Advancement of phonetics in the 21st century: Exemplar models of speech production

Matthew Goldrick[a]*, Jennifer Cole[a]

*Corresponding author: matt-goldrick@northwestern.edu, Tel: +1 847 491 8053 Fax: +1 847 491 3770
[a]Department of Linguistics, Northwestern University, Evanston, 60208 Illinois, USA

Abstract
In the first decades of the 21st century, exemplar theory has fueled an explosion of theoretical and empirical work in speech production. We review the foundations for this framework in linguistics and cognitive science, and examine how recent empirical findings challenge core principles of exemplar theory. While theoretical advances in hybrid exemplar models address some of these issues, accounting for the emergence of structure, the incorporation of structure into exemplar updating, and the non-uniformity of phonetic variation and convergence (among other phenomena), remain major challenges for current models. We discuss future directions for developing exemplar theories as comprehensive accounts of speech production.

Keywords
Usage based models, exemplar models, connectionism, hybrid models

Advancement of phonetics in the twenty-first century: Exemplar models of speech production

1. Introduction

At the start of the twenty-first century, a sea change was under way in phonological and phonetic theory. The field-wide consensus of a strict separation between the lexicon, abstract phonological knowledge, and continuous, graded, (perhaps universal) phonetic knowledge was disrupted by a usage-based perspective, *exemplar theory*. This framework was truly radical in that it denied the basic tenets of the dominant generative perspective. Exemplar theory proposed that lexical, phonological, and phonetic knowledge were deeply integrated within richly structured memory representations, along with information previously considered non-linguistic e.g. information related to the social contexts of language use. Abstract mental structures (e.g., features that define phone classes, or syllable structure) were not presupposed, but claimed to emerge over time based on processing and learning.

Over the past two decades, this perspective has been the driving force of much of the empirical and (to a lesser extent) theoretical research in phonology and phonetics. In this paper, we review this work, focusing specifically on the insights it has for the understanding of speech production. This is, by design, limited in scope. There have been many important insights from research into speech perception, the production-perception loop, and their influence on language change (for recent reviews, see Morley, 2019; Pierrehumbert, 2016). While comparatively less work has been done on speech production, as we show below this area of research provides key challenges that future usage-based work must address.

We begin our review by briefly considering the antecedent theoretical landscape and the empirical findings (from linguistics and cognate fields) that fueled discontent with this status quo. We then define the core principles of the exemplar perspective that aimed to capture these findings, discussing this in the context of research in related cognitive science disciplines (an overview of these connections is provided in Figure 1). The following sections review empirical research that challenges these principles. As we then discuss, these challenging results motivated key theoretical advances; however, major challenges remain unaddressed. We conclude by suggesting future avenues for theoretical development that may allow for more comprehensive exemplar models of speech production.

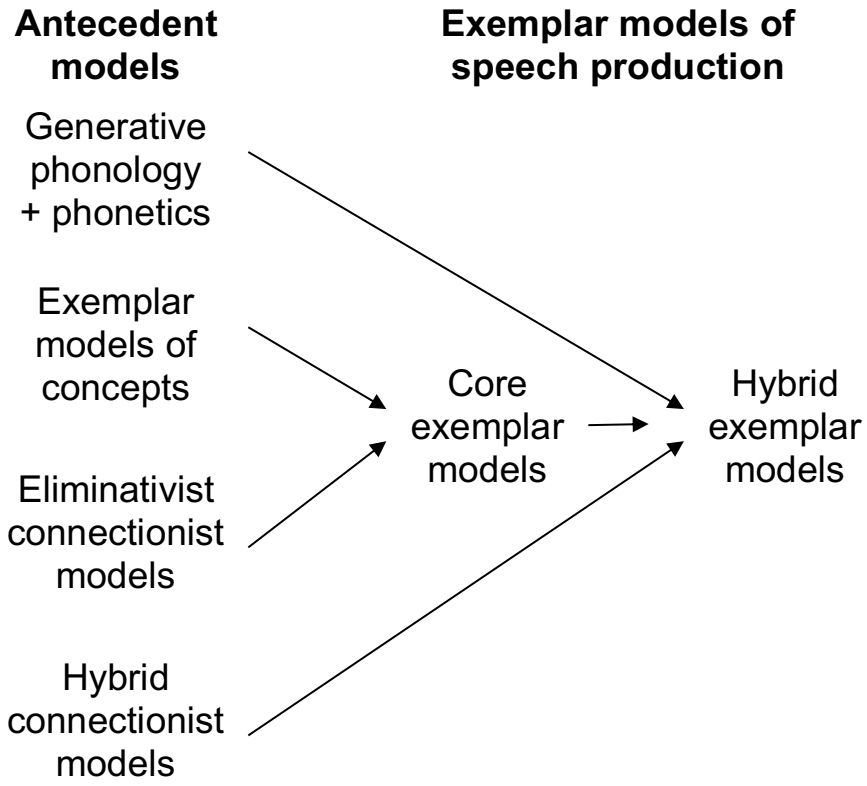


Fig. 1. Exemplar models of speech production and antecedent theoretical proposals. Arrows denote conceptual influences on exemplar proposals.

2. Setting the stage

2.1. Principles of antecedent generative models

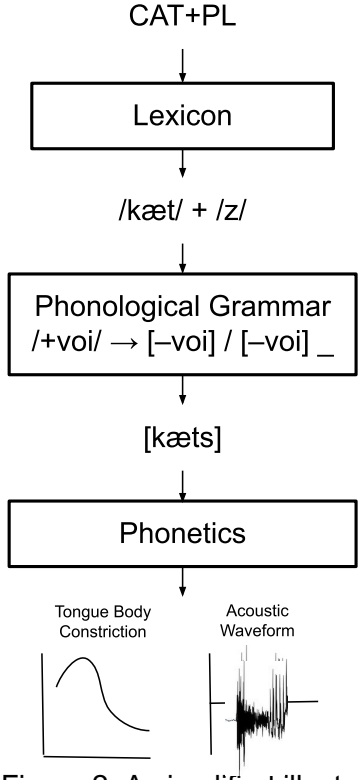


Figure 2. A simplified illustration of a generative grammar architecture. The lexicon associates morphemes with unique underlying forms. A separate module of the cognitive system, the phonological grammar (illustrated here in terms of re-write rules), maps each underlying form to a surface form. This serves as input to a third module, the phonetics; this generates articulatory movements and subsequent acoustic signals.

