi'el*), and the middle-voice verb *hitlamed* "apprentice" (*Hitpa'el*) (Berman, 1987). From a complementary perspective, verbs with different roots share the same *binyan* conjugation, as demonstrated by the causative verbs *higbir* "make stronger," *higdil* "make bigger," *histir* "hide,Tr" and *hiklit* "record," all sharing the *Hif'il* pattern, with different roots. Similarly to roots, the current framework takes a *binyan* (with temporal patterns and agreement inflections, see below) to be a morphological construct which is conceived as a node in a morphological network.

Note that in young children and parental speech, most verbs are based on *Qal*, the most prevalent *binyan* in Hebrew. With age, children are exposed to larger root-pattern networks that highlight the shared vocalic structure of verbs, making it possible for *binyan* conjugations and their syntactic-semantico values to be learned (Levie et al., 2020). The increase in number, size and complexity of networks of root-related derivational verb families is a clear indicator of a growing verb lexicon (Ravid et al., 2016; Levie et al., 2019).

Temporal Patterns Within Binyan ***Conjugations***

In the current framework of analysis, the notion of verb pattern relates the derivational notion of *binyan* to the inflectional paradigm within each *binyan*. Each of the seven conjugations termed *binyanim* actually consists of a phonologically unique bundle of five temporal patterns—past tense, present tense, future tense, imperative, and infinitive forms—as depicted in **Table 1**. Temporal pattern templates determine the basic morpho-phonology of the verb stem, including root radical slots and vowel combinations. This means that temporal shifts within the same *binyan* paradigm require the use of the same root, each time combining with a different *binyan*-unique temporal pattern. For example, *CaCaC, CoCeC,* and *li-CCoC* (where C's stand for root radicals) serve as the respective past, present and infinitive patterns of *Qal*. When combined with root *k-t-b* "write," the stems *katav* "wrote," *kotev* "writes/writing," and *li-xtov* "to-write" are yielded, respectively. In the same way, *hiCCiC, maCCiC, yaCCiC,* and *le-haCCiC* serve as the respective past, present, future and infinitive patterns of *Hif'il*, combining with *k-t-b* to respectively yield *hixtiv* "dictated," *maxtiv* "dictates/dictating," *yaxtiv* "will dictate," and *le-haxtiv* "to-dictate." Given the prominence of the root and *binyan* morphemes in the Hebrew lexicon, this process is critical in the acquisition of verb morphology (Berman, 1987; Ravid, 2003).

As **Table 1** shows, Hebrew speaking children are faced with 31 *binyan*-specific temporal patterns that need to be learned. In the current analysis, when we refer to a *verb pattern*, we actually refer to one of these 31 *binyan*-unique temporal patterns. From a developmental perspective, this construal of verb patterns is a facilitating property of the system, so that for the young child, root-based relations in the verb system can first be learned by attending to the root-pattern temporal shifts within the same *binyan* (Ashkenazi et al., 2016, 2020). **Table 1** shows that, while some temporal patterns are phonologically similar (e.g., the temporal paradigm of *Hitpa'el*), others (e.g., those of *Qal* and *Nif'al*) display more phonological distinctions. This is important, as *Qal*, which occupies about 80% of the verb tokens heard or produced by children up to 3 years of age, has the most phonologically distinct temporal patterns, a boost to the transparency-aided acquisition of root and pattern structure (Ravid, 2019).

To illustrate the central role of this network, think about noting the formal resemblance of verbs sharing the *meCaCeC* present-tense *Pi'el* pattern (e.g., *medaber* "talking," *meshaker* "lying," *melamed* "teaching"), the similarity of their temporal semantics, and their relation to other *Pi'el* patterns such as past-tense *CiCeC* in *diber* "talked," *shiker* "lied," and *limed* "taught" respectively.

Recent research (Ashkenazi et al., 2016, 2020; Ravid et al., 2016) indicates that young Hebrew-speaking children initially learn to manipulate roots and patterns in the inflectional shifts across the temporal stems in the paradigm of a single *binyan* (most often the ubiquitous *Qal*), where semantic coherence of roots is highest. This is in fact the launching pad of non-linear formation in the verb system. Evidence of errors from toddlers and young children acquiring the *binyan*-temporal system indicates that it takes time and linguistic experience for this knowledge to crystallize toward the beginning of elementary school (Berman, 1982; Ravid, 1995). It is only later on, at schoolage, that verb lemmas in different *binyan* conjugations sharing the same root—i.e., derivational families—enrich the young verb lexicon (Levie et al., 2020). The larger, more numerous and varied root-based verb networks in the lexicon of the language learner (both inflectional, across the temporal paradigm of a single *binyan*, and derivational, across different *binyan* conjugations)—the more complex, productive

TABLE 1 | The seven *binyan* conjugations as sets of temporal patterns.

Binyan	Past tense	Present tense	Future tense	Imperative	Infinitive
Qal	*CaCaC*	*CoCeC*	*yiCCoC*	*CCoC*	*liCCoC*
Nif'al	*niCCaC*	*niCCaC*	*yiCaCeC*	*hiCaCeC*	*lehiCaCeC*
Hif'il	*hiCCiC*	*maCCiC*	*yaCCiC*	*haCCeC*	*lehaCCiC*
Huf'al	*huCCaC*	*muCCaC*	*yuCCaC*	—	—
Pi'el	*CiCeC*	*meCaCeC*	*yeCaCeC*	*CaCeC*	*leCaCeC*
Pu'al	*CuCaC*	*meCuCaC*	*yeCuCaC*	—	—
Hitpa'el	*hitCaCeC*	*mitCaCeC*	*yitCaCeC*	*hitCaCeC*	*lehitCaCeC*

TABLE 2 | Subject-verb agreement in Hebrew verbs.

Temporal category	Person	Number	Gender
Infinitive	X	X	X
Imperative	V	V	V
Future tense	V	V	V
Present tense	X	V	V
Past tense	V	V	V

and abstract the organization of the lexical network relying on roots (Levie et al., 2020).

Agreement Inflection

The verb stem created by the non-linear affixation of root plus *binyan*-unique temporal pattern is further inflected for number, gender, and person in agreement with the grammatical subject. Unlike temporal shifts, verb agreement inflection is linear, taking the verb stem rather than the root as its base. For example, *nimsor* "we will deliver" is composed of root *m-s-r* in the future tense pattern of *Qal*, with the prefix *n-* designating the first person plural; and *masru* "they delivered" is composed of root *m-s-r* in the past tense pattern of *Qal*, with the suffix *-u* designating the third person plural. Note, however, that the actual formation of a specific verb (wordform) requires morpho-phonological changes in the stem that are typical of each *binyan*, root type and temporal category. This is not investigated in our current study.

Table 2 presents an overview on agreement marking of Hebrew verbs. In general, it shows that the only temporal category which does not require agreement inflection is the infinitive form; and that present tense verbs are marked for number and gender, but not for person agreement.

Table 3 presents a detailed view of the 25 pattern-inflection categories identified in Ashkenazi's (2015) corpus, which constitutes the database of the current study. Each category represents a temporal pattern (Infinitive, Imperative, Future tense, Present tense, or Past tense) with all possible agreement marking (e.g., past tense 3rd person plural). The actual examples in **Table 3** are the 25 wordforms constituting the temporal category-agreement inflectional paradigm of *Qal* with root *l-q-ħ* "take."[1]

The Current Research

Against this background, the present study has two main objectives: (i) to model the systematic development of the morphology of the Hebrew verb lexicon—that is, the system of relations between roots and inflected patterns; and (ii) to account for various patterns of adaptation between Child Speech and Child Directed Speech (van Geert, 1991; van Dijk et al., 2013), expressed in changes within their respective morphological system structures. The development of the system is shown to be adaptive and complex, such that attributes of the child's system are affected by other attributes within the system, as well as by the parent's system structure, and vice versa. Both the systematic development and the patterns of adaptation are shown to be dynamic, in that the system's structure at one point in time affects its structure in the future. In order to account for verb morphology as a dynamic system, we adopt a network-based framework that allows for measuring complex relations between morphological constructs and their dynamic changes as a function of development and adaptation.

A Dynamic Network Model assumes that higher-order properties are emergent phenomena, such that structure emerges on the basis of the dynamic interactions between lower-level components (Barabasi, 2009; Den Hartigh et al., 2016). This view is compatible with recent usage-based approaches to cognitive representation of language, in which learning is construed as constantly updating connection weights between nodes based on experience (Bybee and McClelland, 2005; Kapatsinski, 2018). The morphological network of the verb lexicon is dynamic in the sense that the values of the constructs it comprises change as a consequence of the interactions with other morphological constructs (among other factors). For example, the importance of a particular root within the verb lexicon can be affected by the importance of the pattern(s) it is linked to (creating specific verb wordforms). Thus, if a low frequency root is linked to a low frequency pattern, it may have consequences for the entrenchment of the verb wordform within cognitive representation, and thus for future usage.

Dynamic network analysis can be helpful in accounting for another facet of dynamicity: over developmental time, morphological constructs may appear or disappear; it is not the case that we use every single root, *binyan* temporal pattern and agreement inflection in our lexicon every single day.

[1] Roots are represented as morphological entities, that is, taking into account their morpho-phonological behavior, as detailed in Ravid (2012). For example, the *k* phoneme in the root *k-t-b* alternates with spirant *x* phoneme, while the *k* phoneme in *q-d-m* does not (Temkin Martinez, 2010). For words (in contrast to roots) we use a broad phonemic transcription.

TABLE 3 | Hebrew verb inflectional categories.

Coding	Inflectional category	Example (root *l-q-h* + *Qal*)
1	Infinitive	*lakáxat'*[a] "to take"
2	Imperative, masculine, singular	*kax* "take.Masc"
3	Imperative, feminine, singular	*kxi* "take.Fm"
4	Imperative, plural	*kxu* "take.Pl"
5	Future, 1st person, singular	*ekax* "I will take"
6	Future, 2nd person, masculine, singular	*tikax* "you.Masc.Sg will take"
7	Future, 2nd person, feminine, singular	*tikxi* "you.Fm.Sg will take"
8	Future, 3rd person, masculine, singular	*yikax* "he will take"
9	Future, 3rd person, feminine, singular	*tikax* "she will take"
10	Future, 1st person, plural	*nikax* "we will take"
11	Future, 2nd person, plural	*tikxu* "you.Pl will take"
12	Future, 3rd person, plural	*yikxu* "they will take"
13	Present, masculine, singular	*loké'ax* "take/s/taking.Masc"
14	Present, feminine, singular	*lokáxat* "take/s/taking.Fm"
15	Present, masculine, plural	*lokxim* "take/taking.Pl"
16	Present, feminine, plural	*lokxot* "take.Pl.Fm"
17	Past, 1st person, singular	*lakáxti* "I took"
18	Past, 2nd person, masculine, singular	*lakáxta* "you.Masc.Sg took"
19	Past, 2nd person, feminine, singular	*lakaxt* "you.Fm.Sg took"
20	Past, 3rd person, masculine, singular	*lakax* "he took"
21	Past, 3rd person, feminine, singular	*lakxa* "she took"
22	Past, 1st person, plural	*lakáxnu* "we took"
23	Past, 2nd person, masculine, plural	*lakáxtem* "you.Masc.Pl took"
24	Past, 2nd person, feminine, plural	*lakáxten* "you.Fm.Pl took"
25	Past, 3rd person, plural	*lakxu* "they took"

[a] *Stress in Hebrew is usually final, thus it is only marked if penultimate.*

Treating development as a dynamically changing set of networks enables us to evaluate such punctuated growth, accounting for accumulated change. For example, the probability of using a particular pattern on a particular day may be higher if that pattern was used the day before (either by the child or by the parent) than if it was not. In the following section we present our data and methods for constructing the network and modeling development.

DATA AND METHOD

Data

The analyses reported below are based on a densely recorded corpus of naturalistic longitudinal interactions of two Hebrew-speaking parent-child dyads—a boy dyad and a girl dyad. The boy dyad was recorded between the ages 1;8.27 (1 year 8 months and 27 days, or 635 days) to 2;2.3 (2 years 2 months and 3 days, or 795 days), yielding 49 recording sessions. The girl dyad was recorded between the ages of 1;9.25 (1 year 9 months and 25 days, or 664 days) to 2;2.19 (2 years 2 months and 19 days, or 810 days), yielding 47 recording sessions. Different child genders were chosen so as to permit analysis of the obligatory gender agreement in Hebrew verb inflection (Schwarzwald, 1998; Ravid and Schiff, 2015). Both families were from mid-high SES background, living in central Israel. The two sets of parents, who did not know each other, were monolingual native-born speakers of Hebrew. They did not receive any monetary remuneration for their voluntary participation.

Both children were first-born and had no siblings at the time of recording. Both had normal cognitive, communicative, and linguistic development according to parental report (including the Hebrew CDI checklist in Maital et al., 2000), periodic assessment at the local neonate and children's health clinic, and assessment by the last author (a certified senior SLP). Neither of them had a history of ear infections or any other major health issues. The boy attended nursery school and the girl did not. **Table 4** summarizes the corpus details (Ashkenazi, 2015).

Data Collection

The children were audio- and video recorded by their parents at home during bath time, play time and meal time using an MP3 recorder and a video camera supplied to the family. Each dyad was audio recorded twice a week and video recorded once a week, for 45–60 min each time, for 6 months between 1;8-2;2 approximately (see details above). The parents were informed that the study concerned early language development in Hebrew. They were asked to record spontaneous, natural interactions. Recordings of both dyads started when each child started producing two word utterances and some verbs, based on parental reports using the Hebrew CDI (Maital et al., 2000).

TABLE 4 | The corpus details.

		Girl (Child 1)	Boy (Child 2)
Age range		1;9.25-2;2.19	1;8.27-2;2.3
# recordings		47	49
Word tokens	CS	39,717	32,369
	CDS	158,679	140,782
Verb types	CS	204	172
= lemmas	CDS	531	503
Verb tokens	CS	4,610	3,101
	CDS	31,283	23,830

Transcriptions of the recordings (see below) ceased when each child produced subject-verb agreement in number and gender in two subsequent recordings, including two different person agreements in past tense, on at least two different verbs. This morphosyntactic criterion indicated that the child was gaining command of the basic components of verb structure and semantics by productively using temporal stems, that is, root-pattern alternations, as well as agreement markers (Berman and Lustigman, 2012; Ravid et al., 2016). All interactions were coded and analyzed, including nursery rhymes and songs in the parental input, as well as speech addressed to the other parent (which consisted <5% the recordings).

Transcription

Dyadic interactions were transcribed in broad phonemic transcription following the CHILDES conventions (MacWhinney, 2005), adapted to take into account Hebrew-specific phonemic, phonological, prosodic, and orthographic features (Albert et al., 2013). The transcriptions were carried out by undergraduate students of an academic SLP program who took a CHILDES course as part of their studies. The recordings were thoroughly checked by the last author and corrected when necessary, with an estimated 5% error rate. Next Hebrew MOR was run over the transcripts. Ambiguous forms and verb forms that were not analyzed by the program were identified and coded manually.

Method

Morphological Variables

Three variables participated in the network analysis described below:

1. Root: The Semitic consonantal construct at the basis of the Hebrew verb, e.g., *s-y-m* "put," *z-h-r* "take care," *r-d-m* "sleep," or *n-g-b* "towel."
2. Pattern + Agreement: This was the complementary construct to the root. In the current analysis, it consisted of (i) the *binyan*-specific temporal pattern (see **Table 1** for the full array of *binyan*-temporal patterns); and (ii) the person-number-gender agreement inflection (see **Table 3** for the full array of agreement inflections). Note that 1 and 2 are the two morphological constructs that participate in the verb structure, rather than actual words.

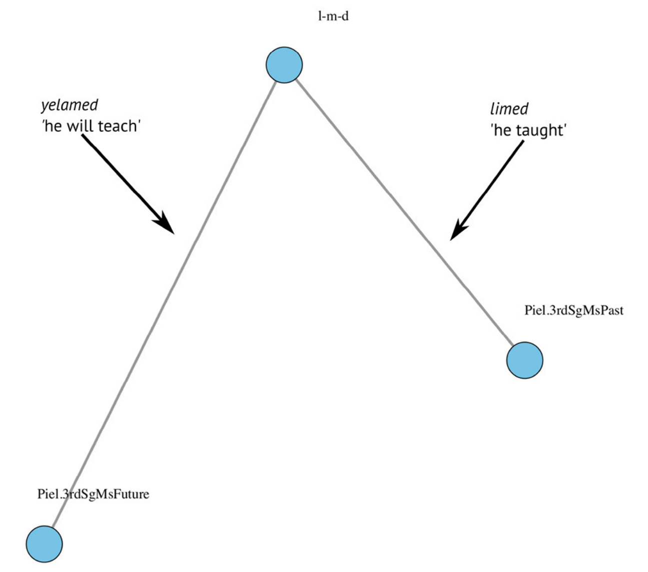

FIGURE 1 | A Hebrew morphological verb network: an illustration. The root *l-m-d* is linked to two *binyan* temporal-inflection patterns, yielding two verb wordforms: *l-m-d*+*Piel*.3rdSgMsPast is *limed* "He taught," and *l-m-d*+*Piel*.3rdSgMsFuture is *yelamed* "He will teach".

3. Verb wordform: The *actual* word as appearing in the transcription: a unique combination of a root and a *binyan*-temporal pattern + agreement marking, as in the following four examples (see also **Figure 1**):

   1. *sámti* "I put" = *s-y-m* + *Qal* past tense, 1st Sg;
   2. *tizahari* "take care!" = *z-h-r* + *Nif'al* imperative, 2nd Sg Fm;
   3. *nirdamim* "are falling asleep" = *r-d-m* + *Nif'al* present tense, Pl Masc;
   4. *yitnagev* "(he will) towel (himself)" = *n-g-b* + *Hitpa'el* future tense, 3rd Sg Masc.

Network Analysis

For each child and parent in the data we constructed a list of all available roots and inflected patterns throughout the entire database, resulting in four lists. These lists formed the basis for the network analysis, such that each participant used a subset of their list on a particular recording. The nodes of the bipartite network of recording *N* are the list of roots and inflected *binyan* patterns that appear in recording *N*, creating links that stand for the actual verb wordforms that were used in recording *N* by participant X. For example, a token of the root *l-m-d* and the inflected pattern <*Pi'el*, masculine, singular, third person, past tense> constitutes one link yielding the wordform *limed* "taught.3.Sg.Ms."; while a token of the same root (*l-m-d*) and the inflected pattern <*Pi'el*, masculine, singular, third person, future tense> constitutes another link, yielding the wordform *yelamed* "will teach.3.Sg.Ms." That is, verb wordforms are *links* between *nodes* in a morphological *network*, as illustrated in **Figure 1**.

Node Level Measure: Degree

Node degree is a centrality measure, corresponding to the number of links a node has with other nodes in a network. Degree (C_D) is calculated as:

$$C_D(j) = \sum_{j=1}^{n} A_{ij} \tag{1}$$

for every node in the data, over its corresponding rows and columns of the matrix A.

A node with a high degree value is more important in the network as it participates in more language events. *Degree* corresponds to the token frequency of each construct. Thus, a network with a few high degree nodes indicates repeated use of particular types, suggesting a less varied network. We hypothesize that the degree level of nodes will increase with age in the CS, and that degree distribution within the CS networks will change with age, indicating usage variation. These changes are not hypothesized to occur in the CDS networks.

Node Level Measure: Eigenvector Centrality

A second centrality measure used here is the *eigenvector centrality* of particular nodes - roots or inflected patterns in our case. To achieve a relevant explanatory assessment of the data, we focus here on eigenvector centrality as reflecting the node's importance (Bonacich, 2007; Lohmann et al., 2010; Oldham et al., 2019). Eigenvector centrality x_i of node i is given by:

$$x_i = \frac{1}{\lambda} \sum_{k} a_{k,i} x_k \tag{2}$$

where $\lambda \neq 0$ is a constant, and k is the node's degree.

A node with high centrality is linked to many other nodes that, in turn, are linked to many other nodes. In non-directed networks, as in the present study, such nodes are said to be in a central, prominent position. For example, an inflected pattern linked to two roots that are themselves linked to three inflected patterns each is higher in centrality than an inflected pattern linked to two roots that are not linked to other patterns. *Centrality* quantifies the significance of a node relative to other nodes in the network. For example, centrality can reveal those morphological patterns that act as centers of gravity for forming verbs, and changes in centrality of a particular pattern can be measured during development. We hypothesize that nodes' centrality will change through development in a non-linear manner, reflecting changes in discourse circumstances, in both the CS and CDS networks. Crucially, these changes are not a matter of mere frequency, but rather of the frequency of links with other frequent nodes.

Network Level Measure: Density

While degree and centrality are measures concerning attributes of the *nodes* of the network, the *density* measure concerns the network as a whole. The density of the network is a measure of fulfilled links between nodes (Wasserman and Faust, 1994). Density is a mathematical notion that measures the proportion of observed links relative to the maximum number of possible links: the closer it is to one, the more possible links are actually manifested, and thus the more interconnected the network. Network density (d) is calculated as:

$$d = \frac{m}{n(n-1)/2} \tag{3}$$

where m is the total number of existing links in the network, and n is the number of nodes in the network. Links within a dense network are more predicted and anticipated. As such, somewhat counter-intuitively, a *sparse* network is taken here to indicate a higher level of potential productivity: In a sparse network, there are more root- and pattern-nodes which are not linked to each other, compared to a dense network in which most of the nodes are already linked. Hence, the potential to link two nodes that have not been linked before, thus creating new verb wordforms, is *higher* in a sparse network, compared to a dense network (Levie et al., 2019). That is, a sparse network means that the pool from which one can choose how to put experience into words, specifically verbs (by linking a root and a pattern) is not exhausted, and new verb wordforms can be created: new links between roots and inflected patterns, which refer to more fine-grained aspects of experience. We hypothesize that network density will decrease with age within the children's networks, but will remain steady through time in the parents' networks.

Network Construction and Model Design

For every recording we calculated two networks, one for each participant. This resulted in 94 networks for Child 1 [47 recordings * (CS + CDS)], and 98 networks for Child 2 [49 recordings * (CS + CDS)]. We account for these networks as consecutive points in a dynamically evolving network, analyzing the development of network measures as obtained in each instance of network. The three measures presented above were extracted for each network, resulting with a time series data of network density for every participant, the changing degree of each node in the networks through development, and the changing centrality of each node through development.

In order to find patterns of adaptation in network structure as representing the verb lexicon, we assessed the development of network measures for each child and parent separately, and modeled the effect of the child's age on each measure, the effect of CDS network measures on CS network measures, and vice versa. Moreover, since time related data are available, we added to the models the level of each measure in the preceding recording, enabling further assessment of adaptation. For example, we could ask whether the density level of the child's network in the preceding recording affects the parent's level of density in the network of the current recording.

Furthermore, for each node that appeared on a particular day in both the child's and the parent's networks, we modeled its degree and centrality (in CS and CDS, in Child 1 and Child 2, separately) as a function of the other measures in the same recording, as well as the levels of the other measures in the preceding recording. For example, we could ask whether the

TABLE 5 | Summary of study variables.

Variable	Interpretation
Situational variables	
Age	Child's age
Speaker	Child Speech (CS) vs. Child Directed Speech (CDS)
Morphological variables	
Verb root	A node in the network
Verb inflected pattern	A node in the network
Verb wordform	A link in the network (linking a root and a pattern)
Network measures	
degree.cs	CS node degree at recording N
prior.degree.cs	CS node degree at recording N-1
degree.cds	CDS node degree at recording N
prior.degree.cds	CDS node degree at recording N-1
centrality.cs	CS node centrality at recording N
prior.centrality.cs	CS node centrality at recording N-1
centrality.cds	CDS node centrality at recording N
prior.centrality.cds	CDS node centrality at recording N-1
density.cs	CS network density at recording N
prior.density.cs	CS network density at recording N-1
density.cds	CDS network density at recording N
prior.density.cds	CDS network density at recording N-1

chance of a root or inflected pattern produced by the child to have high centrality is higher if this root or inflected pattern is central in the parent's network in the preceding recording, and/or is central in the child's preceding recording, and/or has a high degree level in the current recording.

Table 5 summarizes the variables and measures in the study that were part of either the construction of the networks or the model design in analyzing adaptation through development. Each morphological/situational variable and network measure is a part of the four participants design: CS of Child 1, CDS of Child 1, CS of Child 2, and CDS of Child 2. Consequently, the results reported below present four models for each measure. All resulting measurements of the network analysis were centered and scaled before model calculations.

RESULTS

Overall Outlook

We start off the presentation of our results with an overall outlook on the four changing temporal networks (two networks of children's speech, CS 1 and CS 2; and two networks of Child Directed Speech, CDS 1 and CDS 2), from a dynamic perspective that underscores the emergence of the system. These networks are shown in **Figure 2** through four representative time points within the longitudinal corpus: recordings no. 1 (age 1;9.25 for Child 1, 1;8.27 for Child 2); 16 (age 1;11.6 for Child 1, 1;10.7 for Child 2); 31 (age 2;1.1 for Child 1, 2;0.0 for Child 2); and 46 (age 2;2.17 for Child 1, 2;1.23 for Child 2). The nodes of the networks are roots and inflected patterns. Links between nodes represent verb wordforms.

Figure 2 shows that the children's networks go through much more development than the parents' networks, such that more nodes and more links between nodes are shown with time. That is, in morphological terms, we see growth in the number of roots and inflected patterns, and growth in the number of verb wordforms (cf. Ashkenazi, 2015). Growth in number of nodes and links renders a more complex network, as can be seen by the complex structure of the children's networks in older ages. Moreover, we can see that the structures of the parents' networks remain similar throughout the data, such that it is very complex from the very beginning. The view presented by **Figure 2** allows us to observe growth in complexity in a visual manner. The models presented below will add to this view, relating system development to multiple factors. However, before turning to the models results, let us emphasize another facet of dynamic network analysis, that of node activation, as shown in **Figure 3**.

Morphological constructs in a dynamic perspective are portrayed according to their activation patterns. For example, a link between a root and an inflected pattern may appear in recording number 6 (i.e., the link is active), be absent from recording number 7 (i.e., the link is inactive), and reappear in recording number 8. **Figure 3** portrays a timeline of inflected pattern activation throughout development (in the age ranges of the current corpus; root nodes are not represented in order to increase readability of the plot). We can see that the children's networks are characterized by what we term *punctuated development*, such that most of the inflected patterns appear and disappear frequently; while the parents' networks are characterized by more continuous usage of the full array of inflected patterns. This characterization of the development of

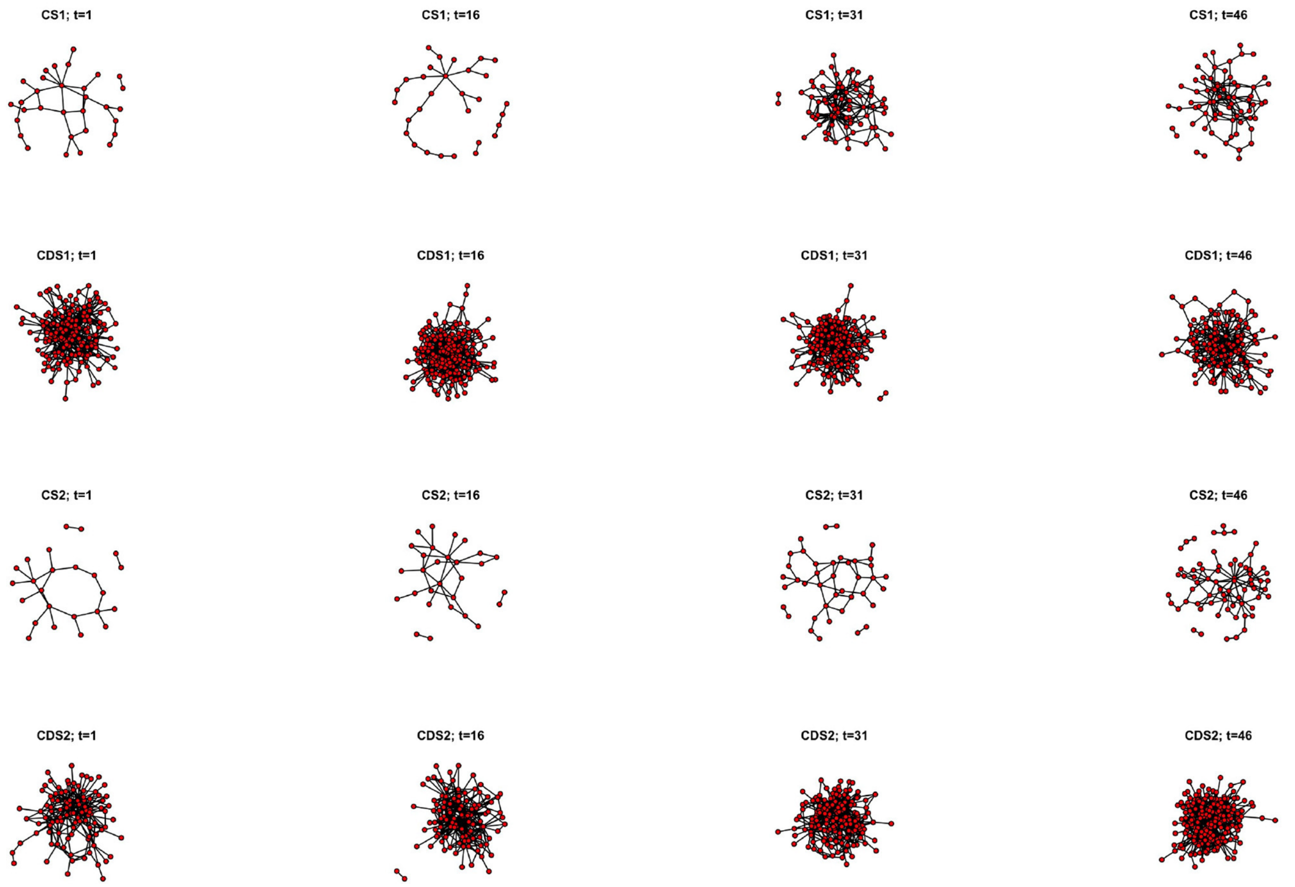

FIGURE 2 | Dynamic view of a developing verb lexicon as a network. Top row: CS of child 1; second row: CDS of child 1; third row: CS of child 2; bottom row: CDS of child 2. Each row portrays four points in the longitudinal corpus: recordings no. 1 (age 1;9.25 for Child 1, 1;8.27 for Child 2); 16 (age 1;11.6 for Child 1, 1;10.7 for Child 2); 31 (age 2;1.1 for Child 1, 2;0.0 for Child 2); and 46 (age 2;2.17 for Child 1, 2;1.23 for Child 2). The nodes of the networks are roots and inflected patterns. Links between nodes represent verb wordforms.

the morphological system is made possible by the framework of dynamic network analysis, and we will return to its implications in the discussion section below. We now turn to the results of the models. First the two node-level measures (degree and centrality), and then the global network measure (density).

Node Level Measure: Degree

Recall that the degree of an inflected pattern within a network is the number of roots linked with it, and the degree of a root within a network is the number of inflected patterns it is linked to. **Figure 4** presents degree distribution through development. Every recording session (the X axis) is a single network within the entire set of networks through time. Each bar represents the degree distribution within a single network as a single point in time.

Figure 4 shows that CDS degree levels are much higher than CS degree levels in both sub-corpora, and that degree level in CS seems to increase with age in both children. That is, parents tend to link more roots to more inflected patterns, and more inflected patterns to more roots, compared with children's linkage distribution.

In order to assess development and adaptation relative to node degree, we fitted our models only on those nodes that appeared in both the child's and the parent's network. Thus, for each participant we fitted a linear mixed model (estimated using REML and nloptwrap optimizer) to predict degree level with the following variables: age, degree of the other party in the dyad in the same recording, and the degrees of both parties of the dyad in the previous recording. We also included eigenvector centrality measures of both parties in current and antecedent networks, and the same for density measures, in order to reveal complex relations within the system and to account for adaptation across time. Each model included the specific root or inflected pattern node as a random effect (coded as *name* in the models below). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. Ninety-five percent Confidence Intervals (CIs) and p-values were computed using the Wald approximation.

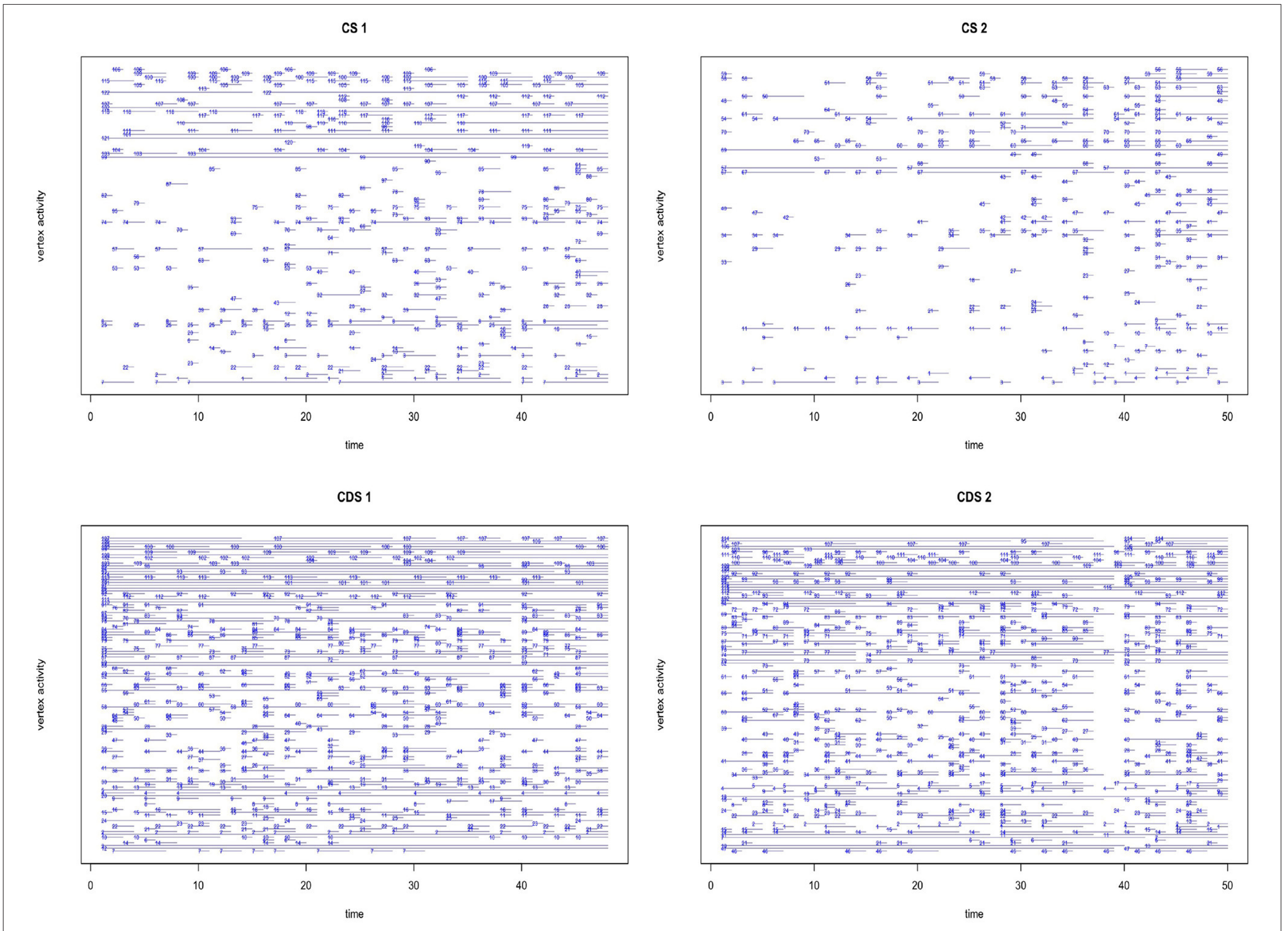

FIGURE 3 | Inflected patterns activity throughout the entire time range in the data. Each line represents the time in which a node is active in the network. Nodes are represented by the numbers at the beginning of each line (root nodes have been removed in order to increase readability).

CS Node Degree

Table 6 shows the results for the linear mixed models for both children, predicting CS degree level. Each model is detailed below.

CS Node Degree: Child 1

The model's total explanatory power is substantial (conditional $R^2 = 0.55$) and the part related to the fixed effects alone (marginal R^2) is of 0.50. Within this model, the following variables have a significant effect on CS1 degree level: child's age (positive effect), CDS degree level (positive effect), CDS degree level in recording N-1 (positive effect), CS degree level in recording N-1 (positive effect), CS centrality level (positive effect), CS density (negative effect), and CS density in recording N-1 (positive effect).

CS Node Degree: Child 2

The model's total explanatory power is substantial (conditional $R^2 = 0.49$) and the part related to the fixed effects alone (marginal R^2) is of 0.40. Within this model, the following variables have a significant effect on CS2 degree level: age (positive effect), CDS degree level (positive effect), CS degree level in recording N-1 (positive effect), CS centrality level (positive effect), CS density (negative effect), and CDS density in recording N-1 (negative effect).

CDS Node Degree

Table 7 shows the results for the linear mixed models for both parents, predicting CDS degree levels. Each model is detailed below.

CDS Node Degree: Child 1

The model's total explanatory power is substantial (conditional $R^2 = 0.78$) and the part related to the fixed effects alone (marginal R^2) is of 0.71. Within this model, the following variables have a significant effect on CDS1 degree level: age (negative effect), CS degree (positive effect), CDS degree in recording N-1 (positive effect), CDS centrality level (positive effect), CS density in recording N-1 (positive effect), CDS density (negative effect), and CDS density in recording N-1 (negative effect).

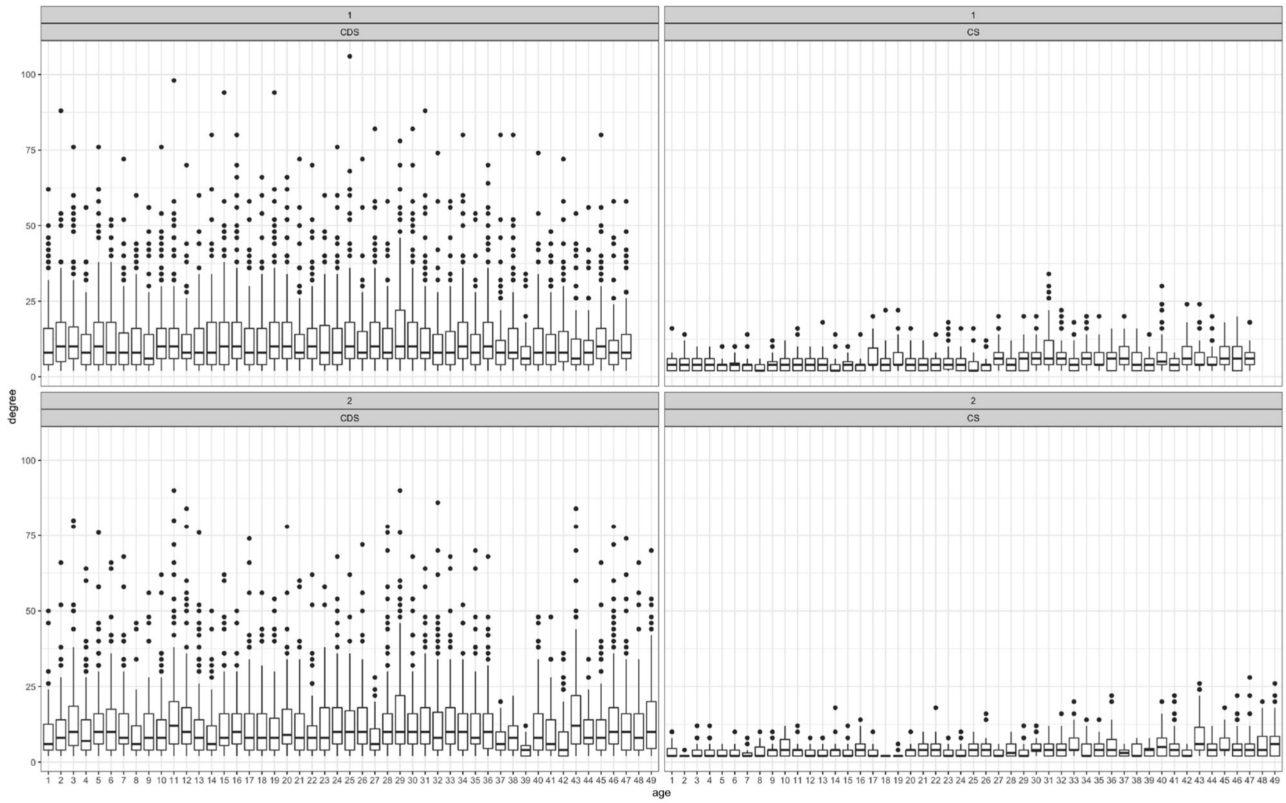

FIGURE 4 | Node degree distribution by session: CS (right panels) and CDS (left panels), child 1 (top panels) and child 2 (bottom panels).

CDS Node Degree: Child 2

The model's total explanatory power is substantial (conditional $R^2 = 0.77$) and the part related to the fixed effects alone (marginal R^2) is of 0.71. Within this model, the following variables have a significant effect on CDS2 degree level: age (positive effect), CS degree (positive effect), CDS degree level in recording N-1 (positive effect), CDS centrality (positive effect), CDS centrality in recording N-1 (positive effect), CS density in recording N-1 (positive effect), CDS density (negative effect), and CDS density in recording N-1 (negative effect).

Node Eigenvector Centrality

Recall that the eigenvector centrality of a node is a measure of importance. For example, an inflected pattern has high centrality if it is linked to many roots that are linked to other inflected patterns, that are linked to other roots in turn. **Figure 5** presents centrality distribution by recording day, showing a mirror image of degree distribution (**Figure 4**): Centrality levels within CS networks are much higher than CDS networks. That is, there are more central nodes within the children's network than there are within the parents' networks, and trends in centrality changes are less apparent in the CDS than in the CS.

The models summarized in **Tables 8**, **9** portray the development of node eigenvector centrality. In a similar model design for the one presented for node degree, we fitted a linear mixed model for each participant (estimated using REML and nloptwrap optimizer) to predict eigenvector centrality with age, centrality of the other party in the dyad in the same recording, and the centralities of both parties of the dyad in the previous recording. We also included degree values of both parties in current and antecedent networks, and the same for density values. Each model included the specific root or inflected pattern node as a random effect (coded as *name* in the models below). Standardized parameters were obtained by fitting the model on a standardized version of the dataset. Ninety-five percent Confidence Intervals (CIs) and *p*-values were computed using the Wald approximation.

CS Node Eigenvector Centrality

Table 8 shows the results for the linear mixed models for both children, predicting CS eigenvector centrality. Each model is detailed below.

CS Centrality: Child 1

The model's total explanatory power is substantial (conditional $R^2 = 0.46$) and the part related to the fixed effects alone (marginal R^2) is of 0.44. Within this model, the following variables have a significant effect on CS1 centrality level: age (negative effect), CDS centrality (positive effect), CS centrality in recording N-1 (positive effect), CDS centrality in recording N-1 (positive effect), CS degree (positive effect), and CS density (positive effect).

TABLE 6 | Linear mixed model: CS node degree.

Predictors	degree.cs: child 1		*p*	degree.cs: child 2		*p*
	Estimates	CI		Estimates	CI	
(Intercept)	−0.73	−0.87 −−0.59	<0.001	−0.7	−0.84 −−0.56	<0.001
age	0.01	0.01 – 0.01	<0.001	0.01	0.00 – 0.01	<0.001
degree.cds	0.07	0.04 – 0.10	<0.001	0.08	0.03 – 0.12	<0.001
prior.degree.cds	0.03	0.00 – 0.06	0.043	−0.01	−0.04 – 0.02	0.566
prior.degree.cs	0.23	0.16 – 0.30	<0.001	0.24	0.15 – 0.32	<0.001
centrality.cs	0.11	0.10 – 0.13	<0.001	0.05	0.03 – 0.06	<0.001
prior.centrality.cs	−0.01	−0.03 – 0.01	0.277	0.01	−0.00 – 0.02	0.241
centrality.cds	−0.03	−0.15 – 0.08	0.596	−0.07	−0.20 – 0.07	0.331
prior.centrality.cds	−0.06	−0.17 – 0.05	0.259	0.07	−0.01 – 0.14	0.078
density.cs	−0.06	−0.10 −−0.03	<0.001	−0.04	−0.06 −−0.02	<0.001
prior.density.cs	0.05	0.01 – 0.10	0.024	0.01	−0.01 – 0.03	0.213
density.cds	0.03	−0.06 – 0.11	0.555	0.04	−0.03 – 0.11	0.28
prior.density.cds	−0.07	−0.16 – 0.02	0.123	−0.1	−0.17 − −0.03	0.005
Random Effects						
$\sigma 2$	0.08			0.08		
$\tau 00$	0.01 name			0.01 name		
ICC	0.1			0.15		
N	147 name			109 name		
Observations	841			541		
Marg.R^2/Cond.R^2	0.502/0.551			0.396/0.486		

TABLE 7 | Linear mixed model: CDS node degree.

Predictors	degree.cds: child 1		*p*	degree.cds: child 2		*p*
	Estimates	CI		Estimates	CI	
(Intercept)	0.59	0.27 – 0.91	<0.001	0.34	0.02 – 0.65	0.038
age	−0.01	−0.01 −−0.00	0.001	0.01	0.00 – 0.01	0.026
degree.cs	0.32	0.18 – 0.46	<0.001	0.32	0.15 – 0.50	<0.001
prior.degree.cs	−0.13	−0.29 – 0.02	0.091	−0.04	−0.21 – 0.14	0.682
prior.degree.cds	0.14	0.07 – 0.20	<0.001	0.2	0.14 – 0.27	<0.001
centrality.cs	0	−0.04 – 0.04	0.979	−0.01	−0.04 – 0.01	0.304
prior.centrality.cs	0.01	−0.03 – 0.05	0.546	−0.02	−0.04 – 0.00	0.095
centrality.cds	1.84	1.63 – 2.06	<0.001	1.7	1.47 – 1.93	<0.001
prior.centrality.cds	0.19	−0.04 – 0.42	0.102	0.18	0.02 – 0.33	0.024
density.cs	−0.07	−0.13 – 0.00	0.051	−0.01	−0.04 – 0.03	0.713
prior.density.cs	0.14	0.04 – 0.24	0.005	0.09	0.06 – 0.12	<0.001
density.cds	−0.57	−0.75 −−0.39	<0.001	−0.58	−0.71 −−0.45	<0.001
prior.density.cds	−0.36	−0.55 −−0.17	<0.001	−0.3	−0.44 −−0.16	<0.001
Random Effects						
$\sigma 2$	0.35			0.33		
$\tau 00$	0.11 name			0.08 name		
ICC	0.24			0.2		
N	147 name			109 name		
Observations	841			541		
Marg.R^2/Cond.R^2	0.711/0.781			0.707/0.766		

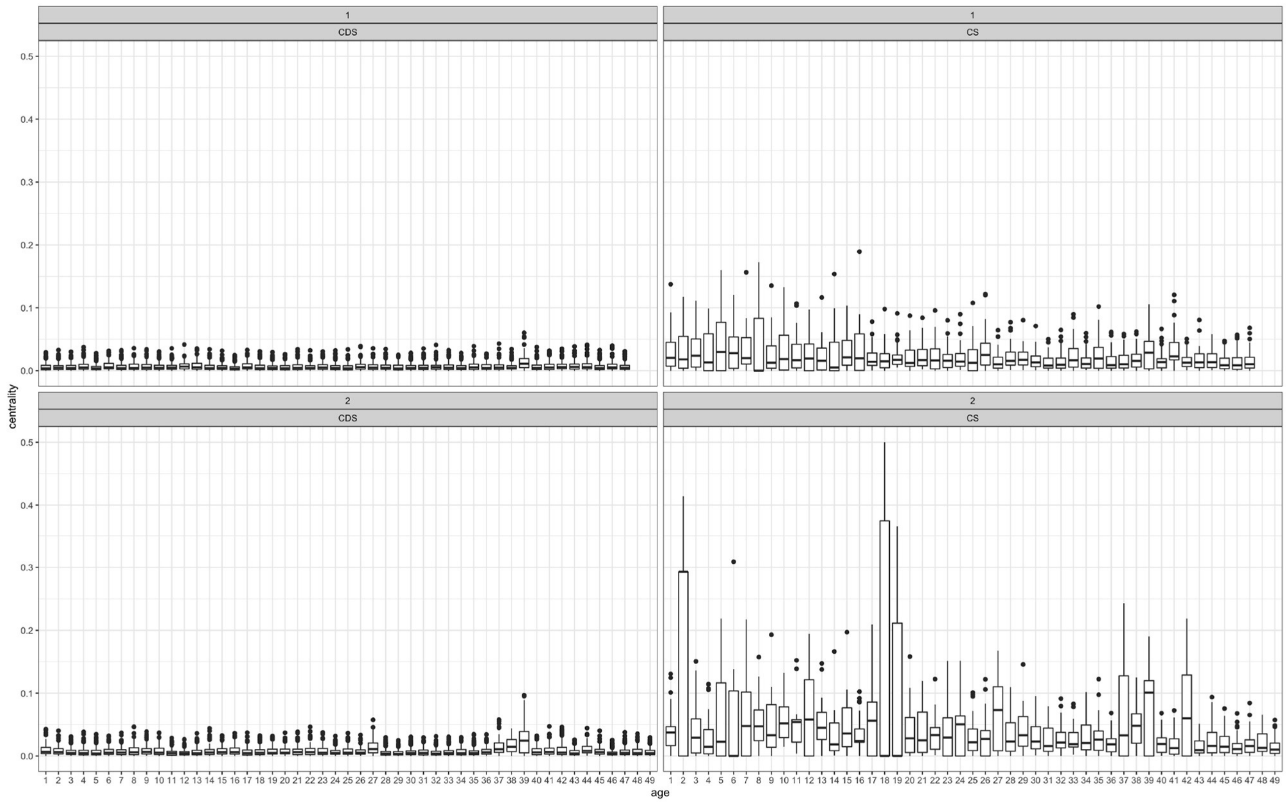

FIGURE 5 | Node centrality distribution by session: CS (right panels) and CDS (left panels), child 1 (top panels) and child 2 (bottom panels).

CS Centrality: Child 2

The model's total explanatory power is substantial (conditional $R^2 = 0.43$) and the part related to the fixed effects alone (marginal R^2) is of 0.43. Within this model, the following variables have a significant effect on CS1 centrality level: age (negative effect), CDS centrality (positive effect), CS degree (positive effect), CS density (positive effect), and CDS density (negative effect).

CDS Centrality

Table 9 shows the results for the linear mixed models for both parents, predicting CDS eigenvector centrality. Each model is detailed below.

CDS Centrality: Child 1

The model's total explanatory power is substantial (conditional $R^2 = 0.81$) and the part related to the fixed effects alone (marginal R^2) is of 0.67. Within this model, the following variables have a significant effects on CDS1 centrality: CS centrality (positive effect), CDS centrality in recording N-1 (positive effect), CS degree in recording N-1 (positive effect), CDS degree (positive effect), CS density (positive effect), CS density in recording N-1 (negative effect), and CDS density (positive effect).

CDS Centrality: Child 2

The model's total explanatory power is substantial (conditional $R^2 = 0.77$) and the part related to the fixed effects alone (marginal R^2) is of 0.60. Within this model, the following variables have a significant effects on CDS2 centrality: age (negative effect), CS centrality (positive effect), CS centrality in recording N-1 (positive effect), CDS centrality in recording N-1 (positive effect), CDS degree (positive effect), CDS degree in recording N-1 (positive effect), CS density in recording N-1 (negative effect), CDS density (positive effect), and CDS density in recording N-1 (positive effect).

Network Density

Recall that network density measures the proportion of active links relative to the maximum number of possible links in the current network, given the active nodes. **Figure 6** depicts the changing densities of networks with age for each of the participants.

Network density seems to decrease with age for both children (although at different rates), while it seems to remain constant for the parents through development. That is, as children grow, they exhaust fewer of their possible links, leaving more room for their network of roots and inflected patterns to grow: a non-exhausted network (i.e., a low density network) is a network with a high growth potential, since new links can be created between existing constructs that have not been linked before.

For each network we fitted a linear model to predict network density with the following variables: age, network density of the other party in the dyad, and network density at the preceding

TABLE 8 | Linear mixed model: CS node centrality.

Predictors	centrality.cs: child 1		p	centrality.cs: child 2		p
	Estimates	CI		Estimates	CI	
(Intercept)	1.58	1.08 – 2.09	<0.001	2.71	1.83 – 3.60	<0.001
age	−0.02	−0.03 – −0.01	<0.001	−0.03	−0.05 – −0.01	0.001
centrality.cds	0.56	0.16 – 0.95	0.006	1.29	0.45 – 2.12	0.003
prior.centrality.cs	0.16	0.09 – 0.22	<0.001	−0.01	−0.08 – 0.07	0.842
prior.centrality.cds	0.39	0.02 – 0.77	0.04	−0.13	−0.66 – 0.39	0.612
degree.cs	1.48	1.26 – 1.70	<0.001	2.68	2.22 – 3.15	<0.001
prior.degree.cs	−0.23	−0.47 – 0.02	0.074			
degree.cds	0	−0.11 – 0.11	0.967	−0.2	−0.47 – 0.08	0.156
prior.degree.cds	−0.08	−0.18 – 0.03	0.146	−0.13	−0.36 – 0.10	0.262
density.cs	0.26	0.15 – 0.37	<0.001	0.6	0.50 – 0.71	<0.001
prior.density.cs	−0.16	−0.33 – 0.01	0.058	−0.06	−0.16 – 0.05	0.304
density.cds	−0.18	−0.48 – 0.13	0.26	−0.58	−1.05 – −0.12	0.014
prior.density.cds	0	−0.32 – 0.32	0.995	0.34	−0.14 – 0.82	0.162
Random Effects						
$\sigma 2$	1.05			4.14		
$\tau 00$	0.04 name			0.0 name		
ICC	0.04			0.0		
N	147 name			109 name		
Observations	841			541		
Marg.R^2/Cond.R^2	0.442/0.462			0.435/0.435		

TABLE 9 | Linear mixed model: CDS node centrality.

Predictors	centrality.cds: child 1		p	centrality.cds: child 2		p
	Estimates	CI		Estimates	CI	
(Intercept)	−0.09	−0.17 – 0.00	0.058	−0.01	−0.11 – 0.10	0.909
age	0	−0.00 – 0.00	0.093	0	−0.00 – −0.00	0.002
centrality.cs	0.01	0.00 – 0.02	0.01	0.01	0.00 – 0.02	0.031
prior.centrality.cs	0	−0.01 – 0.01	0.45	0.01	0.00 – 0.02	0.014
prior.centrality.cds	0.21	0.15 – 0.27	<0.001	0.1	0.05 – 0.14	<0.001
degree.cs	−0.01	−0.05 – 0.03	0.666	−0.02	−0.08 – 0.03	0.374
prior.degree.cs	0.07	0.03 – 0.11	0.001	0.03	−0.03 – 0.08	0.344
degree.cds	0.13	0.11 – 0.14	<0.001	0.15	0.13 – 0.17	<0.001
prior.degree.cds	0	−0.01 – 0.02	0.601	0.03	0.01 – 0.05	0.006
density.cs	0.02	0.00 – 0.04	0.02	0	−0.01 – 0.01	0.709
prior.density.cs	−0.04	−0.07 – −0.02	0.001	−0.02	−0.03 – −0.01	<0.001
density.cds	0.17	0.13 – 0.22	<0.001	0.19	0.15 – 0.22	<0.001
prior.density.cds	0	−0.05 – 0.05	0.888	0.05	0.01 – 0.10	0.011
Random Effects						
$\sigma 2$	0.02			0.03		
$\tau 00$	0.02 name			0.02 name		
ICC	0.44			0.42		
N	147 name			109 name		
Observations	841			541		
Marg.R^2/Cond.R^2	0.668/0.813			0.595/0.767		

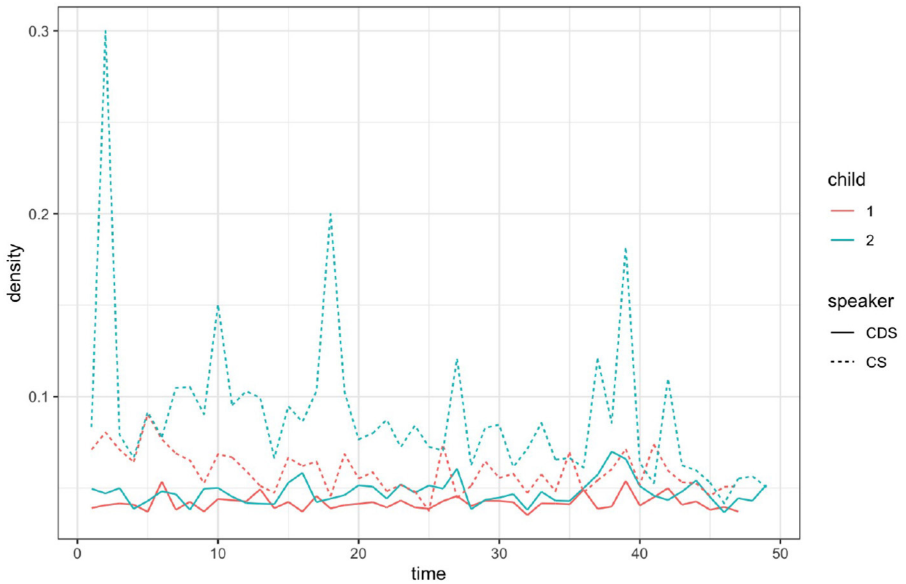

FIGURE 6 | Network density by session, Child 1 and 2, CS and CDS. CS of Child 1 is represented by a dashed red line; CS of Child 2 is represented by a dashed green line; CDS of Child 1 is represented by a solid red line; CDS of Child 2 is represented by a solid green line.

recording for both parties. As network density is a global measure relevant for the entire network, we did not include measures of individual nodes in these models (unlike the models presented above for degree and centrality).

CS Density

Table 10 shows the results for the linear models for both children, predicting CS network density. Each model is detailed below.

CS Density: Child 1

The model's explanatory power is substantial ($R^2 = 0.28$, adj. $R^2 = 0.27$). Within this model, the following variables have a significant effect on CS1 network density: age (negative effect), CDS density (positive effect), and CDS density in recording N-1 (negative effect).

CS Density: Child 2

The model's explanatory power is substantial ($R^2 = 0.40$, adj. $R^2 = 0.39$). Within this model, the following variables have a significant effect on CS2 network density: age (negative effect), CDS density (positive effect), CS density in recording N-1 (negative effect), and CDS density in recording N-1 (positive effect).

CDS Density

Table 11 shows the results for the linear models for both parents, predicting CDS network density. Each model is detailed below.

CDS Density: Child 1

The model's explanatory power is substantial ($R^2 = 0.28$, adj. $R^2 = 0.28$). Within this model, the following variables have a significant effect on CDS1 network density: age (positive effect), CS density (positive effect), CDS density in recording N-1 (negative effect), and CS density in recording N-1 (positive effect).

CDS Density: Child 2

The model's explanatory power is weak ($R^2 = 0.08$, adj. $R^2 = 0.07$). Within this model, the following variables have a significant effect on CDS2 network density: age (positive effect), CS density (positive effect), and CS density in recording N-1 (positive effect).

Interim Summary

The results reported above show that Hebrew verb morphology can be conceptualized as a network linking roots and patterns. This construal sheds new light on the development of this system with respect to patterns of adaptation within and between CS and CDS, as well as tracking small, but meaningful, changes within the system's structure. Looking at node activation in the network, we show that development is punctuated in terms of verb usage. The node degree measure reveals that the CS linkage level (of each root to number of patterns and each pattern to number of roots) is affected by the following factors: age; the CDS linkage level of the same root or pattern; the linkage level the same root or pattern had in the previous recording in CS; the root's or pattern's centrality within the network; and the network's density. The CDS degree (linkage level) is not shown to be affected by age, remaining steady throughout the time range of our data. It is, however, affected by the following: the CS linkage level and by the linkage level of the same pattern/root in previous recordings; the centrality of the pattern/root within the system; the density of the current network; and the density of previous networks of both the child and the parent.

TABLE 10 | Linear model: CS Network density.