2.1.1. Strict modularity

As the twentieth century drew to a close, the dominant theoretical perspective (particularly in the United States) was generative linguistics. (n.b. As discussed below, alternative perspectives were also part of the theoretical landscape, and greatly influenced the development of exemplar models at the dawn of the twenty-first century.) As illustrated in Figure 2, generative phonology and phonetics research adopted a modular view of our knowledge of language, such that distinct types of knowledge were encoded in modularly separated subsystems (see Pierrehumbert, 1994, for a critical review). For example, as shown in Figure 2, lexical (word-specific), phonological, and phonetic information is manipulated by independent modules. These modules are related via a feed-forward (unidirectional) mapping, such that the only interaction consists of outputs from one module forming the inputs to another. Phonology was conceptualized as qualitative, discrete, and symbolic, strictly separated from the quantitative, continuous, and physical phonetics (see Cohn, 1993, for a review; for a contemporaneous dissenting perspective, see Browman & Goldstein, 1992). For example, as

shown in Figure 2, the phonological grammar operates on discrete feature representations (e.g,. [±voice]) while the phonetics operates over continuous articulatory (e.g., constriction of the tongue body to form the velar closure for /k/) and/or acoustic representations (e.g., a burst associated with the release of closure). The role of the lexicon was to provide the input to the phonology: a set of symbolic underlying forms that omitted predictable categorical information that could be specified by (lexical and post-lexical) phonology (for critical discussions, see Cole & Hualde, 2011; Krämer, 2012). Finally, in concert with the general marginalization of variable phenomena (in favor of the ideal speaker-hearer; see Pierrehumbert, 1994, for critical discussion), information about the social identity of the speaker or hearer was viewed as outside of the domain of phonological and phonetic theory (at an extreme, viewed as not in the domain of any scientific theory: Carr, 1999).

2.1.2. The limited role of experience

The highly structured modular architecture of phonological and phonetic knowledge was assumed to reflect the strong biases of the language learner. In this framework, learning involves establishing a unique "underlying" representation for each word (or morpheme) in the language, which is underspecified relative to certain details of phonetic implementation. Learning also involves induction of a phonological grammar that expresses systematic patterns or dependencies among the elements of sounds and their constituent structures in lexical representations, e.g., a distributional pattern in which a certain feature or phone occurs only in certain phonologically specified contexts. Once a lexical representation is established, language use involves mapping between the abstract lexical representation of a word and its variable phonetic instantiations. The type of lexical representations and phonological grammars that are learned are taken to be constrained by Universal Grammar. As an innate endowment of the learner, Universal Grammar not only specifies the overall architecture of lexical representations, but provides strong constraints on the nature of phonological and phonetic knowledge (e.g., the representational primitives, the functional specification of the rules or constraints that make up the grammar; Dresher, 1999; Tesar & Smolensky, 1998). With such strong biases, learners will rapidly converge to a uniform knowledge state (which in the ideal case, consists of the same lexical specifications and phonological grammars across learners), even in the face of substantial variation in experience (Geman et al., 1992). Specifically, learning was assumed to stop at this point because, in this framework, once the target state has been achieved (the parameters set, the constraints ranked) learning is complete. Lexical representations themselves are not affected by usage, except when the mapping to phonetic form fails. For example, in speech perception, when expected phonetic cues to phone identity are not present for a word, a listener may create a new lexical representation.

2.2. Empirical motivations for considering an alternative perspective

While the generative perspective was dominant, a variety of empirical phenomena challenged its core assumptions. Here we briefly review salient data from language production, reported in the final years of the twentieth century (for reviews of influential results from speech perception in this period, see Goldinger, 1998; Tenpenny, 1995). We also point to more recent

work that has directly followed up on these seminal findings from the late twentieth century. Table 1 provides an overview of the issues reviewed in this section, along with a summary of the core exemplar account of these findings (discussed in §3.2).

Table 1. Overview of challenges to generative models of phonetics and phonology with accounts offered by the core principles of exemplar models.

Challenges to generative models	Exemplar account
The plasticity of speech production	Storage of novel exemplars results in changes to subsequent behavior
Lexically conditioned phonetic variation	Storage of exemplars integrating lexical and phonetic information
Lexically conditioned phonetic plasticity	Storage of novel exemplars integrating lexical and phonetic information
Lexically conditioned sociolinguistic variation	Storage of exemplars integrating social, lexical, and phonetic information

2.2.1. The plasticity of speech production

Evidence from a variety of sources suggested that experience after the point at which a language had been acquired plays a central role in the representation and processing of phonological and phonetic structure. One measure of experience is word frequency, an estimation of the unigram probability of a word across a diverse sample of speakers and texts. By the end of the twentieth century, a good deal of research in psycholinguistics (primarily, but not exclusively, conducted in Germanic and Romance languages) had suggested the phonological forms of more vs. less frequent words were retrieved more quickly (Jescheniak & Levelt, 1994) and accurately (Dell, 1990; see Kittredge et al. 2008, for a review of subsequent work). Frequency effects were argued to hold throughout the production system, with high frequency sub-lexical units (e.g., syllables) retrieved more quickly than lower frequency units (perhaps reflecting storage of units in a 'syllabary': Cholin et al., 2006; Levelt & Wheeldon, 1994; see Laganaro, 2019, for a review of subsequent work).

At a shorter time scale, Dell et al. (2000) showed that the speech production system could implicitly adapt within a single experimental session to changes in the distribution of segments, i.e., to phonotactic constraints. English-speaking participants produced tongue twisters in which segments were (in contrast to their typical distribution) confined to a single syllable position (e.g., /f/ was confined to coda for some participants, onset for others). The distribution of speakers' speech errors shifted such that the newly-restricted segments behaved like segments that are always restricted in their distribution. For example, in English /h/ is not found in (word-final) codas and /ŋ/ is not found in (word-initial) onsets. Speech errors rarely violate these constraints (a mis-placed /h/ rarely appears in coda; see Alderete & Tupper, 2018, for a recent review of data from English; Goldrick, 2011, for a review of cross-linguistic data).

Dell et al. found that these newly restricted consonants behaved in the same way; when /f/ was restricted to onset, speakers rarely mis-placed it in coda. Critically, equivalent adaptation was found in counterbalanced conditions (e.g., /f/ restricted to coda), demonstrating that error distributions are not simply avoiding articulatorily complex structures (see Goldrick, 2017, for discussion). This suggests the spoken production system can adapt to very recent experiences (see Dell et al., 2021, for a review of subsequent work in English and Dutch, and Smalle & Szmalec, 2022, for recent work in French).

Other work suggested recent experience could modulate properties of speech at the level of phonetic specification as well. Sancier and Fowler (1997) recorded the speech of a Brazilian-Portuguese - English bilingual who resided in the United States (where English is the dominant language) but regularly traveled to Brazil (where Brazilian Portuguese is the dominant language). They found that native Brazilian Portuguese listeners (in Brazil) could reliably distinguish Portuguese sentences recorded after an extended stay in Brazil vs. after a stay in the United States. Instrumental analysis of the voice onset time (VOT) of phonologically [–voice] stops in each language (unaspirated voiceless stops in Brazilian Portuguese and aspirated voiceless stops in English) showed that VOTs shifted towards English vs. Portuguese norms following immersion in each language. This bi-directional shift provided clear evidence that this speaker's realization of the stop contrast was sensitive to her recent experiences (see Chang, 2019, for a recent review of longitudinal shifts in bilingual production).

2.2.2. Lexically conditioned phonetic variation

The strictly modular generative architecture has no direct mechanism by which the lexicon can influence the phonetic implementation of words for any properties not directly determined by their phonological representations. Words with similar phonological structure should have similar phonetics (assuming factors such as speech rate or speaking style are held constant). Since (at least) the 1980s, data inconsistent with this prediction were documented in a number of languages with phonological processes that neutralize distinctions present in underlying (lexical) representations (e.g., the distinction between /d/ and /t/ in coda position), but nonetheless exhibit a phonetic contrast between the 'neutralized' forms (e.g., although both sounds are pronounced [t], there is a phonetic contrast between forms ending in underlying /d/ vs. underlying /t/). This phenomenon of incomplete neutralization has been found in many languages (Cantonese, Catalan, Dutch, American English, Japanese, Polish, East Andalusian Spanish, Russian; see Braver, 2019, Strycharczuk, 2019, for recent reviews), but is particularly well studied in the context of German word-final devoicing. This process in German affects obstruents in word-final position, eliminating an underlying (lexically specified) contrast between word-final voiced and voiceless obstruents, in favor of the voiceless form. For example, the underlying forms for German *Rad* 'wheel' /ʁa:d/ and *Rat* 'council' /ʁa:t/ differ in the voicing of the final obstruent, but this difference manifests only in suffixed forms, where the root-final obstruents evade the devoicing rule by virtue of not being word-final (*Räder* 'wheels' [ʁɛ:dɐ] *Räte* 'councils' [ʁɛ:tə]). In the absence of a suffix, the devoicing rule applies, eliminating the phonological [voice] distinction (c.f., *Rad* 'wheel' [ʁa:t] and *Rat* 'council' [ʁa:t]). Although the word-final obstruents do not display the typical acoustic distinctions marking the obstruent voicing contrast, small but reliable phonetic differences persist, with vowels preceding underlyingly voiceless stops produced as slightly shorter than those preceding underlyingly

voiced stops (Port & O'Dell, 1985; see Nicenboim et al., 2018, for a meta-analysis confirming the reliability of this effect). This results in lexically conditioned phonetic variation: the phonetic realization of a word-final, phonologically voiceless obstruent varies depending on its source in an underlyingly voiced or voiceless segment in lexical representation. In other words, lexical representations appear to be "reaching" into phonetics, which is inconsistent with the assumption of strict modularity.

In the 1990s, Wright (1997; see also Wright, 2004) examined morphologically unrelated English words that varied in their relationship to other words in the English lexicon. He compared the acoustic properties of vowels in two kinds of words. Words with high neighborhood density differ from many other words only by deletion, addition, or substitution of a single phone (and, for this analysis, these non-target neighboring words have higher frequency than the target). Low density words, in contrast, are related to few, if any words differing by just a single phoneme (and, for this analysis, those few neighbors are lower in frequency than the target). Wright found that, controlling for the phonological/phonetic context of the vowel (c.f. Gahl, 2015), vowels in high density words exhibit a more expanded vowel space than those in low density words. This presents a challenge for a strictly modular architecture which offers no means by which (non-morphological) lexical relationships such as neighborhood density can influence the phonetic realization of words.

2.2.3. Lexically conditioned phonetic plasticity

Bringing together these two strands of research, Goldinger (1998) provided evidence that recent auditory exposure to previously recorded words could shift their phonetic properties – and that this shift was modulated by longer-term experience with the words.

In Goldinger's study, English-speaking participants were first asked to provide a *baseline* pronunciation for a set of words . They were then exposed to talker-specific pronunciations of each word, presented auditorily during a block of listening trials. The words presented during this exposure phase were produced by 10 talkers, with the pairing of individual words and talkers counterbalanced across participants. The listening blocks alternated with shadowing blocks, during which participants heard a word previously encountered in the listening block, and were instructed to repeat it quickly and clearly, as in the baseline session. In different conditions, shadowing was immediate, or following a delay of 3-4s. To assess whether exposure to a specific talker's pronunciation of a word induced changes to the participant's pronunciation from baseline to shadowing, a separate set of listeners were asked to select whether the baseline or shadowed pronunciation was a better imitation of the exposure pronunciation.

The results showed that participants' pronunciations shifted as a result of exposure to a particular talker's pronunciation, with the result that listeners were more likely to select the shadowed pronunciation as a better imitation of the words from the exposure set. The shift in pronunciation is an example of phonetic *convergence* (also termed *accommodation*) to heard speech. Critically, convergence was more likely to occur for low frequency vs. high frequency words. Subsequent experiments with English speakers: confirmed these findings for novel words (with frequency manipulated during the training phase; Goldinger, 1998); showed similar results when post-exposure pronunciations were elicited through reading instead of shadowing

(Goldinger & Azuma, 2004); and provided instrumental evidence that shadowing shifted articulation (i.e., VOTs; Shockley et al., 2004).