Predictors	density.cs: child 1		p	density.cs: child 2		p
	Estimates	CI		Estimates	CI	
(Intercept)	1.31	1.07 – 1.55	<0.001	5.36	4.88 – 5.84	<0.001
age	−0.02	−0.03 – −0.02	<0.001	−0.1	−0.11 – −0.09	<0.001
density.cds	0.51	0.33 – 0.69	<0.001	0.92	0.57 – 1.27	<0.001
prior.density.cs	0.03	−0.07 – 0.14	0.544	−0.14	−0.23 – −0.06	0.001
prior.density.cds	−0.45	−0.64 – −0.25	<0.001	0.76	0.38 – 1.15	<0.001
Observations	841			541		
R^2/R^2 adjusted	0.276/0.273			0.397/0.392		

TABLE 11 | Linear model: CDS Network density.

Predictors	density.cds: child 1		p	density.cds: child 2		p
	Estimates	CI		Estimates	CI	
(Intercept)	−0.95	−1.01 – −0.88	<0.001	−0.49	−0.64 – −0.34	<0.001
age	0.01	0.00 – 0.01	<0.001	0.01	0.00 – 0.01	0.001
density.cs	0.07	0.04 – 0.09	<0.001	0.05	0.03 – 0.07	<0.001
prior.density.cds	−0.36	−0.43 – −0.29	<0.001	−0.07	−0.16 – 0.02	0.141
prior.density.cs	0.27	0.24 – 0.31	<0.001	0.04	0.02 – 0.06	<0.001
Observations	841			541		
R^2 / R^2 adjusted	0.279/0.275			0.078/0.071		

The node centrality measure reveals that within the CS networks there are more central roots/patterns compared with the parents' networks. The centrality of a root/pattern within the CS morphological system (i.e., its centrality within the network) is affected by the following factors: age; the centrality of the same root/pattern in the CDS network; its degree in the CS network; and the CS network's density. The node centrality in the CDS network is not affected by age, but rather by the following: the previous centrality within the CDS network; the degree of the same root/pattern in the CDS network; the density of the CDS network; and the density of the previous CS network.

Finally, the network density measure reveals that CS network density is affected by age as well as by the density of the CDS—of both the previous and the current network. The CDS network density is affected by age as well (contra to the degree and centrality measures) and by the density of the CS network—both the previous and the current. In the following section we discuss each of these results and its implications in detail.

DISCUSSION

Recent studies on Hebrew verb acquisition (Ashkenazi et al., 2016, 2020) have shown that toddlers rely on stable, frequently occurring inflectional verb affixes in maternal input to gain salient information on the opaque, irregular verbs they frequently encounter. Furthermore, children's output greatly resembles and correlates with parental input in terms of structural, semantic and pragmatic features of the verbs used, highlighting the role of CDS in shaping CS verb structure. Previous studies indicated the possible contribution of CDS to CS verb content in the form of parental corrections, reformulations and expansions, children's uptake and imitations in parent-child conversations characterized by mutual attention and responsiveness (Clark and de Marneffe, 2012).

The present study adds to this line of research by accounting for the development of the morphology of Hebrew verbs as a dynamic network of roots and inflected patterns within the interactive domain of Child Speech and Child Directed Speech. The results presented in **Figures 2**–**6** and the models in **Tables 6**–**11** show that the development of the system of roots and inflected patterns between 1;8 and 2;2 is not strictly linear: it is not only a matter of mere frequency of use, but rather that across development, the child links more roots to every available pattern and inflection, and more patterns and inflections to every available root. Development is shown here to be dynamic, complex, and adaptive in several ways.

First, we saw that morphological development in children can be characterized as punctuated rather than continuous. Except for a few frequent nodes, most morphological constructs (i.e., roots and patterns) become active for a specific period of time, and then stop being used, sometimes re-appearing again in later periods. Conversely, as shown in **Figure 3**, the parent's use of morphological constructs is more coherent or continuous, such that each construct is used for a longer period of time, and breaks between uses are shorter. That is, it seems like the child is busy mastering each construct for a certain period, only to move on to another construct. Consider for example the CS1 data (CS of Child 1; top left panel in **Figure 3**): Node number 74 in the CS1 data is the pattern *Pi'el*.Inf; it is active at the first recordings, and then goes inactive and active again throughout the data. Conversely, node number 121 in the same data is the pattern *Qal*.Fut.3.Sg.Ms, and it is active in each and every recording.

Second, the models presented above reveal the dynamic nature of development, underscoring the role of adaptation—between the child and the parent, and relative to past experience. This is explained in the following sections, where we discuss (1) linkage (i.e., node degree) between roots and inflected patterns; (2) importance (eigenvector centrality) of roots and inflected patterns; and (3) systematic growth potential in the network (network density). Since we are interested in the dynamic nature of development through time, we included the level of the various measures at the prior recording as a possible explanatory variable for the measures of a current recording for both CS and CDS.

Root and Inflected Pattern Degree Across Development

The degree of a node in a network measures the number of links it has with other nodes. In our case, links between nodes are verb wordforms created by the affixation of roots to inflected patterns. For example, the inflected pattern node *Qal*.Present.Fm.Sg, has a degree level of 4 in the CS network of Child 2, recording number 3, linked to the roots *r-ʔ-y*, *k-ʔ-b*, *n-w-ħ*, and *b-w-ʔ*, manifested as the verbforms *ro'a* "sees/ing," *ko'évet* "hurts/ing," *náxa* "rests/ing," and *bá'a* "comes/ing," respectively. The same inflected pattern is expanded through development, having a degree level of 8 in recording number 32. This means that this inflected pattern is used in a larger grammatical network, linking other roots to the morphological family of the pattern. That is, it highlights another context in which this pattern can be used in terms of referents and event attributes.

The results of our models show that changes in the degree of roots and inflected patterns are not only a factor of age, but rather that the degree level of a morphological construct is systematically related to other factors within the network and within the dyad. These results are different for the CS and the CDS, and the two models are presented separately. We discuss only those results that were significant for both children, leaving individual differences to future research.

Results for the CS degree models (**Table 6**) show that the number of roots predicted to be linked to a single inflected pattern and the number of inflected patterns a single root is linked to are affected by the age of the child, such that with age, each construct is predicted to be linked to more constructs. The linkage level is also affected by the number of links these constructs have in the previous network of the child speech, and by the number of links they have in the parent's current network. The importance of these constructs in the child's speech also affects their linkage, such that more central nodes are predicted to have more links. Note that the centrality of the nodes in the parent's network does not have a significant effect on the child's degree. That is, for a construct to have more links in the CS, it is crucial that it has more links in the CDS, but not that it has a prominent position in the CDS network. Finally, another factor that is common to both children is the effect of the CS network density on degree level, such that degree levels decrease with higher network density. This is a manifestation of the growth potential interpretation of the density measure, proposed by Levie et al. (2019), since lower density levels indicate a higher potential for the network to grow, realized here as higher degree levels.

Results for the CDS degree model (**Table 7**) show that age has differential effects in the two children participating in the current study: while the degree levels in parental speech rise in the CDS of Child 1, they fall in the CDS of Child 2. However, the degree level in the current CS network and degree level in the previous CDS network have the same effect in both parents: increasing degree levels in the current network of the CS and in the previous CDS network predict increase in the degree levels in the CDS. That is, linkage between roots and inflected patterns within the parent's speech is affected by the current speech of the child, as well as the previous speech of the parents themselves. We may conclude that parents adapt to their children's speech in two ways: first, by relating to their children's current usage. Second, by expanding on previous experience, counting on the usage their children have already been exposed to and building upon it. Interestingly, degree levels in the previous recording in the CS do not have an effect on current degree in the CDS. That is, parents do not build on their children's previous usage, but on their own. This effect, too, should be further investigated in future research.

The importance of nodes in the current network (as realized by eigenvector centrality) affects CDS linkage as well, such that more important nodes are predicted to have more links. Note that this applies only to the importance of the nodes in the current network of the parents, but not to the current network of the child nor to previous networks of the child or the parent. That is, in contrast to previous linkage, which seems to affect current linkage in both children and parents, the importance of morphological constructs has an effect only on the current network. This may be explained by the fact that centrality (i.e., importance) is a more context specific measure than degree (i.e., linkage), and context is changing from one recording session to another. Linkage, however, is a matter of morphological productivity, linking a single root to relatively many inflected patterns, and vice versa. Therefore, it is a systematic measure that grows incrementally.

Root and Pattern Centrality Across Development

The eigenvector centrality of a node is a measure of its importance within the network: It assigns a value to a node based on the number of links it has with other nodes that have many links themselves. For example, in our case, a *binyan*-temporal pattern that is linked to many roots that are linked to other *binyan*-temporal patterns has high centrality. While the measure of degree discussed above underscored the importance of linkage in expanding the network, eigenvector centrality is a relative measure, highlighting network variability. A network with few highly centralized nodes has few hubs through which information in the network can flow. In morphological terms we can think of it as a network with a small number of inflected patterns that are linked to many roots, each of which is linked to other patterns as well. This type of network limits the possibility to link a root to an inflected pattern that is not central in the network: in a given conversation, it is more probable for a new link to be made between a root and a central inflected pattern than with a less central one (resembling diversity situations with high entropy). However, if the network has many low centralized nodes, with no single node (or few nodes) standing out from the crowd, the probability of making new links is higher. **Figure 5** shows that this is indeed the main difference in centrality distribution between the CDS and CS of both dyads in our data.

Results for the CS centrality model (**Table 8**) show that the eigenvector centrality of a root or an inflected pattern is predicted to fall with age. We interpret this result as indicating that networks become more evenly distributed in terms of centrality with age, thus enabling morphological productivity in language use: Since there are no few nodes that stand out from the crowd, the probability to use each node in the network (rather than just a few) is increasing (resembling a low entropy system). Thus, the structure of the system not only reflects productivity, but rather enables it, as more nodes gain usage probability. The centrality of a node in the children's networks is also affected by the centrality of this node in the parent's network, such that rising CDS centrality predicts rising CS centrality. That is, if on a given day a root is used with many patterns that are used with many roots in the parents' speech, than this root is predicted to be used with many patterns that are used with many roots in the children's speech as well. The degree of a node also affects the centrality of the node, but only relative to its linkage in the children's speech. The degree of a node in the parents' speech does not predict its centrality level in CS. Finally, the density of the network has an effect as well, such that rise in density levels (i.e., a less varied network) predicts high centrality. That is, density, as a growth potential measure, affects not only local morphological productivity (as seen above in the discussion of the degree as morphological linkage), but also systematic morphological productivity, such that a network with more growth potential is a less centralized system.

Results for the CDS degree model (**Table 9**) show that eigenvector centrality of roots and inflected patterns in the parents' speech to children is not affected by the age of the child. That is, the system is stable over time in terms of construct importance. However, while it is stable as a system, the network is still adaptive at a more local level: the centrality of roots and inflected patterns in the CDS changes as a function of their centrality in the CS. That is, parents adapt the particular morphological construct they use to the usage of their children, putting the burden of morphological productivity on the same constructs their children already use. The model also shows an effect for the centrality in the previous CDS network on current centrality. We interpret this result as indicating continuation and coherence: If a morphological construct is important in the network, a parent will keep using it in an important manner. This may facilitate learning, as it provides the child with more learning opportunities with the same distributional characteristics. The effect of CDS density on CDS centrality is similar to the effect of CS density on CS centrality discussed above: High CDS density predicts a more centralized network, interpreted as a less variable and less productive one.

Finally, adaptation can also be seen in the effect of CS density in the prior network on CDS centrality, an effect that was absent in the CS model for centrality. The CDS centrality model shows that a rise in density levels in the CS previous network predicts a fall in centrality levels in the CDS current network. High density and low centrality can be thought of as two sides of the same coin, denoting systematic productivity. That is, a network with high density has low growth potential since most of its possible links have already been made. Conversely, a system with many low central nodes is less limited in its potential to form new links, as probability is more evenly distributed. The current model shows that if the previous network of the children had low growth potential, the current network of the parents is systematically more productive. This can be seen as fine-tuned tweaking of the system toward productivity by the parent: when parents experience a limited network in the speech of their children, they will provide them with more opportunities to expand their system in future interactions.

Morphological Network Density Across Development

The density of the network is a measure of fulfilled links among nodes, relative to all possible links. As such, this measure quantifies the level of network exhaustion in terms of how much of the potential of the current network has already actually been fulfilled by the speaker. A speaker that has already fulfilled most of her network's entire potential has nowhere to grow, in the sense that the probability of re-using existing verb wordforms (i.e., links within the networks) is high. In such a network, the main road to expansion is by adding new nodes, and not by creating new links. Thus, a child with a high density network needs to add more roots and/or inflected patterns to her network in order to expand her morphological verb lexicon. On the other hand, a child with a low density network can expand her network also by linking constructs that were not linked before, creating variations on verb wordforms.

Figure 6 and the models summarized in **Tables 10**, **11** show that for both children, CS network density is affected by the child's age, such that networks become less dense with age. CDS network density is affected by age in the opposite direction, with density rising with age. Patterns of adaptation are manifested in the relations revealed here between the density of the CS networks and that of the CDS networks: CDS network density affects CS network density for both children, such that children's network density is adaptive to that of their parents: the higher the density of the CDS network, the higher the density of the CS network. Parents are also adaptive to their children's network density: CDS network density is affected by CS network density, such that higher CS density levels predict higher CDS density levels. The model also shows that CDS network density is affected by the density of the CS in the previous recording.

These results demonstrate the manifestation of adaptation: both parties in the dyad adapt their network structure to that of their interlocutor, taking into account their current and the

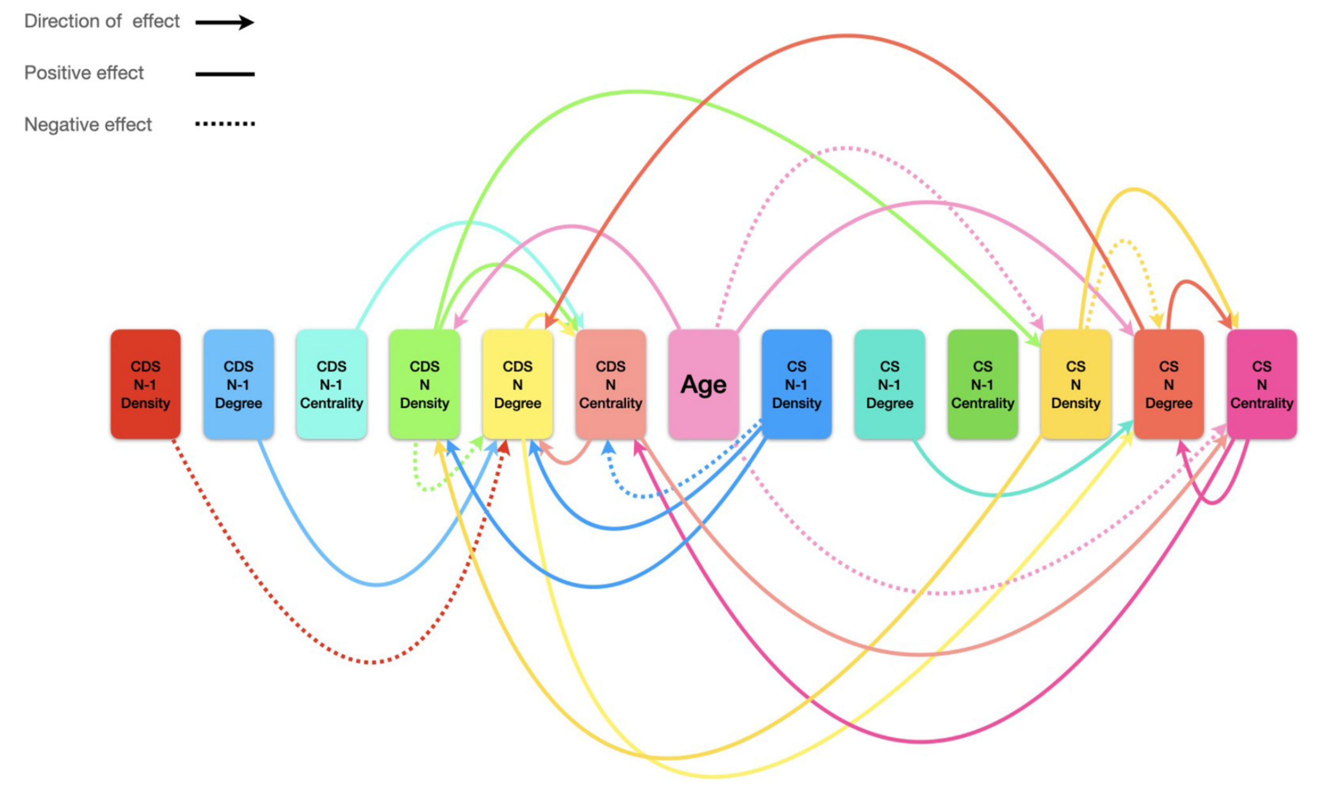

FIGURE 7 | A developmental model of a morphological verb lexicon network: The image shows a dynamic system of adaptation involving the child, the parent, and time. Arrows between affecting and affected network measures are colored according to the affecting variable; solid lines indicate positive effects, and dashed lines indicate negative effects.

previous states (with some individual differences). Going from the statistics to morphology, we interpret density as growth potential (Levie et al., 2019). Thus, we can conclude that the potential to expand the network, forming new links between existing roots and inflected patterns, is a function of more than just the age of the child, with the current and previous morphological network structures of the other party in the developmental tango also having an effect.

An illuminating step in the development of a morphological system is morphological over-generalizations. However, in the present study we do not report on such morphological errors, that are the result of linking an existing root to an existing pattern within the network, creating a link that is not observed in the adult language. The fact that we did not find such errors might seem surprising, but we believe it is due to the nature of our particular data. Specifically, we claim that the age range of our data (1;8–2;2) might be too early for morphological over-generalizations of this type in Hebrew. We might find errors in this age range in terms of word order or agreement marking, but derivational root-*binyan* errors are more typical of children aged 4 years and above, indicating the consolidation of verb morphology (Levie et al., 2020). We assume that recordings of older child-adult dyads may reveal more over-generalizations, since the morphological system would gain more network growth potential, as shown above, until reaching a point of equilibrium in terms of network density, balancing between creativity and conventionality (Tomasello, 2000).

Converging Models

Figure 7 is a visual summary of all significant effects found in the models discussed above, suggesting a complex unified model for patterns of adaptation in Hebrew verb morphology development between the ages of 1;8–2;2 as a network.

Figure 7 shows that many of the variables both affect and are affected by other variables, within the speaker (child or parent), and between speakers. This indicates dynamic relations within the changing system of the morphological verb lexicon involving adaptation within the speaker, between the speakers, and relative to past experience.

CONCLUSION

Language learning is dynamic (van Geert, 1991, 2010), changing as a function of age, individual differences, and input language (CDS), among other factors. Within this growing/emerging system, CDS and CS affect each other in different ways. The current paper shows these relations, modeling the development of Hebrew verb morphology between the ages of 1;8–2;2.

We used network analysis to account for complex relations between morphological constructs, so as to assess changes in network structure over time. Growth in frequency during development is unquestionably apparent. With age, children are shown to produce more roots, more inflected patterns, and more wordforms. Using network analysis we also showed that children's networks are growing with age as well, in terms of node degree and centrality representing linkage level and construct importance respectively, and in terms of the network density as representing network growth potential. However, this method allowed us to go beyond growth to the crucial role of variation, showing that both take part in development. A major finding reported above is that development is not linear, and that children go through periods of punctuated development: We saw that children's use of morphological constructs is not coherent, in the sense that it is not the case that once a child is using a construct she will continue to use it in a productive manner. Rather, the children were shown to use individual constructs for short periods of time. This finding is highlighted by contrasting it with the parents' patterns of usage, which is much more coherent or continuous. This leads us to the conclusion that children between the ages of 1;8–2;2 do not use their entire range of possible constructs based on a cumulative lexicon. Rather, children are attuned to their immediate experience: If a morphological construct was highly linked or very important in the immediate past, or if it is important within the network of their caregiver, it is more likely to be used again by the child. This expresses the variation of the developing dynamic system. The productivity of morphological constructs is varied, and depends on many different (and related) factors.

Another facet of dynamic development revealed by the network analysis concerns the adaptation of the parents to their children's systems. We show that parents adapt to their children's speech patterns in three ways: first, by relating to their children's current usage. Second, by expanding on previous experience, counting on the usage their children have already been exposed to, and building upon it. And third, we show that when parents experience a limited network in the speech of their children, they will provide them with more opportunities to expand their system in future interactions.

A dynamic system goes through changes that are a function of its current state (van Geert, 2010). The analysis suggested in the present paper shows that this is an apt description of the development of Hebrew verb morphology, and that the framework of dynamic network analysis thus provides insights into the complex issue of language development. Given that, we would like to point out two directions for future research. First, the present paper modeled each variable in a separate manner, while the converged model presented in **Figure 7** suggests that an improved model might arise by analyzing the entire array of variables simultaneously so as to expose interactions. Second, the discussion of the results in the present paper focused on those cases where no individual differences were found. As shown in the summaries of the models, there are individual differences between the two children and the two parents in the current database. This suggests that analyzing such data by taking not only the item itself as a random variable (as done in the present paper), but also the speaker, and expanding our sample, might reveal even more interesting relations within the dynamic system.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because it's identifiable data. Requests to access the datasets should be directed to Orit Ashkenazi, orit.ashkenazi@gmail.com.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ED wrote most of the paper, conceptualized the research questions and the method of analysis, conducted the analyses, and provided most of the insights. RL conceptualized and carried out the framework for verb analysis and provided analyses of the data and insights. DR wrote the morphological parts of the paper, participated in conceptualizing the interpretation of the results, and edited the paper. OA collected and analyzed the database, wrote the input-output section, and provided insights. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by ISF grant no. 219/17 to DR.

REFERENCES

Acsu-Koc, A. (1998). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: evidence from Turkish. *First Lang.* 18, 255–280. doi: 10.1177/014272379801805402

Albert, A., MacWhinney, B., Nir, B., and Wintner, S. (2013). The Hebrew CHILDES corpus: transcription and morphological analysis. *Lang. Resour. Eval.* 47, 973–1005. doi: 10.1007/s10579-012-9214-z

Ambridge, B. (2020). Against stored abstractions: a radical exemplar model of language acquisition. *First Lang.* 40, 509–559. doi: 10.1177/0142723719869731

Arunachalam, S., Escovar, E., Hansen, M. A., and Waxman, S. R. (2011). "Verb learning from syntax alone at 21 months," in *Proceedings of the 35th Annual Boston University Conference on Language Development*, eds K. M. Danis and H. Sung (Somerville, MA: Cascadilla Press), 21–24.

Ashkenazi, O. (2015). *Input-Output Relations in the Early Acquisition of Hebrew Verbs*. (Ph.D. thesis), Tel Aviv University, Tel Aviv, Israel.

Ashkenazi, O., Gillis, S., and Ravid, D. (2020). Input–output relations in Hebrew verb acquisition at the morpho-lexical interface. *J. Child Lang.* 47, 509–532. doi: 10.1017/S0305000919000540

Ashkenazi, O., Ravid, D., and Gillis, S. (2016). Breaking into the Hebrew verb system: a learning problem. *First Lang.* 36, 505–524. doi: 10.1177/0142723716648865

Barabasi, A. L. (2009). Scale-free networks: a decade and beyond. *Science* 325, 412–413. doi: 10.1126/science.1173299

Beckage, N., Smith, L., and Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE* 6:e19348. doi: 10.1371/journal.pone.0019348

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59, 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Behrens, H. (2006). The input output relationship in first language acquisition. *Lang. Cogn. Proces.* 21, 2–24. doi: 10.1080/01690960400001721

Ben-Zvi, G., and Levie, R. (2016). "Development of Hebrew derivational morphology from preschool to adolescence," in *Acquisition and Development of Hebrew: From Infancy to Adolescence*, ed R. A. Berman (Amsterdam; Philadelphia, PA: John Benjamins Publishing Company), 135–173. doi: 10.1075/tilar.19.05ben

Berman, R. A. (1982). Verb-pattern alternation: the interface of morphology, syntax, and semantics in Hebrew child language. *J. Child Lang.* 9, 169–191. doi: 10.1017/S030500090000369X

Berman, R. A. (1985). *Acquisition of Hebrew*. Hillsdale, NJ: Lawrence Erlbaum.

Berman, R. A. (1987). Productivity in the lexicon: new-word formation in Modern Hebrew. *Folia Linguist.* 21, 425–461.

Berman, R. A. (1993a). "Developmental perspectives in transitivity: a confluence of cues," in *Other Children, Other Languages: Issues in the Theory of Acquisition*, ed Y. Levy (Hillsdale, NJ: Lawrence Erlbaum), 189–241.

Berman, R. A. (1993b). Marking of verb transitivity by Hebrew-speaking children. *J. Child Lang.* 20, 641–669. doi: 10.1017/S0305000900008527

Berman, R. A. (2000). "Children's innovative verbs vs. nouns: structured elicitations and spontaneous coinages," in *Methods for Studying Language Production*, eds L. Menn and N. B. Ratner (Mahwah, NJ: Lawrence Erlbaum), 69–93.

Berman, R. A. (2012). "Revisiting roots in Hebrew: a multi-faceted view," in *Studies on Modern Hebrew and Jewish languages in Honor of Ora (Rodriguez) Schwarzwald*, eds M. Muchnik and T. Sadan (Jerusalem: Carmel Press), 132–158.

Berman, R. A., and Lustigman, L. (2012). "HARSP: a developmental language profile for Hebrew," in *Assessing Grammar: The Languages of LARSP*, eds M. Ball, D. Crystal, and P. Fletcher (Bristol: Multilingual Matters), 43–76. doi: 10.21832/9781847696397-006

Bolozky, S. (1999). *Measuring Productivity in Word Formation: The Case of Israeli Hebrew* (Leiden: Brill). doi: 10.1163/9789004348431

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Soc. Netw.* 29, 555–564. doi: 10.1016/j.socnet.2007.04.002

Bybee, J., and McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *Linguist. Rev.* 22, 381–410. doi: 10.1515/tlir.2005.22.2-4.381

Cameron-Faulkner, T. (2012). A functional account of verb use in the early stages of English multiword development. *J. Child Lang.* 39, 885–897. doi: 10.1017/S0305000911000328

Chen, H., Chen, X., and Liu, H. (2018). How does language change as a lexical network? an investigation based on written Chinese word co-occurrence networks. *PLoS ONE* 13:e0192545. doi: 10.1371/journal.pone.0192545

Clark, E. V., and de Marneffe, M. C. (2012). Constructing verb paradigms in French: adult construals and emerging grammatical constructs. *Morphology* 22, 89–120. doi: 10.1007/s11525-011-9193-6

Den Hartigh, R. J. R., Van Dijk, M. W. G., Steenbeek, H. W., and Van Geert, P. L. C. (2016). A dynamic network model to explain the development of excellent human performance. *Front. Psychol.* 7:532. doi: 10.3389/fpsyg.2016.00532

Deutsch, A., and Kuperman, V. (2019). Formal and semantic effects of morphological families on word recognition in Hebrew. *Lang. Cogn. Neurosci.* 34, 87–100. doi: 10.1080/23273798.2018.1513541

Gillis, S., and Ravid, D. (2006). Typological effects on spelling development: a crosslinguistic study of Hebrew and Dutch. *J. Child Lang.* 33:621. doi: 10.1017/S0305000906007434

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Hoff-Ginsberg, E. (1985). Some contributions of mothers' speech to their children's syntactic growth. *J. Child Lang.* 12, 367–385. doi: 10.1017/S0305000900006486

Kapatsinski, V. (2018). *Changing Minds Changing Tools: From Learning Theory to Language Acquisition to Language Change*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/11400.001.0001

Kastner, I. (2016). *Form and Meaning in the Hebrew Verb*. (Ph.D. thesis), New York University, New York, United States.

Kastner, I. (2019). Templatic morphology as an emergent property. *Nat. Lang. Linguist. Theor.* 37, 571–619. doi: 10.1007/s11049-018-9419-y

Ko, E. S. (2012). Nonlinear development of speaking rate in child directed speech. *Lingua* 122, 841–857. doi: 10.1016/j.lingua.2012.02.005

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. New York, NY: Springer.

Laks, L. (2013). Why and how do Hebrew verbs change their form? a morpho-thematic account. *Morphology* 23, 351–383. doi: 10.1007/S11525-013-9227-3

Levie, R., Ashkenazi, O., Stanzas, S. E., Zwilling, R. C., Raz, E., Hershkovitz, L., et al. (2020). The route to the derivational verb family in Hebrew: a psycholinguistic study of acquisition and development. *Morphology* 30, 1–60. doi: 10.1007/s11525-020-09348-4

Levie, R., Ben-Zvi, G., and Ravid, D. (2017). Morpho-lexical development in language impaired and typically developing Hebrew-speaking children from two SES backgrounds. *Read. Writ.* 30, 1035–1064. doi: 10.1007/s11145-016-9711-3

Levie, R., Dattner, E., Zwilling, R., Rosenstein, H., Eitan, S., and Ravid, D. (2019). Complexity and density of Hebrew verbs in preschool peer talk: the effect of SES background. *Mental Lexicon* 14, 235–271. doi: 10.1075/ml.00006.lev

Lieven, E. (2008). "Learning the English auxillary: a usage based approach," in *Corpora in Language Acquisition Research – History, Methods, Perspectives*, ed H. Behrens (Amsterdam: Benjamins), 61–98. doi: 10.1075/tilar.6.06lie

Lieven, E., Behrens, H., Speares, J., and Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *J. Child Lang.* 30, 333–370. doi: 10.1017/S0305000903005592

Lohmann, G., Margulies, D. S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., et al. (2010). Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. *PLoS ONE* 5, 1–8. doi: 10.1371/journal.pone.0010232

MacWhinney, B. (2005). The emergence of linguistic form in time. *Connect. Sci.* 17, 191–211. doi: 10.1080/09540090500177687

Maital, S. L., Dromi, E., Sagi, A., and Bornstein, M. H. (2000). The Hebrew communicative development inventory: language specific properties and cross-linguistic generalizations. *J. Child Lang.* 27, 43–67. doi: 10.1017/S0305000999004006

Maslen, R., Theakston, A., Lieven, E., and Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *J. Speech Lang. Hear. Res.* 47, 1319–1333. doi: 10.1044/1092-4388(2004/099)

Montag, J., Jones, M., and Smith, L. (2015). The words children hear: picture books and the statistics for language learning. *Psychol. Sci.* 26, 1489–1496. doi: 10.1177/0956797615594361

Naigles, L., and Hoff-Ginsberg, E. (1998). Why are some verbs learned before other verbs? effects of input frequency and verb structure on children's early verb use. *J. Child Lang.* 25, 95–120. doi: 10.1017/S0305000997003358

Ninio, A. (2014). Variables and values in children's early word combinations. *Psychol. Lang. Commun.* 18, 106–125. doi: 10.2478/plc-2014-0009

Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciute, A., Suo, C., and Fornito, A. (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE* 14, 1–23. doi: 10.1371/journal.pone.0220061

Ravid, D. (1995). *Language Change in Child and Adult Hebrew: A Psycholinguistic Perspective*. Oxford: Oxford University Press.

Ravid, D. (2003). "A developmental perspective on root perception in Hebrew and Palestinian Arabic," in *Language Processing and Acquisition in Languages of Semitic, Root-Based Morphology*, ed Y. Shimron (Amsterdam: John Benjamins), 293–319. doi: 10.1075/lald.28.14rav

Ravid, D. (2012). *Spelling Morphology: The Psycholinguistics of Hebrew Spelling*. New York, NY: Springer. doi: 10.1007/978-1-4419-0588-8

Ravid, D. (2019). "First-Language Acquisition of Morphology," in *Oxford Research Encyclopedia of Linguistics* (Oxford: Oxford University Press), 23. doi: 10.1093/acrefore/9780199384655.013.603

Ravid, D., Ashkenazi, O., Levie, R., Ben Zadok, G., Grunwald, T., Bratslavsky, R., et al. (2016). "Foundations of the early root category: analyses of linguistic input to Hebrew-speaking children," in *Acquisition and Development of Hebrew: From Infancy to Adolescence*, ed R. A. Berman (Amsterdam; Philadephia, PA: John Benjamins), 95–134. doi: 10.1075/tilar.19.04rav

Ravid, D., and Bar-On, A. (2005). Manipulating written Hebrew roots across development: the interface of semantic, phonological and orthographic factors. *Read. Writ.* 18, 231–256. doi: 10.1007/s11145-005-1802-5

Ravid, D., and Schiff, R. (2006). Roots and patterns in Hebrew language development: evidence from written morphological analogies. *Read. Writ.* 19:3. doi: 10.1007/s11145-006-9004-3

Ravid, D., and Schiff, R. (2015). It's all about gender: Hebrew speakers' judgment of adjective plural agreement. *Morphology* 25, 327–343. doi: 10.1007/s11525-015-9259-y

Rowe, M., and Snow, C. (2020). Analyzing input quality along three dimensions: interactive, linguistic, and conceptual. *J. Child Lang.* 47, 5–21. doi: 10.1017/S0305000919000655

Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Curr. Direct. Psychol. Sci.* 12, 110–114. doi: 10.1111/1467-8721.01243

Schiff, R., Raveh, M., and Fighel, A. (2012). The development of the Hebrew mental lexicon: when morphological representations become devoid of their meaning. *Sci. Stud. Read.* 16, 383–403. doi: 10.1080/10888438.2011.571327

Schiff, R., and Ravid, D. (2007). Morphological analogies in Hebrew-speaking university students with dyslexia compared with typically developing gradeschoolers. *J. Psychol. Res.* 36, 237–253. doi: 10.1007/s10936-006-9043-6

Schwarzwald, O. (1998). Inflection and derivation in Hebrew linear word formation. *Folia Linguist.* 32, 265–288. doi: 10.1515/flin.1998.32.3-4.265

Schwarzwald, O. R. (1981). *Grammar and Reality in the Hebrew Verb*. Ramat Gan: Bar Ilan University Press.

Snow, C. (1995). "Issues in the study of input: finetuning, universality, individual and developmental differences, and necessary causes," in *The Handbook of Child Language*, eds P. Fletcher and B. MacWhinney (Oxford: Blackwell), 180–193.

Temkin Martinez, M. (2010). *Sources of Non-Conformity in Phonology: Variation and Exceptionality in Modern Hebrew Spirantization*. (Ph.D. thesis), University of Southern California, Los Angeles, CA, United States.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition* 74, 209–253. doi: 10.1016/S0010-0277(99)00069-4

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Approach to Child Language Acquisition*, Cambridge, CA: Cambridge University Press.

Tomasello, M. (2004). Learning through others. *Dedalus* 133, 51–58. doi: 10.1162/001152604772746693

Tomasello, M. (2006). "Acquiring linguistic constructions," in *Handbook of Child Psychology*, eds R. Siegler and D. Kuhn (New York, NY: Wiley), 256–293. doi: 10.1002/9780470147658.chpsy0206

Tomasello, M. (2009). "The usage-based theory of language acquisition," in *The Cambridge Handbook of Child Language*, ed E. L. Bavin (Cambridge, CA: Cambridge University Press), 69–87. doi: 10.1017/CBO9780511576164.005

van Dijk, M., van Geert, P., Korecky-Kroll, K., Maillochon, I., Laaha, S., Dressler, W. U., et al. (2013). Dynamic adaptation in child–adult language interaction. *Lang. Learn.* 63, 243–270. doi: 10.1111/lang.12002

van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychol. Rev.* 98, 3–53. doi: 10.1037/0033-295X.98.1.3

van Geert, P. (2010). "Dynamic systems methods in the study of language acquisition," in *Language Acquisition Across Linguistic and Cognitive Systems*, eds M. Kail and M. Hickmann (Amsterdam: John Benjamins Publishing Company). doi: 10.1075/lald.52.04gee

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge, CA: Cambridge University Press. doi: 10.1017/CBO9780511815478

Xantos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., et al. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Lang.* 31, 461–479. doi: 10.1177/0142723711409976

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

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fpsyg.2021.720713

Modeling Morphology With Linear Discriminative Learning: Considerations and Design Choices

Maria Heitmeier, Yu-Ying Chuang and R. Harald Baayen*

Department of Linguistics, Eberhard-Karls Universität, Tübingen, Germany

This study addresses a series of methodological questions that arise when modeling inflectional morphology with Linear Discriminative Learning. Taking the semi-productive German noun system as example, we illustrate how decisions made about the representation of form and meaning influence model performance. We clarify that for modeling frequency effects in learning, it is essential to make use of incremental learning rather than the end-state of learning. We also discuss how the model can be set up to approximate the learning of inflected words in context. In addition, we illustrate how in this approach the wug task can be modeled. The model provides an excellent memory for known words, but appropriately shows more limited performance for unseen data, in line with the semi-productivity of German noun inflection and generalization performance of native German speakers.

Keywords: German nouns, linear discriminative learning, semi-productivity, multivariate multiple regression, Widrow-Hoff learning, frequency of occurrence, semantic roles, wug task

Edited by:
Juhani Järvikivi,
University of Alberta, Canada

Reviewed by:
Xiaowei Zhao,
Emmanuel College, United States
Antti Arppe,
University of Alberta, Canada

****Correspondence:***
Maria Heitmeier
maria.heitmeier@uni-tuebingen.de

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

Received: *04 June 2021*
Accepted: *04 October 2021*
Published: *15 November 2021*

Citation:
Heitmeier M, Chuang Y-Y and Baayen RH (2021) Modeling Morphology With Linear Discriminative Learning: Considerations and Design Choices. Front. Psychol. 12:720713. doi: 10.3389/fpsyg.2021.720713

1. INTRODUCTION

Computational models of morphology fall into two broad classes. The first class addresses the question of how to produce a morphologically complex word given a morphologically related form (often a stem, or an identifier of a stem or lexeme) and a set of inflectional or derivational features. We refer to these models as form-oriented models. The second class comprises models seeking to understand the relation between words' forms and their meanings. We refer to these models as meaning-oriented models.

Prominent form-oriented models comprise Analogical Modeling of Language (AML; Skousen, 1989, 2002) and Memory Based Learning (MBL; Daelemans and Van den Bosch, 2005), which are nearest-neighbor classifiers. Input to these models are tables with observations (words) in rows, and factorial predictors and a factorial response in columns. The response specifies an observation's outcome class (e.g., an allomorph), and the model is given the task to predict the outcome classes from the other predictor variables (for allomorphy, specifications of words' phonological make-up). Predictions are based on sets of nearest neighbors, serving as constrained exemplar sets for generalization. These models have clarified morphological phenomena ranging from the allomorphy of the Dutch diminutive (Daelemans et al., 1995) to stress assignment in English (Arndt-Lappe, 2011).

Ernestus and Baayen (2003) compared the performance of the MBL, AML, and Generalized Linear Models (GLM), as well as a recursive partitioning tree (Breiman et al., 1984), on the task of predicting whether word-final obstruents in Dutch alternate with respect to their voicing. They observed similar performance across all models, with the best performance, surprisingly, for the only parameter-free model, AML. Their results suggest that the quantitative structure of morphological datasets may be straightforward to discover for any reasonably decent classifier. The model proposed by Belth et al. (2021) is a recent example of a classifier based on recursive partitioning.

Minimum Generalization Learning (MGL; Albright and Hayes, 2003) offers an algorithm for rule induction (for comparison with nearest neighbor methods, see Keuleers et al., 2007). The model finds rules by an iterative process of minimal generalization that combines specific rules into ever more general rules. Each rule comes with a measure of prediction accuracy, and the rule with the highest accuracy is selected for predicting a word's form.

All models discussed thus far are exemplar-based, in the sense that the input to any of these models consists of a table with exemplars, exemplar features selected on the basis of domain knowledge, and a categorical response variable specifying targeted morphological form changes. In other words, all these models are classifiers that absolve the analyst from hand-engineering lexical entries, rules or constraints operating on these lexical entries, and theoretical constructs such as inflectional classes. In this respect, they differ fundamentally from the second group of the following computational methods.

The DATR language (Evans and Gazdar, 1996) defines non-monotonic inheritance networks for knowledge representation. This language is optimized for removing redundancy from lexical descriptions. A DATR model requires the analyst to set up lexical entries that specify information about, for instance, inflectional class, gender, the forms of exponents, and various kinds of phonological information. The lexicon is designed in such a way that the network is kept as small as possible, while still allowing the model, through its mechanism of inheritance, to correctly predict all inflected variants. Realizational morphology (RM; Stump, 2001) sets up rules for realizing bundles of inflectional and lexical features in phonological form. This theory can also be defined as a formal language (a finite-state transducer) that provides mappings from underlying representations onto their corresponding surface forms and vice versa (Karttunen, 2003). The Gradual Learning Algorithm (GLA; Boersma, 1998; Boersma and Hayes, 2001) works within the framework of optimality theory (Prince and Smolensky, 2008). The algorithm is initialized with a set of constraints and gradually learns an optimal constraint ranking by incrementally moving through the training data, and upgrading or downgrading constraints.

The third group of form-oriented computational models comprises connectionist models. The past-tense model of Rumelhart and McClelland (1986) was trained to produce English past-tense forms given the corresponding present-tense form. An early enhancement of this model was proposed by MacWhinney and Leinbach (1991), for an overview of the many follow-up models, see Kirov and Cotterell (2018). Kirov and Cotterell proposed a sequence-to-sequence deep learning network, the Encoder-Decoder (ED) learner, that they argue does not suffer from the drawbacks noted by Pinker and Prince (1988) for the original paste-tense model. Malouf (2017) introduced a recurrent deep learning model trained to predict upcoming segments, showing that this model has high accuracy for predicting paradigm forms given the lexeme and the inflectional specifications of the desired paradigm cell.

In summary, the class of form-oriented models comprises three subsets: statistical classifiers (AML, MBL, GLM, recursive partitioning), generators based on linguistic knowledge engineering (DATR, RM, GLA), and connectionist models (paste-tense model, ED learner). The models just referenced presuppose that when speakers use a morphologically complex form, this form is derived on the fly from its underlying form. The sole exception is the model of Malouf (2017), which takes the lexeme and its inflectional features as point of departure. As pointed out by Blevins (2016), the focus on how to create one form from another has its origin in pedagogical grammars, which face the task of clarifying to a second language learner how to create inflected variants. Unsurprisingly, applications within natural language processing also have need of systems that can generate inflected and derived words.

However, it is far from self-evident that native speakers of English would create past-tense forms from present-tense forms. Meaning-oriented models argue that in comprehension, the listener or reader can go straight from the auditory or visual input to the intended meaning, without having to go through a pipeline requiring initial identification of underlying forms and exponents. Likewise, speakers are argued to start from meaning, and realize this meaning directly in written or spoken form.

The class of meaning-oriented models comprises both symbolic and subsymbolic models. The symbolic models of Dell (1986) and Levelt et al. (1999) implement a form of realizational morphology. Concepts and inflectional features activate stems and exponents, which are subsequently combined into syllables. Both models hold that the production of morphologically complex words is a compositional process in which units are assembled together and ordered for articulation at various hierarchically ordered levels. These models have been worked out only for English, and to our knowledge have not been applied to languages with richer morphological systems.

The subsymbolic model of Harm and Seidenberg (2004) sets up multi-layer networks between orthographic, phonological, and semantic units. No attempt is made to define morphemes, stems, or exponents. To the extent that such units have any reality, they are assumed to arise, statistically, at the hidden layers. Mirković et al. (2005) argue for Serbian that gender is an emergent property of the network that arises from statistical regularities governing both words' forms and their meanings (see Corbett, 1991, for discussion of semantic motivations for gender systems). The model for auditory comprehension of Gaskell and Marslen-Wilson (1997) uses a three-layer recurrent network to map speech input onto distributed semantic representations, again without attempting to isolate units such as phonemes or morphemes.

The naive discrimination learning (NDL) model proposed by Baayen et al. (2011) represents words' forms sub-symbolically, but words' meanings symbolically. The modeling set-up that we discuss in the remainder of this study, that of linear discriminative learning (LDL, Baayen et al., 2019), replaces the symbolic representation of word meaning in

NDL by sub-symbolic representations building on distributional semantics (Landauer and Dumais, 1997; Mikolov et al., 2013b).

LDL is an implementation of Word and Paradigm Morphology (Matthews, 1974; Blevins, 2016). Sublexical units such as stems and exponents play no role. Semantic representations in LDL, however, are analytical: the semantic vector (word embedding, i.e., a distributed representation of meaning) of an inflected word is constructed from the semantic vector of the lexeme and the semantic vectors of the pertinent inflectional functions. Both NDL and LDL make use of the simplest possible networks: networks with only input and output layers, and no hidden layers.

At this point, the distinction made by Breiman (2001) between statistical models and machine learning is relevant. Statistical models aim to provide insight into the mechanisms that generate the data. Machine learning, on the other hand, aims to optimize prediction accuracy, and it is not an issue whether or not the algorithms are interpretable. LDL is much closer to statistical modeling than to the black boxes of machine learning. All input and output representations can be set up in a theoretically transparent way (Baayen et al., 2019). Furthermore, because LDL implements multivariate multiple regression, its mathematical properties are well-understood. Importantly, modeling results do not depend on the choice of hyper-parameters (e.g., the numbers of LSTM layers and LSTM units), instead, they are completely determined by the representations chosen by the analyst.

The goals of this study are, first, to clarify how such choices of representation affect LDL model performance; second, to illustrate what can be achieved simply with multivariate multiple regression; and third, to call attention to the kind of problems that are encountered when word meaning is integrated into morphology. Our working example is the comprehension and production of German nouns. In what follows, we first introduce the German noun system, and review models that have been proposed for German nouns. We then introduce LDL, after which we present a systematic overview of modeling choices, covering the representation of form, the representation of meaning, and the learning algorithm (incremental learning vs. the regression "end-state of learning" solution).

2. GERMAN NOUN MORPHOLOGY

The German noun system is both highly irregular and semi-productive, featuring three different genders, two numbers and four cases. In this section, we will give an overview over this system, show where irregularity and semi-productivity arise, and which (non-computational) models have been employed to account for it.

Plural forms are marked with one of four suffixes (*-(e)n*, *-er*, *-e*, *-s*) or without adding a suffix [∅; a "zero" morpheme (Köpcke, 1988, p. 306)], three of which can pair with stem vowel fronting [e.g., *a* (/a/) → *ä* (/ɛ/)] (e.g., Köpcke, 1988) (**Table 1**). There are additional suffixes which usually apply to words with foreign origin, such as *-i* (e.g., *Cello* → *Celli*, "cellos") (Cahill and Gazdar, 1999). Cahill and Gazdar (1999) sub-categorize the nouns into 11 classes, based on whether singulars have a different suffix than plurals (*Album* → *Alben*, "albums"). Nakisa and Hahn (1996) distinguish between no less than 60 inflection classes. No plural class is prevalent overall (Köpcke, 1988), and it is impossible to fully predict plural class from gender, syntax, phonology or semantics (Köpcke, 1988; Cahill and Gazdar, 1999; Trommer, 2021). Further complications arise when case is taken into account. German has four cases: *nominative, genitive, dative*, and *accusative*, which are marked with two exponents (applied additional to the plural markers): *-(e)n* and *-(e)s* (Schulz and Griesbach, 1981). Case forms are also not fully predictable from gender, phonology or meaning. Since many forms do not receive a separate marker, the system has been described as "degenerate" (Bierwisch, 2018, p. 245) (see **Table 2**). German speakers do, however, get additional disambiguing information from the definite and indefinite articles which accompany nouns and likewise encode gender, number, and case. **Table 2** shows the definite articles for all genders. Additionally, there are indefinite articles available for singular forms which also express case in their endings (e.g., Gen. sg. m./n./f. *eines*, Dat. sg. m./n. *einem*, Dat. sg. f. *einer*).

Unsurprisingly, it has been the subject of a long-standing debate whether a distinction between regular and irregular nouns is useful for German (the debate has mostly focused on the formation of the nominative plural which we accordingly also

TABLE 1 | Plural classes of German nouns (relative frequencies from Gaeta, 2008).

Plural class	Example	Type frequency (%)
-(e)n	Tasse → Tassen "cup(s)"	56.5
(uml+)-e	Tag → Tage "day(s)"	
	Topf → Töpfe "pot(s)"	23.9
(uml+)-er	Brett → Bretter "board(s)"	
	Glas → Gläser "glass(es)"	2.3
(uml+)∅	Daumen → Daumen "thumb(s)"	
	Apfel → Äpfel "apple(s)"	13.3
-s	Kamera → Kameras "camera(s)"	2.6

Most of the classes can appear with both masculine and neuter nouns. Feminine nouns belong mostly to the -(e)n class (97%; Gaeta, 2008).

TABLE 2 | German noun declension.

Case and number	Masculin I	Masculin II	Neutral	Feminin
Nom. sg.	der Freund	der Mensch	das Kind	die Mutter
Gen. sg.	des Freund**es**	des Mensch**en**	des Kind**es**	der Mutter
Dat. sg.	dem Freund	dem Mensch**en**	dem Kind	der Mutter
Acc. sg.	den Freund	den Mensch**en**	das Kind	die Mutter
Nom. pl.	die Freund**e**	die Mensch**en**	die Kind**er**	die Mütter
Gen. pl.	der Freund**e**	der Mensch**en**	der Kind**er**	der Mütter
Dat. pl.	den Freund**en**	den Mensch**en**	den Kind**ern**	den Mütter**n**
Acc. pl.	die Freund**e**	die Mensch**en**	die Kind**er**	die Mütter

Plural endings vary with declension class. Table adapted from Schulz and Griesbach (1981, p. 105).

focus on here). It is also unsurprising that the system shows limited productivity. Several so-called "wug" studies, where participants are asked to inflect nonce words, have clarified that German native speakers struggle with predicting unseen plurals. Köpcke (1988), Zaretsky et al. (2013), and McCurdy et al. (2020) reported high variability across speakers with respect to the plural forms produced. Köpcke (1988) took this as evidence for a "modified schema model" of German noun inflection, arguing that plural forms are generated based not only on a speaker's experience with the German noun system, but also on the "cue validity" of the plural markers. For example, *-(e)n* is a good cue for plurality, as it does not occur with many singular forms. By contrast, *-er* has low cue validity for plurality, as it occurs with many singulars.

Köpcke (1988) also observed that *-s* is used slightly more in his wug experiments than would be expected from corpus data. Marcus et al. (1995) and Clahsen (1999) therefore argued that *-s* serves as the regular default plural marker in German, contrasting with all other plural markers that are described as irregular. Others, however, have argued that an *-s* default rule does not provide any additional explanatory value (Nakisa and Hahn, 1996; Behrens and Tomasello, 1999; Indefrey, 1999; Zaretsky and Lange, 2015).

Despite the irregularity and variability in the system, some sub-regularities within the German noun system have also been pointed out (Wiese, 1999; Wunderlich, 1999). For instance, Wunderlich (1999, p. 7f.) reports a set of rules that German nouns adhere to, which can be overridden on an item-by-item basis through "lexical storage." For example, he notes that

a. Masculines ending in schwa are weakly inflected (and thus also have n-plurals).
b. Non-umlauting feminine have an n-plural.
c. Non-feminines ending in a consonant have a ə-plural. [...]
e. All atypical nouns have an s-plural. [...]

He also allows for semantics to co-determine class membership. For instance, masculine animate nouns show a tendency to belong to the *-n* plural class (see also Gaeta, 2008). A further remarkable aspect of the German noun system, especially for second language learners, is that whereas it is remarkably difficult to learn to produce the proper case-inflected forms, understanding these forms in context is straightforward.

In the light of these considerations, the challenges for computational modeling of German noun inflection, specifically from a cognitive perspective, are the following:

1. To construct a memory for a highly irregular, "degenerate," semi-productive system,
2. To ensure that this memory shows some moderate productivity for novel forms, but with all the uncertainties that characterize the generalization capacities of German native speakers,
3. To furthermore ensure that the performance of the mappings from form to meaning, and from meaning to form, within the framework of the discriminative lexicon (Baayen et al., 2019), are properly asymmetric with respect to comprehension and production accuracy (see also Chuang et al., 2020a).

2.1. Computational Models for German Nouns

The complexity of the German declension system has inspired many computational models. The DATR model of Cahill and Gazdar (1999) belongs to the class of generating models based on linguistic knowledge engineering. It assigns lexemes to carefully designed hierarchically ordered declension classes. Each class inherits the properties from classes further up in the hierarchy, but will override some of these properties. This model provides a successful and succinct formal model for German noun declension. Other models from this class include GERTWOL which is based on finite-state operations (Haapalainen and Majorin, 1994), as well as the model of Trommer (2021) which draws on Optimality Theory (OT) and likewise requires careful hand-crafting and constraint ranking (but does not currently have a computational implementation).

Belth et al. (2021) propose a statistical classifier based on recursive partitioning, with as response variable the morphological change required to transform a singular into a plural, and as predictors the final segments of the lexeme, number, and case. At each node, nouns are divided by their features, with one branch comprising the most frequent plural ending (which will inevitably include some nouns with a different plural ending, labeled as exceptions), and with the other branch including the remainder of the nouns. Each leaf node of the resulting tree is said to be productive if a criterion for node homogeneity is met. An older model, also a classifier building up rules inductively, was developed 20 years earlier by Albright and Hayes (2003).

Connectionist models for the German noun system include a model using a simple recurrent network (Goebel and Indefrey, 2000), and a deep learning model implementing a sequence-to-sequence encoder-decoder (McCurdy et al., 2020). The latter model takes letter-based representations of German nouns in their nominative singular form as input, together with information on the grammatical gender of the noun. The model is given the task to produce the corresponding nominative plural form. The model learned the task with high accuracy on held out data (close to 90%), but was more locked in on the "correct" forms compared to native speakers, who in a wug task showed substantially more variability in their choices.

The models discussed above also differ with respect to how they generate predictions for novel nouns. The sequence-to-sequence deep learning model of (McCurdy et al., 2020) can do so relatively easily, straight from a word's form and its gender specification but its inner workings are not immediately interpretable (though recent work has started to gain some insights, see e.g., Linzen and Baroni, 2021, for syntactic structure in deep learning). By contrast, the linguistically more transparent DATR model can only generate a novel word's inflectional variants once this word has been assigned to an inflectional class. This may to some extent be possible given its principal parts (Finkel and Stump, 2007), but clearly requires additional mechanisms to be in place.

In what follows, we introduce the LDL model. LDL is a model of human lexical processing, with all its limitations and constraints, rather than an optimized computational system for generating (or understanding) morphologically complex words. It implements a simple linear mapping between form and meaning, where form is represented as a binary vector of sublexical cues, and meaning is represented in a distributed fashion.

By applying LDL to the modeling of the German noun system (including its case forms), we address a question that has thus far not been addressed computationally, namely the incorporation of semantics. Semantic subregularities in the German noun system have been noted by several authors (e.g., Wunderlich, 1999; Gaeta, 2008), and although deep learning models can be set up that incorporate semantics (see e.g., Malouf, 2017), LDL by design must take semantics into account.

The next section introduces the LDL model. The following sections proceed with an overview of the many modeling decisions that have to be made. An important part of this overview is devoted to moving beyond the modeling of isolated words, as words come into their own only in context (Elman, 2009), and case labels do not correspond to contentful semantics, but instead are summary devices for syntactic distribution classes (Blevins, 2016; Baayen et al., 2019).

3. LINEAR DISCRIMINATIVE LEARNING

LDL is the computational engine of the discriminative lexicon model (DLM) proposed by Baayen et al. (2019). The DLM implements mappings between form and meaning for both reading and listening, and mappings from meaning to form for production. It also allows for multiple routes operating in parallel. For reading in English, for instance, it sets up a direct route from form to meaning, in combination with an indirect route from visual input to a phonological representation that in turn is mapped onto the semantics (cf. Coltheart et al., 1993). In what follows, we restrict ourselves to the mappings from form onto meaning (comprehension) and from meaning onto form (production). Mappings can be obtained either with trial-to-trial learning, or by estimating the end-state of learning. In the former case, the model implements incremental regression using the learning rule of Widrow and Hoff (1960); in the latter case, it implements multivariate multiple linear regression, which is mathematically equivalent to a simple network with input units, output units, no hidden layers, and simple summation of incoming activation without using thresholding or squashing functions.

Each word form of interest is represented by a set of cues. For example, *wordform1* might feature the cues `cue1`, `cue2`, and `cue3`, while *wordform2* could be marked by `cue1`, `cue4`, and `cue5`. We can thus express a word form as a binary vector, where 1 denotes presence and 0 absence. This information is coded in the cue matrix **C**:

$$\mathbf{C} = \begin{array}{c} \\ \text{wordform1} \\ \text{wordform2} \end{array} \begin{array}{c} \begin{array}{ccccc} \texttt{cue1} & \texttt{cue2} & \texttt{cue3} & \texttt{cue4} & \texttt{cue5} \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Words' meanings are also represented by numeric vectors. The dimensions of these vectors can have a discrete interpretation, or have a latent interpretation (see section 4.2 below for detailed discussion). In the following example, *wordform1* has strong negative support for semantic dimensions `S3` and `S5`, while *wordform2* has strong positive support for `S4` and `S5`. This information is brought together in a semantic matrix **S**:

$$\mathbf{S} = \begin{array}{c} \\ \text{wordform1} \\ \text{wordform2} \end{array} \begin{array}{c} \begin{array}{ccccc} \texttt{S1} & \texttt{S2} & \texttt{S3} & \texttt{S4} & \texttt{S5} \end{array} \\ \begin{pmatrix} 0.1 & 0.004 & -1.95 & 0.03 & -0.54 \\ -0.49 & -0.32 & 0.03 & 1.06 & 0.98 \end{pmatrix} \end{array}$$

Comprehension and production in LDL are modeled by means of simple linear mappings from the form matrix $\boldsymbol{C}$ to the semantic matrix **S**, and vice versa. These mappings specify how strongly input nodes are associated with output nodes. The weight matrix for a given mapping can be obtained in two ways. First, using the mathematics of multivariate multiple regression, a comprehension weight matrix **F** is obtained by solving **F** from

$$\mathbf{S} = \mathbf{C} \cdot \mathbf{F},$$

and a production weight matrix **G** is obtained by solving **G** from

$$\mathbf{C} = \mathbf{S} \cdot \mathbf{G}.$$

As for linear regression modeling, the predicted row vectors are approximate. Borrowing notation from statistics, we write

$$\hat{\mathbf{S}} = \mathbf{C} \cdot \mathbf{F}$$

for predicted semantic vectors (row vectors of $\hat{\mathbf{S}}$), and

$$\hat{\mathbf{C}} = \mathbf{S} \cdot \mathbf{G}$$

for predicted form vectors (row vectors of $\hat{\mathbf{C}}$).

These equations amount to estimating multiple outcomes from multiple variables, which in statistics is referred to as multivariate multiple regression. In simple linear regression, a single value y is estimated from a value x via an intercept β_0 and a weighing of x with scalar β_1:

$$\hat{y} = \beta_0 + \beta_1 x \tag{1}$$

which can easily be expanded to estimating y from a vector $\mathbf{x}$ (multiple linear regression), using a vector of beta coefficients $\beta_i \in \boldsymbol{\beta}$ to weigh each value $x_i \in \mathbf{x}$:

$$\hat{y} = \beta_0 + x_1\beta_1 + x_2\beta_2 + ... + x_n\beta_n \tag{2}$$

Finally, to estimate a vector $\mathbf{y}$ from a vector $\mathbf{x}$ (multivariate multiple regression), we need an entire matrix of beta coefficients $\beta_{ij} \in \mathbf{B}$. A single value $y_i \in \mathbf{y}$ is then estimated via

$$\hat{y}_i = \beta_{0i} + x_1\beta_{1i} + x_2\beta_{2i} + \ldots + x_p\beta_{pi} \quad (3)$$

Thus, estimating the mappings $\mathbf{F}$ and $\mathbf{G}$ in LDL amounts to computing the coefficients matrix $\mathbf{B}$ for mappings from $\mathbf{C}$ to $\mathbf{S}$ and vice versa. As such, each value in a predicted semantic vector $\hat{\mathbf{s}}$ (form vector $\hat{\mathbf{c}}$) is a linear combination (i.e., weighted sum) of the values in the corresponding form vector $\mathbf{c}$ (semantic vector $\mathbf{s}$) it is predicted from. This means that LDL is mathematically highly constrained: it cannot handle non-linearities that even shallow connectionist models (e.g., Goldsmith and O'Brien, 2006) can take in their stride. Nevertheless, we have found that these simple linear mappings result in high accuracies (e.g., Baayen et al., 2018, 2019) suggesting that morphological systems are surprisingly simple. Cases where model predictions are less precise due to the limitations of linearity become indicative of learning bottlenecks.