These results are difficult to reconcile with an architecture that assumes little or no role for experience within a strictly modular framework. Goldinger's work shows that the phonetic properties of words can shift based on recent auditory experience, and the degree or probability of phonetic shifting is modulated by lexical properties of a word, in this case, word frequency.

2.2.4. Lexically conditioned sociolinguistic variation

The strictly modular generative architecture does not provide a means by which social factors or lexical properties would influence the phonological specification or phonetic realization of words. Hay et al. (1999) examined the monophthongization of /aɪ/, a well-known feature of African American vs. Mainstream U.S. (i.e., White) English. Examining monologues by a popular African American talk show host (Oprah Winfrey), Hay et al. found that monophthongization was more common when Winfrey spoke about African American vs. non-African American individuals. Critically, monophthongization was more common for high vs. low frequency words – an interaction between social and lexical information. This is clearly challenging (along multiple dimensions) for a strictly modular architecture.

3. Turn of the century: The core exemplar perspective

3.1. Inspirations for a new perspective from cognitive science

In spite of the dominance of generative linguistics, alternative frameworks for the study of phonology and phonetics thrived during the twentieth century. Many of these non-exemplar frameworks were centered around the importance of usage or experience and therefore made important contributions to the development of exemplar theories (e.g., functionalist approaches, Natural Phonology, Cognitive Phonology, Construction Grammar). As these contributions have been extensively reviewed (e.g., Bybee, 1999, 2006), here we focus on how work in other cognitive science disciplines outside of linguistics seeded the development of exemplar models.

3.1.1. Exemplar models of categorization

In developing an alternative to the dominant generative perspective, linguists drew heavily on psychological research on categorization – recognizing its relevance to questions about the status of phonological categories and their phonetic implementation. During the latter-half of the twentieth century, this was a central topic of research in the cognitive psychological tradition. How is it that humans have the ability to place stimuli with diverse sensory properties into coherent groups (see Kruschke, 2008, for an overview)? A common experimental approach to this question was to train participants on novel concepts using a set of examples ("exemplars" of the category) and then gauge what information the participants had extracted by asking participants to categorize new exemplars. Some theories modeled the resulting categorization data using mechanisms similar to generative linguistic theories, developing abstract representations that underlie variable exemplars (e.g., prototypes) and/or developing sets of rules (e.g., a summary of category content). Other theories eschewed such abstractions, modeling categorization behavior as a process that compared novel stimuli to the set of

exemplars stored in memory. Stimuli were categorized based solely on their similarity to previously encountered exemplars within each category along specific, quantified features or perceptual dimensions (e.g., Medin & Schaffer, 1978). The 1980s and 1990s saw development of an interrelated set of proposals elaborating on this basic idea, some of which served as direct inspiration for the application of exemplar models in language (e.g., Hintzmann, 1986; Kruschke, 1992; Nosofsky, 1986; see Kruschke, 2008, for a review).

3.1.2. Eliminativist connectionism

In parallel with exemplar models (and in interaction with them; Kruschke, 1992), connectionist models became a highly prominent perspective in the cognitive sciences. The foundational *Parallel Distributed Processing* volumes (Rumelhart & McClelland, 1986; McClelland & Rumelhart, 1986) introduced a framework for modeling cognition as the spread of activation between simple processing units. In this framework, mental representations are realized as distributed patterns of activation over these units. A core concept of connectionist research, shared with exemplar proposals, is *emergence,* the claim that knowledge and behaviors previously attributed to pre-specified symbolic representations and computations instead emerge on the basis of experience. The structure of representation and computation emerge from interactions over non-symbolic elements and processes (McClelland, 2010; see Bybee & McClelland, 2005, for additional discussion of the links between the connectionist and exemplar perspective[1]). The *eliminativist* connectionist perspective claims these emergent representations can support 'symbol-like' behavior while remaining crucially non-symbolic (thus 'eliminating' symbolic representations from the account of cognition). For example, where symbolic representations are inherently discrete (e.g., [±voice] in Figure 2 above), connectionist representations are fundamentally continuous (e.g., defined by the graded activation of simple processing units; Plaut et al., 1996). Effects that might be attributed to a discrete symbolic representation like a distinctive feature would instead arise from the nonlinear interaction of processing units, none of which bear an explicit relationship to any particular distinctive feature.