Furthermore, note that estimating the mappings $\mathbf{F}$ and $\mathbf{G}$ using the matrix algebra of multivariate multiple regression provides optimal estimates, in the least squares sense, of the connection weights (or equivalently, beta coefficients) for datasets that are type-based, in the sense that each pair of row vectors $\boldsymbol{c}$ of $\boldsymbol{C}$ and $\boldsymbol{s}$ of $\boldsymbol{S}$ is unique. Having multiple instances of the same pair of row vectors in the dataset does not make sense, as it renders the input completely singular and does not add any further information. Thus, models based on the regression estimates of $\mathbf{F}$ and $\mathbf{G}$ are comparable to type-based models such as AML, MBL, MGL, and models using recursive partitioning.

Making the estimates of the mappings sensitive to frequency of use requires incremental learning, updating weights after each word token that is presented for learning. Incremental learning is implemented using the learning rule of Widrow and Hoff (1960) and Milin et al. (2020), which defines the matrix $\boldsymbol{W}^{t+1}$ with updated weights at time $t+1$ as the weight matrix $\boldsymbol{W}^t$ at time t, modified as follows:

$$\boldsymbol{W}^{t+1} = \boldsymbol{W}^t + \boldsymbol{c} \cdot (\boldsymbol{o}^T - \boldsymbol{c}^T \cdot \boldsymbol{W}^t) \cdot \eta,$$

where $\boldsymbol{c}$ is the current cue (vector), $\boldsymbol{o}$ the current outcome vector, and η the learning rate. Conceptually, this means that after each newly encountered word token, the weight matrix is changed such that the next time that the same cue vector has to be mapped onto its associated outcome vector, it will be slightly closer to the target outcome vector than it was before. The learning rule of Widrow-Hoff implements incremental regression. As the number of times that a model is trained again and again on a training set increases (training epochs), the network's weights will converge to the matrix of beta coefficients obtained by approaching the estimation problem with multivariate multiple regression (see e.g., Evert and Arppe, 2015; Chuang et al., 2020a; Shafaei-Bajestan et al., 2021). As a consequence, the regression-based estimates pertain to the "end-state of learning," at which the data have been worked through infinitely many times. Unsurprisingly, effects of frequency and order of learning are not reflected in model predictions based on the regression estimates. Such effects do emerge with incremental learning (see section 4.5).

Comprehension accuracy for a given word ω is assessed by comparing its predicted semantic vector $\hat{\boldsymbol{s}}_\omega$ with all gold standard semantic vectors in $\boldsymbol{S}$ (the creation of gold standard semantic vectors will be described in subsequent sections), using either the cosine similarity measure or the Pearson correlation r. In what follows, we use r, and select as the meaning that is recognized that gold standard row vector $\boldsymbol{s}_{\max}$ of $\boldsymbol{S}$ that shows the highest correlation with $\hat{\boldsymbol{s}}_\omega$. If $\boldsymbol{s}_{\max}$ is the targeted semantic vector, the model's prediction is classified as correct, otherwise, it is taken to be incorrect.

For the modeling of production, a supplementary algorithm is required for constructing actual word forms. The predicted vectors $\hat{\boldsymbol{c}}$ provide information about the amount of support that form cues receive from the semantics. However, information about the amount of support received by the full set of cues does not provide information about the order in which a subset of these cues have to be woven together into actual words. Algorithms that construct words from form cues make use of the insight that when form cues are defined as n-grams ($n > 1$), the cues contain implicit information about order. For instance, for digraph cues, cues `ab` and `bc` can be combined into the string `abc`, whereas cues `ab` and `cd` cannot be merged. Therefore, when n-grams are used as cues, directed edges can be set up in a graph with n-grams as vertices, for any pair of n-grams that properly overlap. A word form is uniquely defined by a path in such a graph starting with an initial n-gram (starting with an initial word edge symbol, typically a # is used) and ending at a final n-gram (ending with #). This raises the question of how to find word paths in the graph. This is accomplished by first discarding n-grams with low support from the semantics below a threshold θ[1], then calculating all possible remaining paths, and finally selecting for articulation that path for which the corresponding predicted semantic vector (obtained by mapping its corresponding cue vector $\boldsymbol{c}$ onto $\boldsymbol{s}$ using comprehension matrix $\boldsymbol{F}$) best matches the semantic vector that is the target for articulation. This implements "synthesis by analysis," see Baayen et al. (2018, 2019) for further details and theoretical motivation. For a discussion of the cognitive plausibility of this method, see Chuang et al. (2020b).

The first algorithm that was used to enumerate possible paths made use of a shortest-paths algorithm from graph theory. This works well for small datasets, but becomes prohibitively expensive for large datasets. The **JudiLing** package (Luo et al., 2021) offers a new algorithm that scales up better. This algorithm is first trained to predict, from either the $\hat{\boldsymbol{C}}$ or the $\boldsymbol{S}$ matrix, for each possible word position, which cues are best supported at that position. All possible paths with the top k best-supported cues are then calculated, and subjected to synthesis by analysis. Details about this algorithm, implemented in julia in the **JudiLing** package as the function `learn_paths` can be found in Luo (2021). The `learn_paths` function is used throughout the

[1]This is a simple cut-off point for n-grams with low support, not to be confused with thresholds as often used in deep learning.

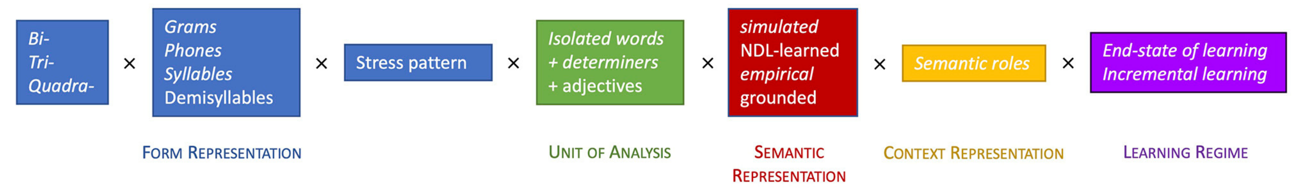

FIGURE 1 | Options when modeling a language's morphology with LDL. Examples with options in *italics* are discussed in the present study.

remainder of this study. A word form is judged to be produced correctly when it exactly matches the targeted word form.

4. MODELING CONSIDERATIONS

When modeling a language's morphology within the framework of the DLM, the analyst is faced with a range of choices, illustrated in **Figure 1**. From left to right, choices are listed for representing form, for the unit of analysis, for the representation of semantics, for the handling of context, and for the learning regime.

With respect to form representations, the kind of n-gram has to be selected, setting *n*, deciding on phonological or orthographic grams, and specifying how stress or lexical tone are represented. With respect to the unit of analysis, the analyst has to decide whether to model isolated words, or words in phrasal contexts. A third set of choices concerns what semantic representations to use: simulated representations, or word embeddings such as *word2vec* (Mikolov et al., 2013b), or grounded vectors (Shahmohammadi et al., in press). A further set of choices for languages with case concerns how to handle case labels, as these typically refer to syntactic distribution classes rather than contentful inflectional features (Blevins, 2016). Finally, a selection needs to be made with respect to whether incremental learning is used, or instead the end-state of learning using regression-based estimation. In what follows, we illustrate several of these choice points using examples addressing the German noun system.

The dataset on German noun inflection that we use for our worked examples was compiled as follows. First, we extracted all monomorphemic nouns and their inflections with a frequency of at least 1 from CELEX (Baayen et al., 1995), resulting in a dataset of about 6,000 word forms. Of these we retained the 5,486 word forms for which we could retrieve grammatical gender from Wiktionary, thus including word forms of 2,732 different lemmas. The resulting data was expanded such that each attested word form was listed once for each possible paradigm cell it could belong to. For instance, *Aal* ("eel") is listed once as singular nominative, once as dative and once as accusative (**Table 3**). This resulted in a dataset with 18,147 entries, with word form frequencies ranging from 1 to 5,828, (M log frequency 2.56, SD 1.77). Word forms are represented in their DISC notation, which represents German phones with single characters[2]. **Table 3** clarifies that there are many homophones. As a consequence, the actual number of distinct word forms in our dataset is only 5,486, which amounts to on average about two word forms per lemma.

TABLE 3 | Representation of the paradigm for *Aal* "eel" in our dataset.

Word form	Pronunciation	Lemma	Case	Number	Frequency	Gender
Aal	`al`	Aal	Nominative	Singular	29	M
Aal	`al`	Aal	Dative	Singular	29	M
Aal	`al`	Aal	Accusative	Singular	29	M
Aale	`al@`	Aal	Nominative	Plural	34	M
Aale	`al@`	Aal	Genitive	Plural	34	M
Aalen	`al@n`	Aal	Dative	Plural	17	M
Aalen	`al@n`	Aal	Accusative	Plural	17	M

Genitive singular (Aals) is not included as it has a frequency of 0 in CELEX.

There are many ways in which model performance can be evaluated. First, we may be interested in how well the model performs as a memory. How well does the model learn to understand and produce words it has encountered before? Note that because the model is not a list of forms, this is not a trivial question. For evaluation of the model as a memory, we consider its performance on the training data (henceforth `train`). Second, we may be interested in the extent to which the memory is productive. Does it generalize so that new forms can be understood or produced? Above, we observed that the German noun system is semi-regular, and that German native speakers are unsure about what the proper plural is of words they have not encountered before (McCurdy et al., 2020). If our modeling approach properly mirrors human limitations on generalization from data with only partial regularities, evaluation on unseen, held-out data of German should not be perfect. At this point, however, several issues arise that require careful thought.

For one, from the perspective of the linguistic system, it seems unreasonable to assume that any held-out form can be properly produced (or understood) if some of the principal parts (Finkel and Stump, 2007) of the lexeme are missing in the training data. In what follows, we will make the simplifying assumption that under cross-validation with sufficient training data, this situation will not arise.

A further question that arises is how to evaluate held-out words that have homophones in the training data. Such homophones present novel combinations of a form vector (shared with another data point in the training data) and a semantic vector (not attested for this form in the training data). We may want to impose a strict evaluation criterion requiring

[2] Data and code are available in the **Supplementary Materials** at https://osf.io/zrw2v/

TABLE 4 | Types of model evaluation.

Evaluation type			
Simple	Blind evaluation of all held-out data		`val_all`
Nuanced	Evaluation on novel forms only		`val_newform`
	Evaluation on homophones	Strict	`val_strict`
		Lenient	`val_lenient`

that the model gets the semantic vector exactly right. However, when presented with a homophone in isolation, a human listener cannot predict which of a potentially large set of paradigm cells is the targeted one (the problem of modeling words in isolation). We may therefore want to use a lenient evaluation criterion for comprehension according to which comprehension is judged to be accurate when the predicted semantic vector $\hat{\boldsymbol{s}}$ is associated with one of a homophonic word's possible semantic interpretations. Yet a further possible evaluation metric is to see how well the model performs on words with forms that have not been encountered in the training data. These possibilities are summarized in **Table 4**. Below, in section 4.3.1, we consider further complications that can arise in the context of testing the model on unseen forms.

For evaluating the productivity of the model, we split the full dataset into 80% training data and 20% validation data, with 14,518 and 3,629 word forms, respectively. In the validation data, 3,309 forms are also present in the training data (i.e., homophones), and 320 are new forms. Among the 320 new forms, 8 have novel lemmas that are absent in the training data. Since it is unrealistic to expect the model to understand or produce inflected forms of completely new words, these 8 words are excluded from the validation dataset for new forms, although they are taken into consideration when calculating the overall accuracy for the validation data. The same training and validation data are used for all the simulations reported below, unless indicated otherwise.

4.1. Representing Words' Forms

Decisions about how to represent words' forms depend on the modality that is to be modeled. For auditory comprehension, Arnold et al. (2017) and Shafaei-Bajestan et al. (2021) explore ways in which form vectors can be derived from the audio signal. Instead of using low-level audio features, one can also use more abstract symbolic representations such as phone n-grams[3]. For visual word recognition, one may use letter n-grams, or, as lower-level visual cues, for instance, features derived from histograms of oriented gradients (Dalal and Triggs, 2005; Linke et al., 2017). In what follows, we use binary vectors indicating the presence or absence of phonological phone or syllable n-grams.

4.1.1. Phone-Based Representations

Sublexical phone cues can be of different granularity, such as biphones and triphones. For the word *Aale* (pronunciation `al@`), the biphone cues are `#a`, `al`, `l@`, and `@#`, and the triphone cues are `#al`, `al@`, and `l@#`. The number of unique cues (and hence the dimensionality of the form vectors) increases as granularity decreases. For the present dataset, there are 931 unique biphone cues, but 4,656 triphone cues. For quadraphones, there are no less than 9,068 unique cues. Although model performance tends to become better with more unique cues, we also run the risk of overfitting. That is, the model does not generalize and thus performs worse on validation data. The choice of granularity therefore determines the balance of having a precise memory on the one hand and a productive memory on the other hand. In the simulation examples with n-phones that follow, we made use of simulated semantic vectors. Details on the different kinds of semantic vectors that can be used are presented in section 4.2.1.

Accuracy for n-phones is presented in the first three rows of **Table 5**. For the training data, comprehension accuracy is high with both triphones and quadraphones. For biphones, the small number of unique cues clearly does not offer sufficient discriminatory power to distinguish word meanings. Under strict evaluation, unsurprisingly given the large number of homophones in German noun paradigms, comprehension accuracy plummets substantially to 8, 33, and 35% for biphone, triphone, and quadraphone models, respectively. Given that there is no way to tell the meanings of homophones apart without further contextual information, we do not provide further details for strict evaluation. However, in section 4.1.1 we will address the problem of homophony by incorporating further contextual information into the model.

With regards to model accuracy for validation data, we see that overall accuracy (`val_all`) is quite low for biphones, while it remains high for both triphones and quadraphones. Closer inspection reveals that this high accuracy is mainly contributed by homophones (`val_lenient`). Since these forms are already present in the training data, a high comprehension accuracy under lenient evaluation is unsurprising. As for unseen forms (i.e., `val_newform`), quadraphones perform slightly better than triphones.

Production accuracy, presented in the right half of **Table 5**, is highly sensitive to the threshold θ used by the `learn_paths` algorithm. Given that usually only a relatively small number of cues receive strong support from a given meaning, we therefore set the threshold such that the algorithm does not need to take into account large numbers of irrelevant cues. Depending on the form and meaning representations selected, some fine-tuning is generally required to obtain a threshold value that optimally balances both accuracy and computation time. Once the threshold is fine-tuned for the training data, the same threshold is used for the validation data.

Production accuracy is similar to comprehension accuracy, albeit systematically slightly lower. Triphones and quadraphones again outperform biphones by a large margin. For the training data, triphones are somewhat less accurate than quadraphones. Interestingly, in order to predict new forms in the validation data, triphones outperform quadraphones. Clearly, triphones offer better generalizability compared to quadraphones, suggesting that we are overfitting when modeling with quadraphones as cues. Accuracy under the `val_newform` criterion is quite low,

[3] Other work (e.g., Joanisse and Seidenberg, 1999) has used slot-coding for representing phonology, but we do not think that this representation is optimal, since, for example, we are not sure how prefixation is to be modeled without hand-engineering (details in Heitmeier and Baayen, 2020).

TABLE 5 | Comprehension and production accuracy for train and validation datasets, with biphones, triphones, quadraphones, and bisyllables as cues.

	Comprehension				Production			
	Train (%)	val_all (%)	val_lenient (%)	val_newform (%)	Train (%)	val_all (%)	val_lenient (%)	val_newform (%)
Biphone	22	16	17	8	48	31	33	12
Triphone	93	88	92	51	84	64	68	21
Quadraphone	97	93	97	53	91	67	73	11
Bisyllable	99	93	99	20	95	63	69	0.3
word2vec	87	72	79	0.3	97	88	94	25

For the first four rows, we used simulated semantic vectors. For the last row, cues are triphones, and semantic vectors are word2vec *embeddings (discussed in section 4.2.2). For the* `learn_paths` *algorithm, the threshold θ was set to 0.05, 0.008, 0.005, 0.005, and 0.008, respectively.*

which is perhaps not unexpected given the uncertainty that characterizes native speakers' intuitions about the forms of novel words (McCurdy et al., 2020). In section 4.3.2, we return to this low accuracy, and consider in further detail generated novel forms and the best supported top candidates.

4.1.2. Syllable-Based Representations

Instead of using n-phones, the unit of analysis can be a combination of *n* syllables. The motivation for using syllables is that some suprasegmental features, such as lexical stress in German, are bound to syllables. Although stress information is not considered in the current simulation experiments, suprasegmental cues can be incorporated (see Chuang et al., 2020a, for an implementation).

As for n-phones, when using n-syllables, we have to choose a value for the unit size *n*. For the word *Aale*, the bi-syllable cues are `#-a`, `a-l@`, and `l@-#`, with "-" indicating syllable boundary. When unit size equals two, there are in total 8,401 unique bi-syllable cues. For tri-syllables, the total number of unique cues increases to 10,482. Above, we observed that the model was already overfitting with 9,068 unique quadraphone cues. We therefore do not consider tri-syllable cues, and only present modeling results for bi-syllable cues[4].

As shown in the fourth row of **Table 5**, comprehension accuracy (for bi-syllables) for the training data is almost error-free, 99%, the highest among all the cue representations. For the validation data, the overall accuracy is also high, 93%. This is again due to the high accuracy for the seen forms (`val_lenient` = 99%). Only one fifth of the unseen forms, however, is recognized successfully (`val_newform` = 20%). Production accuracies for the training and validation data are 95 and 63%, respectively. The model again performs well for homophones (`val_lenient` = 69%) but fails to produce unseen forms (`val_newform` = 0.3%). This extremely low accuracy is in part due to the large number of cues that appear only in the validation dataset (325 for bisyllables, but only 23 for triphones). Since such novel cues do not receive any training, words with such cues are less likely to be produced correctly. We will come back to the issue of novel cues in section 4.3.1. For now, we conclude that triphone-based form vectors are a good choice as they show a good balance of comprehension and production accuracy on training and validation data.

[4]Even though the number of bi-syllables is close to that of quadraphones, the fact that quadraphones still outnumber bi-syllables suggests that quadraphones have captured within-syllable phone collocations that are not available in bi-syllable cues. These further fine-grained cues might include, for example, consonant clusters, as in Sprache "language."

4.2. Semantic Representation

There are many ways in which words' meanings can be represented numerically. The simplest method is to use one-hot encoding (i.e., a binary vector where a single value/bit is set to one), as implemented in the naive discriminative learning model proposed by Baayen et al. (2011). One-hot encoding, however, misses out on the semantic similarities between lemmas: all lemmas receive meaning representations that are orthogonal. Instead of using one-hot encoding, semantic vectors can also be derived by turning words' taxonomies in WordNet into binary vectors with multiple bits on (details in Chuang et al., 2020a). In what follows, however, we work with real-valued semantic vectors, known as "word embeddings" in natural language processing. Semantic vectors can either be simulated, or derived from corpora using methods from distributional semantics (see e.g., Landauer and Dumais, 1997; Mikolov et al., 2013b).

4.2.1. Simulated Semantic Vectors

When corpus-based semantic vectors are unavailable, semantic vectors can be simulated. The **JudiLing** package enables the user to simulate such vectors using normally distributed random numbers for content lexemes and for inflectional functions. By default, the dimension of the semantic vectors is set to be identical to that of the form vectors.

The semantic vector for an inflected word is obtained by summing the vector of its lexeme and the vectors of all the pertinent inflectional functions. As a consequence, all vectors sharing a certain inflectional feature are shifted in the same direction in semantic space. By way of example, consider the German plural dative of *Aal* "eel," *Aalen*. We compute its semantic vector by adding the semantic vector for PLURAL and DATIVE to the lemma vector $\overrightarrow{Aal}_{lemma}$:

$$\overrightarrow{Aalen}_{dat.pl} = \overrightarrow{Aal}_{lemma} + \overrightarrow{\text{PLURAL}} + \overrightarrow{\text{DATIVE}}$$

The corresponding singular dative *Aal* can be coded as:

$$\overrightarrow{Aal}_{dat.sg.} = \overrightarrow{Aal}_{lemma} + \overrightarrow{\text{SINGULAR}} + \overrightarrow{\text{DATIVE}}$$

Alternatively, the singular form could be coded as unmarked, following a privative opposition approach:

$$\overrightarrow{Aal}_{dat.sg.} = \overrightarrow{Aal}_{lemma} + \overrightarrow{\text{DATIVE}}$$

For the remainder of the paper, we treat number as an equipollent opposition. Finally, a small amount of random noise is added to each semantic vector (*M* 0, *SD* 1; compare this to *M* 0, *SD* 4 for lexeme and inflectional vectors), as an approximation of further semantic differences in word use other than number and case [see Sinclair (1991, e.g., p.44ff.)[5], Tognini-Bonelli (2001) and further discussion below]. The results reported thus far were all obtained with simulated vectors.

It is worth noting that when working with simulated semantic vectors, the meanings of lexemes will still be orthogonal, and that as a consequence, all similarities between semantic vectors originate exclusively from the semantic structure that comes from the inflectional system.

4.2.2. Empirical Semantic Vectors

A second possibility for obtaining semantic vectors is to derive them from corpora. Baayen et al. (2019) constructed semantic vectors from the TASA corpus (Ivens and Koslin, 1991), in such a way that semantic vectors were obtained not only for lexemes but also for inflectional functions. With their semantic vectors, the semantic vector of *Aalen* can be straightforwardly constructed from the semantic vectors of *Aal*, PLURAL, and DATIVE.

However, semantic vectors that are created with standard methods from machine learning, such as *word2vec* (Mikolov et al., 2013a), *fasttext* (Bojanowski et al., 2017), or *GloVe* (Pennington et al., 2014), can also be used (albeit without semantic vectors for inflectional features; see below). In what follows, we illustrate this for 300-dimensional vectors generated with *word2vec*, trained on the German Wikipedia (Yamada et al., 2020). For representing words' forms, we used triphones.

The model in general performs well for the training data (**Table 5**). For the validation data, while the homophones are easy to recognize and produce, the unseen forms are again prohibitively difficult. Interestingly, if we compare the current results with the results of simulated vectors (cf. second row, **Table 5**), we observe that while the `train` and `val_all` accuracies are fairly comparable for the two vector types, their `val_newform` accuracies nonetheless differ. Specifically, understanding new forms is substantially more accurate with simulated vectors (51 vs. 0.3%), whereas *word2vec* embeddings yield slightly better results for producing new forms (21 vs. 25%).

To understand why these differences arise, we note, first, that lexemes are more similar to each other than is the case for simulated vectors (in which case lexemes are orthogonal), and second, that *word2vec* semantic vectors are exactly the same for each set of homophones within a paradigm, so that inflectional structure is much less precisely represented. This lack of inflectional structure may underlie the inability of the model to understand novel inflected forms correctly. Furthermore, the lack of differentiation between homophones simplifies the mapping from meaning to form, leading to more support from the semantics for the relevant triphones, which in turn facilitates synthesis by analysis.

In addition, we took the *word2vec* vectors, and reconstructed from these vectors the vectors of the lexemes and of the inflectional functions. For a given lexeme, we created its lexeme vector by averaging over the vectors of its inflectional variants[6]. For plurality, we averaged over all vectors of forms that can be plural forms. Using these new vectors, we constructed semantic vectors for a given paradigm cell by adding the semantic vector of the lexeme and the semantic vectors for its number and case values. The mean correlation between the new "analytical" *word2vec* vectors and the original empirical vectors was 0.79 ($sd = 0.076$). Apparently, there is considerable variability in how German inflected words are actually used in texts, a finding that has also emerged from corpus linguistics (Sinclair, 1991; Tognini-Bonelli, 2001). The idiosyncracies in the use of individual inflected forms renders the comprehension of an unseen, but nevertheless also idiosyncratic, inflected word form extremely difficult. From this we conclude that the small amount of noise that we added to the simulated semantic vectors is likely to be unrealistically small compared to real language use.

Interestingly, semantic similarity facilitates the production of unseen forms. A Linear Discriminant Analysis (LDA) predicting nine plural classes (the eight sub-classes presented in **Table 1** plus one "other" class) from the *word2vec* semantic vectors has a prediction accuracy of 62.7% (50.5% under leave-one-out cross validation). Conducting 10-fold cross-validation with Support Vector Machine (SVM) yields an average accuracy of 56.7%, considerably higher than the percentage of the majority choice (the -n plural class, 35.6%). Apparently, semantically similar words tend to inflect similarly. When a novel meaning is encountered in the validation set, it is therefore possible to predict to some extent its general form class. Given the similarities between LDA and regression, the same kind of information is likely captured by LDL.

4.3. Missing Forms and Missing Semantics

Evaluation on held-out data is a means for assessing the productivity of the network. However, it often happens during testing that the model is confronted with novel, unseen cues, or with novel, unseen semantics. Here, linguistically and cognitively motivated choices are required.

4.3.1. Novel Cues

For the cross-validation results presented thus far, the validation data comprise a random selection of words. As a consequence, there often are novel cues in the validation data that the model has never encountered during training. The presence of such novel cues is especially harmful for production. As mentioned in section 4.1.2, the model with bi-syllables as cues fails to produce

[5]Our approach of adding small semantic differences to individual word forms does probably not do justice to Sinclair (1991)'s view that word forms can have completely idiosyncratic meanings, since we still assume commonalities across word forms such as e.g., a shared meaning of plurality. We hope to be able to address this issue in future research.

[6]Note that these vectors are not sense-disambiguated, so that the they can cover homophonous forms from various paradigm cells.

TABLE 6 | Comprehension and production accuracy for train and validation datasets, which are split in such a way that no novel cues are present in the validation set.

	Comprehension				Production			
	Train (%)	val_all (%)	val_lenient (%)	val_newform (%)	Train (%)	val_all (%)	val_lenient (%)	val_newform (%)
Triphone	93	88	91	52	85	63	67	17
Bisyllable	99	99	99	61	95	52	52	12

Both the triphone and bisyllable models make use of simulated semantic vectors.

unseen forms, due to the large number of novel cues in the validation data.

What is the theoretical status of novel cues? To answer this question, first consider that actual speakers rarely encounter new phones or new phone combinations in their native languages. Furthermore, novel sounds encountered in loan words are typically assimilated into the speaker's native phonology[7]. Also, many cues that are novel for the model actually occur not only in the held-out nouns, but also in verbs, adjectives, and compounds that the model has no experience with. Thus, the presence of novel cues is in part a consequence of modeling only part of the German lexicon.

Since novel cues have zero weights on their efferent connections (or, equivalently, zero beta coefficients), they are completely inert for prediction. One way to address this issue is to select the held-out data with care. Instead of randomly holding out words, we make sure that in the validation data all cues are already present in the training data. We therefore split the dataset into 80% training and 20% validation data, but now making sure that there are no novel triphone cues in the validation dataset. Among the 3,629 validation words, 3,331 are homophones, and 298 are unseen forms. We note that changing the kind of cues used typically has consequences for how many datapoints are available for validation. When bi-syllables are used instead of triphones, due to the sparsity of bi-syllable cues, we have to increase the percentage of validation data to include sufficient numbers of unseen forms. Even for 65% training data and 35% validation data, the majority of validation data are homophones (98.5%), and only 76 cases represent unseen forms (with only known cues).

For the triphone model (top row, **Table 6**), for both comprehension and production, the `train`, `val_all`, and `val_lenient` accuracies are similar to the results presented previously (**Table 5**). For the evaluation of unseen forms (`val_newform`), there is only a slight improvement for comprehension (from 51 to 52%); for other datasets, the improvement can be larger. However, for production, `val_newform` becomes worse (decreasing from 21 to 17%). The reason is that even though all triphone cues of the validation words are present in the training data, they obtain insufficient support from the semantics. The solution here is to allow a small number of triphone cues with weak support (below the threshold θ) to be taken into account by the algorithm that orders triphones into words. This requires turning on the `tolerance` mode in the `learn_paths` function of the **JudiLing** package. By allowing at most two weakly supported triphones to be taken into account, production accuracy for unseen forms increases to 57%.

The bi-syllable model benefits more from the removal of novel cues in the validation data. Especially for comprehension, the accuracy of unseen forms reaches 61%, compared to 20% with random selection. For production, we observe a non-negligible improvement as well (from 0.3 to 12%). Further improvements are expected when tolerance mode is used, but given the large number of bi-syllables, this comes at considerable computation costs. In other words, bi-syllables provide a model that is an excellent memory, but a memory with very limited productivity specifically for production.

4.3.2. Unseen Semantics

In real language, speakers seldom encounter words that are completely devoid of meaning: even novel words are typically encountered in contexts which narrow down their interpretation. In the wug task, by contrast, participants are often confronted with novel words presented without any indication of their meaning, as, for instance, in the experiment on German nouns reported by McCurdy et al. (2020). Within the framework of the discriminative lexicon, this raises the question of how to model the wug task, as the model has no way to produce inflected variants without semantics.

For modeling the wug task, and comparing model performance with that of German native speakers, we begin with observing that the comprehension system generates meanings for non-words. Chuang et al. (2020c) showed that measures derived from the semantic vectors of non-words were predictive for both reaction times in an auditory lexical decision task and for non-words' acoustic durations in a reading task. In order to model the wug task, we therefore proceeded as follows:

1. We first simulated a speaker's lexical knowledge prior to the experiment by training a comprehension matrix using all the words described in section 4. Here, we made use of simulated semantic vectors.
2. We then used the resulting comprehension network to obtain semantic vectors $\boldsymbol{s}_{\text{nom.sg}}$ for the nominative singular forms of the non-words by mapping their cue vectors into the semantic space, resulting in semantic vectors $\boldsymbol{s}_{\text{nom.sg}}$.

[7] Note that such assimilation effects could be modeled using real acoustic input (i.e., audio files) with LDL-AURIS (Shafaei-Bajestan et al., 2021). Here, unseen sounds would presumably be assimilated to the closest seen sounds, similar to human performance. Of course, given sufficient training data, such a model would over time also be able to acquire the new sounds. We have, however, restricted ourselves to modeling using letter/phone representations.

3. Next, we created the production mapping from meaning to form, using not only all real words but also the non-words (known only in their nominative singular form).
4. Then, we created the semantic vectors for the plurals ($\boldsymbol{s}_{\text{nom.pl}}$) of the non-words by adding the plural vector to their nominative singular vectors while subtracting the singular vector.
5. Finally, these plural semantic vectors were mapped onto form vectors ($\hat{\boldsymbol{c}}_{\text{nom.pl}}$) using the production matrix, in combination with the `learn_paths` algorithm that orders triphones for articulation.

We applied these modeling steps to a subset of the experimental materials provided by Marcus et al. (1995) (reused by McCurdy et al., 2020), in order to compare model predictions with the results of McCurdy et al. (2020). The full materials of Marcus et al. (1995) contained non-words that were set up such that only half of them had an existing rhyme in German. We restricted ourselves to the non-words with existing rhymes, first, because non-rhyme words have many cues that are not in the training data; and second, because, as noted by Zaretsky and Lange (2015), many of the non-rhyme words have unusual orthography and thus are strange even for German speakers. Furthermore, many of the non-rhyme non-words share endings and therefore do not provide strong data for testing model predictions.

McCurdy et al. (2020) presented non-words visually and asked participants to write down their plural form. To make our simulation more comparable to their experiment, in the following we made use of letter trigrams rather than triphones. We represented words without their articles, as the wug task implemented by McCurdy et al. (2020) presented the plural article as a prompt for the plural form; participants thus produced bare plural forms. For assessing what forms are potential candidates for production, we examined the set of candidate forms, ranked by how well their internally projected meanings (obtained with the synthesis-by-analysis algorithm, see section 3), correlated with the targeted meaning $\boldsymbol{s}_{\text{nom.pl}}$. We then examined the best supported candidates as possible alternative plural forms.

The model provided a plausible plural form as the best candidate in 7 out of 12 cases. Five of these belonged to the *-en* class. A further plausible candidate was also only provided in 5 of the cases. The lack of diversity as well as the bias for *-en* plurals does not correspond to the responses given by German speakers in McCurdy et al. (2020).

Upon closer inspection, it turns out that a more variegated wug performance can be obtained by changing two parameters. First, we replaced letter trigrams by letter bigrams. This substantially reduces the number of n-grams that are present in the non-words but that do not occur in the training data. Second, we made a small but important change to how semantic vectors were simulated. The default parameter settings provided with the **JudiLing** package generate semantic vectors with the same standard deviation for both content words and inflectional features. Therefore, the magnitudes of the values in semantic vectors is very similar for content words and inflectional features. Since words are inflected for case and number, their semantic vectors are numerically dominated by the inflectional vectors. To enhance the importance of the lexemes, and to reduce the dominance of the inflectional functions, we reduced the standard deviation (by a factor of $\frac{1}{10}$) when generating the semantic vectors for number and case. As a consequence, the mean of the absolute values in the plural vector decreased from 3.25 to 0.32. (Technical details are provided in the **Supplementary Materials**.) With these two changes, the model generated a more diverse set of plural non-word candidates (**Table 7**). Model performance is now much closer to the performance of native speakers as reported by Zaretsky et al. (2013); McCurdy et al. (2020).

The model also produces some implausible plural candidates, all of which however are phonotactically legal; these are marked with an asterisk in **Table 7**. Sometimes a plural marker is interfixed instead of suffixed (e.g., Spand, Span-*en*-d; Pund, Pun-*en*-d). Almost all words have a candidate which shows double plural marking (e.g., Bral, Bral-*en*-*en*; Nuhl, Nuhl-*er*-*e*; Pind, Pind-*er*-*n*; cf. Dutch kind-*er*-*en*), or a mixture of both (e.g., Span, Span-*en*-d-*e*; Spert, Sper-*er*-t-*en*). For Klot, doubling of the -*t* can be observed, as this form is presumably more plausible in German [e.g., Motte ("moth"), Gott ("god"), Schrott ("scrap, rubbish")]. One plural has been attracted to an existing singular (Spand, Spaten-d). Apparently, by downgrading the strength (or more precisely, the L1-norm) of the semantic vectors of inflectional functions, the model moves in the direction of interfixation-like changes.

The model does not produce a single plural form with an umlaut, even though in corpora umlauted plurals are relatively frequent (see e.g., Gaeta, 2008). Interestingly, the German speakers in McCurdy (2019) also tended to avoid umlauted forms (with the exception of *Kach* → *Kächer*). Interestingly, children at the age of 5 also tend to avoid umlauts when producing plurals

TABLE 7 | First five candidates for the plural forms of non-words.

Bral	Kach	Klot	Mur	Nuhl	Pind	Pisch	Pund	Raun	Spand	Spert	Vag
Bralen	Kachen	Klot	Muren	Nuhlen	Pinden	Pischen	Punden	Raunen	*Spanend	Sperten	Vag
Bral	Kach	*Klotten	Murn	Nuhl	Pind	Pisch	*Punend	Raun	Spand	Sperte	Vagen
*Bralenen	Kacher	*Klotte	Mur	Nuhle	Pinder	Pischer	Pund	*Raunern	*Spanende	Sperter	Vage
*Bralern	Kache	*Klotter	*Murnen	*Nuhlern	Pinde	Pische	Punde	Rauner	*Spanenden	*Spererten	Vager
Braler	*Kachern	*Klieloten	Murer	*Nuhlere	*Pindern	*Pischern	*Pundene	Raune	*Spatend	*Spererte	*Vagern

Forms that are implausible as plurals are marked with an asterisk. Non-words are taken from Marcus et al. (1995).

for German non-words, but usage increases for 7-year-olds and adults (Van de Vijver and Baer-Henney, 2014).

Finally, most non-words have a plural in *-en* as one of the candidates (10 out of 12 cases), with as runners-up the *-e* plural (8 out of 12 cases), and the *-er* plural (8 out of 12). There is not a single instance of an *-s* plural, which fits well with the low prevalence (around 5%) of *-s* plurals in the experiment of McCurdy et al. (2020).

This simulation study shows that it is possible to make considerable headway with respect to modeling the wug task for German. The model is not perfect, unsurprisingly, given that we have worked with simulated semantic vectors and estimates of non-words' meanings. It is intriguing that a strong weight imposed on the stem shifts model performance in the direction of interfixation-like morphology. However, the model has no access to information about words' frequency of use, and hence is blind to an important factor shaping human learning (see section 4.5 for further discussion). Nevertheless, the model does appear to mirror the uncertainties of German speakers fairly well.

4.4. Words in Context

Thus far, we have modeled words in isolation. However, in German, case and number information is to a large extent carried by preceding determiners. In addition, in actual language use, a given grammatical case denotes one of a wide range of different possible semantic roles. The simplifying assumption that an inflectional function can be represented by a single vector, which may be reasonable for grammatical number, is not at all justified for grammatical case. In this section, we therefore explore how context can be taken into account. We first present modeling results of nouns learned together with their articles. Next, we break down grammatical cases into actual semantic functions, and show how we can begin to model the noun declension system with more informed semantic representations.

4.4.1. Articles

We first consider definite articles. Depending on gender and case, a noun can follow one of the six definite articles in German—*der, die, das, dem, den, des*. We added these articles, transcribed in DISC notation, before the nouns. Although in writing articles and nouns are separated by a space character (e.g., *der Aal*), to model auditory comprehension we removed the space character (e.g., `deral`). By adding the articles to the noun forms, the number of homophones in our dataset was reduced to a substantial extent, whereas the number of unique word forms more than doubled (from 5,427 to 12,798).

In the first set of simulations we used the same semantic vectors as we did previously for modeling isolated words. That is, the meanings of the definite articles are not taken into account in the semantic vectors, as all forms would be shifted in semantic space in the exactly the same way. After including articles, the validation data now only contained 3,982 homophones, but the number of unseen forms increased to 3,260. Using triphones as cues, we ran two models, one with simulated vectors and the other with *word2vec* semantic vectors. For simulated vectors the results (**Table 8**) are generally similar to those obtained without articles (**Table 5**). However, if we look at the evaluation of comprehension with the strict criterion (according to which recognizing a homophone is considered incorrect), without articles `val_strict` is 6%, whereas it is 34% with articles. The generalizability of the model also improves as the number of homophones in the dataset decreases. Even though there are more unseen forms in the current dataset with articles than in the original one without articles, the `val_newform` for comprehension increases by 12% from 51 to 63%.

When using *word2vec* embeddings, adding articles to form representations also improved the comprehension of unseen forms: the `val_newform` astonishingly increased from 0.3 to 58%. Without articles, homophones all shared the same form representations and exactly the same *word2vec* vectors. As a consequence, many triphone cues were superfluous and not well-positioned to discriminate between lemma or inflectional meanings. Now, with the addition of articles, the form space is better discriminated. With an increased number of triphone cues, the model is now able to predict and generalize more accurately for comprehension. However, for production, model performance is generally worse when articles have to be produced. For the training data, for instance, production accuracy drops from 97% (without articles) to 48%. This is of course unsurprising. In the simulation with articles, the semantic representations remain the same, but now identical semantic vectors have to predict more variegated triphone vectors. The learning task has become more challenging, and inevitably resulted in less accurate performance. Replacing the contextually unaware *word2vec* vectors by contextually aware vectors obtained using language models such as BERT (Corbett et al., 2019; Miaschi and Dell'Orletta, 2020) should alleviate this problem.

We can test the model on more challenging data by including indefinite articles *(ein, eine, einem, einen, einer, eines)*, and

TABLE 8 | Comprehension and production accuracy for train and validation datasets with articles.

	Comprehension				Production			
	Train (%)	**val_all (%)**	**val_lenient (%)**	**val_newform (%)**	**Train (%)**	**val_all (%)**	**val_lenient (%)**	**val_newform (%)**
Simulated	94	76	92	63	81	37	57	19
word2vec	91	69	81	58	48	14	28	1
def + indef	94	80	93	64	82	40	60	15

All three simulations use triphones as cues. The first two rows present results with simulated vectors and word2vec *embeddings as semantic representations. The simulation presented in the bottom row also makes use of simulated vectors, but includes both definite and indefinite articles.*

creating two additional semantic vectors, one for definiteness and one for indefiniteness. This doubles the size of our dataset: half of the words are preceded by definite articles, and the other half by indefinite articles. However, because German indefinite articles are restricted to singular forms, only indefinite singular forms are preceded by indefinite articles. On the meaning side, the $\overrightarrow{\text{DEFINITE}}$ vector is added to the semantic vectors of words preceded by definite articles, and the $\overrightarrow{\text{INDEFINITE}}$ vector is added to vectors for words preceded by either indefinite articles in the singular, or no article in the plural.

The validation data of this dataset confronts the model with in total 3,982 homophones and 3,260 unseen forms. Homophones comprise slightly more words with indefinite articles (57%) whereas unseen forms comprise slightly more definite articles (59%). The results, presented in the bottom row of **Table 8**, are very similar to those with only definite articles (top row). Closer inspection of the results for the validation data shows that for comprehension, accuracies do not differ much across definite and indefinite forms. For production, however, especially for unseen forms, the accuracy for definite articles is twice higher than that for indefinite articles (20 and 9%, averaging out to 15%). This is a straightforward consequence of the much more diverse realizations of indefinite nouns. For definite nouns, the possible triphone cues at the first two positions in the word are always limited to the triphone cues of the six definite articles. For indefiniteness, however, in addition to the six indefinite articles, initial triphone cues also originate from words' stems—indefinite plural forms are realized without articles. The mappings for production are thus faced with a more complex task for indefinites, and the model is therefore more likely to fail on indefinite forms.

4.4.2. Semantic Roles

The simulation studies thus far suggest it is not straightforward to correctly comprehend a novel German word form in isolation, even when articles are provided. This is perhaps not that surprising, as in natural language use, inflected words appear in context, and usually realize not some abstract case ending, but a specific semantic role (also called *thematic role*, see e.g., Harley, 2010). For example, a word in the nominative singular might express a theme, as *der Apfel* in *Der Apfel fällt vom Baum.* ("The apple falls from the tree"), or it might express an agent as *der Junge* in *Der Junge isst den Apfel.* ("The boy eats the apple."). Exactly the same lemma, used with exactly the same case and number, may still realize very different semantic roles. Consider the two sentences *Ich bin bei der Freundin* ("I'm at the friend's") and *Ich gebe der Freundin das Buch* ("I give the book to the friend"). *der Freundin* is dative singular in both cases, but in the first sentence, it expresses a location while in the second it represents the beneficiary or receiver. Semantic roles can also be reflected in a word's form, independently of case markers. For example, German nouns ending in *-er* are so-called "Nomina Agentis" (Baeskow, 2011). As pointed out by Blevins (2016), case endings are no more (or less) than markers for the intersection of form variation and a distribution class of semantic roles. Since within the framework of the DLM, the aim is to provide mappings between form and meaning, a case label is not a proper representation of a word's actual meaning. All it does is specify a range of meanings that the form can have, depending on context. Therefore, even though we can get the mechanics of the model to work with case specifications, doing so clashes with the "discriminative modeling" approach. In what follows, we therefore implement mappings with somewhat more realistic semantic representations of German inflected nouns.

Our starting point is that in German, different cases can realize a wide range of semantic roles. For our simulations, we restrict ourselves to some of the most prominent semantic roles for each case (**Table 9**). Even though these clearly do not reflect the full richness of the semantics of German cases, they suffice for a proof-of-concept simulation.

In order to obtain a dataset with variegated semantic roles, we expanded the previous dataset, with each word form (including its article) appearing with a specification of its semantic role, according to the probabilities presented in **Table 9**. The resulting dataset had 45,605 entries, which we randomly split into 80% training data and 20% validation data. For generating the semantic matrix, we again used number, but instead of a case label, we provided the *semantic role* as inflectional feature. Comprehension accuracy on this data is comparable to the previous simulations: 89% for the training data `train`, and 85% `val_lenient`. Comprehension accuracy on the validation set drops dramatically when we use strict evaluation (4% accuracy). This is unsurprising given that it is impossible for the model to know which semantic role is intended when only being exposed to the word form and its article in isolation, without syntactic context. Production accuracy is likewise comparable to previous simulations with `train` at 78% and `val_lenient` at 61% (`val_newform` 25%). This simple result clarifies that in order to properly model German nouns, it is necessary to take the syntactic context in which a noun occurs into account. Future research will also have to face the challenge of integrating words' individual usage profiles into the model (see also section 4.2.1 above).

4.5. Incremental Learning vs. the End-State of Learning

In the simulation studies presented thus far, we made use of the regression method to estimate the mappings between form and meaning. The regression method is strictly type based: the data on which a model is trained and evaluated consists of all unique combinations of form vectors $\boldsymbol{c}$ and semantic vectors $\boldsymbol{s}$. In this

TABLE 9 | Probabilities of semantic roles by cases in the German noun system.

Case	Semantic roles
Nominative	Agent (50%), theme (40%), patient (10%)
Genitive	Possessive (90%), partitive (10%)
Dative	Beneficiary (50%), location (50%)
Accusative	Patient (40%), motion (30%), experiencer (30%)

Semantic roles are informed by Schulz and Griesbach (1981). Percentages are simulated and do not necessarily reflect corpus-frequencies of the respective semantic role.

respect, the regression method is very similar to models such as AML, MBL, MGL, and to statistical analyses with the GLM or recursive partioning methods. However, word types (understood as unique sets $\{\boldsymbol{c}, \boldsymbol{s}\}$) are not uniformly distributed in language, and there is ample evidence that the frequencies with which word types occur co-determines lexical processing (see e.g., Baayen et al., 1997, 2007, 2016; Tomaschek et al., 2018). While some formal theorists flatly deny that word frequency effects exist for inflected words (Yang, 2016), others have argued that there is no problem with integrating frequency of use into formal theories of the lexicon (Jackendoff, 1975; Jackendoff and Audring, 2019), and yet others have argued that it is absolutely essential to incorporate frequency into any meaningful account of language in action (Langacker, 1987; Bybee, 2010).

Within the present approach, effects of frequency of occurrence can be incorporated seamlessly by using incremental learning instead of the end-state of learning as defined by the regression equations (see Danks, 2003; Evert and Arppe, 2015; Shafaei-Bajestan et al., 2021, for the convergence over learning time of incremental learing to the regression end-state of learning). We illustrate this for our German nouns dataset with number and semantic role as crucial constructors of simulated semantic vectors.

We begin with noting that word forms usually do not instantiate all possible semantic roles equally frequently. For instance, a word such as *der Doktor* ("doctor") will presumably occur mostly as *agent* in the nominative singular form, rather than as *theme* or *patient*. If the model is informed about the probability distributions of semantic roles in actual language use (both in language generally, and lexeme-specific), it may be expected to make more informed decisions when coming across new forms, for instance, by opting for the best match given its past experience.

Incremental learning with the learning rule of Widrow-Hoff makes it possible to start approximating human word-to-word learning as a function of experience. As a consequence, the more frequent a word type occurs in language use, the better it can be learned: practice makes perfect. This sets the following simulation study apart from models such as proposed by McCurdy et al. (2020) or Belth et al. (2021), who base their training regimes strictly on types rather than tokens.

In the absence of empirical frequencies with which combinations of semantic roles and German nouns co-occur, we simulated frequencies of use[8]. To do so, we proceeded as follows. First, we collected token frequencies for all our word forms from CELEX. Next, we assigned an equal part $freq_p$ of this frequency count to each case/number cell realizing this word form. Third, for each paradigm cell, we randomly set to zero some semantic roles, drawing from a binomial distribution with $n = 1$, $p = \frac{1}{K}$, with K the number of semantic roles for the paradigm cell (see **Table 9**). In this way, on average, one semantic role was omitted per paradigm cell. Finally, given a proportional frequency count $freq_p$, the semantic roles associated with a paradigm cell received frequencies proportional to the percentages given in **Table 9**. Further details on this procedure are available in the **Supplementary Materials**, a full example can be found in **Table 10**.

Having obtained simulated frequencies, we proceeded by randomly selecting 274 different lemmas (1,274 distinct word forms with definite articles included), in order to keep the size of the simulation down — simulating with the Widrow-Hoff rule is computationally expensive. The total number of tokens in this study was 4,470. For the form vectors, we used triphones. The dimension of the simulated semantic vectors was identical to that of the cue vectors. As before, the data was split into 80% training and 20% validation data. We followed the same procedure as in the previous experiments, but instead of computing the mapping matrices in their closed form (i.e., end-state) solution, we used incremental learning.

While for comprehension, the implementation of the learning algorithm is relatively straightforward, this is not the case for production. The `learn_paths` algorithm calculates the support for each of the n-grams, for each possible position in a word. In the current implementation of **JudiLing**, the calculation of positional support is not implemented for incremental learning. Therefore, we do not consider incremental learning of production here.

Comprehension accuracy was similar to that observed for previous experiments. Training accuracy when taking into account homophones was 85%, validation accuracy on the full data was 79% (`val_lenient`). Without considering homophones, validation accuracy drops substantially (`val_strict` 7%). This is unsurprising given that from the form alone it is impossible to predict the proper semantic role.

The accuracy of the model's predictions is also closely linked to the frequencies with which words' form+role combinations are encountered in the training data. If a word's form+role combination is very frequent, it is learned better. **Figure 2** presents the correlations of words' predicted and targeted semantic vectors against their frequency of occurrence. The left panel presents the results for the incrementally learned

TABLE 10 | Example of simulated frequencies for combinations of case and semantic role for the word form "Adresse."

Word form	Lemma	Case	Number	Semantic role	Form frequency	Form+role frequency
Adresse	Adresse	Nominative	Singular	Agent	137	20
Adresse	Adresse	Nominative	Singular	Theme	137	16
Adresse	Adresse	Nominative	Singular	Patient	137	0
Adresse	Adresse	Genitive	Singular	Possessive	137	0
Adresse	Adresse	Genitive	Singular	Partitive	137	35
Adresse	Adresse	Dative	Singular	Beneficiary	137	18
Adresse	Adresse	Dative	Singular	Location	137	18
Adresse	Adresse	Accusative	Singular	Patient	137	0
Adresse	Adresse	Accusative	Singular	Motion	137	0
Adresse	Adresse	Accusative	Singular	Experiencer	137	35

[8]Though there are several semantic role labelers available for English [e.g., arising from the CoNLL-2004 and 2005 Shared Tasks (https://www.cs.upc.edu/~srlconll/home.html)], there are—to our knowledge—currently no suitable taggers for German.

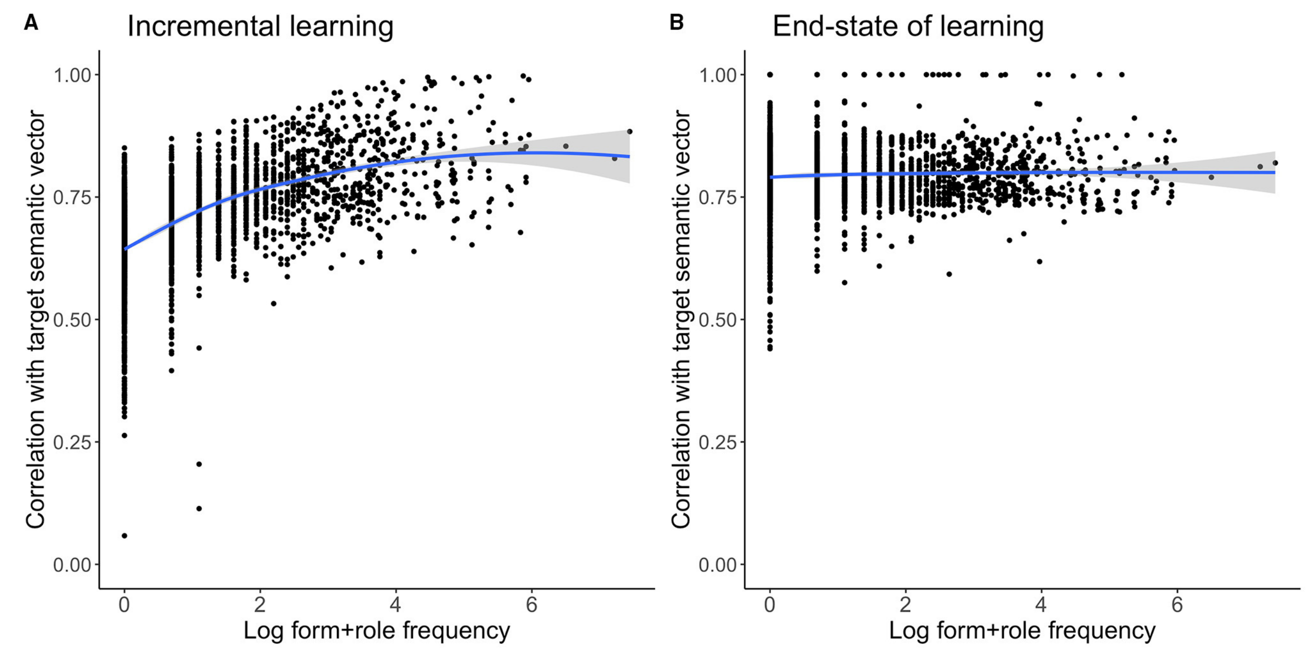

FIGURE 2 | Correlation between the simulated frequency and correlation of the predicted semantic vector with its target. Generally, the more frequent a word form is, the more accurate its semantic vector is predicted. The blue line indicates a loess smooth with a 0.95 confidence interval. **(A)** Incremental learning. **(B)** End-state of learning.

model, the right panel for the end-state of learning. Clearly, after incremental learning the model predicts the semantics of more frequent form+role combinations more accurately. For the end-state of learning on the other hand, no such effect can be observed. These results clearly illustrate the difference between a token-based model and a typed-based model.

The effect of frequency of use on the kind of errors made by the model is also revealing. We zoom in on those cases where the model was able to correctly identify the lemma and paradigm cell of the word form, but did not get the semantic role right. **Figure 3** provides scatterplots graphing the number of times a semantic role was (incorrectly) understood against the frequency of the form's semantic role, cross-classified by training method (incremental, left panels; end-state of learning, right panels) and by evaluation set (top panels: training data, bottom panels: validation data). For incremental learning, there is a positive correlation between the number of times a semantic role was (incorrectly) identified and the frequency of the semantic role in the training data. Note that the relation is not linear, but curvilinear. A linear relation would have implied that a fixed proportion of word forms would be incorrectly recognized, across all semantic roles. What we see, by contrast, is that greater exposure in language use has an increasingly detrimental effect on learning, with more probable semantic roles being over-identified. Importantly, for the end-state of learning, this curvilinear effect of frequency on learning is absent, with the PATIENT role representing an atypical outlier. This outlier status is due to the *patient* semantic role being realized by two cases: *nominative* and *accusative*. As a consequence, it is not only frequent, but it is also predicted by many more different cues (especially cues from the articles) than is the case for other semantic roles.

In other words, with incremental learning, strong frequency effects emerge, hand in hand with overgeneralization of semantic roles (the study by Ramscar et al., 2013 makes the same point for irregular English noun plurals). By contrast, for the end-state of learning, such effects are absent. Mathematically, this makes sense: as experience (i.e., volume of training data) goes to infinity, all forms are learned an infinite number of times, and frequency is no longer distinctive.

With incremental learning, it is also possible to follow the learning trajectory of the model. **Figure 4** presents this trajectory at 10 evaluation points. Learning proceeds rapidly during the first 15,000 learning events and slows down afterwards. Validation accuracy `val_lenient` closely follows training accuracy, which is a straightforward consequence of the large numbers of homophones. `val_newforms` on the other hand stays relatively low, in accordance with the semi-productivity of the German declension system.

Note that in this simulation we only pass through the data once. If a word form has a form+role frequency of 1, it is only seen a single time during training. As such, it is not possible for the model to reach accuracies as high as at the end-state of learning (indicated as dots in **Figure 4**), which would be reached eventually after an infinite number of passes through the data (Danks, 2003; Evert and Arppe, 2015; Shafaei-Bajestan et al., 2021). This sets our approach apart from deep learning, where models are trained on many iterations through the data set until the loss function reaches a local minimum. Whereas such a procedure makes sense

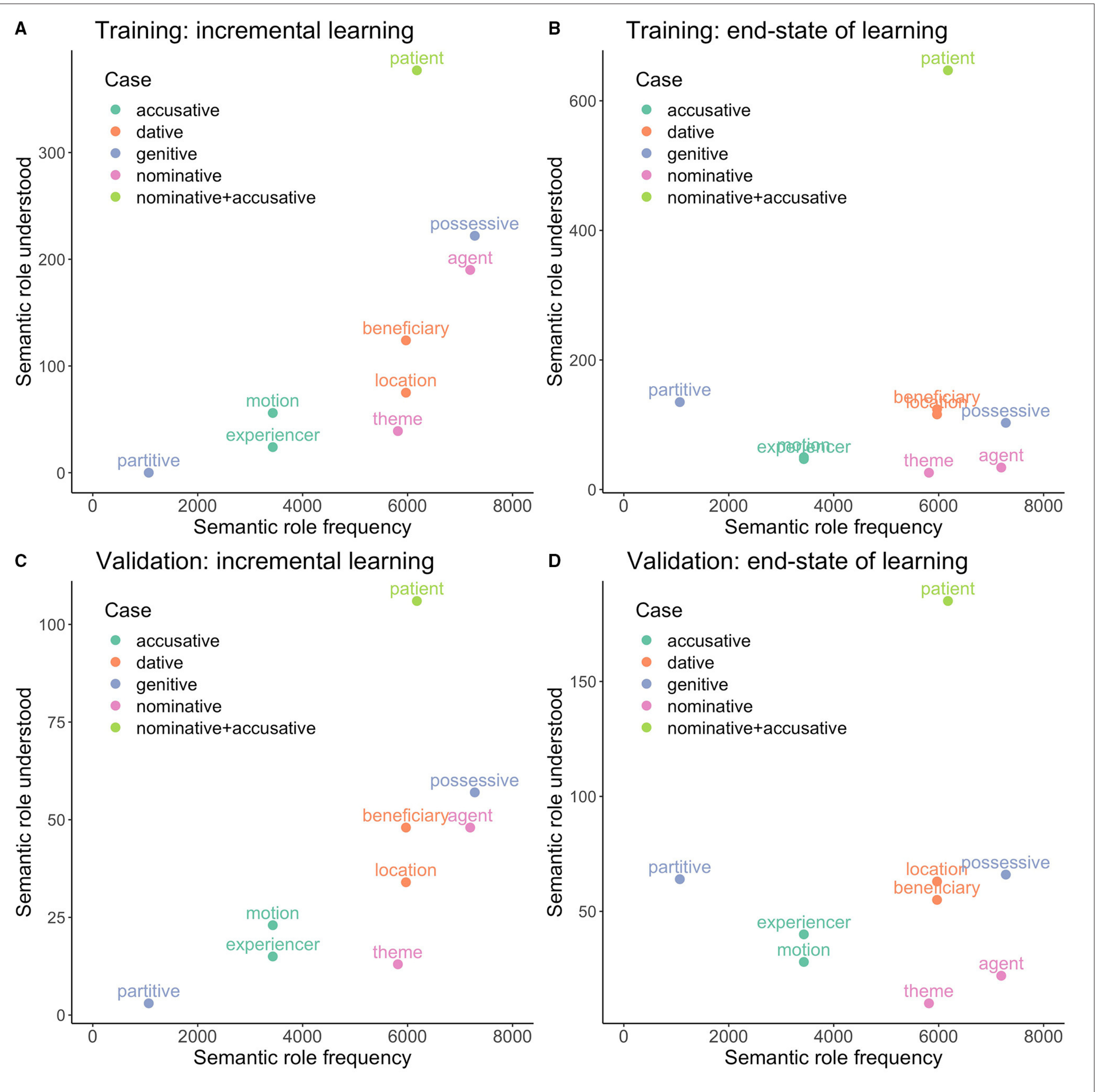

FIGURE 3 | Counts of overgeneralization errors of semantic roles for training (top) and test data (bottom), for incremental learning (left) and the end-state of learning (right), conditional on the model having understood lexeme, number, and case correctly. **(A)** Training: incremental learning. **(B)** Training: end-state of learning. **(C)** Validation: incremental learning. **(D)** Validation: end-state of learning.

for language engineering, it does not make sense for human learning: we don't relive the same exposure to data multiple times, and for healthy people, there is no point in learning after which performance degrades. For instance, vocabulary learning is a continuous process straight into old age (Keuleers et al., 2015).

Note finally, that even though incremental learning is certainly superior for modeling realistic frequency effects, there are also cases where the end-state of learning can be the preferred choice of modeling. Incremental learning is much more computationally expensive which becomes a problem especially if the training set is large and frequencies are high. Moreover, in cases where simulated speakers are expected to have learned a phenomenon well enough, the end-state simulation might be sufficient.

In summary, the present modeling framework offers the possibility to approximate incremental human learning and the consequences of frequency of exposure for learning in a cognitively motivated way (see also Chuang et al., 2020a, for learning in a multilingual setting).

4.6. Model Complexity

LDL mappings are costly in the number of connection weights, or equivalently, the number of beta coefficients. For example, the mapping matrix **F** for the dataset discussed in section 4.4.2 has 35 million weights (5,913 × 5,913 dimensions), rendering it much more costly in terms of the number of weights than deep-learning models, models such as AML, MBL, and recursive partitioning methods.

Inspection of the distribution of weights, however, clarifies that many weights are very close to zero. Apparently, many cues have low discriminative value. This suggests their connections can be pruned without seriously affecting model performance. This can be tested by selecting a threshold ϑ and setting all absolute values in the mapping matrix that fall below this threshold to zero. **Figure 5** shows, for varying ϑ, that up to 40% of the small weights can be pruned without substantially impacting model performance with end-state of learning. As neural pruning is part and parcel of human cortical development (see e.g., Gogtay et al., 2004), an interesting topic for further research is to integrate incremental learning with neural pruning of uninformative connections.

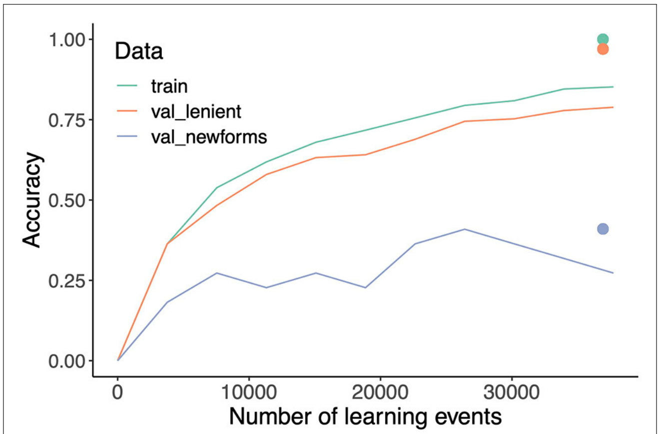

FIGURE 4 | Comprehension accuracy over the course of learning. After a very fast increase in accuracy over the first 15,000 learning events, the amount of learning levels off. Points indicate the accuracy at the end-state of learning which the incremental model would reach eventually after an infinite number of learning events.

5. DISCUSSION

In this study, we illustrated the methodological consequences of the many different choices that have to be made when modeling morphological systems within the discriminative lexicon framework, using LDL as modeling engine. We illustrated these choices for the German noun system. This system is "degenerate," as many of its paradigm cells share the same word forms (homophones). This system is also in many ways irregular: a noun's declension class can often not be fully predicted by its phonology, gender, or semantics (Köpcke, 1988). The results we obtained with LDL reflect this complexity. The model can learn word forms very well, achieving accuracies of more than 90% on both comprehension and production when evaluated on training data. It can also generalize very well to new paradigm cells when it comes to word forms it has already seen, thanks to the ubiquitous homophony that characterizes German noun paradigms. However, it also mirrors the unpredictability of German inflections when it comes to word forms it has not seen before. Accuracies for both comprehension and production suffer. Nevertheless, the model shows some semi-productivity and succeeds in generalizing to many of the sub-regularities found in the German noun system (Wunderlich, 1999), reaching accuracies of 50% on comprehension and 20% on production. Since German speakers encounter similar problems with new German word forms, as has been demonstrated in various wug studies (Zaretsky et al., 2013; McCurdy et al., 2020), our model properly exhibits the limitations that are also characteristic for native speakers.

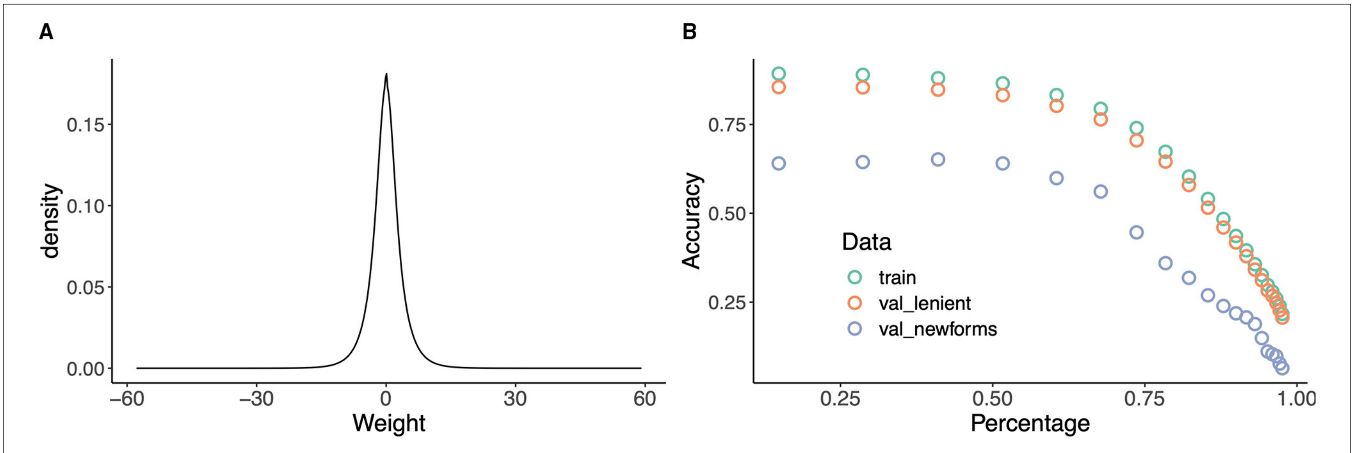

FIGURE 5 | (A) Distribution of weights in the mapping matrix from form to meaning for the dataset with semantic roles. **(B)** Accuracy of the end-state model as a function of the proportion of connection weights close to zero are pruned. About 40% of the weights can be set to zero without seriously affecting the performance of the model.

In this study, we also probed the modeling of German nouns in context. The rampant homophony that characterizes German noun paradigms is a straightforward consequence of considering words in isolation. The amount of homophony can be substantially reduced by including articles, in which case the model still performs well. In context, case-inflected words typically do not realize a specific case meaning, but rather a specific semantic role. As case endings typically do not stand in a one-to-one relation with semantic roles, we also examined to what extent we can make the model more realistic by replacing semantic vectors for cases with semantic vectors for a variety of semantic roles. For the simulated dataset that we constructed, the model again performed well.

For this dataset, we also demonstrated how the consequences of frequency of occurrence can be brought into the model, namely, by moving from the end-state of learning (estimated with regression) to incremental learning using the Widrow-Hoff learning rule.

One limitation of the present approach is that most models have been using very high-level abstract representations. The phone-based representation, for example, involves tremendous simplifications compared to real speech, as variability in pronunciations is enormous (Ernestus et al., 2002; Johnson, 2004; Shafaei-Bajestan et al., 2021). On the meaning side, traditional case labels have no intrinsic semantic content, and although we can replace cases with semantic roles, these too are still too simplistic to be able to capture the full complexity of the semantics of words in context. However, we note that even with the present high-level representations, the model can still generate useful predictions. We note here that various other studies carried out within this framework have successfully modeled a range of aspects of human lexical processing (see Chuang and Baayen, 2021, for further details). In summary, even though the current framework undoubtedly misses out on a great number of nuanced but potentially informative features of forms and meanings in real language use, it can still serve as a useful linguistic tool to explore the strengths and weaknesses of morphological systems.

A question that inevitably arises in the context of computational modeling is how cognitively plausible a model is. In the introduction, we called attention to the distinction made by Breiman (2001) between statistical models and machine learning models. We view LDL primarily as a statistical model that enables us to clarify, at a functional level of analysis, quantitative structure in the lexicon as well as understand the challenges a language processing system faces, without claiming that our model is cognitive reality. However, it is worth noting that LDL helps incorporate biologically and psychologically plausible learning into linguistic theory by making use of the principle of error-driven learning (when training the model incrementally). The very simple learning rules of Widrow-Hoff and Rescorla-Wagner have been shown to excellently explain phenomena from a range of domains in e.g., biology and psychology (see e.g., Rescorla, 1988; Schultz, 1998; Marsolek, 2008; Oppenheim et al., 2010; Trimmer et al., 2012).

It is possible to take the model as point of departure for addressing questions at the level of neural organization in the brain. For instance, Heitmeier and Baayen (2020) were interested in clarifying whether the framework of the discriminative lexicon properly predicts the dissociations of form and meaning observed for aphasic speakers producing English regular and irregular past-tense forms, following Joanisse and Seidenberg (1999). They took the unordered banks of units of form and meaning (the column dimensions of the $\boldsymbol{C}$ and $\boldsymbol{S}$ matrices) and projected them onto two-dimensional surfaces approximating, however crudely, cortical maps. This made it possible to lesion the network in a topologically cohesive way, rather than by randomly taking out connections across the whole network. For projection, they made use of an algorithm from physics (http://www.schmuhl.org/graphopt/) for displaying graphs, but temporal self-organizing maps (TSOMs, Ferro et al., 2011; Chersi et al., 2014) offer a much more fine-grained and principled way for modeling morphological organization that builds on principles of error-driven learning.

Deep learning algorithms provide the analyst with powerful modeling tools, but it seems they are too powerful (see e.g., McCurdy et al., 2020) for understanding not only the strengths but also the weaknesses and the frailties of human lexical memory and lexical processing. However, linguistic models are in a different way also too powerful on the one hand, and too underspecified on the other hand. Paradigms are typically constructed to accommodate any contrast between forms and inflectional functions, even when a contrast is attested only for a few forms in the language. The result is an overabundance of homophones, which are severely underspecified with respect to their real meanings in actual language use (such as their semantic roles). Furthermore, in actual language use, inflected forms can occur at very different frequencies and some are never encountered at all (Karlsson, 1986; Janda and Tyers, 2018), which in turn has demonstrable consequences for lexical processing (Lõo et al., 2018)[9]. An interesting challenge for further research is to clarify how different degrees of paradigm economy (Ackerman and Malouf, 2013) are reflected in the matrices that define mappings between form and meaning within the framework of the discriminative lexicon.

In this study, we have provided an overview of the many choice points that arise in modeling with LDL, each of which requires knowledge of morphology and morphological theory. The implications of our approach to psycho-computational modeling for morphological theory depends on the specifics of a given specific theory of morphology. Our approach is broadly consistent with usage-based approaches to morphology (Bybee, 1985, 2010), and with Word and Paradigm Morphology (Blevins, 2016). It is less clear whether our modeling approach is informative for theories that are only interested in defining possible words. With this methodological study, we have shed some light on the many questions and issues that do not arise in formal theories of morphology, but that have to be addressed

[9]Note that we do not claim that rare inflected word forms cannot be processed. Generally, the more regular a morphological system, the more easily the model can predict new forms (e.g., in Estonian, Chuang et al., 2020b), while in semi-productive cases such as German or Maltese (Nieder et al., 2021) generalization is much more difficult.

in a linguistically informed way when the goal of one's theory is to better understand, and predict, in all its complexity, human lexical processing across comprehension and production.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/zrw2v/.

AUTHOR CONTRIBUTIONS

RB, Y-YC, and MH: conception and design of the study and writing. MH and Y-YC: computational implementation and modeling. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by ERC-WIDE (European Research Council—Wide Incremental learning with Discrimination nEtworks), grant number 742545.

ACKNOWLEDGMENTS

We acknowledge support by Open Access Publishing Fund of University of Tübingen. We thank Ruben van de Vijver and the two reviewers for their helpful comments on earlier versions of this manuscript.

SUPPLEMENTARY MATERIAL

Supplementary Materials are available at https://osf.io/zrw2v/.

REFERENCES

Ackerman, F., and Malouf, R. (2013). Morphological organization: the low conditional entropy conjecture. *Language* 89, 429–464. doi: 10.1353/lan.2013.0054

Albright, A., and Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition* 90, 119–161. doi: 10.1016/S0010-0277(03)00146-X

Arndt-Lappe, S. (2011). Towards an exemplar-based model of stress in English noun-noun compounds. *J. Linguist*. 47, 549–585. doi: 10.1017/S0022226711000028

Arnold, D., Tomaschek, F., Lopez, F., Sering, T., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLoS ONE* 12:e0174623. doi: 10.1371/journal.pone.0174623

Baayen, R. H., Chuang, Y.-Y., and Blevins, J. P. (2018). Inflectional morphology with linear mappings. *Mental Lexicon* 13, 232–270. doi: 10.1075/ml.18010.baa

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019:4895891. doi: 10.1155/2019/4895891

Baayen, R. H., Dijkstra, T., and Schreuder, R. (1997). Singulars and plurals in Dutch: evidence for a parallel dual route model. *J. Mem. Lang*. 36, 94–117. doi: 10.1006/jmla.1997.2509

Baayen, R. H., Milin, P., Filipović Durdević, 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

Baayen, R. H., Milin, P., and Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology* 30, 1174–1220. doi: 10.1080/02687038.2016.1147767

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database [CD ROM]*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baayen, R. H., Wurm, L. H., and Aycock, J. (2007). Lexical dynamics for low-frequency complex words. A regression study across tasks and modalities. *Mental Lexicon* 2, 419–463. doi: 10.1075/ml.2.3.06baa

Baeskow, H. (2011). *Abgeleitete Personenbezeichnungen im Deutschen und Englischen: kontrastive Wortbildungsanalysen im Rahmen des minimalistischen Programms und unter Berücksichtigung sprachhistorischer Aspekte*, Vol. 62. Berlin; Boston, MA: Walter de Gruyter.

Behrens, H., and Tomasello, M. (1999). And what about the chinese? *Behav. Brain Sci*. 22, 1014–1014. doi: 10.1017/S0140525X99222224

Belth, C., Payne, S., Beser, D., Kodner, J., and Yang, C. (2021). The greedy and recursive search for morphological productivity. *arXiv [Preprint]. arXiv:2105.05790*.

Bierwisch, M. (2018). *Syntactic Features in Morphology: General Problems of So-Called Pronominal Inflection in German*. Berlin; Boston, MA: De Gruyter Mouton.

Blevins, J. P. (2016). *Word and Paradigm Morphology*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199593545.001.0001

Boersma, P. (1998). *Functional Phonology*. The Hague: Holland Academic Graphics.

Boersma, P., and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguist. Inq*. 32, 45–86. doi: 10.1162/002438901554586

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist*. 5, 135–146. doi: 10.1162/tacl_a_00051

Breiman, L. (2001). Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat. Sci*. 16, 199–231. doi: 10.1214/ss/1009213726

Breiman, L., Friedman, J. H., Olshen, R., and Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, CA: Wadsworth International Group.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526

Bybee, J. L. (1985). *Morphology: A Study of the Relation Between Meaning and Form*. Amsterdam: Benjamins. doi: 10.1075/tsl.9

Cahill, L., and Gazdar, G. (1999). German noun inflection. *J. Linguist*. 35, 1–42. doi: 10.1017/S0022226798007294

Chersi, F., Ferro, M., Pezzulo, G., and Pirrelli, V. (2014). Topological self-organization and prediction learning support both action and lexical chains in the brain. *Top. Cogn. Sci*. 6, 476–491. doi: 10.1111/tops.12094

Chuang, Y.-Y., and Baayen, R. H. (2021). "Discriminative learning and the lexicon: NDL and LDL," in *Oxford Research Encyclopedia of Linguistics*, ed M. Aronoff (Oxford: Oxford University Press).

Chuang, Y.-Y., Bell, M., Banke, I., and Baayen, R. H. (2020a). Bilingual and multilingual mental lexicon: a modeling study with Linear Discriminative Learning. *Lang. Learn*. 71, 219–292. doi: 10.31234/osf.io/adtyr

Chuang, Y.-Y., Loo, K., Blevins, J. P., and Baayen, R. H. (2020b). "Estonian case inflection made simple: a case study in word and paradigm morphology with linear discriminative learning," in *Complex Words Advances in Morphology*, eds L. Körtvélyessy and P. Štekauer (Cambridge: Cambridge University Press), 119–141. doi: 10.1017/9781108780643.008

Chuang, Y.-Y., Vollmer, M.-l., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., and Baayen, R. H. (2020c). The processing of pseudoword form and meaning in production and comprehension: a computational modeling approach using linear discriminative learning. *Behav. Res. Methods* 53, 945–976. doi: 10.3758/s13428-020-01356-w

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behav. Brain Sci*. 22, 991–1013. doi: 10.1017/S0140525X99002228

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: dual-route and parallel-distributed-processing approaches. *Psychol. Rev*. 100:589. doi: 10.1037/0033-295X.100.4.589

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

Corbett, G. G., Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1* (Minneapolis, MN: Long and Short Papers), 4171–4186. doi: 10.18653/v1/N19-1423

Daelemans, W., Berck, P., and Gillis, S. (1995). "Linguistics as data mining: Dutch diminutives," in *CLIN V, Papers from the 5th CLIN Meeting*, eds T. Andernach, M. Moll, and A. Nijholt (Enschede: Parlevink), 59–71.

Daelemans, W., and Van den Bosch, A. (2005). *Memory-Based Language Processing*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486579

Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (Los Alamitos, CA), 886–893. doi: 10.1109/CVPR.2005.177

Danks, D. (2003). Equilibria of the Rescorla-Wagner model. *J. Math. Psychol*. 47, 109–121. doi: 10.1016/S0022-2496(02)00016-0

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

Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cogn. Sci*. 33, 547–582. doi: 10.1111/j.1551-6709.2009.01023.x

Ernestus, M., and Baayen, R. H. (2003). Predicting the unpredictable: interpreting neutralized segments in Dutch. *Language* 79, 5–38. doi: 10.1353/lan.2003.0076

Ernestus, M., Baayen, R. H., and Schreuder, R. (2002). The recognition of reduced word forms. *Brain Lang*. 81, 162–173. doi: 10.1006/brln.200 1.2514

Evans, R., and Gazdar, G. (1996). DATR: a language for lexical knowledge. *Comput. Linguist*. 22, 167–216.

Evert, S., and Arppe, A. (2015). "Some theoretical and experimental observations on naive discriminative learning," in *Proceedings of the 6th Conference on Quantitative Investigations in Theoretical Linguistics (QITL-6)* (Tübingen).

Ferro, M., Marzi, C., and Pirrelli, V. (2011). A self-organizing model of word storage and processing: implications for morphology learning. *Lingue e Linguaggio* 10, 209–226.

Finkel, R., and Stump, G. (2007). Principal parts and morphological typology. *Morphology* 17, 39–75. doi: 10.1007/s11525-007-9115-9

Gaeta, L. (2008). Die deutsche Pluralbildung zwischen deskriptiver Angemessenheit und Sprachtheorie. *Z. German. Linguist*. 36, 74–108. doi: 10.1515/ZGL.2008.005

Gaskell, M. G., and Marslen-Wilson, W. (1997). Integrating form and meaning: a distributed model of speech perception. *Lang. Cogn. Process*. 12, 613–656. doi: 10.1080/016909697386646

Goebel, R., and Indefrey, P. (2000). "A recurrent network with short-term memory capacity learning the German's plural," in *Models of Language Acquisition: Inductive and Deductive Approaches*, eds P. Broeder and J. Murre (Oxford: Oxford University Press), 177–200.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A*. 101, 8174–8179. doi: 10.1073/pnas.0402680101

Goldsmith, J., and O'Brien, J. (2006). Learning inflectional classes. *Lang. Learn. Dev*. 2, 219–250. doi: 10.1207/s15473341lld0204_1

Haapalainen, M., and Majorin, A. (1994). *Gertwol: Ein System zur Automatischen Wortformerkennung Deutscher Wörter*. Lingsoft, Inc.

Harley, H. (2010). "Thematic roles," in *The Cambridge Encyclopedia of the Language Sciences*, ed P. Hogan (Cambridge: Cambridge University Press), 861–862.

Harm, M. W., and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. Rev*. 111, 662–720. doi: 10.1037/0033-295X.111.3.662

Heitmeier, M., and Baayen, R. H. (2020). Simulating phonological and semantic impairment of English tense inflection with Linear Discriminative Learning. *Mental Lexicon* 15, 385–421. doi: 10.1075/ml.20003.hei

Indefrey, P. (1999). Some problems with the lexical status of nondefault inflection. *Behav. Brain Sci*. 22:1025. doi: 10.1017/S0140525X99342229

Ivens, S., and Koslin, B. (1991). *Demands for Reading Literacy Require New Accountability Methods*. Brewster, NY: Touchstone Applied Science Associates.

Jackendoff, R., and Audring, J. (2019). *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press. doi: 10.1093/oso/9780198827900.001.0001

Jackendoff, R. S. (1975). Morphological and semantic regularities in the lexicon. *Language* 51, 639–671. doi: 10.2307/412891

Janda, A. L., and Tyers, M. F. (2018). Less is more: why all paradigms are defective, and why that is a good thing. *Corpus Linguist. Linguist. Theory*. 17, 109–141. doi: 10.1515/cllt-2018-0031

Joanisse, M. F., and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: a connectionist model. *Proc. Natl. Acad. Sci. U.S.A*. 96, 7592–7597. doi: 10.1073/pnas.96.13.7592

Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (Tokyo: The National International Institute for Japanese Language), 29–54.

Karlsson, F. (1986). Frequency considerations in morphology. *Zeitschrift Phonetik Sprachwissenschaft Kommunikationsforschung* 39, 19–28.

Karttunen, L. (2003). "Computing with realizational morphology," in *International Conference on Intelligent Text Processing and Computational Linguistics* (Berlin; Heidelberg: Springer), 203–214. doi: 10.1007/3-540-36456-0_20

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G., and Martens, E. (2007). Dutch plural inflection: the exception that proves the analogy. *Cogn. Psychol*. 54, 283–318. doi: 10.1016/j.cogpsych.2006.07.002

Keuleers, E., Stevens, M., Mandera, P., and Brysbaert, M. (2015). Word knowledge in the crowd: measuring vocabulary size and word prevalence in a massive online experiment. *Q. J. Exp. Psychol*. 8, 1665–1692. doi: 10.1080/17470218.2015.1022560

Kirov, C., and Cotterell, R. (2018). Recurrent neural networks in linguistic theory: revisiting Pinker and Prince (1988) and the past tense debate. *Trans. Assoc. Comput. Linguist*. 6, 651–665. doi: 10.1162/tacl_a_00247

Köpcke, K.-M. (1988). Schemas in German plural formation. *Lingua* 74, 303–335. doi: 10.1016/0024-3841(88)90064-2

Landauer, T., and Dumais, S. (1997). A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychol. Rev*. 104, 211–240. doi: 10.1037/0033-295X.104.2.211

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*, Vol. 1. Stanford, CA: Stanford University Press.

Levelt, W., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci*. 22, 1–38. doi: 10.1017/S0140525X99451775

Linke, M., Broeker, F., Ramscar, M., and Baayen, R. H. (2017). Are baboons learning "orthographic" representations? Probably not. *PLoS ONE* 12:e0183876. doi: 10.1371/journal.pone.0183876

Linzen, T., and Baroni, M. (2021). Syntactic structure from deep learning. *Annu. Rev. Linguist*. 7, 195–212. doi: 10.1146/annurev-linguistics-032020-051035

Lõo, K., Jaervikivi, J., Tomaschek, F., Tucker, B., and Baayen, R. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 1, 71–97. doi: 10.1007/s11525-017-9318-7

Luo, X. (2021). *JudiLing: an implementation for Linear Discriminative Learning in JudiLing* (unpublished Master's thesis).

Luo, X., Chuang, Y.-Y., and Baayen, R. H. (2021). Judiling: an implementation in Julia of Linear Discriminative Learning algorithms for language modeling. Available online at: https://github.com/MegamindHenry/JudiLing.jl

MacWhinney, B., and Leinbach, J. (1991). Implementations are not conceptualizations: revising the verb learning model. *Cognition* 40, 121–157. doi: 10.1016/0010-0277(91)90048-9

Malouf, R. (2017). Abstractive morphological learning with a recurrent neural network. *Morphology* 27, 431–458. doi: 10.1007/s11525-017-9307-x

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: the exception that proves the rule. *Cogn. Psychol*. 29, 189–256. doi: 10.1006/cogp.1995.1015

Marsolek, C. J. (2008). What antipriming reveals about priming. *Trends Cogn. Sci*. 12, 176–181. doi: 10.1016/j.tics.2008.02.005

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. Cambridge: Cambridge University Press.

McCurdy, K. (2019). *Neural networks don't learn default rules for German plurals, but that's okay, neither do Germans* (Master's Thesis). University of Edinburgh, Edinburgh, United Kingdom.

McCurdy, K., Goldwater, S., and Lopez, A. (2020). "Inflecting when there's no majority: limitations of encoder-decoder neural networks as cognitive models for German plurals," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Stroudsburg, PA: Association for Computational Linguistics), 1745–1756. doi: 10.18653/v1/2020.acl-main.159

Miaschi, A., and Dell'Orletta, F. (2020). "Contextual and non-contextual word embeddings: an in-depth linguistic investigation," in *Proceedings of the 5th Workshop on Representation Learning for NLP* (Stroudsburg, PA), 110–119. doi: 10.18653/v1/2020.repl4nlp-1.15

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. *arXiv [Preprint]. arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, eds C. J. C. Burges and L. Bottou and M. Welling and Z. Ghahramani and K.Q. Weinberger (Red Hook, NY: Curran Associates Inc.), 3111–3119.

Milin, P., Madabushi, H. T., Croucher, M., and Divjak, D. (2020). Keeping it simple: Implementation and performance of the proto-principle of adaptation and learning in the language sciences. *arXiv [Preprint]. arXiv:2003.03813*.

Mirković, J., MacDonald, M. C., and Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Lang. Cogn. Process*. 20, 139–167. doi: 10.1080/01690960444000205

Nakisa, R. C., and Hahn, U. (1996). "Where defaults don't help: the case of the German plural system," in *Proc. 18th Annu. Conf. Cogn. Sci. Soc* (Mahwah, NJ), 177–182.

Nieder, J., Chuang, Y.-Y., van de Vijver, R., and Baayen, R. (2021). Comprehension, production and processing of maltese plurals in the discriminative lexicon. doi: 10.31234/osf.io/rkath

Oppenheim, G. M., Dell, G. S., and Schwartz, M. F. (2010). The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. *Cognition* 114, 227–252. doi: 10.1016/j.cognition.2009.09.007

Pennington, J., Socher, R., and Manning, C. D. (2014). "Glove: global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)* (Stroudsburg, PA), 1532–1543. doi: 10.3115/v1/D14-1162

Pinker, S., and Prince, A. (1988). On language and connectionism. *Cognition* 28, 73–193. doi: 10.1016/0010-0277(88)90032-7

Prince, A., and Smolensky, P. (2008). *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: John Wiley & Sons.

Ramscar, M., Dye, M., and McCauley, S. M. (2013). Error and expectation in language learning: the curious absence of mouses in adult speech. *Language* 89, 760–793. doi: 10.1353/lan.2013.0068

Rescorla, R. A. (1988). Pavlovian conditioning. It's not what you think it is. *Am. Psychol*. 43, 151–160. doi: 10.1037/0003-066X.43.3.151

Rumelhart, D. E., and McClelland, J. L. (1986). "On learning the past tenses of English verbs," in *Parallel Distributed Processing. Explorations in the Microstructure of Cognition. Vol. 2: Psychological and Biological Models*, eds J. L. McClelland and D. E. Rumelhart (Cambridge, MA: The MIT Press), 216–271. doi: 10.7551/mitpress/5237.001.0001

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol*. 80, 1–27. doi: 10.1152/jn.1998.80.1.1

Schulz, D., and Griesbach, H. (1981). *Grammatik der deutschen Sprache*. Munich: Max Hueber Verlag. doi: 10.1515/infodaf-1981-080512

Shafaei-Bajestan, E., Tari, M. M., and Baayen, R. H. (2021). LDL-AURIS: error-driven learning in modeling spoken word recognition. *Lang. Cogn. Neurosci*. 1–28. doi: 10.1080/23273798.2021.1954207

Shahmohammadi, H., Lensch, H. and Baayen, R. H. (in press). "Learning zero-shot multifaceted visually grounded word embeddings via multi-task training," in *Proceedings of the 25th Conference on Computational Natural Language Learning*.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Skousen, R. (1989). *Analogical Modeling of Language*. Dordrecht: Kluwer. doi: 10.1007/978-94-009-1906-8

Skousen, R. (2002). *Analogical Modeling*. Amsterdam: Benjamins. doi: 10.1075/hcp.10

Stump, G. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486333

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*, Vol. 6. Amsterdam: John Benjamins Publishing. doi: 10.1075/scl.6

Tomaschek, F., Tucker, B. V., Fasiolo, M., and Baayen, R. H. (2018). Practice makes perfect: the consequences of lexical proficiency for articulation. *Linguist. Vanguard* 4:s2. doi: 10.1515/lingvan-2017-0018

Trimmer, P. C., McNamara, J. M., Houston, A. I., and Marshall, J. A. R. (2012). Does natural selection favour the Rescorla-Wagner rule? *J. Theoret. Biol*. 302, 39–52. doi: 10.1016/j.jtbi.2012.02.014

Trommer, J. (2021). The subsegmental structure of German plural allomorphy. *Nat. Lang. Linguist. Theory* 39, 601–656. doi: 10.1007/s11049-020-09479-7

Van de Vijver, R., and Baer-Henney, D. (2014). Developing biases. *Front. Psychol*. 5:634. doi: 10.3389/fpsyg.2014.00634

Widrow, B., and Hoff, M. E. (1960). "Adaptive switching circuits," in *1960 WESCON Convention Record Part IV* (New York, NY: IRE), 96–104. doi: 10.21236/AD0241531

Wiese, R. (1999). On default rules and other rules. *Behav. Brain Sci*. 22, 1043–1044. doi: 10.1017/S0140525X99532226

Wunderlich, D. (1999). German noun plural reconsidered. *Behav. Brain Sci*. 22, 1044–1045. doi: 10.1017/S0140525X99542222

Yamada, I., Asai, A., Sakuma, J., Shindo, H., Takeda, H., Takefuji, Y., et al. (2020). "Wikipedia2Vec: an efficient toolkit for learning and visualizing the embeddings of words and entities from Wikipedia," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (Stroudsburg, PA: Association for Computational Linguistics), 23–30. doi: 10.18653/v1/2020.emnlp-demos.4

Yang, C. (2016). *The Price of Linguistic Productivity*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/9780262035323.001.0001

Zaretsky, E., and Lange, B. P. (2015). "No matter how hard we try: still no default plural marker in nonce nouns in modern high German," in *A lend of MaLT: Selected Contributions from the Methods and Linguistic Theories Symposium* (Bamberg), 153–178.

Zaretsky, E., Lange, B. P., Euler, H. A., and Neumann, K. (2013). Acquisition of German pluralization rules in monolingual and multilingual children. *Stud. Second Lang. Learn. Teach*. 3, 551–580. doi: 10.14746/ssllt.2013.3.4.6

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

frontiers in Psychology

ORIGINAL RESEARCH
published: 29 November 2021
doi: 10.3389/fpsyg.2021.701802

Check for updates

What Language Disorders Reveal About the Mechanisms of Morphological Processing

Christina Manouilidou[1]*, Michaela Nerantzini[2], Brianne M. Chiappetta[3], M. Marsel Mesulam[4] and Cynthia K. Thompson[3,4,5]

[1] Department of Comparative and General Linguistics, University of Ljubljana, Ljubljana, Slovenia, [2] Department of Philology, University of Ioannina, Ioannina, Greece, [3] Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, United States, [4] Mesulam Center for Cognitive Neurology and Alzheimer's Disease, Northwestern University, Chicago, IL, United States, [5] Department of Neurology, Northwestern University, Chicago, IL, United States

OPEN ACCESS

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
Shinri Ohta,
Kyushu University, Japan
Fanny Elise Meunier,
Centre National de la Recherche Scientifique (CNRS), France

***Correspondence:**
Christina Manouilidou
Christina.Manouilidou@ff.uni-lj.si

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

Received: 28 April 2021
Accepted: 27 October 2021
Published: 29 November 2021

Citation:
Manouilidou C, Nerantzini M, Chiappetta BM, Mesulam MM and Thompson CK (2021) What Language Disorders Reveal About the Mechanisms of Morphological Processing.
Front. Psychol. 12:701802.
doi: 10.3389/fpsyg.2021.701802

We addressed an understudied topic in the literature of language disorders, that is, processing of derivational morphology, a domain which requires integration of semantic and syntactic knowledge. Current psycholinguistic literature suggests that word processing involves morpheme recognition, which occurs immediately upon encountering a complex word. Subsequent processes take place in order to interpret the combination of stem and affix. We investigated the abilities of individuals with agrammatic (PPA-G) and logopenic (PPA-L) variants of primary progressive aphasia (PPA) and individuals with stroke-induced agrammatic aphasia (StrAg) to process pseudowords which violate either the syntactic (word class) rules (**reheavy*) or the semantic compatibility (argument structure specifications of the base form) rules (**reswim*). To this end, we quantified aspects of word knowledge and explored how the distinct deficits of the populations under investigation affect their performance. Thirty brain-damaged individuals and 10 healthy controls participated in a lexical decision task. We hypothesized that the two agrammatic groups (PPA-G and StrAg) would have difficulties detecting syntactic violations, while no difficulties were expected for PPA-L. Accuracy and Reaction Time (RT) patterns indicated: the PPA-L group made fewer errors but yielded slower RTs compared to the two agrammatic groups which did not differ from one another. Accuracy rates suggest that individuals with PPA-L distinguish **reheavy* from **reswim*, reflecting access to and differential processing of syntactic vs. semantic violations. In contrast, the two agrammatic groups do not distinguish between **reheavy* and **reswim*. The lack of difference stems from a particularly impaired performance in detecting syntactic violations, as they were equally unsuccessful at detecting **reheavy* and **reswim*. Reduced grammatical abilities assessed through language measures are a significant predictor for this performance, suggesting that the "hardware" to process syntactic information is impaired. Therefore, they can only judge violations semantically where both **reheavy* and **reswim* fail to pass as semantically ill-formed. This finding further suggests that impaired grammatical knowledge can affect word level processing as well. Results are in line with the psycholinguistic literature which postulates the existence of various stages in accessing complex pseudowords, highlighting the contribution of syntactic/grammatical knowledge. Further, it points to the worth of studying impaired language performance for informing normal language processes.

Keywords: derivational morphology, morphological processing, pseudowords, primary progressive aphasia, stroke-induced aphasia, agrammatism

INTRODUCTION

Morphological Processing in Healthy Adults

An important dimension of word knowledge which has been found to affect lexical processing is *morphological structure.* Morphology changes a word's form either to denote some grammatical function, e.g., *boy* > *boys* (singular > plural) or to create new lexical items with related (most of the time) meanings, e.g., *boy* > *boyish, boy* > *boyfriend.* The first operation is referred to as *inflection* (*boy-s*) while the other two are known as *derivation* (*boy-ish*) and *compounding* (*boy-friend*). In case of *inflection* and *derivation,* an *inflectional* or *derivational* morpheme attaches to a lexical stem, *boy* + *-s, boy* + *-ish* while in *compounding* two lexical stems merge together, *boy* + *friend.* These are highly productive operations in many languages.

The effects of complex structure on word processing have been studied extensively (see Amenta and Crepaldi, 2012 for a review). Till recently, the main issue among them was how morphologically complex words are accessed and how they are stored in the Mental Lexicon. In other words, the major question was whether we need to decompose them into their parts to access their meaning, e.g., *boy* + *ish* or whether we access them as one unit, e.g., *boyish.* This debate has had, and it still has proponents on both sides. Some researchers have claimed that morphological structure plays no role and that morphologically complex words are fully listed in the memory (Butterworth, 1983; Lukatela et al., 1987). Connectionist models are also against the representation of morphological structure in the mental lexicon (Elman et al., 1996; Sereno and Jongman, 1997). In contrast, other researchers have argued that complex words are obligatorily decomposed into their constituents and that the mental lexicon comprises only stems and affixes and not affixed words (Taft and Forster, 1975; Taft, 1988). Finally, several models of morphological processing have combined whole word access with affix-stripping, suggesting dual route processing for complex words (Frauenfelder and Schreuder, 1991; Chialant and Caramazza, 1995).

In the current literature, this remains an open debate, with recent papers providing evidence for both directions. This is certainly an important issue, nonetheless, its thorough discussion is beyond the scope of this paper. In our view, the balance might be turning in favor of decomposition route, as most recent neuroimaging studies suggests. Indeed, besides the numerous behavioral studies, a variety of neuroimaging studies from unimpaired populations about complex word processing have provided converging evidence that the human language processor has immediate access to constituent morphemes (see Münte et al., 1999; Lehtonen et al., 2011; Royle et al., 2012; Fruchter et al., 2013; Fruchter and Marantz, 2015; Leminen et al., 2019 for a review of neuroimaging data; Diependaele et al., 2009; Dominguez et al., 2010; Rastle and Davis, 2008 for a review of behavioral studies). This appears to be a process which operates in an early, automatic and semantic-blind way in both prefixed and suffixed words. Thus, in healthy populations, morphological decomposition takes place immediately after word viewing, resulting in the activation of both stem and affix, e.g., *teach* + *er,* for *teacher*.

If we assume that this is a property of the human language processing system, then it should be universal and it should operate independently of modality, i.e., visual vs. auditory. Lexical access takes place when sensory information is matched to lexical information. In auditory lexical access what activates lexical information is the first few phonemes (regardless of syllable structure), whereas in visual lexical access is the first (orthographically defined) syllable (Taft, 2004). Each modality is subject to additional restrictions related to the physical properties of input. For instance, in visual lexical decision factors such as frequency (Balota et al., 2004), family size (Bertram et al., 2000), derivational family entropy (del Prado Martín et al., 2004) facilitate lexical access. Similarly, in a relevant study about auditory recognition of prefixed words, Wurm et al. (2006) showed that cohort entropies, conditional root uniqueness points and morphological family size influenced lexical access of prefixed words. A general finding is that participants usually respond faster in visual lexical access compared to auditory but, importantly, there is no qualitative difference between their responses both behaviorally and in terms of EEG (Zunini et al., 2020) and MEG components (Brennan et al., 2014) which is suggestive of common underlying and universal ways of dealing with complex lexical items.

A big bulk of research regarding lexical access of complex words comes from pseudowords This spans from the early days of psycholinguistics (Caramazza et al., 1988; Laudanna et al., 1992; Burani et al., 1999) to the era of neuroimaging (Leinonen et al., 2009; Leminen et al., 2013; Kim et al., 2015) and it covers both auditory and visual lexical access. The reason for this choice is because pseudowords are devoid of lexical representations, and therefore, are supposed to be accessed through decomposition into their constituents. Furthermore, pseudowords are particularly useful for the exploration of prelexical effects of morphological segmentation, without lexical interference from the whole word. The main contribution of studies using pseudowords is that they robustly support multi-stage processing models for morphologically complex lexical items (such as the one adopted for the current study) ranging from behavioral experimental and modeling studies to neural evidence (for a review see Ripamonti et al., 2015). The use of pseudowords has also drawn criticism, mostly on the assumption that pseudowords lack semantics and that they are detached from the mental lexicon (Chuang et al., 2021). As we will see later, these lines of criticism are not directly relevant for the current study, as we do not treat pseudowords as meaningless units.

Several questions remain unsolved, however, pertaining to what happens once we have decomposed a complex word or pseudoword. Schreuder and Baayen (1995) in their meta-model which is designed to account for both visual and auditory processing of complex words[1] described a two-stage post–decomposition process, consisting of *licensing,* during which each activated morpheme is validated through its subcategorization specifications (syntactic checking) and *composition,* during which we check whether the lexical representation of the whole word can be computed based on the semantic representations of the activated morphemes. In other words, syntactic licensing checks whether we are allowed to combine *teach* + *er* in terms of their syntactic properties and composition checks whether it makes sense to combine *teach* + *-er* on semantic grounds. With this in mind, we can postulate that all formations that do not respect either syntactic or semantic restrictions will fail to be recognized as real words, and their rejection will take place at different stage and timeframe.

Based on this and considering the latest advancements in lexical processing we can outline the architecture of complex pseudoword recognition. This would include a first stage, where obligatory decomposition occurs, and all lexicalized substrings are exposed. It is during this stage that pure non-words with the form of *stem* (non-existing) + *affix,* such as **pearn-able* are rejected. The second stage includes *syntactic licensing,* during which all formations which violate the syntactic specifications of the base (grammatical class), are processed, and rejected. It is in this stage that a pseudoword of the type of *inappropriate stem + affix,* such as **river-able* would be rejected. The third stage includes *semantic composition,* where formations such as **danceable* would be rejected. Although both the *stem* (*dance*) and the *affix* (*-able*) are already activated in stage 2, it is not until this stage that semantic processing occurs, and participants decide on the well-formedness of semantic violations.

This architectural model has been confirmed in a variety of behavioral and neuroimaging studies, by using data from various languages and by employing either the violation paradigm described above which distinguishes between syntactic and semantic information or existing words. Several studies have investigated the temporal and spatial dynamics of grammatical category (*licensing*), showing that information associated with the syntactic category elicit an early left anterior negative ERP component (ELAN) peaking at about 250 ms after stimulus presentation (e.g., Hahne and Friederici, 1999; Hahne and Jescheniak, 2001). This response is identified usually at the inferior portion of the superior temporal gyrus. Data from MEG (Dikker et al., 2009; Linzen et al., 2013) confirm the early processing of grammatical category. Similarly, a separate body of studies advocate the existence of semantic composition as taking place at a later stage and at distinct brain-areas. Fruchter and Marantz (2015) were the first ones to establish the distinction between *lexeme lookup* and *semantic composition* by using derivational family entropy targeting *lexeme lookup* and a variable they called *derived semantic coherence* targeting *semantic composition.* The first variable elicited early activation (between 241 and 387 ms) in middle temporal gyrus, while the second elicited activation in orbitofrontal cortex at a later stage (between 431 and 500 ms). Finally, Whiting et al. (2014) compared complex words such as *teacher* to pseudo-complex words such as *corner* in a MEG experiment. While at early stages both types of words evoked same type of responses (inferior temporal gyrus and fusiform between 150 and 230 ms after stimulus presentation), there was greater activation for the pseudo-derived words at a later stage (between 300 and 360 ms) in the middle temporal gyrus. This later effect was interpreted by the authors as *lexicality effect* which amounts to semantic composition of an initially erroneously decomposed item.

The effects of *syntactic licensing* and *semantic composition* within one single experiment were first addressed by Manouilidou (2006, 2007) in a series of experiments. For instance, Manouilidou (2006, 2007) tested the ability of Greek-speaking individuals to detect violations of word formation with the aim to detect what kind of information is available after initial decomposition and morpheme recognition. The ultimate goal was to tease apart the contribution of *syntactic* and *semantic* information in deverbal structures. A variety of suffixes creating deverbal formations, nouns, and adjectives, was used. For instance, by using the Greek suffix *-tis* (equivalent to the English *-er*) which creates agentive nominalizations such as *pezo 'play'* > *pex-tis 'player'* we created syntactic violations, such as **potiri-tis* (Noun + *-tis*) 'glass-er', and also semantic violations not respecting the argument structure specifications of the base such as **diaferistis* 'differ-er.' The main finding of these behavioral experiments was that participants were faster and more accurate in detecting syntactic violations (**potiritis* 'glasser') compared to semantic violations (**diaferistis* '*differ-er'). These studies were later replicated by using stimuli from other, typologically quite distinct languages, such as English (Manouilidou and Stockall, 2014) and Slovenian (Manouilidou et al., 2016). Findings of these later studies are in complete agreement with the original studies conducted in Greek, confirming the architectural model of complex word recognition outlined above.

Moreover, subsequent neuroimaging studies (Neophytou et al., 2018; Stockall et al., 2019) confirm the existence of these two stages and the involvement of syntactic and semantic processing in post-decomposition processes. Specifically, Neophytou et al. (2018), using a subset of the Greek stimuli used in Manouilidou (2007), provided Magnetoencephalography (MEG) evidence for

[1] Schreuder and Baayen (1995) acknowledge the complications of auditory lexical access when it comes to segmentation and mapping of speech input on form-based access representations. "*Prosodic information, resyllabification, stress shifts, tone sandhi, and other phonological mutations may complicate this mapping operation*" (p. 133).

the distinction between these two types of pseudowords which violated syntactic and semantic rules of word formation, in terms of distinct timeframes and brain correlates. In their study, syntactic violations evoked more activity than semantic violations in the temporal lobe in the 200–300 ms time-window, while semantic violations evoked more activity than syntactic violations in the orbitofrontal cortex in the 425–500 ms window. This finding clearly differentiates the two types of information (syntactic vs. semantic) which needs to get processed when dealing with a complex word. This new piece of evidence adds to the growing body of research, which highlights the involvement of various sub-processes during lexical access of complex words and distinguishes *syntactic* vs. *semantic* processing in complex word recognition. Specifically, evidence advocates toward the idea that *syntactic licensing* and *semantic composition* occur at two distinct stages, whereby the former precedes the latter. The same pattern was later replicated in another MEG study by Stockall et al. (2019) by using data from English prefixation. As with Neophytou et al. (2018) and by using the same type of violations, the study targeted the spatial organization and temporal dynamics of morphological processing in the human brain. Results were identical with Neophytou et al. (2018) reinforcing the idea of the existence of stages in lexical processing and their sequence in time.

Thus, taken all this together we have credence to the existence of a consistent architectural map on the complete process of processing morphologically complex words, from initial form-based decomposition to syntactic licensing, and semantic interpretation. Interestingly, combined results of MEG studies from Greek (Neophytou et al., 2018) and English (Stockall et al., 2019) further suggest that the spatial and temporal dynamics of this process are very similar across different languages.

Morphology in Primary Progressive Aphasia and Stroke-Induced Agrammatic Aphasia

Primary Progressive Aphasia is a neurodegenerative disease which slowly and progressively disrupts the language regions of the brain, resulting in a gradual, and initially isolated, decline in language function (Mesulam, 1982, 2013). Other mental faculties such as memory remain intact, at least at initial stages. According to recent guidelines, PPA can be subdivided into three main variants based on clinical and imaging criteria (Gorno-Tempini et al., 2011; Maruta et al., 2015). What appears to be a common feature is *impaired word knowledge,* mainly manifested as *anomia.* However, there are specific deficits associated with each variant. The main characteristics of the *logopenic variant* (PPA-L) is intermittent word-finding hesitations, impaired phonological memory and problems with repetition (Gorno-Tempini et al., 2011) while their grammatical ability is, in general preserved. However, several recent studies have brought into light various difficulties with grammatical domains, especially in connected speech, such as difficulties with verbal morphology and avoidance of complex structures in production (e.g., Knibb et al., 2009; Ash et al., 2013; Fraser et al., 2014; Marcotte et al., 2017; Mack et al., 2021), albeit these difficulties appear to stem from a general word retrieval and verbal working memory deficit, rather than from an underlying grammatical impairment (Mack et al., 2021). The *agrammatic variant* (PPA-G)[2] is characterized by impairments of grammar (syntax and morphology) but not of word comprehension. Non-fluent speech, production of grammatically impoverished sentences, verb production difficulties, difficulties with complex syntactic structure (production and comprehension), difficulties in producing function words and bound morphemes, and in general, impaired processing of morphosyntactic structure (e.g., Thompson et al., 1997, 2013; Thompson and Mack, 2014 for a review) also complete the PPA-G profile. Finally, the *semantic variant* (PPA-S) is characterized by impairments of word comprehension, and more specifically by difficulty in processing lexical–semantic information (i.e., word meaning) in both production and comprehension, with associated neural atrophy in the left anterior temporal lobe. Given that no PPA-S patients participated in the study, we will not further elaborate on this condition.

Despite neuropathological differences, similar language deficits, mainly in the syntactic domain, can be found in stroke-induced agrammatic aphasia (StrAg) and in PPA-G as well (Thompson et al., 2012c, 2013). In particular, the processing of *argument structure* is a domain for which difficulties have been reported for both populations. For instance, difficulties with processing complex argument structure or violations of verb argument structure in narrative speech have been described for both PPA-G and StrAg (Thompson et al., 2012c for PPA-G, Bastiaanse and Jonkers, 1998; Thompson and Bastiaanse, 2012, for StrAg). Thus, it is not uncommon for researchers to compare the two conditions, in order to gain insights about the nature of agrammatism as manifested in two different conditions after brain damage. Even though argument structure difficulties are mostly a result of sentence processing, a comparison of the two groups at the lexical level is valid, given that *inflectional morphology* has been found to be compromised in both PPA-G and StrAg. Interestingly, a recent study by Kordouli et al. (2018) has brought into light interesting dissociations in compound naming between PPA-G and StrAg, with PPA-G performing significantly worse. At the same time, less is known about how patients with PPA and stroke-induced aphasia process *derivational morphology*, which is the topic of the current study.

Derivational morphology, that is the production of a new lexical item from another lexical stem, e.g., *happy* > *unhappy, emerge* > *reemerge* is usually better preserved than inflectional morphology in brain-damaged populations. For instance, Miceli and Caramazza (1988) report on agrammatic aphasic patients' ability to use derivational affixes as relatively intact. However, subsequent studies brought into light various interesting facts about the processing of derivational morphology by brain-damaged populations. For example, the study by

[2]In some patients, grammar and comprehension are jointly impaired early in the disease. These patients can be said to have a fourth '*mixed*' *variant* (PPA-m) (Mesulam, 2013).

Miceli et al. (2004) reports morphological errors in association with phonological errors, while Faroqi-Shah and Thompson (2010), focusing on complexity as manifested in past tense forms (*-ed*) and progressive aspectual forms (*-ing*), did not find any effects of morphological complexity. On the other hand, Semenza et al. (2002) studied the performance of two Slovenian-speaking patients, one diagnosed with agrammatic aphasia and the other with transcortical motor aphasia. The study showed that while prefixes (e.g., *re-* in *reappear*) are well-preserved in the grammar of both patients, with no phonological distortions on them, at the same time, they can be omitted or substituted. This fact suggests that prefixation, as a morphological operation, and the structure of a prefixed word are preserved in these two types of aphasia. However, the fact that patients do not always succeed in producing the right form of the derived verb suggests certain difficulties with this operation, for both individuals with aphasia (agrammatic and transcortical).

An interesting study by Marangolo et al. (2003) reports on two patients with comparable right hemisphere lesions which involved the gray and white matter of the right temporal and parietal lobes and the right centrum semiovale, who showed a selective deficit in the processing of derived words without any other linguistic deficit. This study was the first one to show that derivational morphology can be selectively impaired and that its processing can be mediated by the right hemisphere. Patients were tested in a picture naming task where they had to name either an action verb or the corresponding derived nouns. They were also asked to produce derived nouns that corresponded to verbs presented to them orally and to produce the verb that corresponded to the nouns they heard. Both patients were unsuccessful in naming derived nouns from verbs (e.g., *liberare* 'to free' > *liberazione* 'freedom') but they could name verbs from derived nouns (e.g., *liberazione* 'freedom' > *liberare* 'to free'). This study highlights in the best way that derivational morphology can be selectively impaired and that it can have ties with the right hemisphere as well and not necessarily with typical language areas. Finally, not only overt derivation but also zero-derivation (Lukic et al., 2016) appears to be affected in StrAg, especially in cases when aphasic individuals with verb impairments had to "derive" verbs from nouns (*brush* > *to brush*), stressing the crucial role of the grammatical category of the base (i.e., verbs) in performing morphological processes.

Taking all the above into consideration, it appears that derivational morphology leads its own life when it comes to language disorders. On the one hand, it appears better preserved than inflectional morphology. On the other hand, it appears to engage different brain areas, since derived words exhibit a variety of properties that are not found in inflected forms, such as a distinct semantic component, given that the derived word is a separate concept. However, it remains an understudied area in the field of language disorders, thus, calling upon further investigation.

The Current Study – Research Questions

In the present study we analyze data of complex pseudoword processing from English-speaking individuals diagnosed with two variants of PPA and with StrAg. Given that PPA is a condition which mostly affects lexical processing and given that pseudoword processing touches upon many issues (see Section "Morphological Processing in Healthy Adults"), it appears to be an appropriate domain of investigation in order to see how the underlying deficits of these conditions might affect it. At the same time, a secondary goal is to inform morphological theory by providing independent evidence about a linguistic phenomenon which has occupied the psycholinguistic literature for decades, that is, complex word recognition.

Thus, the overarching aim of the study is to investigate processing of complex pseudowords in these populations and to contribute new data to the literature of lexical/morphological processing by PPA individuals. Within this general frame, we also seek to shed light to related issues, with respect to the type of stimuli investigated and the specific populations that participated in the study. *First* and foremost, given that there is no evidence about complex pseudoword processing, the main aim of the study is to fill this gap of knowledge, thus, making it the first study to bring into light evidence about word-structure building in PPA, a productive operation across languages. *Second,* the specific types of pseudowords used in the current study allow us to investigate the contribution of finer-grained types of information necessary in word-structure building, that is information that pertains to knowledge of the grammatical category of the base and to argument structure specifications. This is particularly important given that no previous study has looked at the influence of both syntactic and semantic properties in the processing of word-building in PPA. Finally, given that both StrAg and PPA-G are characterized by agrammatism, the *third* aim of the study is to compare the two conditions and examine whether agrammatism affects pseudoword processing in the same way.

MATERIALS AND METHODS

Participants

Thirty brain-damaged individuals diagnosed with PPA and meeting the criteria for logopenic ($n = 12$) and agrammatic ($n = 8$) variant or stroke-induced agrammatic aphasia ($n = 10$) were recruited to participate in the study. An additional group of 10 healthy volunteers, aged-matched controls (AM) (5 males and 5 females) were also selected. All participants were monolingual native English speakers with self-reported normal vision and hearing. One healthy AM control was excluded due to poor performance on the lexical decision task, and thus all analyses were done on the remaining 9 AM controls. The participant groups were matched on age [$t(37) = 1.118$, $p = 0.271$] and years of education [$t(37) = 0.917$, $p = 0.365$] although StrAg participants were marginally younger than the participants with PPA [$t(4) = -2.629$, $p = 0.058$].

Individuals with PPA were recruited from the Mesulam Center for Cognitive Neurology and Alzheimer's Disease in Chicago, IL, United States. All patients were clinically diagnosed with PPA based on neurological examination and related test results [i.e., magnetic resonance imaging and were further categorized by PPA variant based on language and neuropsychological testing, and their magnetic resonance images based on the criteria

TABLE 1 | Participants' demographic information.

AM	AM01	AM02	AM03	AM04	AM05	AM06	AM07	AM08	AM09	Group average
Age	56	67	60	53	76	64	75	68	63	64.7
Gender	F	F	M	M	F	M	F	M	F	4 M (5 F)
Handedness	R	R	R	R	R	R	R	R	R	all R
Education (years)	18	18	18	18	14	20	21	18	17	18.0

StrAg	SA01	SA02	SA03	SA04	SA05	SA06	SA07	SA08	SA09	SA10	Group average
Age	41	64	29	46	42	22	48	38	67	51	44.8
Gender	M	M	F	M	M	F	M	F	M	M	7 M (3 F)
Months post-stroke	94	13	28	27	20	31	18	98	306	37	67.2
Handedness	R	R	R	R	L	R	R	R	L	R	8 R (2 L)
Education (years)	16	18	19	18	16	14	16	18	20	20	17.5

PPA participants	P1	P2	P3	P4	P5	P6	P7	P8	PPA-G ave	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	PPA-L ave	Total group ave
PPA-type	**G**	**G**	**G**	**G**	**G**	**G**	**G**	**G**		**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**	**L**		
Age	76	64	69	74	65	63	66	53	66.25	69	64	80	58	65	70	73	51	60	63	75	67	66.25	66.25
Gender	M	F	F	M	F	M	F	F		F	M	M	F	M	M	M	M	M	M	F	F		11 M (9 F)
Symptom duration (years)	3.6	2.5	2.6	4.9	2.5	7.0	1.5	6.0	3.8	2.6	3.8	6.8	2.0	3.5	2.0	4.0	3.5	4.0	3.5	4.0	4.5	3.7	3.7
Handedness	R	R	R	R	R	R	R	R		R	R	R	R	R	R	R	R	R	R	R	R		20 R
Education (years)	18	16	16	16	14	20	18	16	16.75	16	16	18	19.5	19	17	20	12	18	18	19	18	17.54	17.22

AM, age-matched; StrAg, stroke agrammatic; SA01, SA02, SA03 etc., stroke agrammatic; PPA, Primary progressive aphasia; G, agrammatic; L, logopenic; S, semantic; P1, P2, P3...etc. Patient; F, female; M, male; R, right-handed.

discussed in Mesulam (2001, 2003, 2013) and Gorno-Tempini et al. (2011)]. Demographics for all participants are presented in **Table 1**; scores on language measures across participants are provided in **Supplementary Appendix 1**. None of the PPA patients showed evidence of stroke or other neurological disorder, while all presented a history of progressive language deficits in the face of relatively spared abilities in other cognitive domains. The study was approved by the Institutional Review Board at Northwestern University and informed consent was obtained from all participants.

The stroke-induced agrammatic aphasic individuals suffered a single left-hemisphere stroke at least 1 year prior to the study with no history of other speech and language impairments prior to stroke. Participants were selected for inclusion based on neuropsychological assessments and according to the criteria of the *Western Aphasia Battery-Revised (WAB-R)* (Kertesz, 2006). Participants exhibited mild-to-moderate aphasia (WAB-AQ mean: 75.4, range: 53.5–89), with non-fluent agrammatic features, such as (a) slow and effortful spontaneous speech (WAB fluency mean: 10.9, range 2–20), (b) impaired comprehension and production of non-canonical sentences, as indicated by performance on the Sentence Comprehension Test (SCT) and the Sentence Production Priming Test (SPPT) of the *Northwestern Assessment of Verbs and Sentences* (NAVS) (Thompson, 2011): for comprehension: non-canonical range: 33.3–86.7% correct; canonical range: 46.7–93.3% correct; for production: non-canonical range: 0–73.3% correct; canonical range: 33.3–100% correct, (c) unimpaired noun production and preserved single-word comprehension of both nouns and verbs, as illustrated by scores $\geq$ 50% correct on the Confrontation Naming subtest of the *Northwestern Naming Battery* (NNB) (Thompson and Weintraub, 2014; experimental version) and by scores $\geq$ 60% on the Auditory Comprehension subtest of the NNB, respectively[3]. Details are listed in **Supplementary Appendix 1**.

Experimental Conditions and Materials

Four experimental conditions with 40 items each (39 for non-words) and one filler condition were included in the experiment. Specifically, the experimental conditions included one group of non-words (#1 below), two groups of words violating certain constraints of word formation in English (see #2 and #3 below) and one group of real words (#4 below). All were formed with the prefix *re-*. The filler conditions (#5 below) consisted of 80 well-formed words that contained different decomposable affixes (e.g., *unable*). Fillers were used exclusively to distract the participants and to balance the ratio of grammatical vs. ungrammatical words and were not further analyzed. Materials were based on Manouilidou and Stockall (2014). They were modified to comply with the requirements of American English participants i.e., word frequencies of existing items were recalculated based on CELEX English database (Baayen et al., 1995) and a set of new real words were selected. All experimental items were matched for CELEX spoken and written stem/root frequency [for spoken: $F(3) = 1.095$, $p = 0.353$; for written: $F(3) = 0.049$, $p = 0.986$), and for length, apart from real words, which were slightly longer (mean: 7.575, $p = 0.006$ when compared to SynViol and SemViol)]. Finally, durations of auditory files were also calculated. There is no significant difference between durations in the two critical conditions ($t = -1.056$, $p = 0.297$) but they both differ significantly when compared to fillers ($p = 0.000$ in both comparisons). **Table 2** presents details on the experimental stimuli.

The stimulus set comprised the following experimental conditions:

(1) Non-words (NWs): pseudowords stems + *re-* (e.g., **repearn*; $n = 39$).
(2) SynViol: real word base + *re-*, forming a grammatical category constraint violation (SynViol) (e.g., **resimple*; $n = 40$).
(3) SemViol: real word base + *re-*, forming an argument structure/thematic constraint violation (SemViol) (e.g., **rescream*; $n = 40$).
(4) Real words with *re-* and no base form violations (e.g., *resubmit*; $n = 40$).
(5) Fillers: real words without *re-* (e.g., *acceptable*; $n = 80$).

In total, the stimuli included 239 words and the ratio between well-formed and ill-formed was 50:50.

Procedure

An auditory lexical decision task was conducted, running on an IBM computer using E-prime 2.0 professional software (Psychology Software Tools, Pittsburgh, PA, United States), which collected and recorded response time and accuracy data. Initially, participants were given detailed instructions about the experiment and 10 practice trials were provided to familiarize participants with the task. All stimuli were recorded by a native speaker of American English and were presented to the participants *via* headphones. Participants first saw a cross " +" in the middle of the screen for 1,000 ms and then they heard the stimulus. Participants had to decide as quickly and as accurately as possible whether the word that they heard was a word of English. Participants had 3,000 ms to press with their left hand one of two pre-specified color-coded buttons (either the YES "s" or the NO "a" key), on the left side of the QWERTY keyboard. Participants could pause the task and have a break at any point during the experiment.