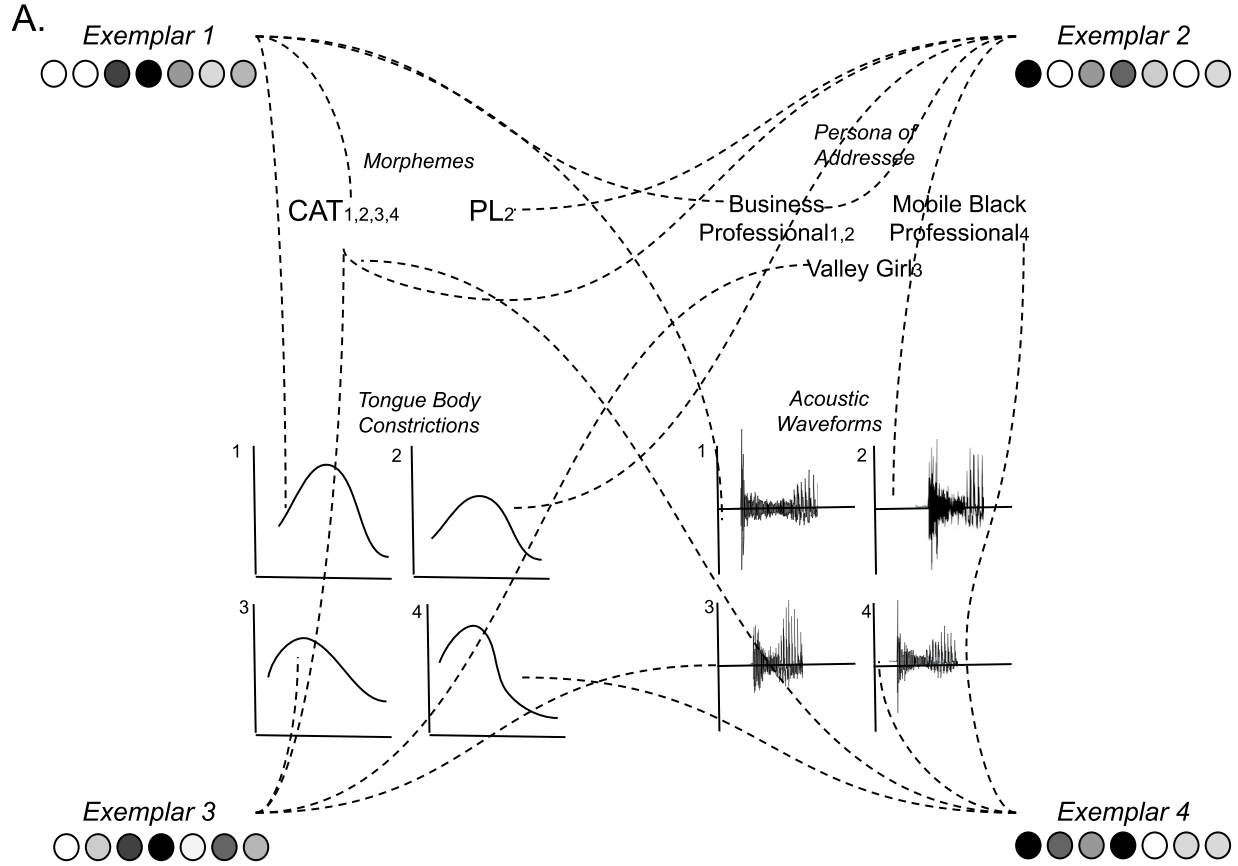
Note that, unlike the exemplar accounts discussed in the preceding section, connectionist models do not typically involve explicit storage of or reference to exemplars; the internal structure of the network reflects the combined influence of all patterns encountered during training (see e.g., Plaut et al., 1996, for discussion; but see Kruschke, 1992, for discussion of an integrated exemplar-connectionist framework).

The eliminativist approach can be compared to psycholinguistic theories of speech production at the time, which typically assumed that symbolic representations (e.g., syllable structure) play a key role in the production of speech sounds (e.g., Shattuck-Hufnagel, 1979; Dell, 1986; Levelt et al., 1999). Dell et al. (1993), building on the work of Jordan (1986), examined an alternative, eliminativist account of the data used to motivate such models (for related work, see Anderson et al., 1998; Dell & Kim, 2005; Plaut & Kello, 1999). Dell et al. trained a network to map a static vector representing a word (i.e., a unique identifier for each lexical item, unrelated to its phonological or phonetic structure) to a sequence of phonological feature representations (one feature representation per segment in the target word). Training

[1] Naive discriminative learning models of production (Tomaschek et al., 2021; for a general overview, see Baayen & Ramscar, 2019) represent an eliminativist framework within linguistics with links to these connectionist accounts.

involved adjusting the weights that linked the lexical representation to a *hidden* representation (with structure that was not pre-specified) and then from this hidden representation to the phonological feature outputs. Hidden units also had trained self-connections, allowing previous hidden representational states to influence future ones (providing a 'memory' for the network). The network was able to successfully learn this task. Critically, when it made errors, the errors respected phonotactic constraints (see above for discussion of related empirical data) – even though such constraints are typically assumed to require symbolic representations (e.g., constraints often make reference to syllable structure). Successes such as these supported the claim that symbol-like behavior could instead reflect non-symbolic representations and computations that emerge as the result of experience with language processing.

3.2 The core exemplar perspective



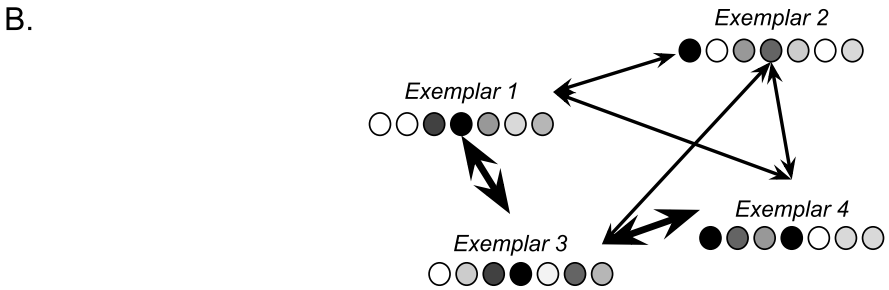


Figure 3. A simplified illustration of the core exemplar architecture showing four exemplars: three instances of the word *cat,* and one of the word *cats*. A. Memory representations (here, a seven-dimensional vector, analogous to representations in eliminativist connectionist or deep learning systems) are particular speech experiences or *exemplars*. These are associated (as shown by dotted lines and subscripts) with various dimensions of information: morphological, phonetic, and social information (here, personae; see D'Onofrio, 2021). B. Exemplars interact during processing based on their similarity (shown by the arrows of varying thicknesses) along the combined information dimensions and the emergent structures in exemplars.

In this interdisciplinary context, at the turn of the twenty-first century exemplar models of production arose to prominence within linguistics (e.g., Bybee, 2001; Goldinger, 1998;

Pierrehumbert, 2001, 2002). Building on exemplar models of concepts, the core principles of these models are:

> Principle I. Storage of the details of every experience. All aspects of each experience with speech (exemplars) are stored in long-term memory (i.e., the lexicon). This includes detailed representations of the acoustic and articulatory properties of a spoken word, its lexical specification (i.e., the morpheme(s) associated with the production), as well as information about the linguistic context of the word (e.g., its phrasal context), and "non-linguistic" social, indexical, and environmental information.
>
> Principle II. Similarity-based processing. Stored exemplars contribute to processing based on their similarity to the current target form. In the context of speech production, *target* could be the acoustic form of a heard word in a repetition/shadowing task (Goldinger, 1998) or a token from the cloud of previously experienced exemplars associated with the target word, selected at random or based on its status in relation to other exemplars (e.g., as a centroid in phonetic space; Kirchner et al., 2010; Pierrehumbert, 2001a).
>
> Principle III. Emergent phonological structure. Effects attributed in earlier work to symbolic phonological structure emerge from non-symbolic representations and processes during learning and processing (for discussion, see Goldinger, 2007; Johnson, 2007; Kirchner et al., 2010; Pierrehumbert, 2003).