ANALYSIS AND RESULTS

A mixed-effects logistic regression was performed on the item-level data for accuracy and a linear mixed-effects regression was performed on the item-level data for reaction times (RT) using the lme4 package in R Studio version 1.2.1335 (Bates et al., 2015; Kuznetsova et al., 2015; R Core Team, 2015; Team, 2018). Participants' accuracy and the logarithmic transform of their reaction times (logRT) were used as the dependent variables in separate analyses. For both accuracy and reaction time analyses, group (PPA-G, PPA-L, StrAg, and AM), condition (pseudowords with SynViol, pseudowords with SemViol, Non-Words, and real

[3]NNB scores are missing for one participant, SA09, due to time constraints when testing.

TABLE 2 | Characteristics of experimental stimuli.

	SynViol	SemViol	Real	NWs	fillers
Mean length (letters)	6.85	6.85	7.57	7.35	7.39
Mean Audio file duration (sec)	0.75	0.77	0.73	0.76	0.66
Mean stem/root frequencies (CELEX_log)	0.98	1.16	1.18		

words), and their interaction were entered as fixed factors with age as a covariate, and random intercepts for participants and trial items were entered as crossed random factors in the full model. Models with and without each fixed factor were compared using the anova function in R [see (a) – (e) below for formulas of compared models] to identify the best-fit model for accuracy and RT data separately. In the presence of significant effects, *post hoc* planned comparisons were run, and *p*-values were corrected for multiple comparisons using a single-step method in the multcomp package (Bretz et al., 2010) in R.

Model formulas: For accuracy data, DV = accuracy (1/0), while for RT data, DV = logRT. Formula *d* was the best-fit model for all analyses.

(a) Intercept and random factors only:
DV ∼ 1 + (1| participant) + (1| item)
(b) Intercept, group, and random factors:
DV ∼ 1 + group + (1| participant) + (1| item)
(c) Intercept, group, condition, and random factors:
DV ∼ 1 + group + condition + (1| participant) + (1| item)
(d) Intercept, group, condition, and their interaction, and random factors:
DV ∼ 1 + group*condition + (1| participant) + (1| item)
(e) Full model: Intercept, group, condition and their interaction, and age (covariate), and random factors:
DV ∼ 1 + group*condition + age + (1| participant) + (1| item).

With respect to accuracy, group means and standard deviations are presented in **Table 3**.

For accuracy data, the best-fit model was the one that included the interaction term [formula (*d*) above; $\chi^2(9) = 61.71$, $p < 0.001$]. Results from the mixed-effects logistic regression analyses showed a significant group*condition interaction. *Post hoc* comparisons of participant groups indicated that for non-words and real words, the AM group performed better than all patient groups, although this was only significant when comparing the AM group to the PPA groups for real words: (PPA-G: $z = -3.56$, $p = 0.002$; PPA-L: $z = -2.73$, $p = 0.03$). None of the patient groups differed significantly from each other for real words or non-words. Comparisons between groups with respect to the two critical conditions (SynViol and SemViol) revealed the following: For the SynViol condition, the AM group stands out yielding, on average, significantly more accurate rates compared to the PPA-G group ($z = -2.92$, $p = 0.018$) and compared to the StrAg group ($z = -3.31$, $p = 0.005$). None of the patient groups significantly differed from each other for the SynViol condition. For the SemViol condition, the AM group was, on average, only marginally significantly more accurate than the StrAg group ($z = -2.51$, $p = 0.059$). There were no other significant comparisons for the SemViol condition.

TABLE 3 | Average percent correct (SD) scores for each condition and group.

	SynViol	SemViol	NW	real
StrAg	63 (26.7)*	56 (23.3)	76 (23.9)	84 (9.7)
PPA-G	63 (31.0)*	59 (33.2)	78 (17.9)	71 (18.6)*
PPA-L	77 (13.1)*	65 (13.2)*	81 (9.8)	79 (14.0)*
AM	90 (3.3)*	78 (13.1)*	86 (8.4)	92 (5.6)

*Blue asterisk (*) indicates significant difference between critical conditions *reheavy vs. *reswim and red asterisk (*) indicates significant difference between groups of patients and control group.*

Post hoc comparisons of conditions for each group indicated no reliable differences between the SynViol and SemViol conditions for the PPA-G group ($z = -0.82$, $p = 0.85$) or for the StrAg group ($z = -1.75$, $p = 0.30$). Interestingly, both the PPA-G and StrAg groups performed significantly better for NWs compared to pseudowords with SynViol (PPA-G: $z = 4.46$, $p < 0.001$; StrAg: $z = 3.83$, $p < 0.001$) and compared to pseudowords with SemViol (PPA-G: $z = -5.37$, $p < 0.001$; StrAg: $z = -5.81$, $p < 0.001$). This suggests that for the two agrammatic groups, the two types of violations (SynViol and SemViol words) are clearly distinguishable from non-words, even though they do not differ between each other. At the same time, both the PPA-L and healthy AM participants produced distinct rates of accuracy for SynViol and SemViol conditions, with significantly better performance for SynViol words (PPA-L: $z = -3.01$, $p = 0.014$; AM: $z = -3.25$, $p = 0.006$). Notably though, the PPA-L group showed no distinct performance between NWs and pseudowords with SynViol ($z = 1.22$, $p = 0.61$).

Looking at individual responses at **Table 4**, we see that there is within group variability in the data also illustrated in **Figures 1A,B**, which is mostly manifested in the two agrammatic groups (for PPA-G, SynViol range: 18–97.5% SemViol range: 10–95%; StrAg, SynViol range: 10–92.5%; SemViol 10–80%). The PPA-L group appears to be less variable (SynViol range: 52.5–90%; SemViol range: 40–80%). Since the PPA classification does not necessarily control for the extent of sentence comprehension/production deficits, we also ran separate models using performance on language measures of non-canonical sentence comprehension (nc*SCT*) and production (nc*SPPT*) instead of group as a fixed factor. These two tasks are not related to the lexical decision task used in the current study, as they tap into participants' grammatical knowledge, as a broader domain of language knowledge. However, they can provide valuable information with respect to the underlying language deficits of the populations under investigation which can possibly affects participants'

TABLE 4 | Individual responses (% correct responses) per experimental condition.

SynViol accuracy	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
StrAg	65%	55%	80%	87.50%	78%	92.50%	65%	75%	25%	10%		
PPA-G	97.50%	75%	23%	45%	68%	18%	85%	93%				
PPA-L	62.50%	80%	57.50%	77.50%	90%	78%	75%	83%	52.50%	88%	95%	85%
SemViol accuracy	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**
StrAg	57.50%	62.50%	80%	77.50%	50%	77.50%	55%	65%	22.50%	10%		
PPA-G	95%	80%	10%	32.50%	70%	20%	82.50%	85%				
PPA-L	55%	80%	52.50%	67.50%	75%	75%	80%	70%	52.50%	40%	57.50%	77.50%

performance in the lexical decision task as well. The same procedure was used to determine the best-fit model using the same formulas listed above (a) – (e), with the only difference of replacing the fixed factor of group with a (continuous) fixed factor for performance on nc*SCT* (percent correct), and separately with a (continuous) fixed factor for performance on nc*SPPT* (percent correct). For the model including sentence comprehension of non-canonical structures (nc*SCT*) as a fixed factor, the best-fit model was the one that included the interaction term [$\chi^2(3) = 35.84$, $p < 0.001$]. Results from the logistic regression analysis showed a significant interaction between condition and performance on nc*SCT*. *Post hoc* comparisons revealed that performance on nc*SCT* was not a significant predictor of accuracy for any of the conditions. For the model including sentence production of non-canonical structures (nc*SPPT*) as a fixed factor, the best-fit model was the one that included the interaction term [$\chi^2(3) = 50.73$, $p < 0.001$]. Results from the logistic regression analysis showed a significant interaction between condition and performance on nc*SPPT*. *Post hoc* comparisons indicated that the *ncSPPT* language measure was a significant predictor of SynViol accuracy ($z = 2.52$, $p = 0.01$), but not of accuracy for the other conditions. As shown in **Figure 2**, the degree of impairment of grammatical abilities agrees with accuracy in detecting SynViol, while no significant interactions were found for SemViol.

Comparisons between groups revealed that the PPA-L group performed significantly better than the StrAg group for these two language measures (nc*SCT*: $z = -4.86$, $p < 0.001$; ncSPPT: $z = -3.16$, $p = 0.005$) and significantly better than the PPA-G group for nc*SPPT* ($z = 2.41$, $p = 0.04$). The StrAg group performed worse than the PPA-G group for nc*SCT* (marginal significance: $z = -2.32$, $p = 0.053$), but not for ncSPPT ($z = -0.53$, $p = 0.86$; see **Table 5** for SCT and **Table 6** for SPPT). Participants' percentages of correct responses in these tasks can be found in **Supplementary Appendix 1**, the relevant part repeated here for convenience in **Table 7**.

With respect to RTs, only response latencies corresponding to correct trials were analyzed, and RTs smaller than 300ms were eliminated for all participants (less than 1% of the data). Although the model was run using log-transformed data, raw values are presented in **Table 8** for easier interpretability. For RT data, the best-fit model was the one that included the interaction term [formula (*d*) above; $\chi^2(9) = 59.86$, $p < 0.001$]. Results from the linear mixed-effects regression analyses showed a significant group*condition interaction. *Post hoc* comparisons of participant groups indicated no significant differences for SynViol or SemViol words. *Post hoc* comparisons of conditions for each participant group revealed that all groups performed faster for both NW and real words compared to the critical conditions (SynViol and SemViol) ($p < 0.001$)[4], but only the AM group differentiated between the two critical types of violation (syntactic and semantic) by presenting significantly faster responses in the former type of violation ($z = 3.279$, $p = 0.005$). We also modeled for interactions with performance on language measures (*ncSCT*/*ncSPPT*), but there were no significant outcomes.

In sum, the current pattern of results can be summarized as follows. Healthy controls distinguished the two critical conditions, both in terms of accuracy and RTs, with SynViol being easier and faster to reject compared to SemViol. With respect to patient groups, closer to AM was the PPA-L group, as they were the only group which did tell apart the two critical conditions based on error rates, however, RTs did not indicate distinct timeframes in terms of processing. PPA-G and StrAg were comparable to each other, not being able to tell apart the two critical conditions, but clearly isolating them from both real words and non-words. Finally, participants' performance on *ncSPPT* was a strong predictor for their accuracy rates on SynViol. Based on this summary, we will discuss our data in the following section.

DISCUSSION

The current investigation aimed at: (a) examining the ability of PPA and StrAg individuals to process pseudowords and more specifically to detect violations in deverbal word formation, (b) isolating the contribution of each type of relevant information (e.g., syntactic vs. semantic) in deverbal word structure building and (c) comparing the performance of PPA-G and StrAg, two conditions characterized by agrammatism, in order to detect its effect in pseudoword processing. For the investigation of the above questions, we will focus on the data obtained for the two critical conditions, that is SynViol and SemViol, and we will consider participants' scores on both *accuracy* of response as well as *reaction times*.

[4]For PPA-G the comparison between NW vs. SynViol was only marginally significant ($p = 0.0535$) and for the AM group, $p = 0.0162$ for NW vs. SynViol.

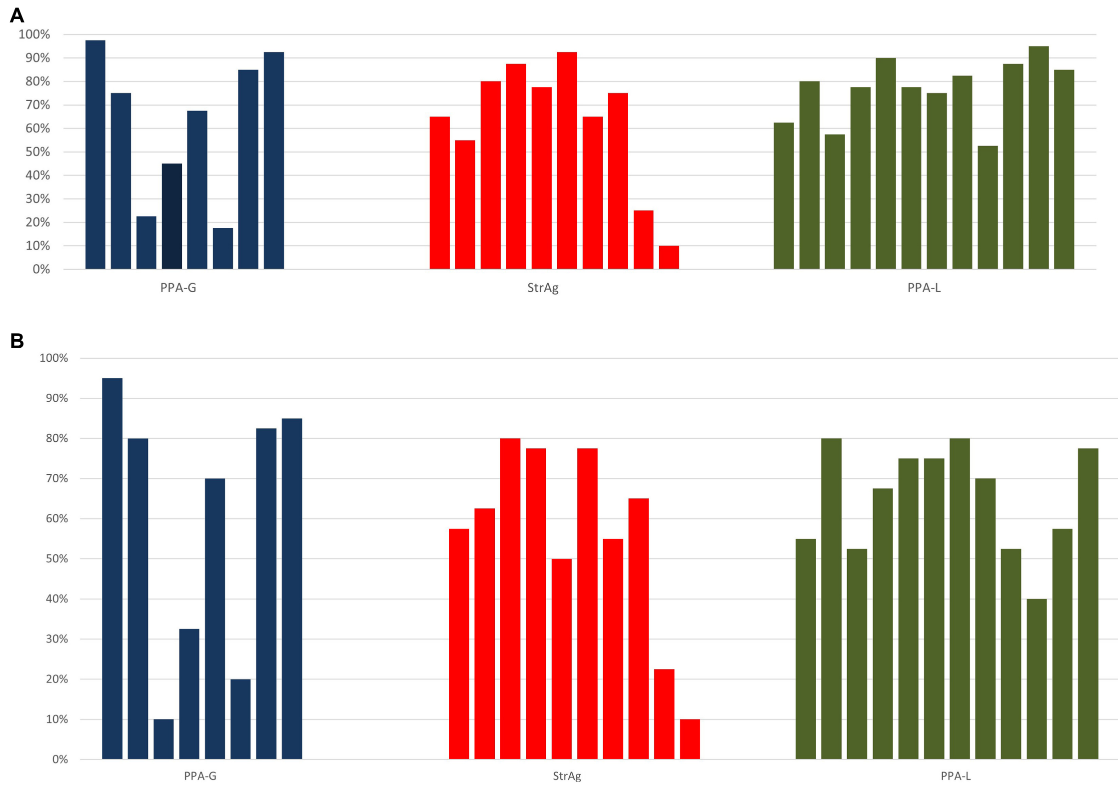

FIGURE 1 | Within group variability across participants [% correct responses for pseudowords with SynViol **(A)** and SemViol **(B)**].

Looking at accuracy data, when it comes to the first question, the PPA-L group appears to be the group with the best performance, as it does not differ significantly from AM in either the detection of SynViol or SemViol. The two agrammatic groups clearly have difficulties with the detection of SynViol, as they both differ significantly from AM, while for SemViol, the PPA-G group did not differ from AM while there was only a marginally significant difference between StrAg and AM[5]. This is a first indication that the two agrammatic groups have an increased difficulty in accessing information of syntactic nature within a complex word. A comparison among experimental conditions within groups reveals further important dissociations. Addressing the second question will shed more light into the source of these differences.

The second question aimed at investigating whether the groups of participants can process separately the two types of information (syntactic vs. semantic) in pseudoword lexical access. In other words, we seek to examine whether they can tell apart the two critical conditions which will also help us further investigate the source of their difficulties. For this purpose, we are looking for distinct accuracy rates for SynViol and SemViol across participant groups. It seems that while all participants had higher accuracy rates for SynViol, this difference reached significance only for PPA-L but not for the two agrammatic groups, PPA-G and StrAg, which appear to treat them alike (**resmile* = **rehappy*). If this is the case, then we would have to assume that the PPA-L group is able to process separately information associated with the grammatical category of the base and information associated with argument structure. The group of PPA-L performed as expected, that is, they processed the two types of information, and they did not differ from AM controls, suggesting a better preserved morphological and lexical system than the two agrammatic groups.

Finally, there is a dichotomy between the two agrammatic groups (PPA-G and StrAg) and the PPA-L group. PPA-G and StrAg groups did not differ significantly, neither with respect to overall accuracy rates nor with respect to accuracy rates regarding the two types of violations. This lack of difference is in line with previous studies comparing the two conditions in various grammatical tasks and it suggests a unified effect of agrammatism in detecting word-formation violations. Of particular interest is the performance of these groups when it comes to language measures which target the investigation of their grammatical abilities (see **Figure 1**). Specifically, grammatical abilities (as modeled through complex sentence production), turned out to be a significant predictor for SynViol accuracy. Several studies have shown that individuals with acquired aphasia often present with sentence comprehension and production deficits in sentences with non-canonical word order, such as passives and object relative clauses,

[5]Following Olsson-Collentine et al. (2019), we will interpret this difference as "insignificant." According to the study values between 0.05 and 0.10 are known to have low evidential value and they should be treated as insignificant.

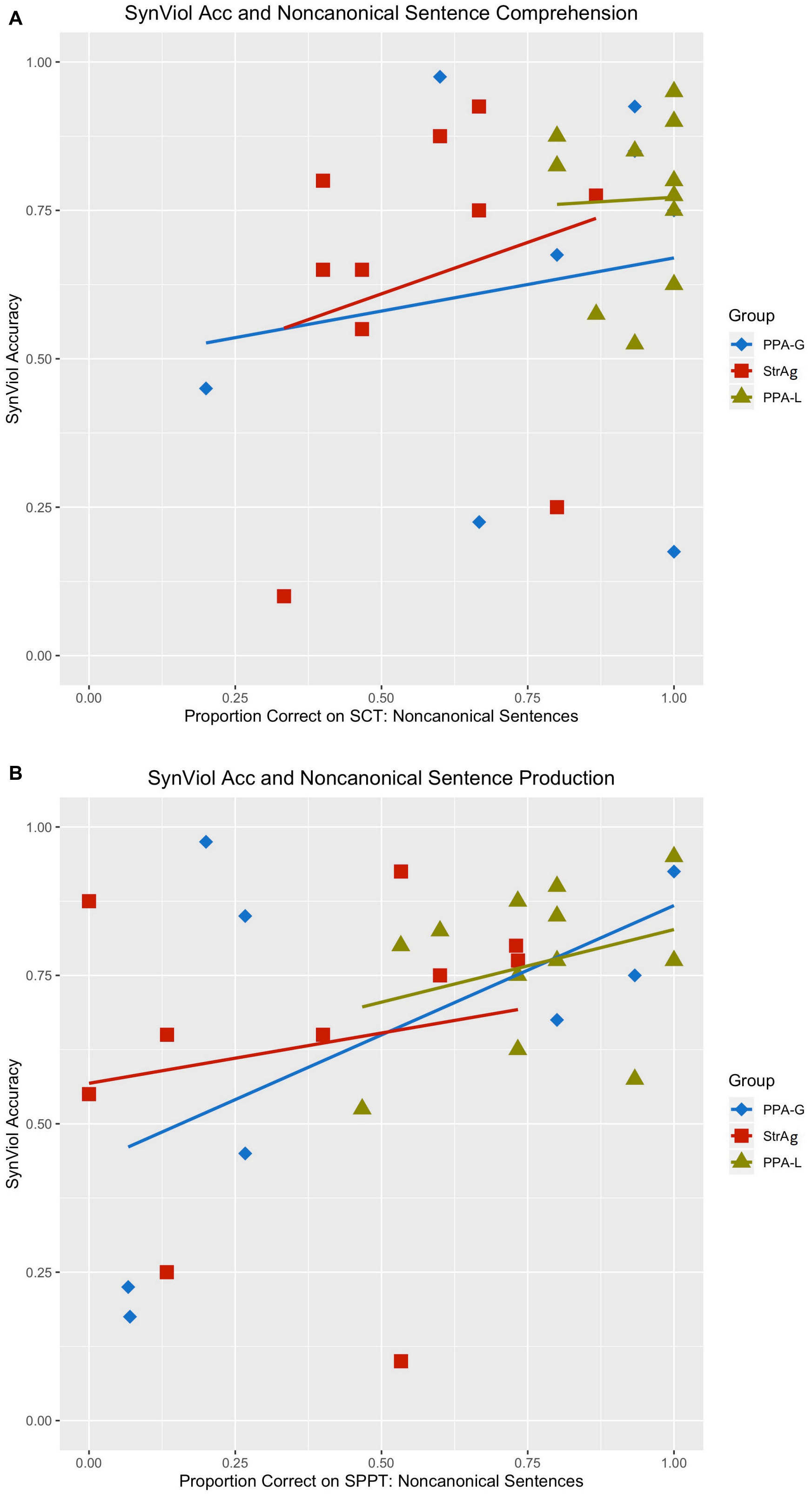

FIGURE 2 | Interaction between accuracy rates for SynViol and language measures as fixed factors. Language measure used: Northwestern Assessment of Verbs and Sentences (NAVS*) Sentence Comprehension Task (SCT) for non-canonical constructions **(A)** and NAVS Sentence Production Priming Task (SPPT) for non-canonical constructions **(B)** (*Thompson, 2011).

TABLE 5 | Between group comparisons for SCT.

Comparison	Estimate	Standard error	*Z*-value	*P*-value
PPAL – PPAG	0.178	0.083	2.15	0.081
StrAg – PPAG	−0.200	0.086	−2.32	0.053
StrAg – PPAL	−0.378	0.078	−4.86	< 0.001***

*Significance level: *p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 6 | Between group comparisons for SPPT.

Comparison	Estimate	Standard error	*Z*-value	*P*-value
PPAL – PPAG	0.311	0.129	2.41	0.042*
StrAg – PPAG	−0.071	0.134	−0.53	0.86
StrAg – PPAL	−0.381	0.121	−3.16	0.005**

*Significance level: *p < 0.05; **p < 0.01; ***p < 0.001.*

TABLE 7 | Percentages of correct responses (standard deviations) per group for language measures of comprehension (SCT) and production (SPPT) of non-canonical sentence structures.

	SCT	SPPT
StrAg	56.7 (18.12)	38.0 (28.94)
PPA-G	76.7 (27.36)	45.0 (39.26)
PPA-L	94.4 (7.95)	76.1 (16.92)

TABLE 8 | Mean RTs (SD) in milliseconds for each condition and group for only correct responses > 300 ms.

	SynViol	SemViol	NW	real
StrAg	1379 (270)	1416 (270)	1279 (268)	1280 (208)
PPA-G	1488 (218)	1551 (217)	1420 (220)	1340 (182)
PPA-L	1596 (173)*	1636 (143)*	1463 (150)	1415 (123)
AM	1389 (105)*	1477 (103)*	1320 (97)	1159 (73)

*Blue asterisk (*) indicates significant difference between critical conditions *reheavy vs. reswim and red asterisk (*) indicates significant difference between groups of patients and control group.*

compared to those with a basic, canonical Subject, Verb, Object (SVO) order (for production: Schwartz et al., 1994; Friedmann and Grodzinsky, 1997; Caplan and Hanna, 1998; for comprehension: Caramazza and Zurif, 1976; Schwartz et al., 1980; Caplan and Futter, 1986; Grodzinsky and Finkel, 1998; Friedmann and Shapiro, 2003; Thompson and Shapiro, 2005). Thus, what unites these two groups is their compromised grammatical knowledge, which appears to be a decisive factor for the detection of SynViol. Similarly, while both populations have clear and well-documented difficulties with identifying violations of verb argument structure at the sentence level (Thompson et al., 2012c), the image, as emerged in the current study, is not that clear at the lexical level as there is no statistical difference with the control group when detecting SemViol, that is argument structure violations at the lexical level. If further research establishes this finding, it could suggest that the source of their sentence deficit does not have to do with a loss of argument structure knowledge but with a difficulty processing it at the sentence level, as also suggested in Thompson and Mack (2019).

Results on RTs add a different dimension to the current investigation. First, results of AM controls were in line with previous studies dealing with these types of violations. That is, healthy participants produced distinct RTs for each type of unattested pseudowords, and most importantly, they distinguished SynViol from SemViol. This piece of evidence suggests that speakers are selectively sensitive to levels of linguistic analysis when it comes to lexical processing. This on its own is an important piece of information. However, the interesting issue to be addressed is what this selective sensitivity reflects. It can reflect a qualitative difference between the types of information one needs to evaluate during lexical processing, suggesting an "ease" of detecting a violation of syntactic type in word formation. That is, speakers need more time to evaluate semantic information compared to syntactic information. This is a plausible interpretation, given the nature of the pseudowords used in the current study, as the processor seeks interpretable situations for SemViol, and this "search" could be easily reflected in RTs. On the other hand, the observed pattern might also reflect a deeper architectural mechanism in word-building structure. We will discuss this possibility in the following paragraphs.

This pattern further supports the argument put forward by Manouilidou (2006, 2007), Manouilidou and Stockall (2014), and further validated by Neophytou et al. (2018) and Stockall et al. (2019), that the processing of the grammatical category information temporally precedes the processing of the argument structure information. Looking at the broader picture, these results support the idea that syntactic licensing and semantic composition occur at two distinct stages, the former preceding the latter. With respect to our first question, the two types of violations did not produce distinct RTs for any group of pathological populations. In contrast, RT patterns obtained from PPA-L, PPA-G and StrAg suggest that participants from these groups did not process these two types of critical stimuli at distinct timeframes. This could mean that their overall approach to these types of pseudowords was not to process them at distinct stages but altogether, in a more holistic way. However, they all process them at distinct timeframes compared to NWs and real words, suggesting that for each group SynViol and SemViol are not pure NWs, and that participants tried to interpret them but failed to tell them apart.

Taken together, the results from accuracy and RTs as well as our previous knowledge about the processing for these pseudowords by healthy participants, one can make the following observations. Let us assume a staged lexical access, as outlined in Schreuder and Baayen (1995), Burani et al. (1999) and Fruchter and Marantz (2015). At an initial stage, decomposition occurs, and all lexicalized substrings are exposed. This is when NWs (**repearn*) are processed and rejected as bearers of a non-existent stem. The second stage is where syntactic licensing occurs, and the stage during which SynViol (**recomplex*) are dealt with. The third stage is dedicated to semantic processing or recombination, and it is the stage where SemViol (**reswear*) are processed.

The AM controls follow this pattern as reflected in distinct RTs produced for each category. The lack of difference at the RTs between SynViol and SemViol for all patient groups is suggestive of the following scenarios which should be considered with caution, given the variability among our participants in pathological groups and the confounding effect it might have. First, either stages 2 and 3 are unified as one stage (where both syntactic and semantic information are being processed) or one of them (either syntactic or semantic) is eliminated or skipped depending on the deficit of the specific population. On the grounds of this, let us examine the performance of all groups of participants. Accuracy rates suggest that the two agrammatic groups (PPA-G and StrAg) do not distinguish between SynViol and SemViol. Thus, the first thought would be to assume that agrammatic speakers have one single stage (a combination of stages 2 and 3) where any kind of information is being processed. However, given that the reduced grammatical abilities of these two groups are a strong predictor for accuracy rates, it is plausible to assume that what they miss is the "hardware" to perform syntactic licensing (stage 2), thus judging pseudowords with violations only at the semantic level, where both SynViol and SemViol fail to pass. This pattern explains both the lack of distinct RTs and the lack of distinct accuracy rates for these two groups.

On the other hand, accuracy rates suggest that PPA-L distinguish SynViol from SemViol. Thus, they must have access to the different kinds of information that are violated in each formation. The PPA-L group yielded the highest accuracy rates, and it was the only group which did not differ from controls. It is the group that demonstrates the most consistent (smallest variability) and best-preserved performance when it comes to detecting violations and for telling them apart. This is in accordance with their profile as demonstrated in the literature (Thompson et al., 2012a; Thompson and Mack, 2014). That is, while derivational morphology has not been examined in PPA-L, evidence from inflectional morphology suggests that patients do not have difficulties in the production of morphology. In other words, their performance in processing pseudowords is compatible with their manifested lexical difficulties stemming mostly from the phonological component of lexical knowledge (Mack et al., 2013, 2021), a deficit that could not have interfered with the nature of a lexical decision task. However, their high RTs (overall significantly slower than StrAg and AM) suggest a processing slowdown which could also be responsible for the lack of RT difference between the critical conditions (SynViol vs. SemViol), possibly as a speed-to-accuracy trade-off[6].

Before we conclude anything along the previous lines about PPA-L, an important piece of information that we should consider is the fact that accuracy rates for SynViol do not differ from NWs in this group. This suggests a robust rejection of these formations as pure non-words, possibly by applying a coarse structural well-formedness criterion, rejecting them without hesitation and being unsure about finer-grained distinctions such as SemViol. Even though semantic impairments are not the main feature of PPA-L, there have been studies in the literature, suggesting faulty semantic processing as well (Rogalski et al., 2008; Thompson et al., 2012b; Barbieri et al., 2021). Specifically, in Barbieri et al. (2021), individuals with PPA-L failed to detect violations of argument structure (which constitute the basis of our SemViol) in an EEG sentence processing experiment. Hence, one could claim that the difference between the two (SynViol and SemViol) appears to stem from a sensitivity to what is being violated at the syntactic level and a slight disturbance at the semantic level. Thus, it seems that there is a dichotomy between the two agrammatic groups on the one hand and the PPA-L group on the other hand, with the first ones judging the pseudowords under investigation at a semantic level and the latter ones, judging them at a structural well-formedness level. In fact, such a dichotomy, agrammatic groups on the one hand and PPA-L on the other, has already been manifested in previous studies (Thompson and Mack, 2014) examining grammatical impairments in all variants of PPA.

Thus, if we indeed accept a staged lexical access as outlined in Section "Morphological Processing in Healthy Adults," we will have to assume a two-way performance for our groups of participants. Specifically, agrammatic groups fail to fully apply the syntactic licensing criterion – they judge them at the semantic level (different from AM when it comes to SynViol), a judgment which produces similar accuracy at similar timeframes for both SynViol and SemViol. Ultimately, the two agrammatic groups are not selectively sensitive to various levels of linguistic analysis, as they treat both violations as semantic. Finally, PPA-L demonstrates performance with the highest accuracy rates (like AM controls), an indication of a preserved ability to process morphologically complex words, albeit with the slight interference of a possible semantic disturbance [as in Barbieri et al. (2021)].

Finally, we will conclude this section with a comment on the issue of variability. Variability among participants has been a feature of many pathological conditions and it is very well manifested in aphasia. It has also been one of the methodological challenges in group studies. Genuine individual differences exist in every aspect of human existence. It is the challenge for the researcher to pin down their source, to the extent that this is possible. In our study, within group variability is undeniable and it is mostly manifested within the two agrammatic groups. However, when controlling for this individual variation by using participants as a random factor in our mixed models, a uniform pattern emerges, and it is in accordance with the patients' clinical and cognitive profile. Furthermore, by modeling for language measures, we have shown how variation (in dealing with these pseudowords that deviate from canonicity) can be understood and we have identified its source.

GENERAL DISCUSSION AND CONCLUSION

Language research on brain-damaged populations is informative for two main reasons; first it contributes to the understanding of the pathology; second it allows us to learn more about the normal process. Before we conclude, we will address these two points having in mind the findings of the present study.

The hallmark of PPA is impaired *word knowledge*. Given the vastness of this, the current study is the first one attempting to

[6]Similar speed-to-accuracy trade-off effects were also reported for a combined PPA-L/PPA-G group in a word comprehension eye-tracking study (Seckin et al., 2016).

shed light onto finer aspects of word knowledge in PPA, by using a linguistically informed approach in order to provide detailed profiles of linguistic strengths and weaknesses of the populations under investigation. We focused on complex pseudowords, aiming at investigating morphological processing, an under-studied domain when it comes to language disorders. As outlined in the introduction, morphological processing requires the combination of knowledge of various linguistic domains, such as syntax and semantics. With this in mind, we aimed at examining how the specifics of each PPA variant under consideration could be affected.

This study allowed us to confirm some facts about the different variants and it also brought into light new insights. First, the study provides evidence for a unified effect of agrammatism, resulting from stroke and from a neurodegenerative disease, at the lexical level. What we knew up until now is that the two populations demonstrate similar performance at the sentence level and in syntactic tasks (Thompson et al., 2012c; Thompson and Mack, 2014). The current study brought into light striking similarities at the lexical level as well suggesting that both groups operate in the same way when judging pseudowords as well. Given that their performance correlates with their weak grammatical abilities altogether, we have evidence that they rely on their semantic knowledge rather than on anything else in order to process these pseudowords.

The current dataset also brought into light a dichotomy between the two agrammatic groups and PPA-L, as it is also reported in Thompson and Mack (2014). Results are in line with the profiles of PPA-L, as manifested in the literature, that is, a relatively good performance of PPA-L at detecting violations at the lexical level (no difference compared to the AM group). Given the scarcity of chronometrized studies when it comes to PPA, what we did not know before is that PPA-L shows a speed-accuracy trade-off effect, suggestive of their strategy in dealing with these pseudowords. In other words, this group approaches with caution the lexical decision task, taking time in using their relatively preserved abilities.

Overall, the novelty of the current study with respect to PPA is that it provides an explanation for what "impaired word knowledge" could mean by revealing the different strategies of these populations when confronted with pseudowords, thus allowing a window to our understanding on how these populations treat any complex lexical item. Therefore, when we say that PPA affects word knowledge, the current study offers an account as to what might be the underlying reason for failing word knowledge for the variants under consideration.

Looking at the other side of the coin, the present study offered an alternative way of looking at morphological operations. Most psycholinguistic literature postulates the existence of various stages in accessing complex pseudowords, each stage being devoted to the processing of specific types of information. The present study confirms this procedure, albeit in an alternative way. The lack of time differences in the processing of SynViol vs. SemViol does not allow us to clearly talk about temporal stages. However, combined results from RTs and accuracy confirm the different types of information that are involved in these types of structures.

First, looking at the performance of PPA-G and StrAg when it comes to SynViol and the fact that this performance is predicted by their weak grammatical abilities altogether, we have a first-hand piece of evidence that grammatical knowledge is at stake when it comes to processing these pseudowords. Alternatively seen, syntactic licensing is an obligatory step in complex word recognition, a step which is being compromised by agrammatism. Taken together with their control-like performance for the SemViol condition, we have the second piece of evidence that although SemViol words result from violating argument structure specifications, they are ultimately processed at a semantic level, as semantic recomposition suggests (Fruchter and Marantz, 2015). This distinction between the types of information being processed is further reinforced by the performance of the PPA-L group.

Thus, the current study evidently and inevitably provides further input to our knowledge about morphological processing of complex words in a totally innovative way. Empirical evidence of this type constitutes a contribution to our perception of morphology which is beyond the theoretical level. Given the increase of linguistically informed research in language disorders, the role of this type of study to our understanding of normal language may turn out to be vital, in a way that, until recently, might have looked unimaginable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB) of Northwestern University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

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

FUNDING

This work was supported by the NIH-NIDCD, Clinical Research Center Grant P50DC012283 (PI: CT), R01DC01948 (PI: CT), R01DC008552 (PI: MM) and R01AG13854 (PI: MM). Writing of the chapter was partly supported by grant J6-1806 from ARRS (Slovenian Research Agency) awarded to CM. The Fulbright Foundation was also acknowledged for funding CM's visiting professorship at Northwestern University.

ACKNOWLEDGMENTS

We would like to thank the research participants and their families and caregivers for their contributions to this work, as well as colleagues at the Mesulam Cognitive Neurology and Alzheimer's Disease Center and the Aphasia and Neurolinguistics Research Laboratory at Northwestern University. Control participants were recruited in part through the Communication Research Registry at Northwestern University. We are grateful to Elena Barbieri for her help and guidance with the statistical analysis and interpretation of the results, to Davide Crepaldi for commenting thoroughly on the manuscript and to Jennifer Mack for useful discussions on various parts of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.701802/full#supplementary-material

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

Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., et al. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology* 81, 329–336. doi: 10.1212/WNL.0b013e31829c5d0e

Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX Lexical Database*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., and Yap, M. J. (2004). Visual word recognition of single-syllable words. *J. Exp. Psychol. General* 133, 283–316. doi: 10.1037/0096-3445.133.2.283

Barbieri, E., Litcofsky, K. A., Walenski, M., Chiappetta, B., Mesulam, M. M., and Thompson, C. K. (2021). Online sentence processing impairments in agrammatic and logopenic primary progressive aphasia: evidence from ERP. *Neuropsychologia* 151:107728. doi: 10.1016/j.neuropsychologia.2020.107728

Bastiaanse, R., and Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology* 12, 951–969. doi: 10.1080/02687039808249463

Bates, D. M., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed effects models using lme4. *J. Statistical Software* 67, 1–48.

Bertram, R., Schreuder, R., and Baayen, R. H. (2000). The balance of storage and computation in morphological processing: the role of word formation type, affixal homonymy, and productivity. *J. Exp. Psychol. Learn. Memory Cogn.* 26, 489–511. doi: 10.1037/0278-7393.26.2.489

Brennan, J., Lignos, C., Embick, D., and Roberts, T. P. (2014). Spectro-temporal correlates of lexical access during auditory lexical decision. *Brain Lang.* 133, 39–46. doi: 10.1016/j.bandl.2014.03.006

Bretz, F., Hothorn, T., and Westfall, P. (2010). *Multiple Comparisons Using R*. Boca Raton, FL: CRC Press.

Burani, C., Dovetto, F. M., Spuntarelli, A., and Thornton, A. M. (1999). Morpholexical access and naming: the semantic interpretability of new root–suffix combinations. *Brain Lang.* 68, 333–339. doi: 10.1006/brln.1999.2073

Butterworth, B. L. (1983). "Lexical representation," in *Language Production*, ed. B. L. Butterworth (London: Academic Press).

Caplan, D., and Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain Lang.* 27, 117–134. doi: 10.1016/0093-934x(86)90008-8

Caplan, D., and Hanna, J. (1998). Sentence production by aphasic patients in a constrained task. *Brain Lang.* 63, 184–218. doi: 10.1006/brln.1998.1930

Caramazza, A., Laudanna, A., and Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*. 28, 297–332. doi: 10.1016/0010-0277(88)90017-0

Caramazza, A., and Zurif, E. (1976). Dissociations of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain Lang.* 3, 572–582. doi: 10.1016/0093-934x(76)90048-1

Chialant, D., and Caramazza, A. (1995). "Where is morphology and how is it processed? the case of written word recognition," in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Erlbaum Press).

Chuang, Y. Y., Vollmer, M. L., Shafaei-Bajestan, E., Gahl, S., Hendrix, P., Baayen, R. H., et al. (2021). The processing of pseudoword form and meaning in production and comprehension: a computational modeling approach using linear discriminative learning. *Behav. Res.* 53, 945–976. doi: 10.3758/s13428-020-01356-w

del Prado, Martín, F. M., Kostić, A., and Baayen, R. H. (2004). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94, 1–18. doi: 10.1016/j.cognition.2003.10.015

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

Dikker, S., Rabagliati, H., and Pylkkänen, L. (2009). Sensitivity to syntax in visual cortex. *Cognition* 110, 293–321. doi: 10.1016/j.cognition.2008.09.008

Dominguez, A., Alija, M., Rodriguez-Ferreiro, J., and Cuetos, F. (2010). The contribution of prefixes to morphological processing of Spanish words. *Eur. J. Cogn. Psychol.* 22, 569–595. doi: 10.1080/09541440903007792

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness. a Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Faroqi-Shah, Y., and Thompson, C. K. (2010). Production latencies of morphologically simple and complex verbs in aphasia. *Clin. Linguist. Phon.* 24, 963–979. doi: 10.3109/02699206.2010.488314

Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., et al. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex* 55, 43–60. doi: 10.1016/j.cortex.2012.12.006

Frauenfelder, U., and Schreuder, R. (1991). "Constraining psycholinguistic models of morphological processing and representation: the role of productivity," in *Yearbook of Morphology 1991. Yearbook of Morphology*, eds G. Booij and J. van Marle (Dordrecht: Springer).

Friedmann, N., and Grodzinsky, Y. (1997). Tense and agreement in agrammatic production pruning the syntactic tree. *Brain Lang.* 56, 397–425. doi: 10.1006/brln.1997.1795

Friedmann, N., and Shapiro, L. (2003). Agrammatic comprehension of simple active sentences with moved constituents. *J. Speech Lang. Hear Res.* 46, 288–297. doi: 10.1044/1092-4388(2003/023)

Fruchter, J., and Marantz, A. (2015). Decomposition, lookup, and recombination: MEG evidence for the full decomposition model of complex visual word recognition. *Brain Lang.* 143, 81–96. doi: 10.1016/j.bandl.2015.03.001

Fruchter, J., Stockall, L., and Marantz, A. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. *Front. Hum. Neurosci.* 22:798. doi: 10.3389/fnhum.2013.00798

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. E. E. A., et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology* 76, 1006–1014.

Grodzinsky, Y., and Finkel, L. (1998). The neurology of empty categories: Aphasics' failure to detect ungrammaticality. *J. Cogn. Neurosci.* 10, 281–291. doi: 10.1162/089892998562708

Hahne, A., and Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *J. Cogn. Neurosci.* 11, 194–205. doi: 10.1162/089892999563328

Hahne, A., and Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? an ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cogn. Brain Res.* 11, 199–212. doi: 10.1016/s0926-6410(00)00071-9

Kertesz, A. (2006). *Western Aphasia Battery-Revised*. San Antonio, TX: Pearson.

Kim, S. Y., Wang, M., and Taft, M. (2015). Morphological decomposition in the recognition of prefixed and suffixed words: evidence from Korean. *Sci. Stud. Read.* 19, 183–203. doi: 10.1080/10888438.2014.991019

Knibb, J. A., Woollams, A. M., Hodges, J. R., and Patterson, K. (2009). Making sense of progressive non-fluent aphasia: an analysis of conversational speech. *Brain* 132(Pt 10), 2734–2746. doi: 10.1093/brain/awp207

Kordouli, K., Manouilidou, C., Stavrakaki, S., Mamouli, D., and Ioannidis, P. (2018). Compound production in agrammatism: evidence from stroke-induced and primary progressive Aphasia. *J. Neurolinguist.* 47, 71–90. doi: 10.1016/j.jneuroling.2018.02.009

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). lmerTest: tests in linear mixed effects models. *J. Statistical Software* 82, 1–26.

Laudanna, A., Badecker, W., and Caramazza, A. (1992). Processing inflectional and derivational morphology. *J. Mem. Lang.* 31, 333–348. doi: 10.1016/0749-596X(92)90017-R

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

Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappi, O., Laine, M., et al. (2009). Neurocognitive processing of auditory and visually presented inflected words and pseudowords: evidence from a morphologically rich language. *Brain Res.* 1275, 54–66. doi: 10.1016/j.brainres.2009.03.057

Leminen, A., Lehtonen, M., Leminen, M., Nevalainen, P., Mäkelä, J., and Kujala, T. (2013). The role of attention in processing morphologically complex spoken words: an EEG/MEG study. *Front. Hum. Neurosci.* 6:353. doi: 10.3389/fnhum.2012.00353

Leminen, A., Smolka, E., Dunabeitia, J. A., and Pliatsikas, C. (2019). Morphological processing in the brain: the good (inflection), the bad (derivation) and the ugly (compounding). *Cortex* 116, 4–44. doi: 10.1016/j.cortex.2018.08.016

Linzen, T., Marantz, A., and Pylkkanen, L. (2013). Syntactic context effects in visual word recognition: an MEG study. *Mental Lexicon* 8, 117–139. doi: 10.1075/ml.8.2.01lin

Lukatela, G., Carello, C., and Turvey, M. T. (1987). Lexical representation of regular and irregular inflected nouns. *Lang. Cogn. Proc.* 2, 1–17. doi: 10.1080/01690968708406349

Lukic, S. B., Yoshida, M., and Thompson, C. K. (2016). The role of zero-derivation in lexical processing. *Front. Psychol.* 7:137. doi: 10.3389/conf.fpsyg.2016.68.00137

Mack, J. E., Barbieri, E., Weintraub, S., Mesulam, M. M., and Thompson, C. K. (2021). Quantifying grammatical impairments in primary progressive aphasia: structured language tests and narrative language production. *Neuropsychologia* 151:107713. doi: 10.1016/j.neuropsychologia.2020.107713

Mack, J. E., Cho-Reyes, S., Kloet, J. D., Weintraub, S., Mesulam, M. M., and Thompson, C. K. (2013). Phonological facilitation of object naming in agrammatic and logopenic primary progressive aphasia (PPA). *Cogn. Neuropsychol.* 30, 172–193. doi: 10.1080/02643294.2013.835717

Manouilidou, C. (2006). *Processing of Thematic Features in Deverbal Nominals*. PhD Dissertation, Ottawa, ON: University of Ottawa.

Manouilidou, C. (2007). "Thematic constraints in deverbal word formation: psycholinguistic evidence from pseudowords," in *Proceedings of the 7th International Conference on Greek Linguistics*, (London: University of York).

Manouilidou, C., Dolenc, B., Marvin, T., and Pirtošek, Z. (2016). Processing complex pseudowords in Mild cognitive impairment: the interaction of preserved morphological rule knowledge with compromised cognitive ability. *Clin. Linguist. Phon* 30, 49–67. doi: 10.3109/02699206.2015.1102970

Manouilidou, C., and Stockall, L. (2014). Teasing apart syntactic category vs. argument structure information in deverbal word formation: a comparative psycholinguistic study. *Italian J. Linguistics* 26, 71–98.

Marangolo, P., Incoccia, C., Pizzamiglio, L., Sabatini, U., Castriota-Scanderbeg, A., and Burani, C. (2003). The right hemisphere involvement in the processing of morphologically derived words. *J. Cogn. Neurosci.* 15, 364–371. doi: 10.1162/089892903321593090

Marcotte, K., Graham, N. L., Fraser, K. C., Meltzer, J. A., Tang-Wai, D. F., Chow, T. W., et al. (2017). White matter disruption and connected speech in non-fluent and semantic variants of primary progressive Aphasia. *Dementia Geriatric Cogn. Disord. Extra* 7, 52–73. doi: 10.1159/000456710

Maruta, C., Pereira, T., Madeira, S. C., De Mendonça, A., and Guerreiro, M. (2015). Classification of primary progressive aphasia: do unsupervised data mining methods support a logopenic variant? *Amyotrophic Lateral Sclerosis Frontotemporal Degenerat.* 16, 147–159. doi: 10.3109/21678421.2015.1026266

Mesulam, M. M. (1982). Slowly progressive aphasia without generalized dementia. *Ann. Neurol.* 11, 592–598.

Mesulam, M. M. (2001). Primary progressive aphasia. *Ann. Neurol.* 49, 425–432.

Mesulam, M. M. (2003). Primary progressive aphasia – a language-based dementia. *N. Engl. J. Med.* 349, 1535–1542. doi: 10.1056/nejmra022435

Mesulam, M. M. (2013). Primary progressive aphasia: a dementia of the language network. *Dementia Neuropsychol.* 7, 2–9. doi: 10.1590/s1980-57642013dn70100002

Miceli, G., Capasso, R., and Caramazza, A. (2004). The relationships between morphological and phonological errors in aphasic speech: data from a word repetition task. *Neuropsychologia* 42, 273–287. doi: 10.1016/j.neuropsychologia.2003.09.002

Miceli, G., and Caramazza, A. (1988). Dissociation of inflectional and derivational morphology. *Brain Lang.* 35, 24–65. doi: 10.1016/0093-934x(88)90100-9

Münte, F., Say, T., Clahsen, H., Schiltz, K., and Kutas, M. (1999). Decomposition of morphologically complex words in English: evidence from event-related brain potentials. *Cogn. Brain Res.* 7, 241–253. doi: 10.1016/s0926-6410(98)00028-7

Neophytou, K., Manouilidou, C., Stockall, L., and Marantz, A. (2018). Processing of lexical category and argument structure information in deverbal adjectives: an MEG study on Greek. *Brain Lang.* 183, 11–20.

Olsson-Collentine, A., van Assen, M. A. L. M., and Hartgerink, C. H. J. (2019). The prevalence of marginally significant results in psychology over time. *Psychol. Sci.* 30, 576–586. doi: 10.1177/0956797619830326

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rastle, K., and Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Lang. Cogn. Processes* 23, 942–971. doi: 10.1080/01690960802069730

Ripamonti, E., Traficante, D., and Luzzatti, E. C. (2015). L'effetto di superiorità della parola e della pseudoparola: una rassegna della letteratura e prospettive future [Word and pseudoword superiority effects: a literature review and future perspectives]. *Giornale Italiano Psicol.* 42, 711–751.

Rogalski, E., Rademaker, A., Mesulam, M. M., and Weintraub, S. (2008). Covert processing of words and pictures in non semantic variants of primary progressive aphasia. *Alzheimer's Dis. Assoc. Disord.* 22, 343–351. doi: 10.1097/WAD.0b013e31816c92f7

Royle, P., Drury, J., Bourguignon, N., and Steinhauer, K. (2012). The temporal dynamics of inflected word recognition: a masked ERP priming study of French verbs. *Neuropsychologia* 50, 3542–3553. doi: 10.1016/j.neuropsychologia.2012.09.007

Schreuder, R., and Baayen, H. (1995). "Modeling morphological processing," in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Lawrence Erlbaum Associates).

Schwartz, M. F., Saffran, E. M., Fink, R. B., Myers, J. L., and Martin, N. (1994). Mapping therapy: a treatment programme for agrammatism. *Aphasiology* 8, 19–54. doi: 10.1080/02687039408248639

Schwartz, M. F., Saffran, E. M., and Marin, O. S. M. (1980). The word order problem in Agrammatism: I. *Comprehension. Brain Lang.* 10, 249–262.

Seckin, M., Mesulam, M. M., Voss, J. L., Huang, W., Rogalski, E. J., and Hurley, R. S. (2016). Am I looking at a cat or a dog? gaze in the semantic variant of primary progressive aphasia is subject to excessive taxonomic capture. *J. Neurolinguistics* 37, 68–81. doi: 10.1016/j.jneuroling.2015.09.003

Semenza, C., Girelli, L., Spacal, M., Kobal, J., and Mesec, A. (2002). Derivation by prefixation in Slovenian: two aphasia case studies. *Brain Lang.* 81, 242–249. doi: 10.1006/brln.2001.2521

Sereno, J. A., and Jongman, A. (1997). Processing of English inflectional morphology. *Mem. Cognit.* 25, 425–437. doi: 10.3758/bf03201119

Stockall, L., Manouilidou, Ch, Gwilliams, L., Neophytou, K., and Marantz, A. (2019). Prefix stripping re-re-revisited: MEG investigations of morphological decomposition and recomposition. *Front. Psychol.* 10:1964. doi: 10.3389/fpsyg.2019.01964

Taft, M. (1988). A morphological decomposition model of lexical representation. *Linguistics* 26, 657–667.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quar. J. Exp. Psychol. Section A* 57, 745–765. doi: 10.1080/02724980343000477

Taft, M., and Forster, K. (1975). Lexical storage and retrieval of prefixed words. *J. Verb. Learn. Verbal Behav.* 14, 638–647. doi: 10.1016/s0022-5371(75)80051-x

Team, R. (2018). *RStudio: Integrated Development*. Boston, MA: R. RStudio, Inc.

Thompson, C. K. (2011). *Northwestern Assessment of Verbs and Sentences*. Evanston, IL: Northwestern University.

Thompson, C. K., and Bastiaanse, R. (2012). "Introduction to agrammatism: brain, behaviour and cognition," in *Perspectives on Agrammatism: Brain, Behaviour and Cognition*, eds R. Bastiaanse and C. K. Thompson (Hove: Psychology Press).

Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B. B., et al. (2012a). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology* 26, 20–43.

Thompson, C. K., Cho, S., Price, C., Wieneke, C., Bonakdarpour, B., Rogalski, E., et al. (2012b). Semantic interference during object naming in agrammatic and logopenic primary progressive aphasia (PPA). *Brain Lang.* 120, 237–250. doi: 10.1016/j.bandl.2011.11.003

Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., and Weintraub, S. (2012c). Verb and noun deficits in stroke-induced and primary progressive aphasia: the Northwestern Naming Battery. *Aphasiology* 26, 632–655. doi: 10.1080/02687038.2012.676852

Thompson, C. K., Lange, K. L., Schneider, S. L., and Shapiro, L. P. (1997). Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. *Aphasiology* 11, 473–490. doi: 10.4102/sajcd.v62i1.89

Thompson, C. K., and Mack, J. E. (2014). Grammatical impairments in PPA. *Aphasiology* 28, 1018–1037. doi: 10.1080/02687038.2014.912744

Thompson, C. K., and Mack, J. E. (2019). "Neurocognitive mechanisms of agrammatism," in *Oxford Handbook of Neurolinguistics*, eds G. Zubicaray and N. O. Schiller (Oxford: Oxford University Press).

Thompson, C. K., Meltzer-Asscher, A., Cho, S., Lee, J., Wieneke, C., and Weintraub, S. (2013). Syntactic and morphosyntactic processing in stroke-induced and primary progressive aphasia. *Behav. Neurol.* 26, 35–54. doi: 10.3233/BEN-2012-110220

Thompson, C. K., and Shapiro, L. (2005). Treating agrammatic aphasia within a linguistic framework: treatment of underlying forms. *Aphasiology* 19, 1021–1036.

Thompson, C. K., and Weintraub, S. (2014). *The Northwestern Naming Battery (NNB)*. Evanston, IL: Northwestern University.

Whiting, C., Shtyrov, Y., and Marslen-Wilson, W. (2014). Real-time functional architecture of visual word recognition. *J. Cogn. Neurosci.* 27, 246–265. doi: 10.1162/jocn_a_00699

Wurm, L., Ernestus, M., Schreuder, R., and Baayen, R. H. (2006). Dynamics of the auditory comprehension of prefixed words. *Mental Lexicon* 1, 125–146.

Zunini, R. A. L., Baart, M., Samuel, A. G., and Armstrong, B. C. (2020). Lexical access versus lexical decision processes for auditory, visual, and audiovisual items: insights from behavioral and neural measures. *Neuropsychologia* 137:107305. doi: 10.1016/j.neuropsychologia.2019.107305

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

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fpsyg.2022.754395

Check for updates

Understanding the Phonetic Characteristics of Speech Under Uncertainty—Implications of the Representation of Linguistic Knowledge in Learning and Processing

Fabian Tomaschek* and Michael Ramscar

Quantitative Linguistics Lab, Department of General Linguistics, University of Tübingen, Tübingen, Germany

The uncertainty associated with paradigmatic families has been shown to correlate with their phonetic characteristics in speech, suggesting that representations of complex sublexical relations between words are part of speaker knowledge. To better understand this, recent studies have used two-layer neural network models to examine the way paradigmatic uncertainty emerges in learning. However, to date this work has largely ignored the way choices about the representation of inflectional and grammatical functions (IFS) in models strongly influence what they subsequently learn. To explore the consequences of this, we investigate how representations of IFS in the input-output structures of learning models affect the capacity of uncertainty estimates derived from them to account for phonetic variability in speech. Specifically, we examine whether IFS are best represented as outputs to neural networks (as in previous studies) or as inputs by building models that embody both choices and examining their capacity to account for uncertainty effects in the formant trajectories of word final [ɐ], which in German discriminates around sixty different IFS. Overall, we find that formants are enhanced as the uncertainty associated with IFS decreases. This result dovetails with a growing number of studies of morphological and inflectional families that have shown that enhancement is associated with lower uncertainty in context. Importantly, we also find that in models where IFS serve as inputs—as our theoretical analysis suggests they ought to—its uncertainty measures provide better fits to the empirical variance observed in [ɐ] formants than models where IFS serve as outputs. This supports our suggestion that IFS serve as cognitive cues during speech production, and should be treated as such in modeling. It is also consistent with the idea that when IFS serve as inputs to a learning network. This maintains the distinction between those parts of the network that represent message and those that represent signal. We conclude by describing how maintaining a "signal-message-uncertainty distinction" can allow us to reconcile a range of apparently contradictory findings about the relationship between articulation and uncertainty in context.

Keywords: linguistic knowledge, discriminative learning, cue-to-outcome structure, morphological structure, phonetic characteristics, reduction, enhancement, context

OPEN ACCESS

Edited by:
Robert Malouf, San Diego State University, United States

Reviewed by:
Themis Karaminis, Edge Hill University, United Kingdom
Claudia Marzi, Istituto di linguistica computazionale "Antonio Zampolli" (ILC), Italy

****Correspondence:***
Fabian Tomaschek
fabian.tomaschek@uni-tuebingen.de

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

Received: *06 August 2021*
Accepted: *24 March 2022*
Published: *25 April 2022*

Citation:
Tomaschek F and Ramscar M (2022) Understanding the Phonetic Characteristics of Speech Under Uncertainty—Implications of the Representation of Linguistic Knowledge in Learning and Processing. Front. Psychol. 13:754395. doi: 10.3389/fpsyg.2022.754395

1. INTRODUCTION

The phonetic characteristics of speech signals are highly variable. Separating the variability that is simply noise from that which is informative is central to our understanding of speech. Some parts of this problem have been solved. It is known that variability occurs in relation to coarticulation (e.g., Öhman, 1966; Zsiga, 1992; Magen, 1997), speaking rate (e.g., Lindblom, 1963; Gay, 1978), syllable position (Pouplier and Hoole, 2016), prosody (Mooshammer and Fuchs, 2002; Mücke et al., 2009) and even the idiosyncrasies of speakers (e.g., Tomaschek and Leeman, 2018; Gittelson et al., 2021). By contrast, there is still much debate about the way that representations of linguistic knowledge—and the differing levels of uncertainty associated with this knowledge—serve to co-determine articulation, and in turn the phonetic characteristics of speech. This is especially the case when it comes to the representation of words within inflectional paradigms and the way that the uncertainty associated with different word-forms correlates with fine phonetic detail in the speech signal. Some studies report effects of reduction associated with lower *paradigmatic uncertainty*—mirroring findings within the information theoretic and the *Smooth Signal Redundancy Hypothesis* framework. By contrast, work within the *Paradigmatic Signal Enhancement Hypothesis* framework reports the enhancement of phonetic characteristics (these findings are discussed in detail below).

In what follows, we investigate these effects by addressing the relationship between the uncertainty associated with the inflectional functions of German word-final [ɐ], as in the word *Lehrer* [ˈleːʁɐ] "teacher", and the phonetic characteristics of [ɐ]. This phone discriminates roughly sixty different grammatical and inflectional functions in German, in morphologically simple and complex words, making it an ideal test bed for this research.

One potential confound in the earliest studies investigating the effects of sublexical relationships on articulation lies in their operationalizations of paradigmatic relations, which were based on theoretically motivated definitions of word-internal structure. To avoid having to make these kinds of assumptions, we follow the approach of Tucker et al. (2019) and Tomaschek et al. (2019) who investigated these phenomena from a discriminative learning perspective. In this approach, which employs a simple neural network trained with an error-driven learning algorithm (widely known as the *delta-rule*), paradigmatic uncertainty is an emergent property within lexical systems, which develops as the individual items it comprises are learned. In doing this, we shall also address some often neglected questions that this approach raises. Psycholinguistic studies using neural networks have typically ignored the way that implementational choices concerning the relationships between inputs and outputs in a network can shape its performance. However, as Bröker and Ramscar (2020) demonstrate, decisions about the input-output structure of computational learning models serve to co-determine what these models actually learn. This in turn affects researchers' interpretations of the performance of models in relation to their theoretical contribution. Accordingly—and in line with the topic of this special issue—a further aim of this work will be the investigation of the kind of input-output structure that is most appropriate for the representation of morphological and inflectional paradigms. Specifically, we shall examine whether inflectional functions of [ɐ] are best characterized as serving as inputs to neural networks or as their outputs, as implemented in Tucker et al. (2019) and Tomaschek et al. (2019).

To analyze the performance of our network models (which we also describe in detail below), we use simulated activations as a measure of the uncertainty associated with each inflectional function. These are regressed against the phonetic characteristics of [ɐ] in order to assess their capacity to predict the phonetic characteristics of the speech signal. We show an enhancement of [ɐ]'s phonetic characteristics associated with lower paradigmatic uncertainty. Critically, we find that when inflectional functions of [ɐ] serve as inputs to the learning network, uncertainty associated with these functions obtained from the network is a better statistical predictor for [ɐ]'s phonetic characteristics than when inflectional functions serve as outputs. Accordingly, the present study contributes to a line of research that investigates how uncertainty affects speech production through a combination of computational modeling of learning and an examination of the predictions of these models for the phonetic characteristics of actual speech (for example Baayen et al., 2019; Tomaschek et al., 2019; Tucker et al., 2019; Stein and Plag, 2021; Schmitz et al., 2021b in the present special issue).

We begin by discussing the empirical and theoretical background of this study, as well as previous work by Tucker et al. (2019) and Tomaschek et al. (2019) that we seek to further examine. We then describe our simulations and analyses before discussing the theoretical and computational implications of our results.

2. BACKGROUND

2.1. Phonetic Characteristics and Paradigmatic Probability

It is well-established that phonetic reductions occur in contexts where syntagmatic uncertainty is low. Lower uncertainty has been shown to be associated with shorter words, syllables and segments (Aylett and Turk, 2004; Cohen Priva, 2015) and more centralized vowels (Wright, 2004; Aylett and Turk, 2006; Munson, 2007; Malisz et al., 2018; Brandt et al., 2019). This has been demonstrated by studies that operationalized uncertainty by means of word frequency (Wright, 1979, 2004; Fosler-Lussier and Morgan, 1999; Bybee, 2002), conditional probability (Jurafsky et al., 2001a,b; Aylett and Turk, 2004; Bell et al., 2009), or informativity (Cohen Priva, 2015; Schulz et al., 2016; Malisz et al., 2018; Brandt et al., 2019, 2021). Aylett and Turk (2004, 2006)'s *Smooth Signal Redundancy Hypothesis* explains these *reduction* phenomena from an information theoretic perspective (Shannon, 1948), arguing that the amount of information in the speech signal is balanced against the amount of information conveyed at the syntagmatic level. These systematic findings sparked a line of research that investigated whether equivalent changes in phonetic characteristics can be found when uncertainty is operationalized within other contexts, such as morphological and paradigmatic families.

However, while there is an abundance of evidence showing a systematic relation between uncertainty within these contexts and the phonetic characteristics of speech, when it comes to uncertainty within morphological families, the effects of this relationship seems to run in the *opposite* direction to those reported at the syntagmatic level. Numerous studies have shown lower uncertainty within morphological families to be associated with *enhancement*. This is reflected in longer word durations (Lõo et al., 2018) and consonant durations at compound boundaries (Bell et al., 2019), in longer interfixes in Dutch compounds (Kuperman et al., 2007), in more enhanced articulatory positions in stem vowels of English verbs (Tomaschek et al., 2021), in lower deletion probabilities of the word final [t] in Dutch words (Schuppler et al., 2012) and in Dutch regular past-participles (Hanique and Ernestus, 2011), and in less centralized vowel articulations in Russian verbal suffixes (Cohen, 2015). Kuperman et al. (2007) have proposed the *Paradigmatic Signal Enhancement Hypothesis* to provide a theoretical formalization of these patterns of findings, arguing that phonetic enhancements are a consequence of the greater levels of paradigmatic support that these voicings receive. However, while it might may seem that the findings just discussed appear to contradict one another, it is not entirely clear whether they actually do.

This is because although the studies just described do appear to support the *Paradigmatic Signal Enhancement Hypothesis*, other studies have found an opposite effect, demonstrating an association between lower uncertainty in morphological and paradigmatic families and *reduction*. This is reflected, for example, in higher deletion probability of [t] in derived adverbs (e.g., *swiftly*) (Hay, 2004) and in Dutch irregular past-participles (Hanique and Ernestus, 2011), in shorter [ə] durations in Dutch prefixes (Hanique and Ernestus, 2011), in shorter duration of English prefixes and their consonants (Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2017), and finally, in more centralized [-i] and [-o] when they serve as suffixes in Russian (Cohen, 2015). The different effects associated with paradigmatic uncertainty—enhancement or reduction—emerge independently of the kind of probabilistic measure used to operationalize uncertainty in the domain of morphological and paradigmatic families. That is, regardless of whether paradigmatic uncertainty is operationalized as family size, as word frequency divided by the summed frequency of all the words in a paradigm, or as the frequency of a morphologically complex word divided by its base frequency.

Thus far in this discussion, we have treated the idea of uncertainty in linguistic knowledge as if it is an objective matter of fact. There are, however, good reasons to believe this is not the case. First, because all of the measures used to operationalize the uncertainty associated with different kinds of knowledge are based on theoretical assumptions. Second, because these theoretical assumptions typically disregard the fact that all morphological knowledge is *learned*. Since languages are learned, it necessarily follows that the word-internal structures and distinctions posited by any given theory are unlikely to correspond exactly to the structures and distinctions that have actually been learned by a given speaker at any given point in time.

Tucker et al. (2019) and Tomaschek et al. (2019)'s solution to this problem was to model learning by means of a two-layer neural network that was trained with an error-driven learning rule (the delta rule Rescorla and Wagner (1972), Rumelhart and McCelland (1987), provided by the Naive Discriminative Learner package in R, Arppe et al., 2018). If trained in a naive way, the neural network does not explicitly embody the structures of linguistic knowledge that are typically assumed in psycholinguistic theories. Rather, the model's representation of these structures emerges in bottom-up fashion, as a result of training the network. As a consequence, knowledge in the model is represented by the distribution of its connection weights such that "morphological structure" emerges gradually, in gradient fashion, as the model is trained[1].

Tucker et al. (2019) and Tomaschek et al. (2019) used network measures to operationalize uncertainty within a morphological paradigm. The results of these studies showed lower uncertainty to be associated with longer stem vowel duration in regular and irregular English verbs and longer duration of word final [s] that encodes multiple inflectional functions (plural noun, genitive, second person singular verbs, etc.). Accordingly, these results provided evidence to corroborate the claim that phonetic enhancement is associated with lower paradigmatic uncertainty.

Because the present study builds on the work by Tucker et al. (2019) and Tomaschek et al. (2019), we shall need to discuss their models and input-output structures in detail. However, before we can do so, it is first important that we flesh out the theoretical background to this work. This is because, as we noted above, we do not only aim to examine the relation between paradigmatic uncertainty and articulation here. Our goal is also to provide a theoretical examination of the way that the various factors that contribute and provide evidence for these effects are best represented in neural network models (see also Bröker and Ramscar, 2020; Ramscar, 2021a).

Accordingly, we shall begin by discussing how previous computational models of speech production have addressed these issues, and how they were used to make predictions about the phonetic characteristics of speech. Then, since both Tucker et al. (2019) and Tomaschek et al. (2019) are rooted in the theory of *discriminative learning*, a cognitive theory of how language (and actually any kind of behavior) is learned (Ramscar and Yarlett, 2007; Ramscar et al., 2010, 2013b; Ramscar, 2019, 2021b), we shall examine the constraints that this theory imposes on the way the input-output structure of models is configured.

2.2. Computational Models of Speech Production

Researchers in the twentieth century collected a great deal of information in the form of speech errors and data from controlled psycho-linguistic experiments. This information then informed theoretical speculations about the nature of the speech production process (e.g., Fromkin, 1971; Levelt et al., 1999). While these psycho-linguistic theories are useful at a general

[1]The way that knowledge about input-output structures is represented in a network trained by the error-driven learning rule is neatly demonstrated by Hoppe et al. (2022).

level, they are subject to the standard limitations of all verbal theories. One of the limitations is that they are open to interpretation and that they are often vague when it comes to the specific details of processing. Computational models, such as those presented by Dell (1986) and Roelofs (1997) ameliorate these problems of vagueness. These models force language researchers to make definitive commitments regarding the detailed structure of processes, regarding the kinds of algorithms involved and, of importance to the present study, regarding the structure of the representations that are required to model speech production. In return for these commitments, researchers are not only able to eliminate some of the vaguenesses in theory, they are also able to obtain quantitatively testable predictions. While most research on computational models of speech production has focused on the structure of models at an algorithmic level, the structure of the input and output to/from these models has been largely taken for granted. However, the performance of computational models does not only depend on their individual architectures and algorithms. The representation of knowledge in the model can also have a critical bearing on its behavior. That is, the structure of its inputs (on which its predictions are based) and its outputs (what it predicts) can systematically change how a model performs. Indeed, as Bröker and Ramscar (2020) recently demonstrated, depending on the representational assumptions made, different models of the same empirical result can provide support for psycholinguistic theories that make opposing claims about the nature of learning and processing.

The relation between input-output structures and the subsequent interpretation of performance become further apparent when we consider computational models such as WEAVER++ (e.g., Roelofs, 1997) or the Spreading-Activation Theory of Retrieval (Dell, 1986; Dell et al., 2007, and follow-up models). These models use a network framework that reflects a common conceptualization of speech production in psycholinguistics, assuming it to be a sequential, transformational process. At the highest level, the production of spoken words is initiated by information that represents the semantics of the words to be uttered. These in turn activate discrete information at lower levels of processing such as morphemes, syllables, and finally phonemes[2]. In terms of the representation of linguistic knowledge, this means that the complexity of information within these models fans out into more and more fine grained units. This situation is illustrated in **Figure 1B** where "label" can be taken as a placeholder for any kind of higher level units of information—e.g., inflectional functions or morphological contrast—and "feature 1", "feature 2", etc. can be regarded as a placeholder for lower level units—e.g., phones. This raises a question: How reasonable is this flow of information from the perspective of learning theory? We address this in the next section.

[2]Both models stop at the phonological representation and outsource the problem of articulatory movements to theories of articulation and their computational implementations such as Articulatory Phonology/Task Dynamic framework (Browman and Goldstein, 1986; Saltzman and Kelso, 1987) or DIVA (Guenther, 2016).

2.3. Linear Order and Discriminative Learning

It seems clear that where systematic patterns of variance in production have been seen to relate to morphological and paradigmatic structure, these effects must be a product of what speakers have learned. The mechanisms that support this learning thus offer an obvious source of explanation for the patterns of behavior observed. While different kinds of mechanisms have been proposed for language learning (see e.g., Ellis, 2006), research has revealed that the majority of human (and animal) learning mechanisms are based on prediction and prediction-error, i.e., error-driven learning (O'Doherty et al., 2003; Schultz, 2006).

Rescorla and Wagner (1972)'s implementation of the delta rule defines a simple error-driven learning algorithm that is often used in psychological research, and was used by Tucker et al. (2019) and Tomaschek et al. (2019) to train their two-layer networks (a detailed description is provided in their Appendix)[3]. Its algorithm implements a systematic learning process that aims to produce a set of mappings that best discriminate the informative, predictive relationships between a set of inputs and a set of outputs given a training schedule. Because of this, Ramscar et al. (2010) suggest that from a computational perspective the algorithm is best understood as describing a discriminative learning mechanism[4].

Because prediction is a time-sensitive process, the order in which experiences occur is a strong determinant of the kind information that can be learned about cue-outcome relationships through error-driven learning (Ramscar et al., 2010; Arnon and Ramscar, 2012; Hoppe et al., 2020; Vujovic et al., 2021). Speech comprises an ordered series of gestures. These yield an ordered series of phonetic contrasts (Nixon and Tomaschek, 2020) that represent an ordered series of linguistic events (Dell et al., 1997; Grodner and Gibson, 2005). Given that it seems clear that language is learned through an error-driven mechanisms it follows that speech production is likely to be particularly sensitive to these sequential/time-sensitive effects.

However, although speech is clearly ordered, in its use in communication it supports "displaced reference" (Hockett and Hockett, 1960). That is, it allows for reference to things that are not present in the here and now. One consequence of this is that the constraints that are imposed by predictive relationships in language use are not always obvious. This is especially the case when it comes to the relations between form and meaning in linguistic morphology (Ramscar et al., 2010; Ramscar, 2013; see also Ramscar, 2021a for a general review of this issue in relation to morphology).

To explain these constraints, it is first important to note that because prediction and prediction error modulate the values of

[3]The algorithm Rescorla and Wagner (1972)'s implementation of the delta rule is simply the linear form of an earlier rule proposed by Widrow and Hoff (1960), Rumelhart and McCelland (1987), and this in turn is formally equivalent to the delta-rule used in connectionist networks (Sutton and Barto, 1981).

[4]This point also applies to the error-driven learning algorithms found at the heart of most connectionist/neural network model (Jordan et al., 2002), and Bayesian models of learning (e.g., Daw et al., 2008).

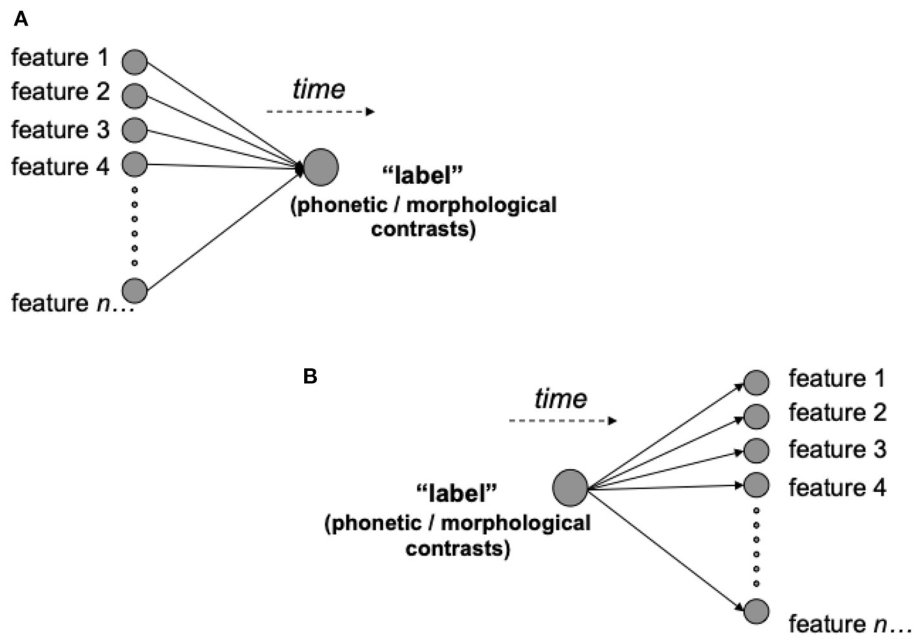

FIGURE 1 | The possible predictive relationships labels (in morphological terms, series of words and affixes) can enter into with the other features of the world (or other elements of a code). A feature-to-label relationship **(A)** will facilitate cue competition between features, and the abstraction of the informative dimensions that predict morphological contrasts (e.g., nouns and plural affixes) in learning. By contrast, a label-to-feature relationship **(B)** will be constrained to simply learning the probability of each feature given the label.

cue-outcome relationships, these values are not determined by simple co-occurrence. Rather, when multiple cues to an outcome are present, a given cue's value will depend on a competitive process that weighs the informativity of each cue in relation to the current uncertainty of a learner. This situation is illustrated in **Figure 1A**, where multiple present features compete for the prediction of an outcome or a label. Informativity thus takes into account both co-occurrences between a cue and an outcome and the non-occurrence of the outcome given the cue. Because uncertainty is finite, more informative cues gain value at the expense of less informative cues. In other words, cues compete for predictive value, a process that leads to the discovery of reliable cues through the discriminatory weakening and elimination of other cues (Ramscar et al., 2010; Nixon, 2020).

While this mechanism is simple in principle, in practice it is an extremely efficient method for extracting predictive structures. For example, in English morphology, plurality is typically marked on nouns by a final sibilant /s/ (whose voicing depends on phonetic context).

The existence of this predictable regularity has implications for the informativity of cues about inflectional structures. Someone learning to predict the form of English nouns will be presented with a large number of cues to the wide range of articulatory events that English nouns comprise. Most of the plural nouns that children encounter will tend to provide evidence for the highly informative cue-outcome relationship between plurality and the presence of a final sibilant at the end of the noun's form. Because of this, it follows that once children have begun to learn the cues to nouns, the relationship between plurality and a final sibilant at the end of nouns can be expected to be reliably learned. However, because this relationship is not informative about the subset of irregular plurals, children will have to learn to ignore this cue in irregular contexts, and learn the more specific cues to these nouns instead. It follows from this that until children have learned to ignore the more general cue to regular plurals, the intermediate representation they acquired may cause them to over-regularize irregulars (Ramscar and Yarlett, 2007; Ramscar and Dye, 2009; Ramscar et al., 2013b). In the same way that children learn to ignore the erroneous cues to irregulars, they will also learn that the other, less informative cues associated with regular plurals should also either be ignored, or associated with other parts of the signal (Ramscar et al., 2010, 2013a). Accordingly, as speech unfolds in time, similar forms of this process will allow for the many abstract features associated with verbs and their suffixes (e.g., tense, aspect etc.) to be learned and extracted in much the same way.

In addition, because learning happens in time, and because the events signaled in speech occur serially, it follows that linguistic regularities (or "units") can serve as both cues and outcomes in learning. For example, in the sentence 'The girl plays football', "girl" predicts "plays" which in turn predicts "football". It thus follows that, when all of these considerations are taken together, determining exactly what counts as a cue and what counts as an outcome in speech production is not always obvious. Moreover, when it comes to modeling, these matters will often be determined by the specific goals of the model.

2.4. Cue-to-Outcome Structure in Speech Production and Implications for Input-Output Structures

With cue competition, prediction and prediction error in mind, we can conceptualize speech production and articulation from the perspective of discriminative learning. As we discussed earlier, in existing psycho-linguistic theories of speech production, semantics, inflectional and morpho-syntactic information should serve as cues for articulation. In addition to these high level sources of information, there is evidence that articulation is also driven by articulatory, sensory and acoustic targets ("articulatory target cues", cf. Hickok, 2014; Guenther, 2016). From a discriminative perspective, all these cues will compete simultaneously for informativity about the executed articulatory gestures during learning. As a consequence, it follows that during production, these cues will serve to activate the execution of articulatory gestures. Note that we do not make any statements about the size of gestural chunks. Following Guenther (2016), we assume that their size can range between a single phone, and sequences of multiple phones. Moreover, even the size of the "same chunk" might vary, depending, for example, on the amount of practice a particular speaker has with them (see Tomaschek et al., 2018a,c, 2020; Saito et al., 2020a,b, for electromagnetic articulography and ultrasound studies on practice).

It thus follows from the above that when it comes to the computational modeling of speech, it is these semantic, morpho-syntactic, inflectional and articulatory target cues that should serve as the *inputs* to neural network learning models. In the same vein, the articulatory gestures that will be activated by these cues should serve as the *outputs* of these models.

However, Tucker et al. (2019) and Tomaschek et al. (2019) did not employ this input-output structure to train the networks described earlier. Rather, following the approach taken by Baayen et al. (e.g., 2011, 2016b), in the model of Tucker et al. (2019) the target gestures served as the only inputs—reflected by diphones of words in the Buckeye Corpus (Pitt et al., 2007). The outputs of the model then consisted of the tense of the verbs under investigation, in addition to inflected word forms. This meant that, from the perspective of our analysis above, the outputs of this models contained information that actually serves as inputs when speakers learn to articulate inflections.

Tomaschek et al. (2019) followed Tucker et al. (2019)'s example regarding the input-output structure, but extended the input to a five-word window around the targeted word in the Buckeye corpus. From this five-word window, two kinds of inputs for the network were extracted. First, diphones from all words that served as an approximation of acoustic and sensory targets that serve to initiate articulation in models of speech production (Hickok, 2014; Guenther, 2016). Second, the word forms preceding and following the target word. These word form inputs were assumed to capture the target word's semantic embedding—in the same way that studies of distributional semantics counted the number of co-occurrences between words within a specific context (Lund and Burgess, 1996; Landauer et al., 1997; Shaoul and Westbury, 2010; Mikolov et al., 2013), and in the same way that studies within the framework of "naive discriminative learning" used word forms to discriminate word meanings (Baayen et al., 2016a,b). As outcomes, the inflectional functions encoded by word final [-s] in English were used. In summary, this meant that the input-output structure provided to the neural networks in both of these studies did not reflect the cue-to-outcome structure that actually seems appropriate to speech production. Instead, some of the information that was represented as outputs in these models actually appears to serve as inputs when production is analyzed from a learning perspective. With this theoretical and empirical background in mind, we turn to the specific aims of the present study.

2.5. The Present Study

The general aims of the present study are: (a) to train a two-layer neural network with an input-output structure that contains the inflectional information relevant to German word final [ɐ]; (b) to use the resulting network measures to predict the phonetic characteristics of [ɐ]. Since findings are contradictory regarding the relationship between uncertainty within the morphological and paradigmatic context and phonetic characteristics, it followed that at the outset, the expected direction of this relationship was unclear.

The network measures might be associated with enhancement, as predicted by the *Paradigmatic Signal Enhancement Hypothesis* (Kuperman et al., 2007) and demonstrated by previous studies using two-layer network models (Tomaschek et al., 2019; Tucker et al., 2019); or they might be associated with reduction, as predicted by the *Smooth Signal Redundancy Hypothesis* (Aylett and Turk, 2004; Cohen Priva, 2015). Accordingly, another aim of this study was to empirically determine which of these hypotheses is supported by a model that accurately captures the dynamics of morphological learning.

Accordingly, the study also aimed to compare the performance of a two-layer learning network employing the input-output structure used by Tucker et al. (2019) and Tomaschek et al. (2019)—where inflectional functions served as outputs—to one in which these functions were represented appropriately: as inputs to the output gestures that represent their realization in speech. We will refer to these learning networks as the *functional output network* and *functional input network*, respectively. We expected that measures extracted from the *functional input network* would be a better predictor of phonetic characteristics than measures computed on the basis of the *functional output network*.

3. METHODS

3.1. Material

The materials for the present study were extracted from the Karl-Eberhards-Corpus of spontaneously spoken southern German (KEC, Arnold and Tomaschek, 2016). The KEC contains recordings of two acquainted speakers having a spontaneous conversation for 1 h about a topic of their own choosing. Speakers were seated in two separate recording booths and their audio signal was recorded on individual channels so that the audio of each speaker can be analyzed without the interference from the other. The KEC contains manually corrected word boundary annotations and forced-aligned segment annotations obtained using the Rapp forced aligner (version 2015, Rapp, 1995).

The corpus contains a total of roughly 23,100 word tokens (1,360 types) that contain a word-final [ɐ]. To make sure the segment annotations are correct, we manually corrected all [ɐ] instances in the corpus for which the aligner provided an annotation. We excluded all instances for which the aligner failed to perform the annotation. This was the case when there was too big a mismatch between the expected and actual duration of the word. In these cases, it was also very hard to annotate the [ɐ] as it was unclear, due to the strong reduction of the [ɐ]-bearing word, where to place segment boundaries. We also excluded the article *der* from the analysis since its annotation is complicated: its pronunciation ranges between [deːɐ], [dɛːɐ], [dɐ], etc. and it is at times unclear at what point the boundary between the two vowels, if present, should be made.

The final data set for the analysis in the present study contained 10,320 word tokens (870 types). It contained 4,944 content words (e.g., nouns, adjectives), 4,463 morphologically simple function words (e.g., adverbs) and 913 morphologically complex function words (e.g., demonstrative pronouns).

The inflectional functions encoded by [ɐ] in these words was manually classified. In total, 60 inflectional functions were obtained, based on combinations of grammatical functions (nouns, articles, pronouns, etc.), numerus (singular, plural), gender (feminine, masculine, neuter) and case (nominative, genitive, dative, accusative). A list of all functions can be found in the **Supplementary Material** (https://osf.io/8jf5s/).

As a measure of spectral characteristics, we investigated the time courses of the first and second formant (F1, F2). We used the LPC algorithm provided by Praat (Boersma and Weenink, 2015, standard settings) to compute the time courses of F1 and F2 in each vowel. For analysis, we excluded vowels shorter than 0.018 s ($\log = -4$) due to sparse data. In addition, we excluded formant measurements for which F1 was outside a range between 250 and 1,000 Hz, and F2 was outside a range between 1,000 and 2,000 Hz. As a result of this exclusion, additional 112 word tokens were excluded, yielding a data set of 11,018 word tokens (871 types) with word final [ɐ] for the analysis. Words with word final [ɐ] will be called *[ɐ]-word* from now on. In order for higher tongue positions to be associated with higher F1 values, thus making F1 frequencies straightforwardly interpretable, F1 frequencies were inverted by being multiplied by −1. Prior to analysis, formant frequencies were centered and normalized by speaker.

3.2. Assessing Uncertainty

In this section, we discuss the details of the input-output structures discussed in the introduction and how we implemented them in the *functional output network* and the *functional input network*. We used the entire KEC to construct the learning events on the basis of which we trained the two network models. Learning was simulated using the Rescorla and Wagner (1972)'s delta-rule [as implemented in the *Naive Discriminative Learner* package 2, Shaoul et al. (2014)]. An explanation of the delta-rule can be found in the Appendix of Tomaschek et al. (2019). As noted above, apart from information about inflectional function, several other sources of information serve as cues to speech production. To operationalize these other cues, we followed Tomaschek et al. (2019)'s approach. Accordingly, both models described below used cues derived from a five-word sliding window that iterated across all learning events. Keeping the rest of the cue structure consistent across the models (and studies) ensured comparability between both the two models implemented here and the previous studies.

3.2.1. Knowledge Representation in the Functional Output Network

The input-output structure used to train the *functional output network* was essentially the same as that employed by Tomaschek et al. (2019). Inputs consisted of the word forms preceding and following the target word in the five-word sliding window. The target word itself never served as an input to avoid direct mappings between inputs and outputs. In addition, inputs contained the diphones of all words in the sliding window including the target word. Diphones were based on the phonetic transcription provided by the Rapp forced aligner used to annotate the corpus (Rapp, 1995).

As in the Tucker et al. and Tomaschek et al. studies, the outcomes in the *functional output network* were the morphological and inflectional functions of the [ɐ]-words. Recall that the network iterated across all word events in the KEC corpus. This means that it also encountered numerous words that did not have word-final [ɐ], and accordingly no inflectional function of interest. In this case, a simple place holder was used to ensure cue competition. To summarize, the *functional output network* was trained to predict inflectional functions of [ɐ]-word on the basis of word and diphone cues.

To obtain a predictor of phonetic characteristics of [ɐ], we computed *functional output activation* on the basis of the trained network. The measure can be regarded as a measure of the uncertainty about the inflectional functions that emerges within the five-word sliding window. *Functional output activation* was computed by summing the weights between all word and diphone inputs in the five-word window around the [ɐ]-word and the inflectional functions of the target word.

3.2.2. Knowledge Representation in the Functional Input Network

The input-output structure in the *functional input network* followed the logic of our analysis in the introduction, where we argued that inflectional functions are learned to serve as cues in speech production and hence should actually serve as inputs to the learning process simulated in the network (Ramscar et al., 2013b; Ramscar, 2021a, see also). Also consistent with this analysis, the outcome of the articulation process, [ɐ], functioned as the output of the network. Accordingly, in addition to diphones and words within the five-word window (the same as in the previous structure), we used the inflectional functions of the words with final [ɐ] as inputs. The output of the network was [ɐ], whenever it was in word-final position of [ɐ]-bearing words. In line with the interpretation by Tomaschek et al. (2019), we regard the outcome [ɐ] to function as an abstract placeholder for potential articulatory gestures representing the articulation of [ɐ] in context. In other words, this network was trained to predict the occurrence of [ɐ] on the basis of word forms, diphones and the inflectional functions. To ensure cue competition, we also used the word forms of the target words in the center of the sliding window as outputs. As a predictor of phonetic characteristics, we computed *functional input activation* by summing the weights between all word, diphone and inflectional function inputs in the five-word window and the [ɐ] output. An introduction to training such a two-layer network and coding the calculation of activations can be found in Tomaschek (2020).

3.2.3. Example

To explain the way training proceeded in the two models, consider the following sentence: *Das ist dieser große Mann* "This is the big man". In the *functional output network*, the word inputs in the five-word sliding window centered on *dieser* "this" were DAS IST DIESER GROßE MANN (we ignored major case). The acoustic diphone inputs in this windows are `#d da as sI Is st td di iz z5 5g gr ro os s@ @m ma an n#`, with # representing boundary cues. The outputs would be the combination of grammatical and inflectional functions of *dieser*: DEMONSTRATIVPRONOMEN MASKULIN NOMINATIV "demonstrative pronoun masculine nominative". Note that grammatical and inflectional functions were used as separate entries and hence, each of them served as an individual output in a learning event (called multiple-hot encoding in the machine learning community). In the *functional input network*, the inputs in the five-word sliding window centered on *dieser* "this" are the words DAS IST GROßE MANN, the acoustic diphones `#d da as sI Is st td di iz z5 5g gr ro os s@ @m ma an n#`, and the inflection functions DEMONSTRATIVPRONOMEN MASKULIN NOMINATIV (multiple-hot encoding). The articulated forms such as *dieser ER*, including a "gestural placeholder" representing the [ɐ]-gesture, served as outputs.

4. ANALYSIS AND RESULTS

4.1. Statistical Analysis of Formant Trajectories

4.1.1. Creating a Baseline Statistical Model

In this section, we describe our statistical approach to analyzing the time course of F1 and F2. We employed generalized additive mixed models (GAMM in the mgcv package, Hastie and Tibshirani, 1990; Wood, 2006, 2011) to investigate how the time course of F1 and F2 in [ɐ] was co-determined by uncertainty in the two models. GAMM uses spline-based smoothing functions to model non-linear functional relations between a response and one or more covariates, modeling wiggly curves using spline smooths as well as wiggly (hyper)surfaces using tensor product smooths (see Wieling et al., 2016; Baayen and Linke, 2020, for an introduction to spline smooths and their use). All model comparisons (and visualization) reported in the following paragraphs were performed with the help of functions provided by the *itsadug* package (van Rij et al., 2015). All analyses can be found in the **Supplementary Material**.

We constructed a model that contained a smooth "s()" for *time* to model the time course of F1 and F2. Time contained the time points at which formant frequencies were measured. Since vowels vary in duration, time points were normalized to a [0, 1] interval, with 0 linked to vowel onset and 1 to vowel offset. We fitted F1 and F2 simultaneously in one model. Accordingly, we needed a predictor to differentiate between the shapes of F1 and F2 trajectories using a factorial predictor *dimension* with the levels F1 and F2. This predictor interacted with the smooth for *time*. To control for speaker dependent formant trajectories, we fitted by-speaker random factor smooths for time, i.e., the non-linear equivalent of a combination between random intercepts and random slopes from standard mixed-effects regression.

The inclusion of words as random effects caused high concurvity in our models[5]. Accordingly, following the suggestion presented in Baayen and Linke (2020), we did not include words as an random effect. Instead, we controlled for effects of coarticulation with the context by fitting by-place-of-articulation random factor smooths for time for the preceding and for the following segment. To allow random factor smooths to vary depending on dimension, all by-factor smooths included an interaction with *dimension* (F1/F2). We controlled for autocorrelation among residuals using the rho parameter ($\rho = 0.8$).

In a bottom-up fitting procedure, we tested whether the inclusion of additional predictors improved the model fit. The first additional predictor we tested was *vowel duration*, log-transformed to obtain normally distributed values. *Vowel duration* served as a control variable as it accounted for undershoot and overshoot associated with temporal variation (Gay, 1978). The inclusion of *vowel duration* as a main effect interacting with *dimension* significantly improved model fit (Δ ML $= -1106$, Δ edf $= +4$, $p < 0.0001$). Allowing *vowel duration* to interact with *time* and *dimension* (by means of a tensor product smooth "te()") further improved model fit (Δ ML $= -845$, Δ edf $= +6$, $p < 0.0001$). The tensor thus accounts for systematic changes in the shape of the trajectory as a function of *vowel duration*.

German word-final [ɐ] discriminates inflectional and grammatical function in content words (e.g., nouns, adjectives), morphologically complex function words (e.g., demonstrative pronouns) and morphologically simple function words (e.g., adverbs). Numerous studies have reported that higher level information such as inflectional function (Plag et al., 2017; Seyfarth et al., 2018; Schmitz et al., 2021a) or pragmatic function (Drager, 2011; Podlubny et al., 2015) correlate with phonetic characteristics. Similar effects have been demonstrated for word class (e.g., Johnson, 2004; Bell et al., 2009; Fuchs, 2016; Lohmann, 2018; Linke and Ramscar, 2020), for which also processing differences during perception (Neville et al., 1992; Pulvermüller, 1999; Brusini et al., 2017) and production

[5]Concurvity is the non-linear equivalent of collinearity that, when high, can render model terms uninterpretable. See the Appendix of Tomaschek et al. (2018b) and Baayen and Linke (2020) for more information on concurvity).

(Fox et al., 2009; Juste et al., 2012) have been demonstrated. Given these systematic differences in perception and production due to higher level information, especially those for word class, we also expect [ɐ] to vary with word class.

This prediction was tested with the predictor *word class*, allowing for potential differences in formant trajectories depending on content words, morphologically complex function words and morphologically simple function words. In order to allow formant trajectories to vary independently in the two dimensions F1 and F2 as well as *word class*, we constructed the factorial predictor "dimension-by-class" (*dbc*) with six levels: one level for each of the six combinations of *dimension* by *word class*. The inclusion of *dbc* as a main effect significantly improved model fit (Δ ML = −405, Δ edf = +4, $p < 0.0001$), as was the case when it was allowed to interact with the *time* by *vowel duration* tensor (Δ ML = −736, Δ edf = +20, $p < 0.0001$). We also tested whether the three levels in *word class* were indeed necessary. We accomplished this by collapsing two levels and refitting the model (e.g., morphologically simple and complex function words were collapsed into one level, and so forth). Collapsing two levels never yielded a better model fit than using *word class* with the three levels. Accordingly, it appears that [ɐ] does indeed vary systematically depending on *word class*. This conclusion is supported by the visualization of the formant trajectories, which are further discussed below. We shall consider this our baseline model, whose formula is illustrated below (with POA = place of articulation):

```
m0 = formant frequency ∼ dbc
+ te(time, vowel duration by = dbc)
+ s(time, speaker, bs="fs", m = 1, by =
dimension)
+ s(time, preceding POA, bs="fs", m = 1,
by = dimension)
+ s(time, following POA, bs="fs", m = 1,
by = dimension)
```

(The random effects structure, indicated by `bs="fs"`, was the same in all models which is why we will not display it anymore in the following formulas).

4.1.2. Testing Activations

In the next analytic stage, we tested the degree to which the inclusion of *functional output activation* and *functional input activation* improved the model fit. The following formula illustrates the model (where *activation* represents both kinds of *activation*):

```
m1 = formant frequency ∼ dbc
+ te(time, vowel duration by = dbc)
+ s(activation, by = dimension)
```

The question thus arises of whether there are also systematic differences of *activation* depending on *word class*. The following model tested this interaction between *activation* and *dbc*.

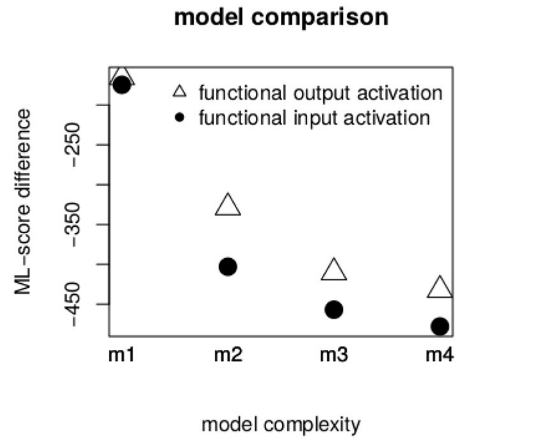

FIGURE 2 | ML-score difference between model m0 and models m1 to m4. The larger the difference, the better the model's goodness of fit.

```
m2 = formant frequency ∼ dbc
+ te(time, vowel duration by = dbc)
+ s(activation, by = dbc)
```

We also tested whether the shape of the trajectory was modulated by *activation*. This was accomplished by fitting an interaction between *time* and *activation* and *dimension* using a partial tensor product smooth "ti()"[6]:

```
m3 = formant frequency ∼ dbc
+ te(time, vowel duration, by = dbc)
+ s(activation, by = dbc)
+ ti(time, network measure, by =
dimension)
```

The final model tested to what degree both the intercept and the shape of the formant trajectories varied in relation to *activation* and *dbc*:

```
m4 = formant frequency ∼ dbc
+ te(time, vowel duration, by = dbc)
+ s(activation, by = dbc)
+ ti(time, network measure, by = dbc)
```

Figure 2 illustrates the difference in ML-scores between our baseline model *m0* and models *m1* to *m4*. The inclusion of both types of activation improved model fit, as can be seen by means of the large negative ML-score difference for model *m1*. Nevertheless, there was no large difference between the gam model containing *functional output activation* (triangles) and the one containing *functional input activation* (circles)

[6]Since main effects are already fitted by means of s(), partial tensor product smooths are used to fit the interaction between two predictors but not the main effects.

(indeed the difference in ML-score between models with the two types was only 1.5). The goodness of fit depending on the two types of activation changed in more complicated models. In models *m2* to *m4*, *functional input activation* provided systematically better model fits, as indicated by larger difference in ML-score to *m0*. In other words, a network that was trained to predict the articulatory gesture of [ɐ] on the basis of semantic, phonological and inflectional functions provided better predictions about [ɐ]'s phonetic characteristics than a network trained to predict the inflectional function itself. We also tested to what degree the inflectional function in the input structure is necessary. We found that activations computed on the basis of network trained without inflectional functions as inputs provided a significantly worse model fit than *functional input activation* (on average, they had an ML-score lower by 200). Accordingly, we regard inflectional functions to be necessary in the input structure (model comparisons can be found in the **Supplementary Materials**).

An inspection of concurvity indicated that the smooths and tensor product smooths for both types of *activation* for morphologically simple function words suffered from high concurvity. Further inspection indicated that this problem was alleviated when individual models were fitted for each level of *word class*. Since the significant interaction with word class (by means of *dbc*) indicated that formant trajectories differ systematically between word classes, fitting individual models for each *word class* was fully supported. Accordingly, below we report the results for models in which formant trajectories were fitted for each of the three levels of *word class* individually. Once models were obtained, data points with residuals larger than 2.5 standard deviations away from the mean were excluded and models were refitted. The following formula illustrates the final model structure:

```
  m.final = formant frequency ~ dimension
+ te(time, vowel duration, by = dimension)
+ s(activation, by = dimension)
+ ti(time, activation, by = dimension)
```

4.2. Modulation of Formant Trajectories

4.2.1. Summaries

Even though *functional output activation* performed worse than *functional input activation*, we will report the estimated trajectories for both of them to allow for a direct comparison. Summaries of all the statistical models indicated that all the tensor product smooths for the *time* by *vowel duration* in both dimensions (F1, F2) were significant ($p < 0.001$) in all statistical models for all activation types. The same result was found for random factor smooths for participants and for place of articulation of the preceding and following vowel. Since these effects are not of primary interest for the present study, and the summaries use up a lot of space, we provide their summaries only in the **Supplementary Material**. Here, we report the summaries for the effect of interest, *functional input activation* and *functional output activation*. **Table 1** illustrates that all but one smooth and tensor terms for *functional input activation* are significant. Only the partial tensor in the F1 dimension in the model fitting morphologically simple function words failed to be significant. Accordingly, the amplitude of the F1 time course was not significantly modulated. A similar result can be see for *functional output activation*. Here, only the partial tensor product smooth for F1 in morphologically complex function words failed to be significant.

4.2.2. Modulation of Formant Trajectory

Figure 3 provides a visualization of the summed effects of the models presented in **Table 1** by means of estimated trajectories. The x-axes represent inverted z-scaled F2 frequencies such that the left edge points toward the front of the vowel space and the right edge points toward the back of the vowel space. Y-axes represent inverted z-scaled F1 frequencies such that the top points to the top of the vowels space and the bottom points toward the bottom of the vowel space. The onset of the trajectories is indicated with a filled star, its center with a circle. Columns represent different word classes (from left to right: content words, morphologically simple function words and morphologically complex function words). Rows represent different numeric predictors.

The onset of the formant trajectories in all three word classes is located at a high fronted position, followed by a fall. Roughly at the mid point of the vowel trajectory (indicated by the black circle), the trajectory makes a turn that results in raised positions. Focusing on the differences between word classes reveals that formant trajectories in morphologically complex function words (left column) are produced at the most fronted position; those in content words are relatively centered (mid column); the trajectories in morphologically simple function words are produced at the most retracted position (right column).

Formant trajectories further differ in their shapes. [ɐ] vowels in morphologically complex word forms have, on average, a relatively wide u-shaped trajectory, while morphologically simple function words have a very narrow trajectory. Moreover, it seems that the differences in shape between word classes is mirrored by the relative horizontal position in the vowel space (ignoring the effect of *vowel duration*): more fronted trajectories have wide trajectories than more retracted trajectories. In conclusion, we observe systematically different formant trajectories in relation to word class. These shapes are further modulated by *vowel duration* and *activation*.

Before we discuss the effects of the *vowel duration* and *activation* predictors, it will be first necessary to discuss how reduction and enhancement can be expected to be reflected in [ɐ]. Typically, reduction of vowels is reflected by more centralized formant trajectories. However, since [ɐ] is already located in the center of the vowel space in a very dense vocalic environment surrounded by [ə] and [a] in the vertical dimension and by [ɪ], [ʏ] and [ɔ] in the horizontal dimension, the specific direction enhancement will take is unclear. Enhancing [ɐ] in any direction and dimension may result in potential competition with its neighboring vowels.

TABLE 1 | Summary of the statistical models using functional input activation and functional output activation as a predictor of formant trajectories.

	edf	Ref.df	*F*-value	*p*-value
FUNCTIONAL INPUT ACTIVATION				
Complex function words				
s(functional input activation):dimension = F1	3.7482	3.9577	39.0716	< 0.0001
s(functional input activation):dimension = F2	3.2589	3.7180	44.7998	< 0.0001
ti(time,functional input activation):dimension = F1	7.6804	9.7289	2.3730	0.0079
ti(time,functional input activation):dimension = F2	4.6737	5.8764	4.3388	0.0002
Content words				
s(functional input activation):dimension = F1	3.3729	3.7829	10.0274	< 0.0001
s(functional input activation):dimension = F2	3.8460	3.9845	94.0980	< 0.0001
ti(time,functional input activation):dimension = F1	10.4838	12.7473	5.2548	< 0.0001
ti(time,functional input activation):dimension = F2	7.7378	9.4625	14.1532	< 0.0001
Simple function words				
s(functional input activation):dimension = F1	3.8933	3.9921	20.9012	< 0.0001
s(functional input activation):dimension = F2	3.7390	3.9562	27.3247	< 0.0001
ti(time,functional input activation):dimension = F1	7.3833	9.7127	1.4650	0.1497
ti(time,functional input activation):dimension = F2	10.8514	12.8554	4.6229	< 0.0001
FUNCTIONAL OUTPUT ACTIVATION				
Complex function words				
s(functional output activation):dimension = F1	1.0020	1.0038	115.2282	< 0.0001
s(functional output activation):dimension = F2	3.8720	3.9862	12.4216	< 0.0001
ti(time,functional output activation):dimension = F1	4.6471	6.5973	0.5412	0.7934
ti(time,functional output activation):dimension = F2	3.6281	4.2364	6.7708	< 0.0001
Content words				
s(functional output activation):dimension = F1	3.7248	3.9528	5.1275	0.0011
s(functional output activation):dimension = F2	3.9479	3.9976	106.9967	< 0.0001
ti(time,functional output activation):dimension = F1	9.6920	12.3538	3.8719	< 0.0001
ti(time,functional output activation):dimension = F2	9.9943	12.3734	8.4570	< 0.0001
Simple function words				
s(functional output activation):dimension = F1	3.2277	3.6812	21.2965	< 0.0001
s(functional output activation):dimension = F2	3.9082	3.9942	39.8523	< 0.0001
ti(time,functional output activation):dimension = F1	8.0654	9.6352	8.6645	< 0.0001
ti(time,functional output activation):dimension = F2	4.9361	6.8905	3.0888	0.0027

Summaries of control variables and random effect structure can be found in the ***Supplementary Material****.*

To establish how enhancement and reduction are manifested in [ɐ], we shall first inspect how they are manifested in relation to hyperarticulation and hypoarticulation in long and short vowels. The top row of **Figure 3** illustrates the effect of *vowel duration* (from the *functional input activation* models). Shades of red represent the 10th, 30th, 50th, 70th, and 90*th* percentile of *vowel duration* with darker shades of red representing longer vowels. Longer vowels are associated with longer formant trajectories, and lower and more retracted vocalic centers in all three word classes. This is a typical effect on the continuum between hypoarticulation and hyperarticulation associated with phonetic duration (Gay, 1978; Lindblom, 1990). Additionally these results show that longer vowels have stronger fronted offsets than shorter vowels. As a result, trajectories for longer vowels are "crossed". How might one account for this effect? First, the offset of the trajectory tends to be located roughly in the center of the vowel space. Second, [ɐ] should not be retracted too far to the back as it may enter into a vowel space where it would compete with the mid low vowel [ɔ]. In order to apply both constraints, long [ɐ] result in narrower trajectories, even though when they are hyper articulated.

4.2.3. Effects of Functional Output Activation

The effect of *functional output activation* is illustrated in the mid row of **Figure 3**. Higher percentiles of *functional output activation* are represented by means of darker shades of red. In morphologically complex function words, higher *functional output activation* is associated with lower, slightly more fronted positions. Comparing the effect to that of *vowel duration*, the lowering could be regarded as an enhancement effect. In content words, there is no observable effect apart from very high percentiles that are associated with more retracted positions. Finally, even though the main effect for *functional output activation* is significant in both dimensions in morphologically simple words, there are comparatively little changes across the activation continuum. In other words,

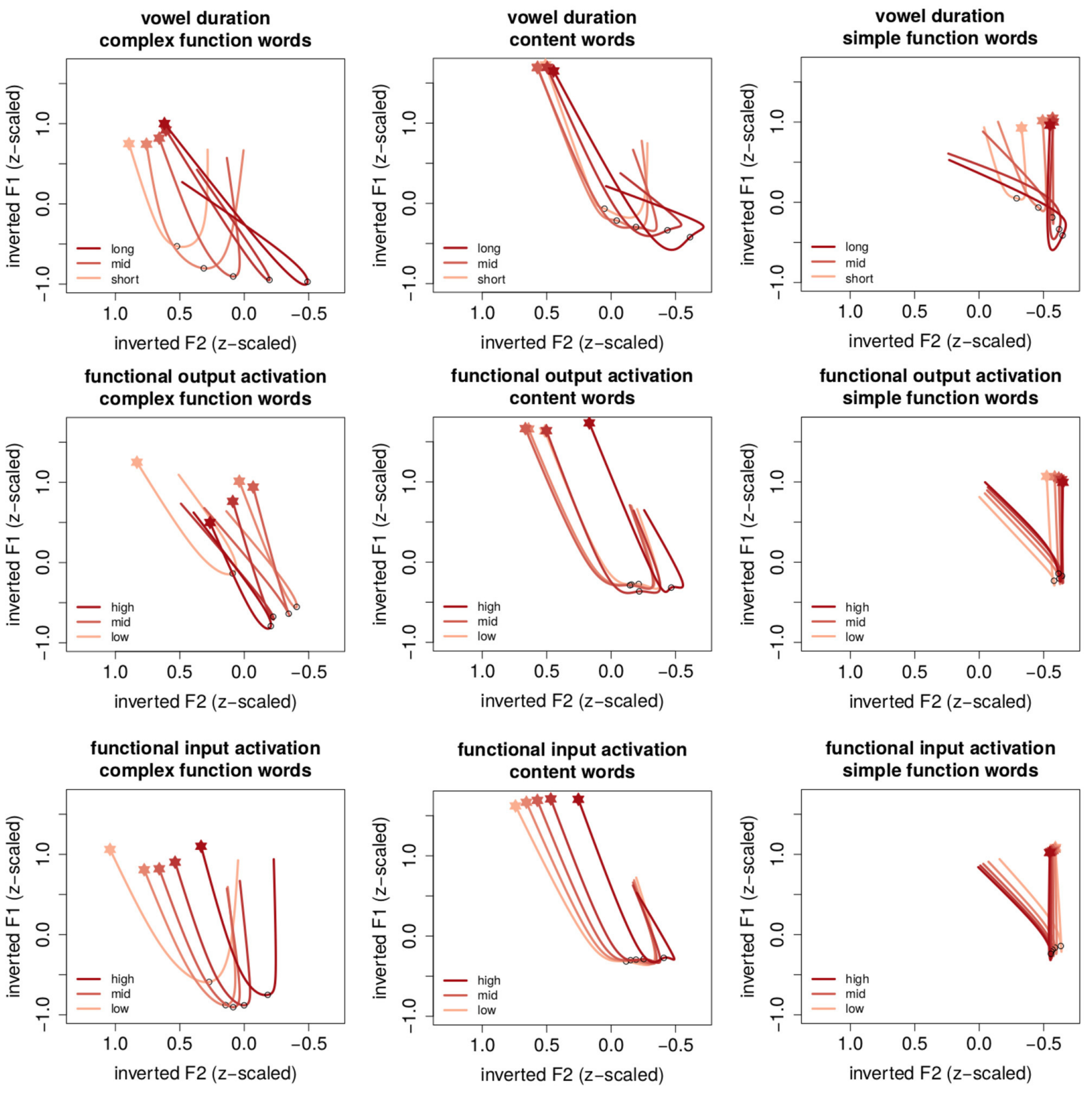

FIGURE 3 | Estimated trajectories for different word classes (columns) in relation to vowel duration **(top)**, functional output activation obtained from a network with inflectional functions of [ɐ] in the output **(middle)** and functional input activation obtained from a network with inflectional functions of [ɐ] in the input **(bottom)**. The x-axes represent inverted z-scaled F2 frequencies such that the left edge points toward the front of the vowel space and the right edge points toward the back of the vowel space. Y-axes represent inverted z-scaled F1 frequencies such that the top points to the top of the vowels space and the bottom points toward the bottom of the vowel space. Shades of red represent percentiles for different predictors (optimized for color blindness). Onset of the time course is located at the filled star, the circle in the trajectory represents the center of the vowel.

functional output activation co-determines the [ɐ] trajectory only in morphologically complex function words.

4.2.4. Effects of Functional Input Activation

Next, we turn our attention to how *functional input activation* modulates the [ɐ] trajectory. In both morphologically complex function words and content words, higher *functional input activation* is associated with stronger retracted formant trajectories. Using the effect of vowel duration as a baseline, we thus observe more enhancement under lower uncertainty, and reduction under higher uncertainty about [ɐ]. The way *functional input activation* co-determines formant trajectories points in the same way as the effect of vowel duration. The effect of *functional input activation* for content and morphologically complex function words is thus consistent in both the temporal and spectral domains.

However, in morphologically simple function words the effect seems to be reversed. Higher *functional input activation* produces slightly more fronted trajectories[7]. Since this effect is only minimal, we refrain from interpreting it to indicate reduction under lower uncertainty. Rather, we conclude that, perhaps unsurprisingly, *functional input activation* has no effect for morphologically simple words.

4.3. Vowel Duration

Even though we controlled for vowel duration during our investigation of formant trajectories, it is still possible that it is also correlated with *functional output activation* and *functional input activation*. Recall that Tucker et al. (2019) and Tomaschek et al. (2019) reported that lower uncertainty about inflectional functions was associated with longer phonetic duration. A Spearman's rank correlation indicated that vowel duration has a correlation of $\rho = -0.01$ (Pearson's $r = -0.03$) with *functional output activation* and $\rho = 0.06$ (Pearson's $r = 0.07$) with *functional input activation*. Thus, the correlation between our activation measures and [ɐ] duration is very small. To statistically evaluate these effects, we fitted log-transformed [ɐ] duration as a function of *functional output activation* and *functional input activation*. We performed a linear mixed-effect regression, controlling for local speaking rate and the number of segments in the word, including random intercepts for speakers and words. The model further indicated that *functional output activation* did not significantly correlate with vowel duration ($\beta = -0.018$, se $=$ 0.04, $t = -0.434$), while with *functional input activation* did (β $= 0.36$, se $= 0.16$, $t = 2.81$). Visual inspection indicated that the difference between low and high *functional input activation* was roughly an increase of 10 ms in vowel duration. We also tested *word class* as a predictor but found no significant effect.

Thus, in the *functional output network* we did not observe a correlation between vowel duration and activation. By contrast, the *functional input network* did yield a small, but significant effect of enhancement.

5. DISCUSSION

This study sought to investigate how the uncertainty associated with inflectional functions influences the phonetic characteristics of speech. It was motivated by the contradictory findings that have been reported regarding the effects of uncertainty on production in relation to paradigmatic and morphological families, where some studies found lower uncertainty to be associated with reduction (e.g., Hay, 2004; Hanique and Ernestus, 2011; Plag and Ben Hedia, 2017), whereas others reported enhancement (e.g., Kuperman et al., 2007; Schuppler et al., 2012; Cohen, 2015; Tomaschek et al., 2021). To assess the degree to which these findings reflected differing assumptions regarding word-internal structures, we followed Tucker et al. (2019) and Tomaschek et al. (2019)'s approach and sought to allow these structures to emerge naturally, in learning. We trained two two-layer networks employing two different representations of the predictive relations relevant to learning in speech production. From these we extracted network measures that we used to gauge the uncertainty associated with the inflectional functions of German word final aschwa [ɐ] (which discriminates around sixty different inflectional functions). We used these models to investigate how the inputs and outputs presented to learning networks should be implemented so as to most appropriately represent the structure of linguistic knowledge. To this end, we tested how accurately the measures of uncertainty derived from different implementations served to predict the phonetic characteristics of [ɐ] in the speech signal.

We observed that formant trajectories of [ɐ] were enhanced in relation to decreased uncertainty in those word classes that were morphologically complex. Below we discuss this finding in more detail in relation to the two questions that guided our study: (1) What is the relation between uncertainty within the context of morphological families and phonetic characteristics and how can it be explained? (2) What kind of input-output structure most appropriately represents linguistic knowledge in speech production models?

5.1. Effects of Word Class

Our analyses revealed that the formant trajectories of [ɐ] systematically differed between the three word classes investigated. These systematic differences emerged independently of the uncertainty measures obtained from the learning networks. Accordingly, this finding supports the assumption that fine phonetic detail is co-determined by lexical information. In phonological theories, definitions of phones and phonemes are typically based on a mixture of impressionistic judgments and theoretical considerations. These definitions thus not only ignore differences in fine phonetic detail, they also ignore potential differences that can arise from the influence of other levels of linguistic description, such as morphology or word class. By contrast, in keeping with other studies showing that the phonetic characteristics of supposedly homophonous "phones" vary systematically according to their morphological or grammatical status (e.g., Drager, 2011; Plag et al., 2017, and references in the introduction), these results raise questions about the adequacy of the "sound units" phonological theories suppose. In particular, it appears that the phonetic detail of speech signals contains fine grained difference that are far more systematic than traditional theories have tended to assume. Moreover, it appears that these differences may actually be informative about word class in communication. Studies have demonstrated that listeners are sensitive to changes at this level of phonetic detail, and that they use them not only to discriminate phonetic (e.g., Whalen, 1983; Beddor et al., 2013)

[7] When all word classes were fitted in one joint model, i.e., m4, this effect was strongly amplified such that the difference between low and high *functional input activation* was in the range of that for content and morphologically complex function words. However, comparing individual models with the joint model indicated that this amplification was most likely due to concurvity in the joint model.

but also morphological contrasts (Kemps et al., 2005; Tomaschek and Tucker, 2021). This suggests that the whole idea that speech signals comprise phonological realizations of words that are somehow analogous to orthography may be fundamentally misguided (Port and Leary, 2005; Ramscar and Port, 2016).

5.2. What Kind of Input-Output Structure Should Speech Production Models Employ?

Theoretically, the network simulations reported in our study were rooted in discriminative learning (Ramscar and Yarlett, 2007; Ramscar et al., 2011, 2013a,b; Ramscar, 2019, 2021b). This framework conceptualizes learning—during perception and production—as a process that serves to discriminate informative relationships between a set of cues and a set of outcomes in a cognitive system. When it comes to modeling, this in turn raises the question of how inputs (representing cognitive cues) and outputs (representing behavioral outcomes) should be implemented so as to most appropriately capture the cognitive process in question: in this case, speech production?

This question is further complicated by the fact that computational modeling inevitably constrains the way that relevant information is represented in a simple set of inputs and outputs (Bröker and Ramscar, 2020). This problem of abstraction is particularly apparent in simple two-layer networks of the kind employed here. This is because these models do not have the hidden layers that can enable multi-layer networks to learn abstractions from data. This is both a strength and a weakness. On one hand, it limits the ability of these models to discover abstract structures—such as inflectional functions—that may be present in a set of training data. On the other hand, simply because of their simplicity, they constrain modelers to utilizing input and output structures that explicitly code for the cues and outcomes that they believe to be important to the process being modeled (see Ramscar, 2021b, for a more detailed discussion of this point).

A similar point applies to most early computational models of speech production, such as Weaver++ (e.g., Roelofs, 1997) or the Spreading-Activation Theory of Retrieval (Dell, 1986; Dell et al., 2007, and follow-up models). While they did not explicitly address learning, these models were based on traditional linguistic and psycho-linguistic theories (e.g., by Fromkin, 1971, 1973; Levelt et al., 1999) that assumed an idealize speech process in which any abstractions posited by the theory had already been learned (and hence existed as discrete elements). Accordingly, in these models the 'lexical semantics' of a word served as an input for lemma selection, which in turn served as an input for the selection of discrete morphological structures. These then activated the abstract phoneme sequences that explicitly represented the words to be pronounced. These abstract phoneme sequences, once syllabified, could then be used to compute the execution of articulatory gestures in a high dimensional acoustic-spatio-temporal space (Browman and Goldstein, 1986; Guenther, 2016; Turk and Shattuck-Hufnagel, 2020).

The *functional input network* presented in this study shares the same general conceptualization of the role semantics as traditional models. It assumes that intended meanings serve as the (main) cues to the initiation of articulations. It thus also shares with these older models the representation of articulation as the outcome of a process that is initiated semantically. Since our model is grounded in learning—which is always subject to experience—the input structure assumed in our model is less discrete. Rather than assuming that morphological functions and lexical meanings are somehow separate dimensions of experience, we assume that learning is required to separate them. That is, we assume that discriminating lexical from morphological features is a function of exposure and learning. Further, given the skewed distribution of linguistic forms, it follows that the degree to which these dimensions are discriminated in a given item or context will vary across the lexicon (Ramscar et al., 2013b).

Accordingly, many of the simplifying assumptions embodied in these earlier models make little sense in a learning model. For example, Levelt et al. (1999)'s theory assumes that "higher level" information is forgotten once it is transformed into a representation at a "lower level". However, this is clearly inconsistent with learning, and the idea of abstraction being a product of the learning process. Rather, from a learning perspective, it is competition between cues representing information at lower levels that enables abstractions at higher levels to form. Finally, if the simplifying assumptions made in earlier models were true, there ought to be no correlation between semantic and morphological information and the phonetic characteristics. Yet, again consistent with the idea of all of this information being discriminated/shaped in learning, the present results, along with many of the other studies we have reviewed, contradict this assumption. Semantic and morphological information clearly does correlate with acoustic characteristics.

It further follows that if the cues to semantic and morphological information must be discriminated and abstracted in order to learn speech, they must play a similar role in speech production. That is, the semantic information that was discriminated into different levels of abstraction—lexical, morphological, inflectional—in learning will then serve as the cues to executed articulatory outcomes. Once again, which cues are informative about which articulations will depend on learning; and learning will be shaped by individual experience, the distributional structure of the language and context. In an actual speaker, this learning will be continuous both in time and across the lifespan (see e.g., Ramscar et al., 2014), and will be then processed by the multiple learning mechanisms contained within the complex architecture of the human brain.

By contrast—and critically—when it comes to modeling these learning processes, a great deal of this abstract information must be simplified and discretized in order to make the learning process tractable. Moreover, depending upon the goal of the modeling exercise, the goal of making the outcomes of the learning process interpretable raises further considerations. If our goal had been to emulate human performance as accurately as possible, there exists a range of more powerful models—multi-layered, deep learning networks (Graves, 2012; LeCun et al., 2015) that are far more capable of learning to capture the many complex factors that seems to drive speech and language (Hannun et al., 2014; Jozefowicz et al., 2016). However, this same complexity inevitably leads to Bonini's paradox (Bonini, 1963), in that understanding exactly how they actually learn their functions can be as challenging as understanding children's learning itself[8].

It is in this regard, as we noted above, that the apparent shortcoming of two-layer networks can actually be an advantage. Because these simple networks lack the hidden layers that would typically be responsible for learning complex abstractions, they require that any implementation be simplified so as to include only the information thought necessary to learning. It furthermore requires that abstractions that are assumed to be necessary to this process be made explicit, and represented in the input-output structure.

Accordingly, by employing simple two-layer network models, we were able to explicitly examine the way that abstract information such as inflectional functions ought to be represented in models of articulatory learning. This was accomplished by configuring two networks with the two different input-output structures, and then testing which of them was the better predictor of phonetic characteristics. Our results showed that the activations from the network trained with inputs that included inflectional functions served to predict the phonetic characteristics of [ɐ] better than activations from the network trained on an input structure in which these functions were outputs.

One question about these models that remains to be answered is why the *functional output model* that successfully predicted phonetic characteristics in Tucker et al. (2019) and Tomaschek et al. (2019) almost failed to do so in the present study, while the *functional input model* succeeded. The data and analyses at hand only allow for speculations. One possible answer lies in the difference between the types of acoustic signals investigated in the previous studies and in the present study. Like the majority of studies investigating effects of uncertainty associated with paradigmatic families, Tucker et al. and Tomaschek et al. focused on durations. By contrast, the present study investigated a higher dimensional spectral signal. Another possible explanation may be the amount of inflectional functions under investigation. Tucker et al. focused on two inflectional functions; Tomaschek et al. investigated nine. By contrast, here, we investigated 60 different inflectional functions. It is of course impossible to draw firm conclusions from these considerations, however it seems likely that the results of these previous studies may have been particularly dependent on the specifics of their approach. It thus follows that any conclusions one might draw from this previous work will be more limited in its generalizability than those one might draw from the current study.