These principles are reflected in Figure 3: Each exemplar representation is associated with multiple sources of information (Principle I). This includes phonetic, social, morphological, and emergent representations (Principle III). As shown by weighted arrows and the relative placement of exemplars, these exemplar representations interact based on similarity (Principle II). For example, exemplars sharing lexical or social information are closer together and have stronger links.

As summarized in Table 1, this perspective provides an account of the data that challenged the core assumptions of the generative perspective:

- *The plasticity of production*: Follows from Principle I; as new experiences are stored, they can then influence subsequent behavior in the same manner as older exemplars. For example, when living in the US, a Portuguese-English bilingual speaker will have gained exemplars reflecting the VOT norms of English. However, after immersion in Portuguese during a visit to Brazil, the speaker will now gain many new exemplars reflecting the VOT norms of Portuguese. These shifting distributions of exemplars can account for shifts in the VOT of such bilingual speakers.
- *Lexically conditioned phonetic variation*: As exemplars encode lexical and phonetic information simultaneously (Principle I), they allow for storage of word-specific phonetics. The pervasive influence of similarity on processing (Principle II) provides a mechanism by which lexical neighbors can influence processing of

target words, including (due to Principle I) their phonetic properties. For example, German speakers can maintain subtle phonetic differences between 'neutralized' words *Rad* 'wheel' /ʁa:d/ and *Rat* 'council' /ʁa:t/ based on the phonetic distinctions between suffixed forms *Räder* 'wheels' [ʁɛ:dɐ] *Räte* 'councils' [ʁɛ:tə] (Ernestus & Baayen, 2006).

- *Lexically conditioned phonetic plasticity*: Recent experiences are stored (Principle I), allowing them to exert an influence on subsequent productions of a target form. Because high frequency words have many more stored exemplars than low frequency words, these new experiences contribute less to the phonetic properties of high vs. low frequency words – reducing the degree to which the addition of new exemplars of high vs. low frequency words influences subsequent productions of the same words. In the context of a shadowing experiment, this accounts for the different effects of high vs. low frequency words on shifting productions during the shadowing experiment.
- *Lexically conditioned sociolinguistic variation*: As exemplars encode lexical, phonetic, and social information simultaneously (Principle I), they provide a means for expressing the interaction of each of these dimensions (see Clopper & Turnbull, 2018, for discussion). For instance, explicit associations between exemplars with a monophthongal vowel variant and a referent identified as African American could reinforce the selection of monophthongal exemplars in subsequent discourse contexts that activate that identity.

3.2.1. The nature of emergent representations

The core exemplar architecture claims the symbolic representations that are an essential component of generative models should not be presupposed. In contrast, exemplar representations are assumed to arise from non-symbolic representations and processes. However, in early exemplar work, no strong commitment was made to the structure and content of these emergent representations. The focus of early studies was explaining plasticity and lexically-conditioned effects (the very phenomena that had escaped generative models); Principles I and II therefore received far more attention than Principle III.

There is a wide range of theoretical possibilities that are consistent with this claim. The representations that emerge may share many properties of symbolic representations (see e.g., Plaut et al., 1996; Smolensky et al., 2022; for examples in the context of connectionist models). Alternatively, we may adopt an eliminativist exemplar account, which completely erases symbolic representational properties from the model of production. In the next section, following the spirit of early exemplar papers that explored the potential of a "'pure' model…[that] takes episodic storage to a logical extreme (Goldinger, 1998:254)", we focus on the eliminativist end of this theoretical spectrum. As we evaluate this extreme possibility, it is important to keep in mind (following those same researchers) that "if it fails, less extreme models are available (*ibid*.)." We consider these possibilities in more detail in §5 and §6.

4. Challenges for an eliminativist exemplar perspective

This non-modular, plastic, emergentist perspective vastly expanded the range of theoretical possibilities for the investigation of phonetics and phonology, providing intellectual fuel for a variety of research projects in the twenty-first century. A wide range of results provided additional empirical support for the plasticity of speech production, e.g., in laboratory word-shadowing tasks (e.g., Pardo 2013; see review in Pardo et al. 2017; and later works e.g., McLeod 2021), in the context of social interactions (e.g., Babel, 2010; Pardo, 2006; Pardo et al. 2018;) and morphologically- or lexically-conditioned phonetic variation (e.g., Arnon & Priva, 2014; Gahl, 2008; Munson, 2007; Scarborough & Zellou, 2013; Seyfarth, 2014; Tang & Bennett, 2018; Tomaschek et al., 2021). A highly active area of research was exemplar-based sociophonetics, exploiting the non-modularity of exemplar representations to model the links between social information, sound structure, and the lexicon (Babel, 2012; Drager & Kirtley, 2016; Foulkes & Docherty, 2006). However, other results have proved to be more challenging for an eliminativist exemplar perspective. Table 2 provides an overview of these challenges (detailed in this section) as well as potential solutions (discussed in §5 and §6).

Table 2. Overview of challenges to eliminativist exemplar models, implemented solutions, and (in parentheses) areas of potential theoretical development of new solutions.

Challenges for eliminativist exemplar models of production	(Potential) Solutions
Generalization over sub-lexical structures	Hybrid models with symbolic sub-lexical structure (emerging on the basis of experience)
Variable phonetic effects of multiple types of lexical relationships	(Integration of exemplars with other speech production mechanisms)
Non-uniformity of phonetic convergence	(Nature of processes underlying exemplar storage)

4.1 Generalization over sub-lexical structures

4.1.1. Lenition reflects experience with sub-lexical units

A key property of symbolic representations is *compositionality.* Complex structures are constructed from smaller elements or building blocks (Pierrehumbert, 2006), which can freely recombine (compose) with other elements (see Smolensky et al., 2022, for discussion). To take an oversimplified example, suppose we claim that the 'underlying' representation of the word *dog* is a symbolic structure composed of three segments /d/, /ɑ/, /g/. Each of the segments that compose *dog* have an *independent* representational status from the word as a whole. They can combine with other segments to form different words. The /d/ in *dog* is the same segment as the /d/ in *dip*, the /g/ the same in *dog* and *pig*.