5.3. Enhancement vs. Reduction

As we noted at the outset, the results of studies of the association between the statistical characteristics of word forms within morphological and inflectional paradigms and their phonetic characteristics in speech show an inconsistent pattern. Some studies demonstrate that higher probability of words and segments is associated with phonetic enhancement (Kuperman et al., 2007; Hanique and Ernestus, 2011; Schuppler et al., 2012; Cohen, 2015; Lõo et al., 2018; Bell et al., 2019; Tomaschek et al., 2021), others find that it is associated with phonetic reduction (Hay, 2004; Hanique and Ernestus, 2011; Cohen, 2015; Ben Hedia and Plag, 2017; Plag and Ben Hedia, 2017). As we have argued, one reason why these contradictory patterns may have emerged is because these studies often disregarded how words and their paradigms are learned. Moreover, even where learning has been taken into account, they have often disregarded the assumptions one makes about the representation of linguistic knowledge and how it can influence learning (Bröker and Ramscar, 2020).

Addressing this last problem enabled us to provide a better account of our data. By taking into account how the distributional characteristics in language are learned, we were able to show that the phonetic characteristics of [ɐ] appear to be enhanced in relation to lower uncertainty associated with inflectional functions. These results support the findings within the framework of the *Paradigmatic Signal Enhancement Hypothesis* (Kuperman et al., 2007; Hanique and Ernestus, 2011; Schuppler et al., 2012; Cohen, 2015; Lõo et al., 2018; Bell et al., 2019; Tomaschek et al., 2021). Since these findings contradict the consistent effects of reduction in syntagmatic context demonstrated in the framework of the *Smooth Signal Redundancy Hypothesis* (Aylett and Turk, 2004), the question arises how the different effects in context of syntagmatic and morphological information are to be explained.

Kuperman et al. (2007) argue that enhancement in the paradigmatic context ought to be expected, because it reflects speaker confidence about the selection of a specific word form. The more confident speakers are (i.e., their speech production systems are) about a selection, the more time they can take to actually produce it. By contrast, Cohen (2015) argued that this effect should be expected for very different reasons. Arguing from within the framework of Exemplar theory, she suggests an alternative explanation: the phonetic characteristics of less frequent word forms will be shifted toward the characteristics of a competitor in the inflectional paradigm. This has the effect of reducing these less probable forms and making more probable form seem to be more enhanced.

While both explanations have their merits, it nevertheless remains the case that they are unable to fully explain all of the effects of enhancement and reduction in relation to uncertainty that have been observed. With regards to

[8]At present it is unclear how the complexities of learning at multiple levels of abstraction that underlie the performance of these models is to be translated into theoretical insight. This is highlighted by recent attempts to understand the performance of multiplayer networks in language processing tasks by treating them as experimental subjects (McCloskey, 1991; Linzen et al., 2016; Wilcox et al., 2018; Futrell et al., 2019).

the confidence account, it is unclear why the effects of increased confidence are not observed within syntagmatic contexts (as pointed out by Cohen, 2015). With regards to the Exemplar theory account, exactly how it accounts for other word forms in the paradigm and how they contribute to systematic changes of phonetic characteristics (as demonstrated by e.g., Kuperman et al., 2007; Tomaschek et al., 2021) remains unclear.

5.4. The Signal-Message-Uncertainty Distinction

So how are the different influences of uncertainty on articulation in context—syntagmatic and paradigmatic—to be reconciled? It seems clear that in some sense both the *Smooth Signal Redundancy Hypothesis* and the *Paradigmatic Signal Enhancement Hypothesis* are true, at least in context. What is needed is an explanation of what this context is and how it applies. We suggest that the answer lies in the contribution of two very different aspects of speech production: The signal and the message, and the very different way that these interact with context.

Accordingly, it is important that we be clear about what it is that we mean when we talk about the "signal". Every type of human communication is rooted in kinematic behavior. In acoustic communication, this behavior involves the movement of the articulators, the vocal cords and all other organs necessary to produce the acoustic speech signal (see Tucker and Tomaschek, forthcoming, for an overview). In another modality, say the visual modality in sign languages or gestures, it involves the movement of the body and the limbs. By "signal", we therefore mean both the execution of kinematic behavior to create the acoustic or visual signals and the contrasts embodied in the different signals themselves, whose properties will of course vary in context.

It is important to stress that our conceptualization of speech production contrasts with the traditional, linguistic conceptualization of communication. This means that we do not assume that speaker messages convey or contain meanings. Rather, speakers produce a signal that listeners use to discriminate the meaning intended by the speakers. The discrimination process is based on a code that has been learned in much the same way as the discriminative models described above. It follows that this code serves to condition meanings onto signals: Language users learn the relationships between the world and the speech contrasts that encode their language's representation of various states of affairs in that world. To do this, they must learn to discriminate the semantic (in its broadest sense) cues to phonetic and articulatory contrasts in context. This in turn allows speakers to use these articulatory/phonetic contrasts in context to construct messages that serve to discriminate the meanings that they have learned to condition onto the same contrasts in similar contexts.

That is, in order for two speakers to have a conversation, they must share the same "source code" (Ramscar, 2019, 2021b) that underlies the language they are using. A listener uses what they have learned about the shared code to predict the messages intended by speakers. These messages will be produced by a speaker who has learned the same—or at least sufficiently the same—shared code. From this it also follows that speakers can use this code to predict when listeners have been provided sufficient cues to discriminate the intended message. In this sense, the relationship between the signal and the message is a function of the speaker's predictions about the meaning that a listener can be expected to be able to discriminate using the signal produced by the speaker in context. With this characterization of the communication process that speech serves to underpin in mind, we now turn our attention to the way these factors influence enhancement and reduction in speech production.

We propose that the different levels of uncertainty that are associated with the signal and the message are critical to explaining why the different kinds of uncertainty that occur in different contexts have such a very different effect on articulation. Moreover, we suggest that the *signal-message-uncertainty distinction* not only explains why these two different sources of uncertainty in speech lead to these apparently contradictory effects, we further suggest that once this distinction is recognized, these effects do not appear to be contradictory at all. Rather, these two different sources of uncertainty simultaneously exert a consistent, if contrastive, influence on articulation:

(1) Lower uncertainty about the message discriminated by the signal leads to reduction.
(2) Lower uncertainty about the signal leads to enhancement.

What is more, once the importance of the *signal-message-uncertainty distinction* is recognized, it becomes clear why two seemingly sensible accounts of effects of uncertainty could nevertheless appear to contradict one another.

This is because from the perspective of this distinction, (1) can be seen as a reformulation of the many insights that led to the hypotheses put forward in the information theoretic framework by Aylett and Turk (2004), Jaeger (2010), and Cohen Priva (2015). Speakers reduce, or even delete word forms or segments when they predict that listeners can discriminate an intended message in context from the signal. This means that under the wrong assumptions about uncertainty about the message, speakers might actually reduce articulations even though the correct strategy would be to enhance them. By contrast, we suggest that when speakers expect that the message will not be fully discriminated, they enhance the signal. This may occur because of the context, because they get appropriate feedback from the listener, or because they find themselves in a noisy environment, (Lindblom, 1990; Junqua, 1993; Buschmeier and Kopp, 2012; Hay et al., 2017).

At the same time, not only is (2) consistent with the present findings, it also captures the theoretical insights captured in the *Paradigmatic Signal Enhancement Hypothesis*. Moreover, in contrast to the *Paradigmatic Signal Enhancement Hypothesis*, the scope of our hypothesis is not constrained to morphological paradigms. Rather, its scope expands to predict potential enhancement effects in all instances in which a signal has to be produced in contexts where its form will be uncertain (see also Linke and Ramscar, 2020; Tomaschek et al., 2020, for enhanced variability associated with uncertainty).

Most importantly, whether a measure—be it activations based on an artificial neural network or probabilistic measures based on information theoretic considerations—operationalizes uncertainty about the signal or the message will ultimately depend on the input-output structure provided to a model—and critically, whether that structure maintains the important distinction between signals and messages. Only when the input-output structure appropriately reflects the relevant cue-outcome relations in a given process can we draw the correct conclusions from the statistical analyses involving these measures. As we have sought to show here, establishing what the appropriate input-output structure to any given process requires detailed analysis and empirical testing. Accordingly, we suggest that questions concerning the way that uncertainty about the message and uncertainty about the signal are to be modeled across the full range of contexts in which speech is produced can only be answered by detailed future research.

6. CONCLUSIONS

We have investigated how uncertainty in the context of inflectional paradigms is associated to phonetic enhancement and reduction of signals discriminating the corresponding inflectional functions. To do so, we trained two learning networks and extracted measures of uncertainty from them. We found that lower uncertainty is associated to phonetic enhancement—supporting work performed within the *Paradigmatic Signal Enhancement Hypothesis* framework. This is only the case when the network was trained on the cognitively appropriate input-output structure, where inputs represent the cognitive cues discriminating articulatory gestures and outputs represent the articulatory gesture at hand. We propose a distinction based on differences in *signal-vs.-message-uncertainty* to account for an apparent contradiction in previous research looking at the effects of uncertainty on the phonetic characteristics of speech.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/8jf5s/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FT conceptualized the study, retrieved the data, and performed the modeling and statistical analysis. FT and MR wrote the manuscript and designed the study. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a collaborative grant from the Deutsche Forschungsgemeinschaft (German Research Foundation; Research Unit FOR2373 Spoken Morphology, Project Articulation of morphologically complex words, BA 3080/3-1 and BA 3080/3-2).

ACKNOWLEDGMENTS

A previous version of the present study was presented at World of Words 2020 online conference and Interfaces of Phonetics 2021. We are thankful for insightful comments from the audience. We thank Ann-Kathrin Reutter for her help tagging the [ɐ]-words, and Maria Heitmeier for her feedback on the manuscript. Likewise, our thanks go to three reviewers for their constructive comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://osf.io/8jf5s/

REFERENCES

Arnold, D., and Tomaschek, F. (2016). "The Karl Eberhards Corpus of spontaneously spoken Southern German in dialogues - audio and articulatory recordings," in *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, eds C. Draxler and F. Kleber (München: Ludwig-Maximilians-Universität München), 9–11.

Arnon, I., and Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition* 122, 292–305. doi: 10.1016/j.cognition.2011.10.009

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., and Shaoul, C. (2018). *ndl: Naive Discriminative Learning.* Available online at: https://CRAN.R-project.org/package=ndl

Aylett, M., and Turk, A. (2004). The Smooth Signal Redundancy Hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47, 31–56. doi: 10.1177/00238309040470010201

Aylett, M., and Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J. Acoust. Soc. Am.* 119, 3048–3058. doi: 10.1121/1.2188331

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: a unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019, 4895891. doi: 10.1155/2019/4895891

Baayen, R. H., and Linke, M. (2020). "Generalized additive mixed models," in *A Practical Handbook of Corpus Linguistics*, eds M. Paquot and S. T. Gries (Cham: Springer International Publishing), 563–591. doi: 10.1007/978-3-030-46216-1_23

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychol. Rev.* 118, 438–481. doi: 10.1037/a0023851

Baayen, R. H., Milin, P., and Ramscar, M. (2016a). Frequency in lexical processing. *Aphasiology* 30, 1174–1220. doi: 10.1080/02687038.2016.1147767

Baayen, R. H., Shaoul, C., Willits, J., and Ramscar, M. (2016b). Comprehension without segmentation: a proof of concept with naive discriminative learning. *Lang. Cogn. Neurosci.* 31, 106–128. doi: 10.1080/23273798.2015.1065336

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013). The time course of perception of coarticulation. *J. Acoust. Soc. Am.* 133, 2350–2366. doi: 10.1121/1.4794366

Bell, A., Brenier, J. M., Gregory, M., Girand, C., and Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *J. Mem. Lang.* 60, 92–111. doi: 10.1016/j.jml.2008.06.003

Bell, M. J., Ben Hedia, S., and Plag, I. (2019). How morphological structure affects phonetic realization in English compound nouns. *Morphology* 31, 87–120. doi: 10.1007/s11525-020-09346-6

Ben Hedia, S., and Plag, I. (2017). Gemination and degemination in English prefixation: phonetic evidence for morphological organization. *J. Phonet.* 62, 34–49. doi: 10.1016/j.wocn.2017.02.002

Boersma, P., and Weenink, P. (2015). *Praat: Doing Phonetics by Computer [Computer Program], Version 5.3.41*. Retrieved from http://www.praat.org/

Bonini, C. P. (1963). *Simulation of Information and Decision Systems in the Firm*. Englewood Cliffs, NJ: Prentice Hall.

Brandt, E., Andreeva, B., and Möbius, B. (2019). "Information density and vowel dispersion in the productions of Bulgarian L2 speakers of German," in *Proceedings of the 19th International Congress of Phonetic Sciences* (Melbourne, NSW), 3165–3169.

Brandt, E., Möbius, B., and Andreeva, B. (2021). Dynamic formant trajectories in German read speech: impact of predictability and prominence. *Front. Commun.* 6, 643528. doi: 10.3389/fcomm.2021.643528

Bröker, F., and Ramscar, M. (2020). Representing absence of evidence: why algorithms and representations matter in models of language and cognition. *Lang. Cogn. Neurosci.* 1–24. doi: 10.1080/23273798.2020.1862257

Browman, C., and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology* 3, 219–252.

Brusini, P., Dehaene-Lambertz, G., Heugten, M. V., Carvalho, A. D., Goffinet, F., Fiovet, A.-C., et al. (2017). Ambiguous function words do not prevent 18-month-olds from building accurate syntactic category expectations: an ERP study. *Neuropsychologia* 98, 4–12. doi: 10.1016/j.neuropsychologia.2016.08.015

Buschmeier, H., and Kopp, S. (2012). "Adapting language production to listener feedback behavior," in *Proceedings of the Interdisciplinary Workshop on Feedback Behaviors in Dialogue* (Portland, OR: Interspeech).

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Lang. Variat. Change* 14, 261–290. doi: 10.1017/S0954394502143018

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Lab. Phonol.* 6, 243–278. doi: 10.1515/lp-2015-0008

Cohen, C. (2015). Context and paradigms: two patterns of probabilistic pronunciation variation in Russian agreement suffixes. *Mental Lexicon* 10, 313–338. doi: 10.1075/ml.10.3.01coh

Daw, N. D., Courville, A. C., and Dayan, P. (2008). "Semi-rational models of conditioning: the case of trial order," in *The Probabilistic Mind*, eds N. Chater and M. Oaksfort (Oxford: Oxford University Press), 431–452.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* 93, 283–321.

Dell, G. S., Burger, L. K., and Svec, W. R. (1997). Language production and serial order: a functional analysis and a model. *Psychol. Rev.* 104, 123.

Dell, G. S., Martin, N., and Schwartz, M. F. (2007). A case-series test of the interactive two-step model of lexical access: predicting word repetition from picture naming. *J. Mem. Lang.* 56, 490–520. doi: 10.1016/j.jml.2006.05.007

Drager, K. K. (2011). Sociophonetic variation and the lemma. *J. Phonet.* 39, 694–707. doi: 10.1016/j.wocn.2011.08.005

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Appl. Linguist.* 27, 1–24. doi: 10.1093/applin/ami038

Fosler-Lussier, E., and Morgan, N. (1999). Effects of speaking rate and word frequency on pronunciations in conversational speech. *Speech Commun.* 29, 137–158.

Fox, B. A., Maschler, Y., and Uhmann, S. (2009). Morpho-syntactic resources for the organization of same-turn self-repair: cross-linguistic variation in English, German and Hebrew. *Z. Verbalen Interaktion* 10, 245–291.

Fromkin, V. (1973). *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language* 47, 27–52. doi: 10.2307/412187

Fuchs, R. (2016). "The acoustic correlates of stress and accent in english content and function words," in *Proceedings of Speech Prosody* (Boston, MA: Boston University), 290–294.

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., and Levy, R. (2019). Neural language models as psycholinguistic subjects: representations of syntactic state. *arXiv[Preprint]*. arXiv:1903.03260. Available online at: https://arxiv.org/pdf/1903.03260.pdf

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *J. Acoust. Soc. Am.* 63, 223–230.

Gittelson, B., Leemann, A., and Tomaschek, F. (2021). Using crowd-sourced speech data to study socially constrained variation in nonmodal phonation. *Front. Artif. Intell.* 3, 565682. doi: 10.3389/frai.2020.565682

Graves, A. (2012). "Supervised sequence labelling with recurrent neural networks" in *Studies in Computational Intelligence* (Berlin; Heidelberg: Springer), 5–13. doi: 10.1007/978-3-642-24797-2

Grodner, D., and Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentenial complexity. *Cogn. Sci.* 29, 261–290. doi: 10.1207/s15516709cog0000_7

Guenther, F. H. (2016). *Neural Control of Speech*. Cambridge, MA: MIT Press.

Hanique, I., and Ernestus, M. (2011). "Final /t/ reduction in Dutch past-participles: the role of word predictability and morphological decomposability," in *Interspeech 2011: 12th Annual Conference of the International Speech Communication Association* (Florence), 2849–2852.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., et al. (2014). Deep speech: scaling up end-to-end speech recognition. *arXiv[Preprint]*. arXiv:1412.5567. Available online at: https://arxiv.org/pdf/1412.5567.pdf

Hastie, T., and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman & Hall.

Hay, J. (2004). *Causes and Consequences of Word Structure*. New York, NY: Routledge.

Hay, J., Podlubny, R., Drager, K., and McAuliffe, M. (2017). Car-talk: location-specific speech production and perception. *J. Phonet.* 65, 94–109. doi: 10.1016/j.wocn.2017.06.005

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Lang. Cogn. Neurosci.* 29, 2–20. doi: 10.1080/01690965.2013.834370

Hockett, C. F., and Hockett, C. D. (1960). The origin of speech. *Sci. Am.* 203, 88–97.

Hoppe, D. B., Hendriks, P., Ramscar, M., and van Rij, J. (2022). An exploration of error-driven learning in simple two-layer networks from a discriminative learning perspective. *Behav. Res. Methods.* (2022). doi: 10.3758/s13428-021-01711-5. [Epub ahead of print].

Hoppe, D. B., van Rij, J., Hendriks, P., and Ramscar, M. (2020). Order matters! influences of linear order on linguistic category learning. *Cogn. Sci.* 44, e12910. doi: 10.1111/cogs.12910

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cogn. Psychol.* 61, 23–62. doi: 10.1016/j.cogpsych.2010.02.002

Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (Tokyo: The National International Institute for Japanese Language), 29–54.

Jordan, M. I., LeCun, Y., and Solla, S. A. (2002). "Neural information processing systems conferences from 1988 to 1999 (CDROM)," in *Advances in Neural Information Processing Systems*.

Jozefowicz, R., Vinyals, O., Schuster, M., Shazeer, N., and Wu, Y. (2016). Exploring the limits of language modeling. *arXiv[Preprint]*. arXiv:1602.02410. Available online at: https://arxiv.org/pdf/1602.02410.pdf

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. *J. Acoust. Soc. Am.* 93, 510–524.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (2001a). "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: John Benjamins), 229–254.

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001b). "The effect of language model probability on pronunciation reduction," in *Proceedings of the 2001 IEEE Conference on Acoustics, Speech, and Signal Processing* (Salt Lake City, UT: IEEE), 801–804.

Juste, F. S., Sassi, F. C., and de Andrade, C. R. F. (2012). Exchange of disfluency with age from function to content words in Brazilian Portuguese speakers who do and do not stutter. *Clin. Linguist. Phonet.* 26, 946–961. doi: 10.3109/02699206.2012.728278

Kemps, R. J., Wurm, L. H., Ernestus, M., Schreuder, R., and Baayen, R. H. (2005). Prosodic cues for morphological complexity in Dutch and English. *Lang. Cogn. Process.* 20, 43–73. doi: 10.1080/01690960444000223

Kuperman, V., Pluymaekers, M., Ernestus, M., and Baayen, H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *J. Acoust. Soc. Am.* 121, 2261–2271. doi: 10.1121/1.2537393

Landauer, T. K., Laham, D., Rehder, B., and Schreiner, M. E. (1997). "How well can passage meaning be derived without using word order? A comparison of Latent Semantic Analysis and humans," in *Proceedings of 19th annual meeting of the Cognitive Science Society* (Mawhwah, NJ), 412–417.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38; discussion 38–75.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773–1781.

Lindblom, B. (1990). "Explaining phonetic variation: a sketch of the H&H theory, in *Speech Production and Speech Modelling*, eds A. Marchal and W. Hardcastle (Dordrecht: Springer), 403–439.

Linke, M., and Ramscar, M. (2020). How the probabilistic structure of grammatical context shapes speech. *Entropy* 22, 90. doi: 10.3390/e22010090

Linzen, T., Dupoux, E., and Goldberg, Y. (2016). Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Trans. Assoc. Comput. Linguist.* 4, 521–535. doi: 10.1162/tacl_a_00115

Lohmann, A. (2018). Cut (n) and cut (v) are not homophones: lemma frequency affects the duration of noun-verb conversion pairs. *J. Linguist.* 54, 753–777. doi: 10.1017/S0022226717000378

Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., and Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology* 28, 71–97. doi: 10.1007/s11525-017-9318-7

Lund, K., and Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods Instrum. Comput.* 28, 203–208.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *J. Phonet.* 25, 187–205.

Malisz, Z., Brandt, E., Möbius, B., Oh, Y. M., and Andreeva, B. (2018). Dimensions of segmental variability: interaction of prosody and surprisal in six languages. *Front. Commun.* 325, 1–18. doi: 10.3389/fcomm.2018.00025

McCloskey, M. (1991). Networks and theories: the place of connectionism in cognitive science. *Psychol. Sci.* 2, 387–395. doi: 10.1111/j.1467-9280.1991.tb00173.x

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, eds C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (Lake Tahoe, NV: NIPS), 3111–3119.

Mooshammer, C., and Fuchs, S. (2002). Stress distinction in German: simulating kinematic parameters of tongue-tip gestures. *J. Phonet.* 30, 337–355. doi: 10.1006/jpho.2001.0159

Mücke, D., Grice, M., Becker, J., and Hermes, A. (2009). Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *J. Phonet.* 37, 321–338. doi: 10.1016/j.wocn.2009.03.005

Munson, B. (2007). Lexical access, lexical representation, and vowel production. *Lab. Phonol.* 9, 201–228.

Neville, H. J., Mills, D. L., and Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cereb. Cortex* 2, 244–258.

Nixon, J., and Tomaschek, F. (2020). "Learning from the acoustic signal: error-driven learning of low-level acoustics discriminates vowel and consonant Pairs. in *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*, 585–591.

Nixon, J. S. (2020). Of mice and men: speech sound acquisition as discriminative learning from prediction error, not just statistical tracking. *Cognition* 197, 104081. doi: 10.1016/j.cognition.2019.104081

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337. doi: 10.1016/s0896-6273(03)00169-7

Öhman, S. (1966). Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am.* 39, 151–168.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). *Buckeye Corpus of Conversational Speech (2nd release)*. Columbus, OH: Department of Psychology, Ohio State University.

Plag, I., and Ben Hedia, S. (2017). "The phonetics of newly derived words: testing the effect of morphological segmentability on affix duration," in *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and the Role of Discourse-related Factors*, eds S. Arndt-Lappe, A. Braun, C. Moulin, and E. Winter-Froemel (Berlin, New York, NY: de Gruyter Mouton), p. 93–115.

Plag, I., Homann, J., and Kunter, G. (2017). Homophony and morphology: the acoustics of word-final S in English. *J. Linguist.* 53, 181–216. doi: 10.1017/S0022226715000183

Podlubny, R., Geeraert, K., and Tucker, B. (2015). "It's all about, like, acoustics," in *Proceedings of the ICPHS IIXX*, (Glasgow).

Port, R. F., and Leary, A. P. (2005). Against formal phonology. *Language* 81, 927–964. Available online at: https://www.jstor.org/stable/4490023

Pouplier, M., and Hoole, P. (2016). Articulatory and acoustic characteristics of German fricative clusters. *Phonetica* 73, 52–78. doi: 10.1159/000442590

Pulvermüller, F. (1999). Words in the brain's language. *Behav. Brain Sci.* 22, 253–279.

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija* 46, 377–396. doi: 10.2298/PSI1304377R

Ramscar, M. (2019). Source codes in human communication. *PsyArXiv*. doi: 10.31234/osf.io/e3hps

Ramscar, M. (2021a). A discriminative account of the learning, representation and processing of inflection systems. *Lang. Cogn. Neurosci.* 1–25. doi: 10.1080/23273798.2021.2014062

Ramscar, M. (2021b). How children learn to communicate discriminatively. *J. Child Lang.* 48, 984–1022. doi: 10.1017/S0305000921000544

Ramscar, M., and Dye, M. (2009). "Error and expectation in language learning: an inquiry into the many curious incidences of "mouses" in adult speech," in *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (Amsterdam), 485–490.

Ramscar, M., Dye, M., and Klein, J. (2013a). Children value informativity over logic in word learning. *Psychol. Sci.* 24, 1017–1023. doi: 10.1177/0956797612460691

Ramscar, M., Dye, M., Klein, J., Ruiz, L. D., Aguirre, N., and Sadaat, L. (2011). "Informativity versus logic: children and adults take different approaches to word learning," in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Boston, MA).

Ramscar, M., Dye, M., and McCauley, S. (2013b). Error and expectation in language learning: the curious absence of 'mouses' in adult speech. *Language* 89, 760–793. Available online at: https://www.jstor.org/stable/24671957

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., and Baayen, R. H. (2014). The myth of cognitive decline: non-linear dynamics of lifelong learning. *Top. Cogn. Sci.* 6, 5–42. doi: 10.1111/tops.12078

Ramscar, M., and Port, R. F. (2016). How spoken languages work in the absence of an inventory of discrete units. *Lang. Sci.* 53, 58–74. doi: 10.1016/j.langsci.2015.08.002

Ramscar, M., and Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: a new approach to the logical problem of language acquisition. *Cogn. Sci.* 31, 927–960. doi: 10.1080/03640210701703576

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The Effects of Feature-Label-Order and their implications for symbolic learning. *Cogn. Sci.* 34, 909–957. doi: 10.1111/j.1551-6709.2009.01092.x

Rapp, S. (1995). "Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov Models / An Aligner for German," in *Proceedings of ELSNET goes east and IMACS Workshop* (Moscow).

Rescorla, R., and Wagner, A. (1972). "A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement," in *Classical Conditioning II: Current Research and Theory*, eds A. H. Black and W. Prokasy (New York, NY: Appleton Century Crofts), 64–69.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition* 64, 249–284.

Rumelhart, D. E., and McCelland, J. L. (1987). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. MIT Press.

Saito, M., Tomaschek, F., and Baayen, H. (2020a). "Relative functional load determines co-articulatory movements of the tongue tip," in *12th International Seminar on Speech Production* (Providence, RI). 210–213.

Saito, M., Tomaschek, F., and Baayen, H. (2020b). "An ultrasound study of frequency and coarticulation," in *12th International Seminar on Speech Production* (Providence, RI). 206–209.

Saltzman, E., and Kelso, S. (1987). Skilled actions: a task-dynamic approach. *Psychol. Rev.* 94, 84–106. doi: 10.1037/0033-295X.94.1.84

Schmitz, D., Baer-Henney, D., and Plag, I. (2021a). The duration of word-final /s/ differs across morphological categories in English: evidence from pseudowords. *Phonetica* 78, 571–616. doi: 10.1515/phon-2021-2013

Schmitz, D., Plag, I., Baer-Henney, D., and Stein, S. D. (2021b). Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Front. Psychol.* 12, 680889. doi: 10.3389/fpsyg.2021.680889

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57, 87–115. doi: 10.1146/annurev.psych.56.091103.070229

Schulz, E., Oh, Y. M., Malisz, Z., Andreeva, B., and Mobius, B. (2016). "Impact of prosodic structure and information density on vowel space size," in *Speech Prosody 2016* (Boston, MA), 350–354.

Schuppler, B., Dommelen, W. A. V., Koreman, J., and Ernestus, M. (2012). How linguistic and probabilistic properties of a word affect the realization of its final /t/: studies at the phonemic and sub-phonemic level. *J. Phonet.* 40, 595–607. doi: 10.1016/j.wocn.2012.05.004

Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., and Malouf, R. (2018). Acoustic differences in morphologically-distinct homophones. *Lang. Cogn. Neurosci.* 33, 32–49. doi: 10.1080/23273798.2017.1359634

Shannon, C. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656.

Shaoul, C., Shilling, N., Bitschau, S., Arppe, A., Hendrix, P., and Baayen, R. H. (2014). *NDL2: Naive Discriminative Learning*.

Shaoul, C., and Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behav. Res. Methods* 42, 393–413. doi: 10.3758/BRM.42.2.393

Stein, S. D., and Plag, I. (2021). Morpho-phonetic effects in speech production: Modeling the acoustic duration of English derived words with linear discriminative learning. *Front. Psychol.* 12, 678712. doi: 10.3389/fpsyg.2021.678712

Sutton, R. S., and Barto, A. G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88, 135. doi: 10.1037/0033-295X.88.2.135

Tomaschek, F. (2020). *The Wizard and the Computer: An Introduction to Preprocessing Corpora Using R*. Technical report. PsyArXiv.

Tomaschek, F., Arnold, D., Broeker, F., and Baayen, R. H. (2018a). Lexical frequency co-determines the speed-curvature relation in articulation. *J. Phonet.* 68, 103–116. doi: 10.1016/j.wocn.2018.02.003

Tomaschek, F., Arnold, D., Sering, K., van Rij, J., Tucker, B. V., and Ramscar, M. (2020). Articulatory variability is reduced by repetition and predictability. *Lang. Speech* 64, 654–680. doi: 10.1177/0023830920948552

Tomaschek, F., Hendrix, P., and Baayen, R. H. (2018b). Strategies for managing collinearity in multivariate linguistic data. *J. Phonet.* 71, 249–267. doi: 10.1016/j.wocn.2018.09.004

Tomaschek, F., and Leeman, A. (2018). The size of the tongue movement area affects the temporal coordination of consonants and vowels–a proof of concept on investigating speech rhythm. *J. Acoust. Soc. Am.* 144, EL410–EL416. doi: 10.1121/1.5070139

Tomaschek, F., Plag, I., Ernestus, M., and Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naive discriminative learning. *Journal of Linguistics* 57, 123–161. doi: 10.1017/S0022226719000203

Tomaschek, F., and Tucker, B. V. (2021). The role of coarticulatory acoustic detail in the perception of verbal inflection. *JASA Express Lett.* 1, 085201. doi: 10.1121/10.0005761

Tomaschek, F., Tucker, B. V., Fasiolo, M., and Baayen, R. H. (2018c). Practice makes perfect: the consequences of lexical proficiency for articulation. *Linguist. Vanguard* 4, 1–13. doi: 10.1515/lingvan-2017-0018

Tomaschek, F., Tucker, B. V., Ramscar, M., and Baayen, R. H. (2021). Paradigmatic enhancement of stem vowels in regular English inflected verb forms. *Morphology* 31, 171–199. doi: 10.1007/s11525-021-09374-w

Tucker, B. V., Sims, M., and Baayen, R. H. (2019). *Opposing Forces on Acoustic Duration*. Technical report. PsyArXiv.

Tucker, B. V., and Tomaschek, F. (forthcoming). "Speech production: where does morphology fit?" in *Current Issues in the Psychology of Language*, ed D. Crepaldi (London: Routledge).

Turk, A., and Shattuck-Hufnagel, S. (2020). "Speech timing," *Oxford Studies in Phonology and Phonetics* (Oxford University Press).

van Rij, J., Wieling, M., Baayen, R. H., and van Rijn, H. (2015). *itsadug: Interpreting Time Series, Autocorrelated Data Using GAMMs*. Available online at: https://cran.r-project.org/web/packages/itsadug/index.html

Vujovic, M., Ramscar, M., and Wonnacott, E. (2021). Language learning as uncertainty reduction: the role of prediction error in linguistic generalization and item-learning. *Journal of Memory and Language* 119, 104231. doi: 10.1016/j.jml.2021.104231

Whalen, D. H. (1983). Perceptual effects of phonetic mismatches. Doctoral dissertation, Yale University.

Widrow, B., and Hoff, M. E. (1960). *Adaptive Switching Circuits.* New York, NY: IRE. p. 96–104.

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., et al. (2016). Investigating dialectal differences using articulography. *J. Phonet.* 59, 122–143. doi: 10.1016/j.wocn.2016.09.004

Wilcox, E., Levy, R., Morita, T., and Futrell, R. (2018). "What do RNN language models learn about filler-gap dependencies?" in *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* (Brussels), 211–221.

Wood, S. N. (2006). *Generalized Additive Models*. New York, NY: Chapman & Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. B* 73, 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Mem. Cogn.* 7, 411–419. doi: 10.3758/BF03198257

Wright, R. (2004). "Factors of lexical competition in vowel articulation," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, eds J. Local, R. Ogden, and R. Temple (Cambridge: Cambridge University Press), 75–87.

Zsiga, E. C. (1992). *Acoustic Evidence for Gestural Overlap in Consonant Sequences*. Haskins Laboratories Status Report on Speech Research.

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

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

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fpsyg.2022.841638

Check for updates

Morphosyntactic Skills Influence the Written Decoding Accuracy of Italian Children With and Without Developmental Dyslexia

Emanuele Casani[1]*, Mila Vulchanova[2]* and Anna Cardinaletti[1]

[1] Department of Linguistics and Comparative Cultural Studies, Università Ca' Foscari Venezia, Venice, Italy, [2] Department of Language and Literature, Faculty of Humanities, Norwegian University of Science and Technology, Trondheim, Norway

Some types of developmental dyslexia (DD) are associated with morphology. Deep DD leads to morphological and semantic errors, and possible comorbidity with syntactic deficits; phonological-output-buffer DD causes problems in decoding longer morphologically complex words. In addition, cross-linguistic studies highlight the effects of morphological awareness on reading accuracy and fluency. The role of morphosyntactic abilities on reading is, however, not clear. This study explores the influence of morphosyntactic competence on reading in Italian children with and without DD. A total of 14 children with DD and 28 with Typical Development (TD) attending the Italian primary school were tested on written decoding, syntactic comprehension of different grammatical structures, and syntactic production of direct object clitic pronouns. DD children were significantly less accurate and slower in reading than TD children. Syntactic skills of the two groups did not differ significantly, but some differences in their acquisitional pace emerged. Syntactic comprehension and production of direct-object-clitic pronouns predicted reading accuracy standard scores, thus suggesting that morphosyntactic abilities, beyond clitics' weak phonological status, affect decoding accuracy. Decoding accuracy was influenced by reading errors related to morphology (morphological, semantic, and phonological-output-buffer errors). Decoding speed was a specific weakness of DD children and was rather affected by multi-letter combinations. Consistent with a *dual-route approach to orthographic processing*, we argue that accuracy depends on *fine-grained* decoding strategies maximizing the precise ordering of letters, thus it is more sensitive to morphosyntactic skills. Morphological reading errors were associated with phonologically weak (determiners, clitic pronouns, and prepositions) and salient words (verbs). This suggests that the decoding of function words and morphologically complex words is particularly demanding and related to both phonological and morphosyntactic skills. Age had a negative predictive effect on semantic errors, compatible with the gradual acquisition of lexical decoding strategies, which seemed to be slowed down by DD. We conclude that oral morphosyntactic skills play a role in reading accuracy in the Italian shallow orthography for both DD and TD children. It is then advisable to assess

OPEN ACCESS

Edited by:
Vito Pirrelli,
National Research Council (CNR), Italy

Reviewed by:
Maria Garraffa,
University of East Anglia, United Kingdom
Claudia Marzi,
Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC), Italy

****Correspondence:***
Emanuele Casani
emanuele.casani@unive.it
Mila Vulchanova
mila.vulchanova@ntnu.no

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

Received: *22 December 2021*
Accepted: *25 March 2022*
Published: *27 April 2022*

Citation:
Casani E, Vulchanova M and Cardinaletti A (2022) Morphosyntactic Skills Influence the Written Decoding Accuracy of Italian Children With and Without Developmental Dyslexia. Front. Psychol. 13:841638. doi: 10.3389/fpsyg.2022.841638

children's linguistic profile during DD diagnoses to establish whether some reading errors are related to morphosyntactic weakness. In this case, *ad hoc* morphosyntactic training might support reading accuracy.

Keywords: reading, developmental dyslexia, dual-route model, dual-route approach to orthographic processing, morphology, syntax, clitic pronouns

INTRODUCTION

Reading is a complex activity that involves several underlying abilities including language, metalanguage, and cognitive skills (cf. Nagy and Townsend, 2012). In this study, we bring evidence of the role of morphosyntactic skills on decoding accuracy in Italian children with and without developmental dyslexia (henceforth, DD). We also show that some decoding errors that have to do with morphology to different degrees play a central role in decoding accuracy.

Dual-Route Model of Reading

There is wide consensus on the fact that an accurate model to describe typical reading aloud processes should include a lexical and a sublexical route (Coltheart et al., 2001; for a review, see Castles, 2006; for a case study arguing for the existence of a third route of reading, see Wu et al., 2002). The lexical route allows reading by accessing the lexicon for previously seen written words stored in long-term memory, whereas the sublexical route uses a set of mapping rules to convert graphemes into phonemes, thus allowing the reading of every regular word, both known and unknown, in particular in shallow orthographies. At the top of the dual-route model, there is a common stage of orthographic visual analysis, which is responsible for letter identification, encoding of letter position within the word, and binding of letters to words. In the last stage, the phonological string generated through either the lexical or sublexical route is sent to the phonological output buffer, a short-term interface between phonological representations and articulatory motor programming having the function to keep the information until full production and to assemble phonological strings into larger units (Zoccolotti et al., 2005; Castles, 2006; Friedmann and Coltheart, 2018). Recent studies reveal that different decoding error types can be associated with atypical functioning or non-functioning of specific sections of the dual-route model of reading, which can give rise to different types of DD (Friedmann and Coltheart, 2018), even in a shallow-orthography language like Italian (Traficante et al., 2017). **Figure 1** displays the dual-route model of reading, decoding errors associated with the different sections of the model, and related types of DD.

Some decoding error types have to do with morphology to different extents. In particular, morphological and semantic decoding errors can be associated with a deficit in both the lexical and sublexical route of reading, resulting in deep DD (Stuart and Howard, 1995). In this condition, words that can be read via meaning by resorting to their visual imagery properties are generally preserved, whereas function words and morphologically complex words are particularly problematic. The difficulty with function words might depend on their abstractness, which makes them particularly difficult to be imagined (Friedmann and Coltheart, 2018). Function words can be replaced with visually similar lexical words, other function words, or just omitted. The *imageability effect* could also determine the difficulty with morphologically complex words, which can be decomposed into lexical (bases, stems, or roots) and functional chunks (morphological affixes) requiring the co-activation of different sections of the reading model. In particular, bases, stems, or roots might be read by resorting to the semantic lexicon, whereas affixes via a direct (lexical) route linking the orthographic input buffer to the phonological output buffer (see **Figure 1**). If this route is impaired, morphologically complex words can be simplified through omissions and substitutions of morphological affixes (Friedmann and Coltheart, 2018). Children with deep DD can also present with syntactic deficits, which might make it difficult to resort to the context in reading (Friedmann and Coltheart, 2018).

A deficit in the phonological output buffer (phonological-output-buffer dyslexia) can be responsible for errors in decoding longer morphologically complex words, characterized by omissions, substitutions, and transpositions of some phonemes.

The possible presence of syntactic deficits in children and adults with DD (Bishop, 1991; Muter and Snowling, 1998; Talli et al., 2013; Cardinaletti and Volpato, 2015; Friedmann and Coltheart, 2018) suggests that these deficits might contribute to difficulties in retrieving function words and morphological affixes, which convey morphosyntactic information, during reading. In particular, we wonder whether the three error categories described above (morphological, semantic, and phonological-output-buffer errors) might be influenced by the reader's morphosyntactic skills. Namely, whether adequate morphosyntactic competence might improve children's familiarity with text chunks encoding morphosyntactic information, thus allowing faster retrieval through the activation of the direct (lexical) route from the orthographic input lexicon to the phonological output buffer. At the same time, good morphological knowledge might help individuals with DD in accessing subparts of morphologically complex words, thus benefitting from the cumulated frequency of morphemes to process shorter inputs by co-activation of the direct lexical route beside the semantic lexicon.

A Dual-Route Approach to Orthographic Processing

By applying the dual-route model of reading to a smaller scale of granularity, Grainger and Ziegler (2011) propose a dual-route approach to orthographic processing. They suggest that optimization of print-to-meaning mapping takes place thanks

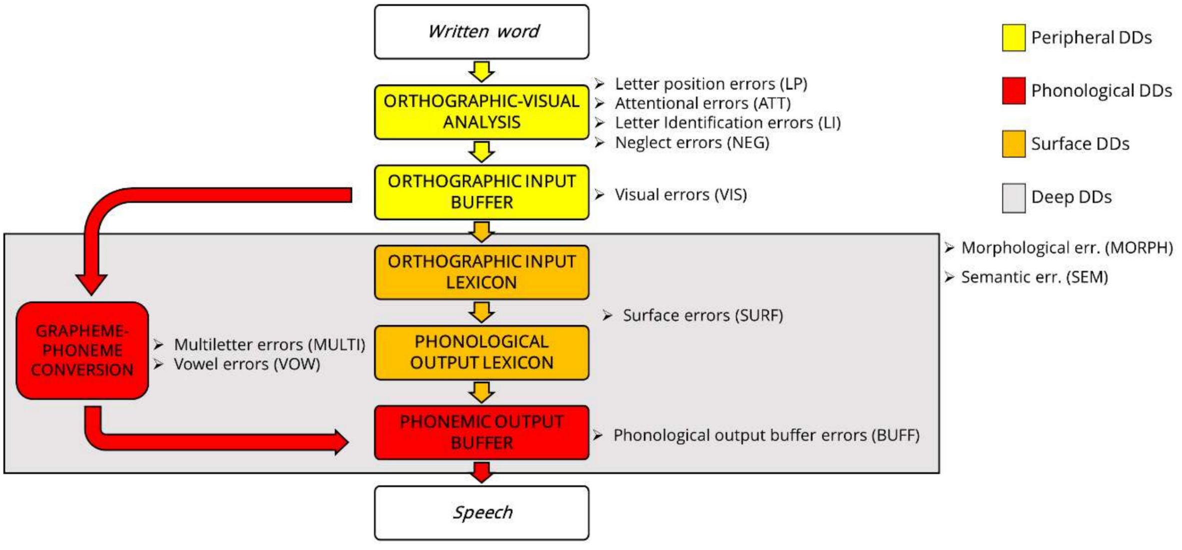

FIGURE 1 | Dual-route model of reading aloud, with different error types associated with each section of the model, and consequential types of DD. Yellow cases display the visual-orthographic step of reading; orange cases display the lexical route; and red cases display the sublexical route of reading, including the phonological output buffer (Casani, 2021, adapted after Friedmann and Coltheart, 2018).

to two distinct learning constraints based, respectively, on the prelexical orthographic coding processes of *diagnosticity* and *chunking*.

Chunking allows the detection of relevant letter combinations corresponding to pre-existing sublexical phonological and morphological representations. It takes place along the *fined-grained* route, which is activated by frequently co-occurring contiguous letter combinations with a precise letter ordering, and precise placement with respect to the beginning and ending of words. These include prefixes and suffixes, which are subject to morpho-orthographic processing.

On the other hand, *diagnosticity* allows for the selection of letter combinations that are informative with respect to word identity. It takes place along the *coarse-grained* route, which codes for approximate letter position within words, irrespective of letter contiguity. This route benefits from the combinations of most visible letters that best constrain word identity. So, it provides a lower precision level in coding letter-position information compared to the fine-grained route, but a higher speed level because it can provide faster top-down activation of whole-word representations. In skilled readers, when most visible letters combined with contextual constraints are not sufficient to activate top-down constraints, the fine-grained route intervenes to disambiguate the information.

Parallel development and smooth integration of the two routes enable the emergence of morpho-semantic and morpho-orthographic representations as well as increased sensitivity to morphological structure, thus reducing the effects of word length and phonological recoding (Grainger and Ziegler, 2011). Since the detection of morphological constituents is the key mechanism of morpho-orthographic chunking (which is performed along the fine-grained route), we suppose that higher morphological competence should facilitate the detection of morphological constituents, thus increasing the reading performance, in particular concerning accuracy. In the following section, we report evidence that morphology influences word decoding through automatic morpho-orthographic segmentation.

Morphological Awareness and Reading

Besides the vast literature around phonological awareness, orthographic competence, and rapid automatized naming (RAN) (for a review, see Casani, in preparation, 2021), cross-linguistic research provides evidence of the role of morphological awareness as a predictor of word reading accuracy (Burani et al., 2008; Traficante et al., 2011), fluency (Fowler and Liberman, 1995; Carlisle and Katz, 2006; Roman et al., 2009), and comprehension (Deacon and Kirby, 2004; Nagy et al., 2006; Tong et al., 2011). Verhoeven and Perfetti (2011) highlight the role of morphology in reading across a wide range of languages, and suggest that morphology, "which is foundational for language knowledge, is universally part of reading, subject to constraints imposed by the language and by how the writing system encodes that language." Besides the universal phonological principle that all writing systems support the activation of phonology at their smallest functional grapheme units (e.g., Perfetti, 2003), they suggest that cross-linguistic research might lead to a universal morphology principle. The ease of word identification and the role of morphology may vary across languages depending on their orthographic depth (Frost et al., 1987) and their morphological richness (Vannest et al., 2002).

According to the *orthographic depth hypothesis* (Frost, 2006), the opaque relationship between phonemes and graphemes in deep orthographies is handled by resorting to lexical mediation. The extent of involvement of lexical mediation is determined by the orthographic depth of the language. Before learning to read, words are stored as holistic phonological units. As literacy is acquired, these bigger phonological representations gradually give way to syllable and then phoneme representations and determine a restructuring of the learner's lexicon granularity. According to the *grain size theory* (Frost et al., 1987), phonology offers a bigger scale and orthography a smaller scale of granularity, which are represented by phonological units and letters, respectively. The degree of consistency between phonemes and letters might determine the speed of reading development (Vulchanova and Farukh, 2018).

In Italian, a shallow orthography with high grapheme-phoneme consistency, morphological information has been shown to influence both reading fluency (Burani, 2010) and accuracy (Angelelli et al., 2014). Children of different reading ages take advantage of morphemic lexical units (Burani et al., 2002). "Morphemes may develop as orthographic and phonological salient reading units" (Burani et al., 2008, p. 254), and these "common letter patterns might become consolidated in lexical memory" (Deacon et al., 2011, p. 476). Masked-priming experiments (e.g., Meunier and Longtin, 2007) confirm that skilled adult readers possess orthographic representations that are structured morphologically and activated before representations of whole words. How this (quasi-)regular trend influences the acquisition of reading across languages is not clear.

There is cross-linguistic evidence (Quémart et al., 2011; Beyersmann et al., 2012; Dawson et al., 2018) that adult skilled readers process complex words and non-words based on morphological structure. Masked priming experiments across English (Beyersmann et al., 2012) and French (Beyersmann et al., 2015) showed robust morphological priming effects on word recognition for child participants, but only when morphological primes had a semantically transparent relationship with targets (e.g., darkness-DARK). Beyersmann et al. (2012) found no evidence that English children aged 8–10 use morpho-orthographic analysis: priming effects for pseudo-morphological pairs (e.g., corner-CORN) could not be distinguished from those based on non-morphological form overlap (e.g., brothel-BROTH). In a related comparison, Beyersmann et al. (2015) could not differentiate masked priming effects for suffixed non-word pairs (e.g., tristerie-TRISTE) and non-suffixed non-word pairs (e.g., tristald-TRISTE) in French readers aged 7–11 (see also Hasenäcker et al., 2016, for a similar study in German). The developmental trajectory observed in English (Beyersmann et al., 2012) also appears in Hebrew, a Semitic language with a very rich morphological structure (Schiff et al., 2012).

In Italian, morphological awareness can affect decoding since the second grade (Burani et al., 2002, 2008; Marcolini et al., 2011; Traficante et al., 2011). Interestingly, a facilitating effect on word reading speed was observed in second-graders and children with DD, whereas in older skilled readers it was limited to low-frequency words. Furthermore, a cross-linguistic study on English and French (Casalis et al., 2015) suggests a higher degree of morphological processing efficiency in French (affecting both accuracy and latencies in a lexical decision task) than in English (affecting accuracy only).

Studies on morphologically productive languages like French (Quémart et al., 2011) and Hebrew (Schiff et al., 2012) provide evidence of morpho-orthographic decomposition in young readers as in adults, differently from English children aged 7–10 years (Beyersmann et al., 2012). Burani et al. (2002) report that Italian children aged 8–10 read aloud morphologically structured non-words more quickly and accurately than non-words without morphological structure. In a similar reading aloud experiment, children aged 9–11 read aloud morphologically complex English words with a high-frequency stem more quickly and accurately than those with a lower-frequency stem (Deacon et al., 2011). In a lexical decision task, Casalis et al. (2015) report that English and French children between the ages of 7 and 10 found morphologically structured non-words (e.g., *gifter*) harder to reject than non-words without a morphological structure (e.g., *curlip*). The same pattern was reported by Burani et al. (2002) for Italian children of similar age, thus replicating the pattern observed in skilled readers (Crepaldi et al., 2010).

In more recent research using this paradigm with three age-groups of developing English readers (ages 7–9, 12–13, and 16–17), the two younger groups showed an effect of morphological structure on accuracy, whereas only older adolescents (16–17 years old) and adults showed this effect on reaction time (Dawson et al., 2018).

Häikiö et al. (2011) examined the role of morphology in Finnish reading development by measuring participants' eye movements while they read sentences containing either a hyphenated (e.g., *ulko-ovi* "front door") or concatenated (e.g., *autopeli* "racing game") compound. The participants were Finnish second, fourth, and sixth graders. Fast second graders and all fourth and sixth graders read concatenated compounds faster than hyphenated compounds. This suggests that they resort to slower morpheme-based processing for hyphenated compounds but prefer to process concatenated compounds via whole-word representations.

Further eye-tracking research showed that eye movements are affected by both whole-word frequency and first-constituent frequency. In processing Dutch (Kuperman et al., 2008, 2009) and Italian (Marelli and Luzzatti, 2012) compounds, frequency in first-fixation duration, namely the time initially spent by the reader fixating the target element, correlates negatively with the frequency of whole-words.

In processing Italian derived words, stem frequency has a facilitating effect on first-fixation duration only within sentences prompting a semantically transparent interpretation of the word, whereas a stem-frequency effect is inhibitory within sentences prompting an opaque interpretation of the target word (Amenta et al., 2015). Word frequency, as well as the amount of information and the size of the morphological family of the suffix, affects the reading times of Dutch derived words with shorter suffixes. This is interpreted as a *relative entropy* effect of morphemes. Affixes occurring more frequently are more salient and processed faster (Kuperman et al., 2010).

In English, root frequency affected fixation times for longer (about eight letters) but not shorter (about six letters) prefixed words, whereas whole-word frequency for shorter but no longer prefixed words (Niswander-Klement and Pollatsek, 2006). In Italian, base and word frequency affected first-fixation duration for nouns derived from noun bases differently: base frequency facilitated first fixation, whereas word frequency had an inhibitory effect (Traficante et al., 2018).

Behavioral data from languages with rich morphology show differences in lexical decision times for nouns, adjectives, and verbs. Kostić and Katz (1987) attribute this effect in Serbo-Croatian to the number of inflectional alternatives available for each grammatical class. Deutsch et al. (1998) ascribe the differences in processing verbs and nouns in Hebrew, beyond semantic and syntactic components, to the distributional properties of constituents, namely to the fact that "when a morpheme is common to more words in the language, its impact on processes of morphological decomposition is prominent" (p. 1,252). Italian skilled adult readers recognized verbs slower than nouns and adjectives. Moreover, latencies for verbs, but not for nouns or adjectives, correlated with their base frequency (Colombo and Burani, 2002; Traficante and Burani, 2003).

Marcolini et al. (2011) showed that Italian children with DD read pseudowords made up of a root and a derivational suffix faster and more accurately than simple pseudowords. However, only dyslexic and reading-matched younger children benefited from morphological structure in reading words aloud. The authors investigated the effects of word frequency and word length on complex-word reading in Italian dyslexic and skilled readers and showed that word frequency affects the probability of morpheme-based reading, interacting with reading ability. Young skilled readers named polymorphemic words faster than simple words only when they were of low frequency, whereas they read high-frequency polymorphemic words as fast as high-frequency simple words. By contrast, poor readers took advantage of polymorphemic words irrespective of word frequency, while adult readers showed no facilitating effect of morphological structure. Similar findings emerged in English (Carlisle and Stone, 2005) and Danish (Elbro and Arnbak, 1996) populations, where only younger and dyslexic children read derived words faster than monomorphemic words, whereas morphological complexity did not affect reading speed in the elder skilled children. This indicates that morpheme-based reading is effective for both poor and skilled young readers when a whole-word representation is not firmly established in the reader's orthographic lexicon, because either the whole word is not familiar to the reader or (s)he has poor reading skills (Marcolini et al., 2011).

Angelelli et al. (2014) found that morphological information in Italian is a useful resource for both reading and spelling, as typically developing children benefit from the presence of morphological structure when they read and spell non-words. In processing low-frequency words, however, morphology facilitates reading, but not spelling. They attribute their results to successful cooperation between lexical and sublexical processes in reading and spelling, which facilitate morpho-lexical access.

These data converge on the fact that morphological awareness facilitates lexical reading for low-frequency words that otherwise would not probably be represented as a whole in the mental lexicon, even in Italian and other shallow orthographies (for Spanish, see Defior et al., 2008; Suárez-Coalla et al., 2017), where the orthography-phonology mapping might be expected as sufficient for correct decoding and spelling. This suggests that morphological competence and its interface with syntactic competence play a role in written decoding.

(Morpho)syntactic Competence and Reading

Syntactic competence has been generally explored in relation to reading comprehension, of which it is deemed as a good predictor (see for instance, Simpson et al., 2020 for Spanish speakers; Morvay, 2012 for speakers of English as a foreign language; Chik et al., 2012 for Chinese speakers). Fewer studies analyzed its effects on decoding. Sana Teixeira et al. (2016) revealed relations of syntactic awareness with both reading accuracy and reading speed in typically developing Brasilian children in primary school (age: 9;0–11;7).

Traficante et al. (2018) analyzed the role of the base word distributional properties on eye-movement behavior and found an inhibitory base frequency effect, but no word frequency effect for nouns derived from verb bases. They suggest that syntactic context, calling for a noun in the target position, is responsible for the inhibitory effect when a verb base is detected, thus hampering the lexical access to the corresponding base-suffix combination.

A recent longitudinal study (Casani, in preparation, 2021), besides confirming a strong predictive role of oral syntactic comprehension on reading comprehension, found that syntactic comprehension, measured in the last year of kindergarten and second grade of primary school through the Italian version of Bishop (2009), predicted both word and text decoding accuracy in second grade. In particular, it was a predictor of surface errors, which are related to the lexical route of reading. This highlights the strict relation between (morpho)syntactic and lexical competence, as confirmed by the correlations of syntactic comprehension with both receptive (ϱ = 0.540, p = 0.000) and productive (ϱ = 0.554, p = 0.000) vocabulary. In the same research, longitudinal syntactic-comprehension skills predicted the emergence of difficulties in word and non-word writing, beyond reading comprehension difficulties, in second and third grade; longitudinal syntactic production of third-person direct object clitic pronouns predicted the emergence of decoding accuracy and decoding speed difficulties.

These data find support in the study of event-related potentials in adult speakers of German (Cantiani et al., 2013), a morphologically rich language with relatively shallow orthography like Italian. Seventeen subjects with DD and seventeen with TD were presented with oral stimuli with morphosyntactic violations. DD participants showed anomalous morphosyntactic processing, especially when morphosyntactic violations were expressed by both lexical and inflectional changes. Furthermore, anomalous morphosyntactic processing was mediated by lexical cues instead of acoustic salience. Several behavioral studies also report the presence of syntactic deficits in subjects with DD (Bishop, 1991; Muter and Snowling, 1998; Talli et al., 2013; Cardinaletti and Volpato, 2015).

The Current Study

The literature mentioned in previous sections shows that some decoding errors related to the central routes of reading (i.e., the lexical and sublexical route) have to do with morphology to different extents. Moreover, morphological awareness and its interface with syntax have a prominent cross-linguistic role in reading accuracy and speed. Yet, only few studies have investigated the influence of general (morpho)syntactic competence on written decoding.

In the present study, we analyze the effects of oral (morpho)syntactic comprehension and production, as well as different reading errors, including morphological ones, on the written decoding of Italian primary-school children with and without DD.

We expected to find effects of (morpho)syntactic competence on decoding accuracy, due to greater familiarity with morpho-lexical chunks and distributional properties, by children with higher morphosyntactic skills. We did not expect the same effects on decoding speed. In fact, according to the dual-route approach to orthographic processing (Grainger and Ziegler, 2011), speeding up reading processes in skilled readers depends on the ability to process not only *fine-grained* orthographic strings preserving the information about letter ordering, as morphemes are, but also *coarse-grained* orthographic representations, which code for the presence of informative letter combinations in the absence of precise positional information (Traficante et al., 2018).

MATERIALS AND METHODS

Participants

A total of 53 Italian monolingual children in primary school were initially tested (for the results of the whole sample, see Casani, 2020a,b). They were recruited in primary schools in the Center and South of Italy. In total, 11 of them were excluded due to the presence of language disorders or the alleged presence of developmental problems based on teachers' reports. Among the 42 participants (24 females + 18 males), one child was in second grade, 16 children were in third grade, 8 were in fourth grade, and 17 were in fifth grade.

A total of 14 children [age 7;5–10;9 (M = 9;9, SD = 0;11)] had a diagnosis of general DD, and 28 [age 8;4–11;3 (M = 9;6; SD = 0;11)] were age-matched Typically Developing (TD) children. Diagnoses were established by the Italian public health system (ASL) or authorized private clinical centers. Children included in the TD group were not reported by teachers for any language or learning problems.

Materials and Procedures

The participants' families signed informed parental consent. Children expressed their willingness to participate in the activities during an exploratory interview. The procedures followed the ethical principles of the Declaration of Helsinki. Children were tested individually in silent and adequately lit rooms in school facilities. Tests were administered by the first author. Different abilities including syntactic comprehension, syntactic production, and reading were tested. Syntactic comprehension was tested through a standardized picture-sentence matching task extracted from the BVN 5–11 (Neuropsychological Assessment Battery for the developmental age) (Bisiacchi et al., 2005). It is a reduced adaptation of Bishop (2009) consisting of 18 items investigating the comprehension of different syntactic structures (for additional information, see **Supplementary Data**).

Syntactic production was tested through a non-standardized elicitation task of third-person direct object clitic pronouns (Arosio et al., 2014). These are complex structures requiring the mastering of phonological, morphosyntactic, syntactic, and pragmatic skills. The test elicits 12 third-person singular direct object clitic pronouns (6 masculine + 6 feminine) under conditions of gender and number match with the sentential subject. Casani and Cardinaletti (2021) recently showed that this morphosyntactic combination is significantly more accessible than combinations including gender mismatch between the clitic pronoun and the sentential subject. Children were shown two-slide cartoons, where the recorded voice of an Italian male native speaker presented the situation through a brief sentence [e.g., *In questa storia c'è un signore che vuole pescare un pesce* (In this story, there is a man who wants to fish a fish)] and then asked a question [*Guarda! Cosa sta facendo al pesce?* (Look! What is Ø doing to the fish? → Look! What is he doing to the fish?)]. The restrictive context should elicit a null-subject sentence containing a third-person direct object clitic pronoun agreeing in gender and number with its antecedent (*Lo*$_{\_3rd.sing.masc.dir.clit}$. *sta pescando.* (Ø it$_{_3rd.sing.masc.dir.clit}$ is fishing. → He is fishing it). Grammatical and pragmatically appropriate responses were assessed as correct. The test was administered through a 15-inch laptop screen with stereo speakers.

Decoding accuracy and speed were tested through standardized texts calibrated to the students' grades. These stem from the MT-2 battery (Cornoldi and Colpo, 2011). They were black printed on A4 white paper.