In rejecting symbolic representations, an eliminativist exemplar model has difficulty modeling effects that reflect properties of segments that are independent of their lexical context. For example, Cohen Priva (2015, 2017) examines the contribution of segment informativity to lenition. Informativity refers to the average predictability of a unit across all the contexts it appears in; in the case of segments, this concerns how probable a particular segment is across many lexical items. Cohen Priva (2015) finds that within American English spontaneous speech, segments with high informativity are longer and less likely to be deleted. Cohen Priva (2017) provides evidence that this extends cross-linguistically, with languages tending to lenite less informative segments. These findings are difficult to reconcile with an eliminativist view that does not represent segments independent of their lexical context.

4.1.2. Type frequency predicts generalization better than token frequency

In an eliminativist exemplar theory, the strength of learning – and therefore the likelihood of generalization to novel word forms – is predicted by the number of experiences a speaker has with a word (i.e., its token frequency). Following the principle of similarity-based processing (II), all exemplars that are similar to a novel form will influence processing. The greater number of similar exemplars– the higher the token frequency of a form – the stronger the response (see Denby et al., 2018, for detailed analyses of predictions based on the MINERVA 2 model examined in Goldinger, 1998). However, as reviewed by Edwards et al. (2015), a number of studies in multiple languages suggest that the diversity of contexts in which the element appears – its type frequency – predicts successful learning and generalization better than token frequency (see Pierrehumbert, 2001b, for discussion of functional motivations of this pattern). For example, in a lab-based learning study Richtsmeier et al. (2011a) familiarized English-speaking four-year old children to the infrequent phonotactic sequence /fp/ in nonwords, varying how many tokens of the sound sequence were presented vs. the diversity of nonword contexts in which the sequence appeared (and the number of model talkers producing the sequence). Accuracy was highest when type, not token, frequency was maximized. Richtsmeier et al. (2011b) shows similar results for adult learners. The dominance of type vs. token frequency in learning is not readily explained by eliminativist principles.

4.1.3. Systematic generalization

In an eliminativist exemplar theory, phenomena that generative theory ascribed to symbolic category representations (e.g., the phone [t^h]; the feature [–voice]) are assumed to reflect non-symbolic representations that emerge during processing (Principle III) on the basis of the activation of similar exemplars (Principle II). This makes strong predictions for learning. After training on some set of exemplars, generalization to novel contexts should be highly sensitive to the similarity between the new context and the exemplars. However, several studies have reported results that show systematic generalization, without sensitivity to similarity. Using a paradigm similar to Goldinger (1998), Nielsen (2011, Experiment 1; see also Lindsay, Clayards, Gennari, & Gaskell, 2022) examined changes to VOT for [p^h]- and [k^h]-initial words after exposing English-speaking participants to [p^h] tokens with lengthened aspiration. Post-test productions showed no sensitivity to similarity; there was equivalent lengthening for trained [p^h] and untrained [k^h] (see Maye et al., 2008, for related results in generalization of vowel shifts). German et al. (2013) asked participants to imitate a novel dialect of English. Over several

training sessions, participants (US college students) were exposed to Glaswegian English example sentences in which intervocalic /t/ (e.g., *sweaty*) was systematically realized as [t^h]. In the participants' baseline productions (and in American English more generally), this phoneme in the same words is typically realized as a flap [ɾ]. Participants were tested on the production of trained as well as novel words. They rapidly and systematically acquired this novel allophone, producing it in over 95% of trials, in novel as well as trained words.

Further evidence of systematic generalization comes from phonetic convergence over lexically unrestricted material. Eliminativist exemplar theory predicts convergence in speech shadowing or delayed repetition tasks, where a speaker produces the same lexical item as previously heard, with a lesser degree of convergence for phonetically similar but not identical lexical items. Yet more broadly generalized patterns of phonetic convergence are attested, suggesting that convergence is not strictly governed by similarity of word forms. For example, Sonderegger et al. (2017) examined convergence in several acoustic correlates of vowel and consonant contrasts (VOT, coronal stop deletion, vowel formants) in samples of conversational speech from speakers participating in a reality TV show, who interacted exclusively with one another over a period of three months. For two of the participants in this study who had a particularly close social bond, these acoustic measures shifted over three months reflecting a converging pattern. Notably, these shifts were evident in acoustic measures drawn from different words, as they occurred over multiple occurrences of spontaneous speech produced by each speaker.

Related findings are reported by Kim et al. (2011). Using a perceptual similarity criterion for the holistic assessment of convergence, Kim et al. reported generalized convergence (i.e., in comparisons of different words and phrases) in goal-oriented spontaneous speech from pairs of interacting speakers. Listeners judged entire intonational phrases of up to 1.5 s in duration in an XAB paradigm, where A and B were speech excerpts taken from the first and third portions, respectively (and counterbalanced with the reverse order), of one speaker's utterances during an interactive task, while X was an excerpt from their partner's speech during the same task. Notably, the excerpts presented on each trial were not matched for lexical or syntactic content. The results showed that the excerpt from the later portion of the interaction was perceived as being more similar to the interlocutor's speech than the excerpt from the earlier portion, indicating phonetic convergence over the course of the interaction. These holistic perceptual similarity judgments may take into account similarity at the level of phrasal prosody (e.g., mean F0, speech rate). Nonetheless, independent work shows a relationship between acoustic correlates of convergence at the level of sub-lexical units (e.g., phones) and holistic perceptual assessments of convergence (Clopper & Dossey, 2020; Pardo et al, 2017), which suggests a role for the systematic generalization of sub-lexical phonetic patterns in the lexically unrestricted, phonetic convergence of sub-lexical material as in the Kim et al. study.

Yet another type of evidence for systematic generalization comes from Neogrammarian sound change (Paul, 1880; discussed e.g., in Labov, 1981; Hale, 2003). This type of sound change is characterized as phonetically motivated and "regular" in its application across the lexicon, wherever the phonetic conditions are present. Examples of this type of sound change abound in research spanning over 100 years (Labov, 1981). Labov (2006) highlights several recent examples from North American English, including the tensing of short-/a/ before a nasal (e.g., *can* vs. *cat*), raising of the diphthong /aɪ/ before voiceless consonants (e.g., *tight* vs. *tide*),

and the fronting of /uʊ/ and /oʊ/ after coronals except when followed by a liquid (*e.g.,* fronting in *do* but not in *boo, tool*). Neogrammarian sound change poses a challenge for eliminativist exemplar theory because it is presumed to operate over discrete sound units, not words. In an eliminativist exemplar account, lacking symbolic (phonological) sound units, phonetically conditioned sound change would affect only those individual words where the phonetic conditions are met. And yet, as Pierrehumbert notes, "*Historical change does not have the character of random drifts of the pronunciation patterns for individual words. If it did [...] each word would be an individual point somewhere in phonetic hyperspace*." [2006: 522]. An eliminativist exemplar account would additionally allow sound change to progress through lexical diffusion, where a change originating in one word may extend to other words through analogical processes operating, e.g., over words related through shared morphosyntactic features, but rapid diffusion across the entire lexicon, as claimed for Neogrammarian sound change, is not an expected outcome (see Pierrehumbert, 2002, 2006, for discussion of related issues in the context of sound change).

Examples such as these are difficult to account for in an eliminativist exemplar system where generalization is governed not by shared, compositional representational structures, but by integrated, non-modular exemplar representations.

4.2 Variable phonetic effects of multiple types of lexical relationships

There have been a large number of studies following up on Wright's (1997, 2004) seminal work examining the influence of neighborhood density on phonetic variation. As reviewed below, neighborhood effects are far more complex than the core exemplar framework would predict. (See also Strycharczuk (2019) for a review of the complex empirical picture of phonetic effects related to morphological structure.)

4.2.1. Enhancement vs. reduction

Wright (1997, 2004) reported that lexical neighbors enhance vowel contrasts in English. Similar[2] results have been reported in subsequent studies of read (Clopper et al., 2017) and spontaneous (Wedel et al., 2018) English speech. However, similarly-powered studies (Gahl et al., 2012; see also Gahl & Strand, 2016) found the opposite – a reduction of vowel contrast (see also Gahl, 2015, for analyses questioning the original conclusions of Wright). A similar reduction was found for cross-linguistic lexical neighbors (cognates – translation equivalents with highly similar forms such as English *telephone* and Spanish *teléfono*; Amengual, 2016). Discrepancies of this sort have been reported for consonantal contrasts. Some studies (read speech: Baese-Berk & Goldrick, 2009; spontaneous speech: Nelson & Wedel, 2017) show that lexical neighbors enhance VOT distinctions for voiceless vs. voiced stops. However, cross-linguistic lexical neighbors reduce VOT contrasts between words (Amengual, 2012; Jacobs et al., 2016); a similar reduction occurs when a target word is preceded by a prime word that is a lexical neighbor (Levi, 2015). While any finding of lexically-conditioned phonetic effects is broadly consistent with the non-modular storage of information in exemplars, the systematic divergence in effects across populations and processing contexts finds no clear account within the core

[2] We use 'enhancement' to refer to both increases in spectral distinctions at vowel midpoint and increased coarticulation (e.g., Scarborough & Zellou, 2013).

exemplar architecture. Accommodating such effects would seem to require changes that directly speak to core principles of the theory (e.g., altering the nature of Principle II by changing how different types of similarity influence processing).

4.2.2. Different patterns of enhancement/reduction for temporal vs. spectral properties of vowels

The preceding section follows the implicit assumption of the core exemplar account – that 'enhancement' or 'reduction' will uniformly impact all phonetic properties. Since the smallest unit of representation is the word, there is no means by which exemplar mechanisms could differentiate between phonetic properties. This account fails to predict any such differences (all else being equal). However, Clopper and Turnbull (2018) review several recent studies, conducted in multiple languages, that suggest the reduction of temporal and spectral properties of vowels are differentially impacted by manipulations of lexical (neighborhood density, lexical frequency) and discourse properties (e.g., predictability, second mention, speaking style).

For example, Burdin et al. (2015) analyze vowel durations and spectral properties in American English read speech. For vowel durations, they find an interaction between neighborhood density and speaking style (plain speech produced without explicit clarity prompts vs. clear speech produced as if talking to someone who is hard of hearing or a non-native speaker). Vowel durations are shorter for words with few vs. many neighbors and for plain vs. clear speech. These interact, such that the effect of neighborhood density is stronger in plain speech. In contrast, this interaction is not found in the analysis of vowel spectral contrasts. There was a main effect of style (shorter vowels in plain vs. clear speech) but the small effect of neighborhood density (reduced contrasts for words with few vs. many neighbors) may have been obscured by large effects of other variables. This, along with other results, lead Clopper and Turnbull (2018) to suggest that reduction/enhancement is not necessarily global. This poses a challenge for an architecture that predicts no differences across different phonetic properties.

4.2.3. The nature of lexical relationships

Wright (1997, 2004), building on work in speech perception, defined neighbors as words related by a single phone substitution, addition, or deletion (e.g., for target *pat*: *bat, spat, at*), weighting their contribution to density measures by frequency. While many other studies in this area have adopted this working definition, an alternative approach has been to define neighborhood relationships solely in terms of minimal pairs differing only by a single phonetic cue (e.g., for target *pat*: *bat*; Baese-Berk & Goldrick, 2009). Recent work explicitly comparing these two measures in English spontaneous speech suggests that the latter better predicts phonetic effects (Nelson & Wedel, 2017; Wedel et al., 2018; but see Fricke et al., 2016, for results favoring a third metric in read speech). Wedel et al. (2013) consider related data from historical change. Phonological processes of contrast neutralization that result in sound change ('mergers') are constrained by functional load. The number of minimal pairs differentiated by a contrast between two phones, *x* and *y,* is inversely correlated with the probability that *x* and *y* will undergo merger. Wedel et al. further show that it is minimal pairs counted over lemmas (root morphemes), not lexemes (surface word forms) that best predicts merger. A successful model of sound change therefore requires lexical encoding of morphological structure *below* the level of the phonetically realized word.

Again, while lexically-conditioned phonetic effects are broadly consistent with the integrated, non-modular storage of phonetic information in exemplars, the core exemplar theory does not explain why certain types of lexical relationships would be more influential than others.

4.3 Non-uniformity of phonetic convergence

One of the advantages of core exemplar theory over generative theory is its capacity to account for lexically conditioned phonetic