Analyses

Syntactic Comprehension

To our aims, we opted to only analyze the effects of grammar-focused items. According to the authors of the test, items 1–8 are focused on lexicon, whereas items 9–18 are focused on grammar. We do not agree with considering item number 5 (*La mucca le sta guardando* ["The cow them is watching" = The cow is watching them)] as a lexicon-focused item because it includes the interpretation of a third-person direct object clitic pronoun, a structure requiring high-level morphosyntactic and syntactic skills (for the difficulties involved in third-person direct object clitic pronouns, see Arosio et al., 2014; Casani, 2020c). We then included it among grammar-focused items and analyzed it as such. The complete list of the structures analyzed is in **Supplementary Data**. Correct responses (pointing to the right picture) were assigned one point; incorrect responses were assigned 0 points. Proportional scores were analyzed.

Syntactic Production

Responses containing a grammatical third-person direct object clitic pronoun were assigned one point. Ungrammatical responses (gender or number errors in clitic agreement, clitic omissions, clitic-position errors) and sentences containing a full determiner phrase instead of the clitic (which is grammatical but pragmatically inappropriate) were assigned 0. Proportional scores were analyzed.

Reading

The total error number and the speed rate (syllables per second) were computed and converted into standard scores. Accuracy and speed standard measures (*Z*) were analyzed.

An analysis of proportional reading errors based on an adaptation of the coding scheme by Friedmann and Coltheart (2018) was performed. Eleven types of decoding errors were detected, as shown below.

1. LP (Letter Position errors), e.g., dispiacere → *despicare; presso → perso.
2. ATT (ATTentional errors), e.g., dal tetto → dal letto; se mi hai letto → se mai hai letto.
3. LI (Letter Identity errors), e.g., due → bue; babbo → *papo.
4. NEGL (NEGLect errors), e.g., rallegrò → allegro; bigi → *bi.
5. VIS (VISual errors), e.g., nipotino → *nipotivino; fradicia → *fraggida.
6. SURF (SURFace errors), e.g., [']*fradicia*–**fra*[']*dicia*; *si* [']*presero* → *si* **pre*[']*sero.*
7. MULTI (MULTIletter errors), e.g., *cresceva*–**crescheva*; *foglio* →**forgerio.*
8. VOW (VOWel errors), e.g., dimissioni–*dimessioni; a lungo → e lungo.
9. MORPH (MORPHological errors), e.g., *dimenticando* → *dimenticato*; *sentì* → *sente.*
10. SEM (SEMantic errors), e.g., *ripeté* → *ribatté*; *a bocca aperta* → *a mano aperta.*
11. BUFF (phonological-output-BUFFer errors), e.g., *ringraziamenti* →**rangrizzamenti*; *sentenziava* → **sensiva*.

Error categories 1, 2, 3, 4, and 5 are related to the orthographic-visual-analysis stage of reading, at the top of the reading model; category 6 is related to the lexical route of reading; categories 7 and 8 are related to the sublexical route of reading; categories 9 and 10 are related to both central routes, namely the lexical and sublexical routes; category 11 is related to the phonological output buffer, at the bottom of the reading model (see **Figure 1**).

Analyses of the elements involved in different error types (adjectives, adverbs, clitic pronouns, conjunctions, determiners, nouns, prepositions, pronouns, verbs, whole phrases) and of superficial errors (additions, omissions, changes, moves, and substitutions with a different part of speech) associated with different error types were performed.

Statistical Analyses

Two generalized mixed models (GMM) were run to analyze the effects of group, syntactic comprehension, and syntactic production on standard measures of decoding accuracy and speed. Decoding errors were entered as random effects. This allowed us to control simultaneously for the effects of all error types.

In addition, we ran 11 GMMs to analyze the effects of the same factors/variables on each error type.

As the distribution of children across grades was not homogeneous (for the number of participants in each grade, see **Supplementary Tables 1, 4**), we opted to enter age in months as a random effect, which allowed us to control for a more analytical measure than the grade variable.

In addition, we ran two GMMs to analyze the combined effect of age and group (age variable nested in the group variable) on syntactic comprehension and production, respectively; and two GMMs, where we replaced the age variable with grade (nested in the group variable).

We finally checked the combined effect of grade and group (grade variable nested in the group variable as a fixed effect) on SEM-error proportions (dependent variable).

Fisher's exact tests with *post hoc Z* (*Bonferroni*) were used to analyze the association between each error type and different parts of speech.

Statistical analyses were run in SPSS-24 and are described in detail in section "Results."

RESULTS

Differences Between Groups

Figure 2 displays the distribution of standard scores obtained in decoding, and of percentage scores obtained in syntactic tests by TD and DD children.

Four distinct GMMs with scores obtained on syntactic (comprehension and production) and decoding (accuracy and speed) tests as respective dependent variables, the double level of group as a fixed effect, and children's age (in months) as a random effect revealed a significant effect of group on decoding accuracy [$F(1, 40) = 8.584$, $p = 0.006$] and speed [$F(1, 40) = 16.727$, $p = 0.000$]. TD children were significantly more accurate and faster than DD-children, as shown in **Table 1**.

No significant effect of group emerged on syntactic comprehension ($p = 0.399$) and syntactic production ($p = 0.535$). The significance of the random effect was also analyzed and revealed no effect of age on any variable ($0.259 < p < 0.697$).

Effects of Grade on Morphosyntactic Skills

As we analyzed (morpho)syntactic proportional scores instead of standard scores,[1] we verified in more depth the absence of effects of age on morphosyntactic skills by running two robust GMMs with syntactic comprehension and production scores as respective dependent variables, and the age variable nested in the group variable as a fixed effect. Age (combined with group) confirmed no predictive effect on syntactic comprehension ($p = 0.184$) and production ($p = 0.187$).

Then, we ran two additional models by replacing the age variable with grade. Grade (nested in the group

[1]The syntactic-comprehension test (Bisiacchi et al., 2005) includes separate standard scores for grammar-focused items and lexicon-focused items, thus allowing to evaluate the two areas independently, at least in the authors' intentions. In our opinion, however, the test should be restandardized by considering item number 5 among grammar-focused items (see the section "Analyses").

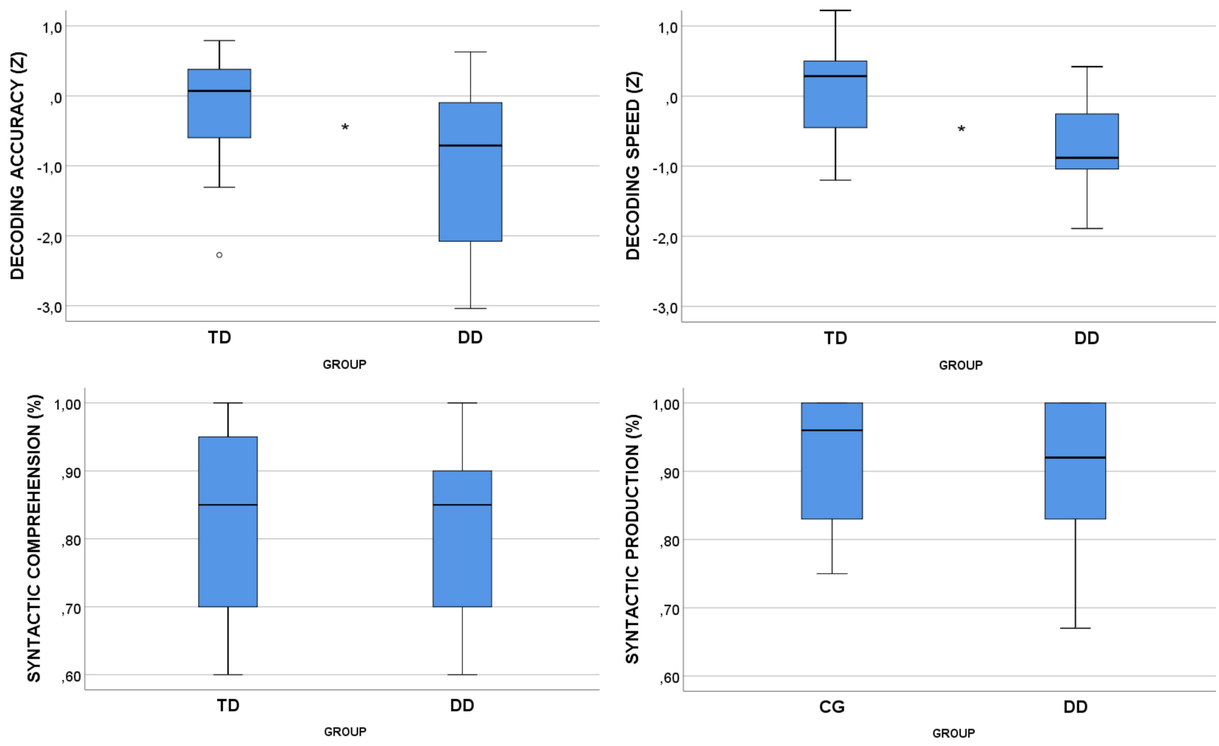

FIGURE 2 | Distribution of decoding standard scores and syntactic percent scores (*$p < 0.05$).

variable) predicted both syntactic comprehension [$F(6, 33) = 95.269$, $p = 0.000$] and production [$F(6, 31) = 14.446$, $p = 0.000$].

As for morphosyntactic comprehension (for complete statistics, see **Supplementary Tables 1–3**), there was a significant score increase in the TD group between third and fourth grade ($Est = 0.182$, $SE = 0.043$, $p = 0.000$).[2] In fourth grade only, the TD-group's score was significantly higher than that of the DD group ($Est = 0.168$, $SE = 0.069$, $p = 0.020$).

As for morphosyntactic production (for complete statistics, see **Supplementary Tables 4, 5**), there was a significant increase in target clitic pronouns between fourth and fifth grade in the DD group only ($Est = 0.154$, $SE = 0.065$, $p = 0.024$). No differences between groups emerged.

[2]We have not commented on the significant increase of morphosyntactic-comprehension scores found in the DD group between second and third grade (see **Supplementary Table 2**) because it is based on one second-grader only and is not comparable with the TD group, where there were no second-graders (see the number of participants in each grade in **Supplementary Table 1**).

TABLE 1 | Differences between TD and DD children in standard measures of decoding accuracy and speed.

					***CI* (95%)**	
Outcome	***Coeff.* (TD)**	***SE***	***t***	***p***	**Lower**	**Upper**
Decoding accuracy (*Z*)	0.820	0.280	2.930	0.006	0.254	1.385
Decoding speed (*Z*)	0.868	0.212	4.090	0.000	0.439	1.297

Predictors of Reading

Two GMMs were run with decoding accuracy and decoding speed as the respective dependent variables, the double level of group (TD and DD), syntactic-comprehension and syntactic-production scores as fixed effects, and age (in months) and the 11 decoding error types as random effects. The models significantly predicted decoding accuracy [$F(3, 38) = 8.477$, $p = 0.000$] and decoding speed [$F(3, 38) = 4.297$, $p = 0.010$].

Significant main effects of group [$F(1, 38) = 4.494$, $p = 0.041$], syntactic comprehension [$F(1, 38) = 14.137$, $p = 0.001$], and syntactic production [$F(1, 38) = 6.716$, $p = 0.013$] emerged on decoding accuracy, whereas only a main effect of group emerged on decoding speed [$F(1, 38) = 12.247$, $p = 0.001$].

Significant fixed effects are reported in **Table 2**.

TD children [$M = 0.604$, $SE = 0.745$, CI (−0.905, 2.113)] were significantly more accurate than DD-children [$M = 0.320$, $SE = 0.754$, CI (−1.206, 1.845)]; TD children [$M = 0.314$, $SE = 0.408$, CI (−0.512, 1.139)] were also faster than DD-children [$M = -0.374$, $SE = 0.432$, CI (−1.248, 0.501)], after controlling for age and decoding errors.

Analyses of random effects revealed that BUFF, MORPH, and SEM errors affect decoding accuracy, whereas MULTI errors affect decoding speed significantly. Significant random effects of decoding errors are reported in **Table 3**.

Morphosyntactic Predictors of Reading Accuracy

A GMM with decoding accuracy as the dependent variable, the 11 syntactic-comprehension items focused on grammar (including

TABLE 2 | Predictive fixed effects of group and syntactic skills on decoding standard scores.

						CI (95%)	
Dependent variable	**Predictor**	***Coeff.***	***SE***	***t***	***p***	**Lower**	**Upper**
Decoding accuracy	Group (TD)	0.284	0.134	2.120	0.041	0.013	0.556
	Syntactic comprehension	1.684	0.448	3.760	0.001	0.777	2.591
	Syntactic production	0.591	0.228	2.591	0.013	0.129	1.052
Decoding speed	Group (TD)	0.687	0.196	3.500	0.001	0.290	1.085

item number 5, which was erroneously coded as a lexicon-focused item by the test authors, as explained in section "Materials and Methods") as fixed effects, and group and age as random effects was run with a stepwise procedure. The model correctly predicted decoding accuracy standard scores [$F(4, 13) = 3.258$, $p = 0.047$]. Item number 5, namely a sentence requiring the interpretation of a third-person feminine plural direct object clitic pronoun (see section "Materials and Methods"), was a significant predictor of decoding accuracy [*Coeff.* = 1.125, $SE = 0.438$, $t = 2.566$, $p = 0.023$, *CI* (0.178, 2.072)].

Reading Errors

Figure 3 displays the percentages of reading errors made by TD and DD children.

LP-errors (7%) were present only in DD-children. Errors were numerically higher in DD-children than in TD children for every category except VIS (TD = 6%; DD = 2%) and NEGL errors (TD = 5%; DD = 4%). These two error types showed high *SD* (VIS = 0.189; NEGL = 0.129). Fisher's exact test revealed a significant association between individuals and error types (*Fisher* = 437.235, $V = 0.393$, $p = 0.000$). *Post hoc* analyses revealed that significant rates of VIS (0.70%; $Z = 4.0$) and NEGL (1.10%; $Z = 4.0$) errors were associated with two different children of the TD group (Bonferroni, $p \leq 0.050$).

A series of 11 robust GMMs with each decoding error type as a dependent variable, group as a fixed effect, and age (in months) as a random effect correctly predicted BUFF errors [$F(1, 40) = 10.104$, $p = 0.003$]. TD children made significantly fewer BUFF errors than DD children [*Coeff.* = −0.137, $SE = 0.043$, $t\ (40) = -3.179$, $p = 0.003$, *CI* (−0.224, −0.050)].

Morphological and Semantic Errors

Fisher's exact test revealed a significant association between error categories and parts of speech ($V = 0.303$, $p = 0.000$). *Post hoc* analyses revealed that MORPH decoding errors are significantly associated with determiners (87.5%, $Z = 4.6$), clitic pronouns (67.9%, $Z = 3.9$), prepositions (63.6%, $Z = 3.0$), and verbs (43%, $Z = 2.1$). Significant rates of MORPH decoding errors (*Fisher* = 165.505, $V = 0.451$, $p = 0.000$) consisted of substitutions with other visually similar parts of speech [e.g., *le mostrò* → *e mostrò*; *ebbe finito* → **ebbe fino*; *in compenso* → *il compenso* (82.4%, $Z = 4.3$)], and morphological changes [e.g., *dovevi* → *devi*; *rimaneva* → *rimane*; *il nipotino* → *il nipote* (40.4%, $Z = 2.1$)] (Bonferroni, $p \leq 0.050$).

Semantic errors, in turn, were significantly associated with nouns (34.9%, $Z = 2.9$) and adverbs (18.6%, $Z = 2.8$) (Bonferroni, $p \leq 0.050$). Noun-errors included substitutions (63%, $Z = 2.2$) with visually similar (*è a metà dell'opera* → **è a mente dell'opera*; *diede le dimissioni* → **diede le dimensioni*) and/or morphologically related words (*un giocatore* → *un gioco*; *parole* → *parlare*). Adverb errors included significant rates of additions (e.g., *ma io sono* → *ma io non-sono*; *sempre chiusa* → *sempre più chiusa*) (30.4%, $Z = 3.0$).

Three distinct GMMs with MORPH, SEM, and BUFF errors as respective dependent variables, syntactic comprehension and production percent scores as fixed effects, and group and age as random effects were run. There was no predictive effect of syntactic skills (fixed effects) on any error type, but age (random effect) had a significant negative effect on SEM errors (*Coeff.* (15.572) = −0.022, $SE = 0.008$, $p = 0.014$, *CI* [−0.039, −0.005].

Effect of Grade on Semantic Errors

We eventually analyzed the effect of grade on semantic errors in each group by running a robust GMM with SEM-error proportions as the dependent variable, and the grade variable nested in the group variable as a fixed effect. The combination of group and grade correctly predicted semantic errors [$F(5, 16) = 2.938$, $p = 0.045$]. In third grade, DD children made significantly more semantic errors than TD children (*Est* = 0.033, $SE = 0.013$, $p = 0.026$). There was a significant decrease of semantic errors between third and fourth grade in the DD group only (*Est* = 0.075, $SE = 0.009$, $p = 0.000$) (for complete statistics, see **Supplementary Tables 6–8**).

TABLE 3 | Significant random effects of percent decoding errors on standard measures of decoding accuracy and speed.

						CI (95%)	
Dependent variable	**Effects**	***Est.***	***SE***	***df***	***p***	**Lower**	**Upper**
Decoding accuracy	BUFF	−0.987	0.337	38	0.006	−1.670	−0.304
	MORPH	−10.929	1.502	38	0.000	−13.970	−7.889
	SEM	−6.259	1.047	38	0.000	−8.379	−4.138
Decoding speed	MULTI	−4.756	1.872	38	0.015	−8.546	−0.966

DISCUSSION

Differences Between Groups

We analyzed the influence of morphosyntactic skills on the reading of Italian primary-school children with DD compared to TD children. As expected, DD-children were significantly

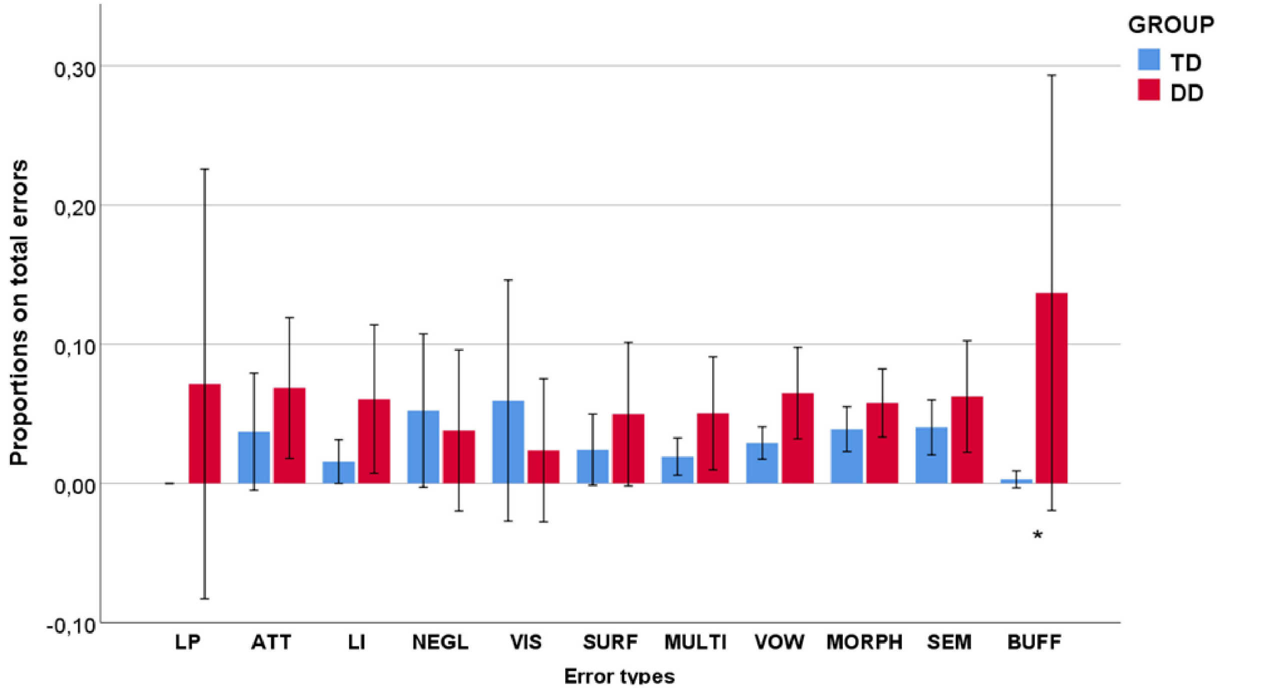

FIGURE 3 | Proportions of decoding errors by TD and DD children (*CI* = 95%) (*p < 0.05).

less accurate and slower in reading than TD children after controlling for decoding errors, irrespective of their age. The two groups obtained similar syntactic comprehension and production results, irrespective of their age (for qualitative response analyses of syntactic tests on a wider sample including the present participants, see Casani, 2020a,b), thus suggesting that DD does not directly affect oral morphosyntactic and syntactic skills in primary school.

The absence of effects of age on syntactic skills raised some doubts. Two additional models, where the age variable combined with the group variable was entered as a fixed effect, confirmed no effect of age on syntactic skills. The use of the grade variable instead of the age variable (combined with the group variable), on the contrary, showed significant effects on both syntactic comprehension and production. This means that children's instructional level rather than their age affects their syntactic skills. As for syntactic comprehension, TD children revealed a significant increase in their performance until the maximum score in fourth grade. Such an increase was not present in DD children, whose score remained around 80%, and this difference between groups (in fourth grade only) was significant. This is partially consistent with studies arguing for the presence of syntactic deficits in children and adults with DD (Bishop, 1991; Muter and Snowling, 1998; Talli et al., 2013; Cardinaletti and Volpato, 2015), which might depend on the presence of undiagnosed language disorders (Guasti, 2013), or differences in DD profiles of participants. For instance, individuals with deep DD are specifically reported for possible syntactic deficits (Friedmann and Coltheart, 2018). The peculiarity of the present study is that these problems seem to be related to a specific developmental phase. Given the small number of participants distributed across grades [in fourth grade, they were 8 (5 with DD + 3 with TD)] this hypothesis should be taken cautiously and checked on larger samples. This outcome, however, stresses the importance of a careful analysis of linguistic profiles of DD children, even in a longitudinal perspective, to reveal possible problems that might emerge in particular developmental steps. A qualitative analysis of different language structures is also advisable, to reveal possible strategies that might be associated with specific developmental conditions (Casani, 2020b). In this regard, third-person direct object clitic pronouns are deemed very sensitive to detect language difficulties (e.g., Tuller et al., 2011; Varlokosta et al., 2016; Casani, 2020c), even in DD (e.g., Guasti, 2013). So, we wonder why there were no differences in their production between our groups.[3] The answer might lie in the test used (Arosio et al., 2014), which elicits clitic pronouns under the most accessible morphosyntactic conditions, namely gender and number[4] match between the clitic pronoun and the sentential subject (Casani and Cardinaletti, 2021). Recent studies report significant difficulties under conditions of subject-object gender mismatch (Arosio and Giustolisi, 2019; Casani and Cardinaletti, 2021). The literature describing clitic pronouns as acquired at 4 or 5 years in typically developing monolingual children (Schaeffer, 2000;

[3]There was only an intra-group difference revealing a significant increase of target clitics between fourth and fifth grade in the DD group, which might suggest a different acquisitional pace between groups.

[4]Subject-object number mismatch showed not to be as problematic as gender mismatch for third-person direct object clitic production (Casani and Cardinaletti, 2021).

Arosio et al., 2014; Belletti and Guasti, 2015; Varlokosta et al., 2016) does not consider these difficulties, so their introduction in the elicitation tasks might make some differences arise. A new test with balanced match/mismatch conditions between subject and object features is in preparation, which will help disentangle this issue (see Casani and Cardinaletti, 2021).

Predictors of Reading

Decoding accuracy was predicted by morphosyntactic comprehension and production, with a stronger effect of comprehension, as well as by three error categories, i.e., MORPH, SEM, and BUFF errors, which have to do with morphology to different extents. MORPH errors had a very strong effect, followed by SEM and, lastly, BUFF errors.

Decoding speed, instead, was predicted by the presence of DD and by MULTI errors, namely errors in decoding multi-letter combinations, which are often non-shallow. The significant (fixed) effect of group only on decoding speed means that speed is a specific problem of DD-children and, differently from accuracy, is not directly mediated by morphosyntactic skills. This is in line with studies considering speed as a more reliable measure of DD than accuracy in shallow-orthography languages (see Zoccolotti et al., 2005). At the same time, the predictive effect of MULTI errors on decoding speed encourages the adoption of a multicomponent approach to reading, in which accuracy and speed interact. In this view, some reading errors might depend on impairments to specific sections of the reading model, which slow down the reading performance by hampering faster processing via the direct route. In this regard, a word decoding assessment battery based on the dual-route model (Friedmann and Gvion, 2003) has been recently adapted to the Italian language by Traficante et al. (2017). Their pilot study found six different types of DD in 52 Italian poor readers compared to 210 typical readers from the second to the fifth grade of primary school, thus showing that it is possible to discriminate several types of reading impairments due to selective segments of the dual-route model of reading even in a shallow-orthography language like Italian. Casani (2019) applied a coding scheme based on Friedmann and Coltheart (2018) to the text reading (Cornoldi and Colpo, 2011) of 21 children with DD, 4 of which with a developmental language disorder, compared to 32 typically developing children from the first to the second grade. That study detected 11 different error types. Children with language disorders presented the most compromised situation, with significant error proportions due to a combined deficit in the sublexical route and the orthographic visual analysis stage. The analyses of individual performances confirmed an impairment in both the lexical and sublexical route in 3 out of 4 children with language disorders. Interestingly, both Traficante et al. (2017) through the word lists, and Casani (2019) through text reading, found 11 and 2 cases of poor readers, respectively, who had not been detected through standard measures. These data encourage a fine-grained decoding error analysis even in shallow-orthography languages. At the same time, the problematic conditions of decoding skills in children with language disorders, who are likely to present with deficits in the production of direct object clitic pronouns, confirm the interrelation between reading and language skills, as well as the importance of outlining an accurate linguistic profile of subjects during DD diagnoses. The predictive role of morphosyntactic skills on reading accuracy emerged in the present study confirms this need.

The present study also revealed that the reading performance is mainly slowed down (in both DD and TD children) by MULTI errors, namely the decoding of multi-letter combinations. This supports a multicomponent approach to reading, in which decoding accuracy and speed interact with each other and with language processes. In light of a dual-route approach to orthographic processing (Grainger and Ziegler, 2011), reading accuracy is increased by familiarity with *fine-grained* representations coding for the presence of frequently co-occurring letter combinations, namely higher-level orthographic representations preserving the information about letter ordering. Morphemic constituents belong to this category. Higher morphosyntactic competence might increase children's familiarity with these *fine-grained* representations. This might not be sufficient, however, to speed up the processes further. To this aim, children need to increase *diagnosticity*, namely to rapidly map orthography to semantics by selecting letter combinations that are most informative with respect to word identity, according to the distributional properties of word features. This is possible by processing *coarse-grained* representations, which code for the presence of informative letter combinations in the absence of precise positional information (Grainger and Ziegler, 2011; Traficante et al., 2018), as in the case of multi-letter combinations.[5] In this interactive view of accuracy and speed, reading would be the product of orthographic, morphosyntactic, and lexico-semantic processes.

Morphosyntactic Predictors of Reading Accuracy

To explore in more depth the morphosyntactic processes involved in reading accuracy, we analyzed which syntactic-comprehension structures predict decoding accuracy. A sentence requiring the interpretation of a third-person direct object clitic pronoun [*La mucca le*$_{\_3rd person\_fem\_plur\_clit}$ *sta guardando* ("The cow them is watching" = The cow is watching them)], in particular, predicted decoding accuracy. Both comprehension and production of third-person direct object clitic pronouns, then, predict decoding accuracy. This might be due to the particular status of third-person direct object clitic pronouns, which match weak phonological salience to a high load of morphosyntactic information, as they are marked for person, gender, number, and case. Moreover, they are subject to syntactic movement and are placed preverbally with finite verbs, which is a non-canonical object position. Finally, they are mandatory in some contexts, forbidden in others, and optional in some others. These characteristics make third-person direct object clitic pronouns particularly sensitive to reveal language difficulties. In this regard, cross-linguistic literature reports clitic pronouns as a clinical marker of atypical language development in several languages including Italian (for Italian, see Bortolini et al., 2006;

[5] Frequent multi-letter compounds can enter the orthographic lexicon and be processed via the *fine-grained* route (Grainger and Ziegler, 2011) in skilled readers. In any case, their processing would not be mediated by morphosyntactic competence as the processing of morphological constituents.

Arosio et al., 2014; for French, see Jakubowicz et al., 1998; Tuller et al., 2011; for a cross-linguistic study on 16 languages, see Varlokosta et al., 2016), as well as vulnerable in bilinguals that are scarcely exposed to the target language (for Italian, see Vender et al., 2016, 2018; Casani, 2020c; Casani and Cardinaletti, 2021). At the same time, clitics' properties might expose them to be easily overlooked for their scarce phonological salience or avoided for their morphosyntactic difficulties during reading.

Reading Errors

As for reading errors made by the two groups, the DD group made significantly more BUFF (phonological-output-buffer) errors than the TD group. This does not necessarily imply the presence of phonological-output-buffer dyslexia in our DD-sample but reveals a specific difficulty of DD-children in decoding longer and morphologically complex words. Other error types did not differ significantly between groups (**Figure 3**). VIS and NEGL errors were numerically higher in TD children, but proportions were small and the difference between groups was non-significant. Given the high standard deviation, we analyzed the association between these error types and participants. VIS and NEGL errors were significantly associated with two different children of the TD group attending the fifth grade (age = 10;4) and third grade (age = 8;10), respectively. This suggests some difficulties in the visual-orthographic analysis stage of reading, at the top of the reading model (see **Figure 1**), which should be carefully evaluated to exclude or confirm the possible presence of peripheral dyslexia in these two children.

Morphological Reading Errors

Analyses of the parts of speech affected by morphological errors revealed significant difficulties in decoding function words, i.e., determiners, clitic pronouns, and prepositions, but also verbs. The presence of verbs as the only lexical part of speech that was significantly affected by MORPH errors is consistent with studies reporting differences in lexical decision times for nouns, adjectives, and verbs (for Italian, see Colombo and Burani, 2002; Traficante and Burani, 2003; for Serbo-Croatian, see Kostić and Katz, 1987; for Hebrew, see Deutsch et al., 1998; for adults with acquired language disorders, see a review in Crepaldi et al., 2011). Deutsch et al. (1998) ascribe these processing differences, besides syntactic and semantic components, to the distributional properties of constituents, namely to the fact that "when a morpheme is common to more words in the language, its impact on processes of morphological decomposition is prominent" (Deutsch et al., 1998, p. 1,252). Traficante et al. (2018) suggest that the processing of Italian verbs might be deemed as more demanding than that of nouns because Italian verbs belong to a larger morphological family, as verb roots are shared by about 50 different inflected forms and several derived words, whereas noun roots are inflected in up to four different ways and are shared by fewer derivations. The authors refer to fMRI studies highlighting stronger activation of the left inferior frontal gyrus associated with longest reaction times during a grammatical-class switching task (for adult skilled readers, see Marangolo et al., 2006; Berlingeri et al., 2008; for subjects with Parkinson's Disease, see Di Tella et al., 2018; Silveri et al., 2018) to conclude that processing difficulties might be due to the complexity of selection and inhibition processes required by the task. These reasons might explain the significant presence of MORPH errors in verbs in our sample.

It is worth noting that both phonologically weak (clitic pronouns, determiners, prepositions) and phonologically salient words (verbs) are significantly associated with decoding MORPH errors. This means that these errors might depend on both phonological and morphosyntactic weakness. The fact that oral comprehension and production of direct object clitic pronouns (which are phonologically weak and morphosyntactically complex) contribute significantly to reading accuracy (see above) supports this idea.

Semantic Errors and Interaction With Morphosyntactic Processes

SEM errors were significantly associated with nouns and adverbs. The co-occurrence of MORPH and SEM errors can be due to deep dyslexia (Friedmann and Coltheart, 2018), namely an impairment in both the lexical and sublexical route of reading. In our sample, MORPH and SEM errors correlate within the DD group (ϱ_s = 0.564, p = 0.036) but not within the TD group (p = 0.538), thus suggesting the possible presence of deep dyslexia in the DD sample. The co-presence of comparable rates of MORPH and SEM errors in the TD group as in the DD group (see **Figure 3**), however, suggests that these errors might be affected by language competence. Generalized mixed analyses showed that syntactic skills do not predict any of these two error types. As for syntactic comprehension, this might be due to the nature of the structures investigated, which mainly involve general syntactic competence, except item number 5, which involves a functionally specific structure as a third-person direct object clitic pronoun. The use of a morphosyntactic-comprehension test specifically built on the structures that revealed a significant association with MORPH and SEM errors might give more informative outcomes.

Higher age predicted fewer SEM errors. This might be because children tend to read via the lexical route as their age (and expertise) increases. The core of the lexical route of reading consists of two lexicon storages containing the orthographic and the phonological lexicon, respectively. A known orthographic string activates the correspondent entry in the input orthographic lexicon. This lexicon is organized by written word frequency. Hence, compared to words with similar orthographic (and phonological) properties, the more frequent, the more accessible words are. The activated lexical representation in the input orthographic storage can be either processed through the semantic system assigning meaning to the read word or directly sent to the phonological output lexicon assigning phonological information to the read known word. The direct connection between the two lexicon storages, possibly mediated by the semantic system, allows for both accurate and faster conversion (Coltheart et al., 2001; Castles, 2006). According to the *self-teaching hypothesis* of phonological decoding (Perry et al., 2013), the activation of preexisting words in the phonological lexicon allows for the creation of orthographic entries, so that phonology works as a self-teaching device (or "built-in teacher") refining and strengthening the network of letter-sound connections.

The teaching signal that is internally generated by phonology contributes to increasing the decoding network (Ziegler et al., 2020). This might explain the predictive effect of age on the decrease of SEM decoding errors, mainly due to the "built-in taught" interactive development of phonology and lexicon thanks to the increased reading experience. We found a significantly higher rate of SEM errors by DD children in third grade, followed by a significant decrease between third and fourth grade, which equated their SEM-error rate and that of TD children. This suggests that DD might be responsible for slower development of the lexical route of reading. Given the small number of participants across grades, however, this result should be verified in larger samples, as it might be differently affected by different types of DD.

Taken together, these data highlight the interactive role of lexical skills and reading skills. In this regard, a recent longitudinal study (Casani, in preparation, 2021) confirmed predictive effects of lexical and syntactic skills on written decoding. Moreover, it found a significant role of school-grade in vocabulary acquisition but no evident effects of school-grade on more complex syntactic abilities, i.e., clitic production under the condition of increasing morphosyntactic difficulties (for information on the test used, see Casani and Cardinaletti, 2021) between the last year of kindergarten and the second grade of primary school in mono and bilingual children. A significant increase in children's receptive vocabulary was evident only between kindergarten and second grade but not in first grade. The author suggests that the instructional input received in primary school, which also entails a certain metalinguistic component, had a role in the development of children's vocabulary.

Given these data, the mentioned self-teaching hypothesis, and the body of research proving the effects of vocabulary on school achievements (for a review, see Elleman et al., 2009), we can motivate an interactive development of vocabulary and reading, with reciprocal effects. In this light, reading is the multicomponent product of interactive processes including phonological, morphological, morphosyntactic, and semantic skills. Similar to what happens in the interactive development of phonemic awareness and reading (see Casani, in preparation, 2021), good morphosyntactic and lexical skills might improve fast recognition of function words as well as morphological strings, such as free and bound morphemes and compound constituents; at the same time, increased abilities to recognize these strings might improve their oral mastering. In this framework, an important role in reading lies in the interaction among morpho-semantic, morpho-orthographic (Feldman et al., 2009), and morpho-syntactic processes, whose successful cooperation should facilitate morpho-lexical access.

CONCLUSION

The present study demonstrates that oral morphosyntactic skills play a role in reading accuracy for both DD and TD children in a shallow-orthography language like Italian. Consistent with the dual-route approach to orthographic processing (Grainger and Ziegler, 2011), reading accuracy was mediated by morphosyntactic competence, which helped process *fine-grained* orthographic representations maximizing precise ordering of letters, such as morphemic constituents. On the other hand, reading speed was mainly affected by familiarity with *coarse-grained* orthographic representations coding for the presence of informative letter combinations without precise positional information (Grainger and Ziegler, 2011; Traficante et al., 2018), such as multi-letter combinations.

Third-person direct object clitic pronouns were confirmed as a sensitive structure not only for oral language but also for written language, as both their comprehension and production predicted decoding accuracy. We attributed this sensitivity to the fact that clitics match weak phonological salience with a heavy load of morphosyntactic information. Direct object clitic pronouns, determiners, and prepositions, as well as verbs, were significantly associated with MORPH decoding errors, thus suggesting that these errors are due to both phonological and morphosyntactic competence.

Age predicted a decrease of semantic decoding errors, meaning that children generally tend to read via the lexical route as their age increases. This is consistent with the development of the orthographic step of reading (Frith, 1985), namely of *fine-grained* chunks (Grainger and Ziegler, 2011) to be processed via the lexical route (Coltheart et al., 2001). We hypothesized that DD might slow down the acquisition of these processes as well as morphosyntactic skills. In any case, chunking the text into meaningful orthographic strings, which include free and bound morphemes, function words, morphological-compound constituents, as well as frequent multi-letter combinations, improves the reading performance by reducing the number of units to be processed (Traficante et al., 2018). Increased familiarity with these units and their distribution, deriving from higher morphosyntactic, lexical, and orthographic competence might improve their decoding.

Since children tend to read via the lexical route as their expertise increases, interventions to enhance their familiarity with functional strings should be planned timely to avoid resorting to inadequate lexical-orthographic compensation strategies, which are required by increasingly demanding texts proposed in school. This might facilitate, in particular, the reading of longer and morphologically complex words, in which DD children revealed particular difficulties, through the decomposition into phonologically and semantically meaningful chunks to be processed via the lexical route.

These data argue in favor of a multicomponent approach to reading, in which linguistic, metalinguistic, orthographic, and cognitive skills interact. In the Italian shallow orthography, morphological competence can affect decoding since the second grade (for studies on the role of morphological awareness in Italian reading, see Burani et al., 2002, 2008; Marcolini et al., 2011; Traficante et al., 2011; for a longitudinal study showing the role of (morpho)syntactic competence in the reading of Italian mono and bilingual children, see Casani, in preparation, 2021). It is then advisable to assess the linguistic profile of children during DD diagnoses to establish whether some reading errors are related to morphosyntactic difficulties. In these cases, in particular, a morphosyntactic training aiming at recognizing function elements, which might be easily mistaken for their morphological complexity and/or overlooked for their phonological weakness, might be useful to increase reading accuracy. Longitudinal intervention studies might support this statement. At the same time, cross-sectional studies are needed to explore the age of impact of morphology on reading in languages with different morphological richness and orthographic depth.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they include participants' sensitive data. Requests to access the datasets should be directed to EC, emanuele.casani@unive.it.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ca' Foscari University of Venice. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

EC collected and analyzed the data, and prepared the first draft and revisions. MV and AC participated in data discussion and edited the manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

Data collection was self-funded by EC during his Master's Degree Thesis project. Data analyses were part of EC's Ph.D. project, funded by the *MIUR* (Italian Ministry of Education, University and Research) at Ca' Foscari University of Venice. Open access publication fees were covered by research funds of MV at NTNU.

ACKNOWLEDGMENTS

We thank Letizia Moretti for her clinical advice during data collection, Maria Teresa Guasti for providing the clitic-production test, and the editor and reviewers for their valuable comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.841638/full#supplementary-material

REFERENCES

Amenta, S., Marelli, M., and Crepaldi, D. (2015). The fruitless effort of growing a fruitless tree: early morpho-orthographic and morpho-semantic effects in sentence reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 41, 1587–1596. doi: 10.1037/xlm0000104

Angelelli, P., Marinelli, C. V., and Burani, C. (2014). The effect of morphology on spelling and reading accuracy: a study on Italian children. *Front. Psychol.* 5:1373. doi: 10.3389/fpsyg.2014.01373

Arosio, F., Branchini, C., Barbieri, L., and Guasti, M. (2014). Failure to produce direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children. *Clin. Linguist. Phonet.* 28, 639–663. doi: 10.3109/02699206.2013.877081

Arosio, F., and Giustolisi, B. (2019). Agreement and Interference in Direct Object Clitic Production in Italian Monolingual Children. *Front. Commun.* 3:66. doi: 10.3389/fcomm.2018.00066

Belletti, A., and Guasti, M. T. (2015). ""The acquisition of pronominal clitics," in *The Acquisition of Italian. Morphosyntax and Its Interfaces in Different Modes of Acquisition*, eds A. Belletti and M. T. Guasti (Amsterdam, NL: John Benjamins Publishin0g Company), 81–128. doi: 10.1075/lald.57

Berlingeri, M., Crepaldi, D., Roberti, R., Scialfa, G., Luzzatti, C., and Paulesu, E. (2008). Nouns and verbs in the brain: grammatical class and task specific effects as revealed by fMRI. *Cogn. Neuropsychol.* 25, 528–558.

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., Grainger, J., Casalis, S., and Ziegler, J. C. (2015). Effects of reading proficiency in embedded stem priming in primary school children. *J. Exp. Child Psychol.* 139, 115–126. doi: 10.1016/j.jecp.2015.06.001

Bishop, D. V. M. (1991). Developmental reading disabilities: the role of phonological processing has been overemphasized. *Mind Lang.* 6, 97–101. doi: 10.1111/j.1468-0017.1991.tb00178.x

Bishop, D. V. M. (2009). *TROG-2. Test for Reception of Grammar - Version 2.* Firenze, IT: Giunti.

Bisiacchi, P. S., Cendron, M., Gugliotta, M., Tressoldi, P. E., and Vio, C. (2005). *BVN 5-11. Batteria di valutazione neuropsicologica per l'età evolutiva*. Trento, IT: Erickson.

Bortolini, U., Arfé, B., Caselli, M. C., Degasperi, L., Deevy, P., and Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *Int. J. Lang. Commun. Disord.* 41, 695–712. doi: 10.1080/13682820600570831

Burani, C. (2010). Word morphology enhances reading fluency in children with developmental dyslexia. *Lang. Lang.* 9, 177–198. doi: 10.1418/33326

Burani, C., Marcolini, S., De Zuca, M., and Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. *Cognition* 108, 243–262. doi: 10.1016/j.cognition.2007.12.010

Burani, C., Marcolini, S., and Stella, G. (2002). How early does morpholexical reading develop in readers of a shallow orthography? *Brain Lang.* 811, 568–586. doi: 10.1006/brln.2001.2548

Cantiani, C., Lorusso, M. L., Guasti, M. T., Sabisch, B., and Männel, C. (2013). Characterizing the morphosyntactic processing deficit and its relationship to phonology in developmental dyslexia. *Neuropsychologia* 51, 1595–1607. doi: 10.1016/j.neuropsychologia.2013.04.009

Cardinaletti, A., and Volpato, F. (2015). ""On the comprehension and production of passive sentences and relative clauses by Italian university students with dyslexia"," in *Structures, Strategies and Beyond*, eds E. Di Domenico, C. Hamann, and S. Matteini (Amsterdam, NL: John Benjamins Linguistik Aktuell Series), 279–302. doi: 10.1075/la.223.13car

Carlisle, J. F., and Katz, L. (2006). Effects of word and morpheme familiarity on reading of derived words. *Read. Writing Interdiscipl. J.* 19, 669–693. doi: 10.1007/s11145-005-5766-2

Carlisle, J. F., and Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Read. Res. Q.* 40, 428–449. doi: 10.1598/RRQ.40.4.3

Casalis, S., Quémart, P., and Duncan, L. G. (2015). How language affects children's use of derivational morphology in visual word and pseudoword processing:

Evidence from a crosslanguage study. *Front. Psychol.* 6:452. doi: 10.3389/fpsyg.2015.00452

Casani, E. (2019). *Distinguishing Developmental Dyslexias. Reading Errors by Italian dyslexic children with and without SLI.* Poster presented at the *DISPOC Conference Brain, Language, and Learning.* September 26 – 28, 2019. Siena, IT: Università di Siena.

Casani, E. (in preparation). *Literacy in Monolinguals and Bilinguals. Acquisition and Preliteracy Intervention in Italian*. Berlin, DE: De Gruyter Mouton.

Casani, E. (2020a). ""Rapid naming, reading, comprehension, and syntactic production by Italian dyslexic children with and without specific language impairment"," in *Strategies and Analyses of Language and Communication in Multilingual and International Contexts*, ed. D. Levey (Newcastle, UK: Cambridge Scholars Publishing), 111–122.

Casani, E. (2020b). Distinguishing DD from SLI. Language profiles of Italian dyslexic children with and without specific language impairment. *Pragmalingüística* 2, 45–69. doi: 10.25267/Pragmalinguistica.2020.iextra2.03

Casani, E. (2020c). ""Production of direct object clitic pronouns by Italian children with different acquisition modes"," in *Lingua e patologia. I Sistemi Instabili (Linguistica delle differenze 5)*, ed. F. M. Dovetto (Roma, IT: Aracne), 449–464. doi: 10.4399/978882552712418

Casani, E. (2021). *Development of Preliteracy and Literacy in Monolingual and Mono-literate Bilingual Children in an Italian School Setting. A Longitudinal Study with Metaphonological, RAN, and Graphic Intervention*. Venezia, IT: Università Ca' Foscari. [Ph.D. thesis].

Casani, E., and Cardinaletti, A. (2021). ""Effects of gender and number agreement on the production of direct object clitic pronouns in Italian monolingual and bilingual children with and without language disorder"," in *APRAR 2021. The Acquisition and Processing of Reference and Anaphora Resolution. Book of abstracts* (Spain: Universidad de Granada and Vrije Universiteit Brussel), 11–13. Available online at: https://drive.google.com/file/d/1--jlT4dx_Tn7IZy1pMzeqjcYJmQifinQ/view

Castles, A. (2006). The Dual route model and the developmental dyslexias. *Lond. Rev. Educ.* 4, 49–61. doi: 10.1080/13603110600574454

Chik, P.P.-m, Ho, C.S.-h, Yeung, P.-s, Chan, D.W.-o, Chung, K.K.-h, Luan, H., et al. (2012). Syntactic skills in sentence reading comprehension among Chinese elementary school children. *Read. Writing Interdiscipl. J.* 25, 679–699. doi: 10.1007/s11145-010-9293-4

Colombo, L., and Burani, C. (2002). The influence of age of acquisition, root frequency and context availability in processing nouns and verbs. *Brain Lang.* 81, 398–411. doi: 10.1006/brln.2001.2533

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295x.108.1.204

Cornoldi, C., and Colpo, G. (2011). *Prove di Lettura MT-2 per la Scuola Primaria*. Firenze, IT: Giunti.

Crepaldi, D., Berlingeri, M., Paulesu, E., and Luzzatti, C. (2011). A place for nouns and a place for verbs? A critical review of neurocognitive data on grammatical class effects. *Brain Lang.* 116, 33–49.

Crepaldi, D., Rastle, K., and Davis, C. J. (2010). Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory Cogn.* 38:312e321. doi: 10.3758/MC.38.3.312

Dawson, N., Rastle, K., and Ricketts, J. (2018). Morphological effects in visual word recognition: Children, adolescents, and adults. *J. Exp. Psychol. Learn., Memory Cogn.* 44, 645–654. doi: 10.1037/xlm0000485

Deacon, H. S., Whalen, R., and Kirby, J. R. (2011). Do children see the "Danger" in "Dangerous"? Grade 4, 6, and 8 children's reading of morphologically complex words. *Appl. Psychol.* 32, 467–481. doi: 10.1017/S0142716411000166

Deacon, S. H., and Kirby, J. R. (2004). Morphological awareness: just "more phonological"? The roles of morphological and phonological awareness in reading development. *Appl. Psycholing.* 25, 223–238. doi: 10.1017/S0142716404001110

Defior, S., Alegría, J., Titos, R., and Martos, F. (2008).). Using morphology when spelling in a shallow orthographic system: the case of Spanish. *Cogn. Dev.* 23, 204–215. doi: 10.1016/j.cogdev.2007.01.003

Deutsch, A., Frost, R., and Forster, K. I. (1998). Verbs and nouns are organized differently in the mental lexicon: evidence from Hebrew. *J. Exp. Psychol. Learn.* 24, 1238–1255. doi: 10.1037/0278-7393.24.5.1238

Di Tella, S., Baglio, F., Cabinio, M., Nemni, R., Traficante, D., and Silveri, M. C. (2018). Selection processing in noun and verb production in left- and right-sided Parkinson disease patients. *Front. Psychol.* 9:1241. doi: 10.3389/fpsyg.2018.01241

Elbro, C., and Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Ann. Dyslex.* 46, 209–240. doi: 10.1007/BF02648177

Elleman, A. M., Lindo, E. J., Morphy, P., and Compton, D. L. (2009). The Impact of Vocabulary Instruction on Passage-Level Comprehension of School-Age Children: A Meta-Analysis. *J. Res. Educ. Effect.* 2, 1–44. doi: 10.1080/19345740802539200

Feldman, L. B., O'Connor, P. A., del Prado, and Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonom. Bull. Rev.* 16, 684–691. doi: 10.3758/PBR.16.4.684

Fowler, A. E., and Liberman, I. Y. (1995). ""The role of phonology and orthography in morphological awareness"," in *Morphological Aspects of Language Processing*, ed. L. B. Feldman (Hillsdale, NJ: Lawrence Erlbaum), 157–188.

Friedmann, N., and Coltheart, M. (2018). ""Types of developmental dyslexia"," in *Handbook of Communication Disorders: Theoretical, Empirical, and Applied Linguistic Perspectives*, eds A. Bar-On and D. Ravid (Berlin, DE: De Gruyter Mouton), 721–751. doi: 10.1515/9781614514909-036

Friedmann, N., and Gvion, A. (2003). *TILTAN: Battery for the Diagnosis of Dyslexias in Hebrew*. Tel Aviv, IL: Tel Aviv University.

Frith, U. (1985). ""Beneath the surface of developmental dyslexia"," in *Surface Dyslexia. Neuropsychological and Cognitive Studies of Phonological Reading*, eds K. Patterson, J. Marshall, and M. Coltheart (Hillsdale, NJ: Lawrence Erlbaum), 301–330.

Frost, R. (2006). Becoming literate in Hebrew: the grain size hypothesis and Semitic orthographic systems. *Dev. Sci.* 9, 439–440. doi: 10.1111/j.1467-7687.2006.00523.x

Frost, R., Katz, L., and Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 13, 104–115. doi: 10.1037/0096-1523.13.1.104

Grainger, J., and Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2:54. doi: 10.3389/fpsyg.2011.00054

Guasti, M. T. (2013). ""Oral skills deficit in children with Developmental Dyslexia"," in *Advances in Language Acquisition*, eds S. Stavrakaki, M. Lalioti, and P. Konstantinopoulou (Newcastle, UK: Cambridge Scholars Press), 416–424.

Häikiö, T., Bertram, R., and Hyönä, J. (2011). The development of whole-word representations in compound word processing: evidence from eye fixation patterns of elementary school children. *Appl. Psychol.* 32, 533–551. doi: 10.1017/S0142716411000208

Hasenäcker, J., Beyersmann, E., and Schroeder, S. (2016). Masked morphological priming in German-speaking adults and children: Evidence from response time distributions. *Front. Psychol.* 7:929. doi: 10.3389/fpsyg.2016.00929

Jakubowicz, C., Nash, L., Rigaut, C., and Gérard, C.-L. (1998). Determiners and Clitic Pronouns in French-Speaking Children with SLI. *Lang. Acquis.* 7, 113–160. doi: 10.1207/s15327817la0702-4_3

Kostić, A., and Katz, L. (1987). Processing differences between nouns, adjectives, and verbs. *Psychol. Res.* 49, 229–236. doi: 10.1007/BF0030903

Kuperman, V., Bertram, R., and Baayen, R. H. (2008). Morphological dynamics in compound processing. *Lang. Cogn. Proc*. 23, 1089–1132. doi: 10.1080/01690960802193688

Kuperman, V., Bertram, R., and Baayen, R. H. (2010). Processing trade-offs in the reading of Dutch derived words. *J. Mem. Lang.* 62, 83–97. doi: 10.1016/j.jml.2009.10.001

Kuperman, V., Schreuder, R., Bertram, R., and Baayen, R. H. (2009). Reading of polymorphemic Dutch compounds: towards a multiple route model of lexical processing. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 876–895. doi: 10.1037/a0013484

Marangolo, P., Piras, F., Galati, G., and Burani, C. (2006). Functional anatomy of derivational morphology. *Cortex* 42, 1093–1106. doi: 10.1016/S0010-9452(08)70221-1

Marcolini, S., Traficante, D., Zoccolotti, P., and Burani, C. (2011). Word frequency modulates morpheme-based reading in poor and skilled Italian readers. *Appl. Psycholing.* 32, 513–532. doi: 10.1017/S0142716411000191

Marelli, M., and Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: modulation of headedness and semantic transparency. *J. Mem. Lang.* 66, 644–664. doi: 10.1016/j.jml.2012.01.003

Meunier, F., and Longtin, C. M. (2007). Morphological decomposition and semantic integration in word processing. *J. Memory Lang.* 56, 457–471. doi: 10.1016/j.jml.2006.11.005

Morvay, G. (2012). The relationship between syntactic knowledge and reading comprehension in EFL learners. *Stud. Second Lang. Learn. Teach.* 2, 415–438. doi: 10.14746/ssllt.2012.2.3.8

Muter, V., and Snowling, M. (1998). Concurrent and longitudinal predictors of reading: the role of metalinguistic and short-term memory skills. *Read. Res. Q.* 33, 320–337. doi: 10.1598/RRQ.33.3.4

Nagy, W., Berninger, V. W., and Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *J. Educ. Psychol.* 98, 134–147. doi: 10.1037/0022-0663.98.1.134

Nagy, W. M., and Townsend, D. (2012). Words as tools: 'learning academic voca-bulary' as language acquisition. *Read. Res. Q.* 47, 91–108. doi: 10.1002/RRQ.011

Niswander-Klement, E., and Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Mem. Cogn.* 34, 685–702. doi: 10.3758/BF03193588

Perfetti, C. A. (2003). The universal grammar of reading. *Sci. Stud. Read.* 7, 3–24. doi: 10.1207/S1532799XSSR0701_02

Perry, C., Ziegler, J. C., and Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cogn. Sci.* 37, 800–828. doi: 10.1111/cogs.12030

Quémart, P., Casalis, S., and Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *J. Exp. Child Psychol.* 109:478e496. doi: 10.1016/j.jecp.2011.02.008

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L. L., and Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in grades 4, 6, and 8. *J. Exp. Child Psychol.* 102, 96–113. doi: 10.1016/j.jecp.2008.01.004

Sana Teixeira, B., Schiefer, A. M., Alves Ferreira, de Carvalho, C., Brandão, and de Àvila, C. R. (2016). Listening and reading comprehension and syntactic awareness in reading and writing disorders. *Rev. CEFAC. 2016 Nov-Dez* 18, 1370–1377. doi: 10.1590/1982-021620161862216

Schaeffer, J. (2000). *). The Acquisition of Direct Object Scrambling and Clitic Placement: Syntax and Pragmatics*. Amstaerdam: John Benjamins Publishing Company, doi: 10.1075/lald.22

Schiff, R., Raveh, M., and Fighel, A. (2012). The development of the Hebrew mental lexicon: When morphological representations become devoid of their meaning. *Sci. Stud. Read.* 16, 383–403. doi: 10.1080/10888438.2011.571327

Silveri, M. C., Traficante, D., Lo Monaco, M. R., Iori, L., Sarchioni, F., and Burani, C. (2018). Word selection processing in Parkinson's Disease: when nouns are more difficult than verbs. *Cortex* 100, 8–20. doi: 10.1016/j.cortex.2017.05.023

Simpson, I. C., Moreno-Pérez, F. J., Rodríguez-Ortiz, I. R., Valdés-Coronel, M., and Saldaña, D. (2020). The effects of morphological and syntactic knowledge on reading comprehension in Spanish speaking children. *Read. Writing Interdiscipl. J.* 33, 329–348. doi: 10.1007/s11145-019-09964-5

Stuart, M., and Howard, D. (1995). KJ: A developmental deep dyslexia. *Cogn. Neuropsychol.* 12, 793–824. doi: 10.1080/02643299508251402

Suárez-Coalla, P., Martínez-García, C., and Cuetos, F. (2017). Morpheme-based reading and writing in Spanish children with dyslexia. *Front. Psychol.* 8:1952. doi: 10.3389/fpsyg.2017.01952

Talli, I., Sprenger-Charolles, L., and Stavrakaki, S. (2013). "Phonological and morpho-syntactic abilities in children with developmental dyslexia and specific language impairment: Evidence from Greek"," in *Advances in Language Acquisition*, eds S. Stavrakaki, P. Konstantinopoulou, and M. Lalioti (Cambridge, UK: Cambridge Scholars Press), 444–453.

Tong, X., Deacon, S., Kirby, J. R., Cain, K., and Parrila, R. (2011). Morphological awareness: a key to understanding poor reading comprehension in English. *J. Educ. Psychol.* 103, 523–534. doi: 10.1037/a0023495

Traficante, D., and Burani, C. (2003). ""Visual processing of Italian verbs and adjectives: the role of inflectional family size,"," in *Morphological Structure in Language Processing*, eds R. H. Baayen and R. Schreuder (Berlin: Mouton de Gruyter), 45–64. doi: 10.1515/9783110910186.45

Traficante, D., Friedmann, N., Rizzi, G., and Luzzatti, C. (2017). "Multiple patterns of dyslexia in a shallow orthography: Principles for a diagnostic screening of developmental dyslexias," in *Poster presented at the Academy of Aphasia 55th Annual Meeting*, November 5-7, 2017 (Baltimore, MD).

Traficante, D., Marcolini, S., Luci, A., Zoccolotti, P., and Burani, C. (2011). How do roots and suffixes influence reading of pseudowords: a study of young Italian readers with and without dyslexia. *Lang. Cogn. Process.* 26, 777–793. doi: 10.1080/01690965.2010.496553

Traficante, D., Marelli, M., and Luzzatti, C. (2018). Effects of Reading Proficiency and of Base and Whole-Word Frequency on Reading Noun- and Verb-Derived Words: An Eye-Tracking Study in Italian Primary School Children. *Front. Psychol.* 9:2335. doi: 10.3389/fpsyg.2018.02335

Tuller, L., Delage, H., Monjauze, C., Piller, A. G., and Barthez, M. A. (2011). Clitic pronoun production as a measure of atypical language development in French. *Lingua* 121, 423–441. doi: 10.1016/j.lingua.2010.10.008

Vannest, J., Bertram, R., Järvikivi, J., and Niemi, J. (2002). Counterintuitive cross-linguistic differences: More morphological computation in English than in Finnish. *J. Psycholing. Res.* 31, 83–106. doi: 10.1023/A:1014934915952

Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K. K., et al. (2016). A cross-linguistic study of the acquisition of clitic and pronoun production. *Lang. Acquis.* 23, 1–26. doi: 10.1080/10489223.2015.1028628

Vender, M., Garraffa, M., Sorace, A., and Guasti, M. T. (2016). How early L2 children perform on Italian clinical markers of SLI: A study of clitic production and nonword repetition. *Clin. Ling. Phonet.* 30, 150–169. doi: 10.3109/02699206.2015.1120346

Vender, M., Hu, S., Mantione, F., Delfitto, D., and Melloni, C. (2018). The production of clitic pronouns: A study on bilingual and monolingual dyslexic children. *Front. Psychol.* 9:2301. doi: 10.3389/fpsyg.2018.02301

Verhoeven, L., and Perfetti, C. A. (2011). Morphological processing in reading acquisition: A crosslinguistic perspective. *Appl. Psycholing.* 32, 457–466. doi: 10.1017/S0142716411000154

Vulchanova, M., and Farukh, A. (2018). Can phonological awareness predict concurrent reading outcomes in a deep orthography? *Vigo. Int. J. Appl. Ling.* 15, 145–166. doi: 10.35869/vial.v0i15.90

Wu, D. H., Martin, R. C., and Damian, M. F. (2002). A Third Route for Reading? Implications from a Case of Phonological Dyslexia. *Neurocase* 8, 274–295. doi: 10.1076/neur.8.3.274.16198

Ziegler, J. C., Perry, C., and Zorzi, M. (2020). Learning to Read and Dyslexia: From Theory to Intervention Through Personalized Computational Models. *Curr. Direct. Psychol. Sci.* 29, 293–300. doi: 10.1177/0963721420915873

Zoccolotti, P., Angelelli, P., Judica, A., and Luzzatti, C. (2005). *I Disturbi Evolutivi di Lettura e Scrittura. Manuale per la Valutazione*. Roma, IT: Carocci.

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

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

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership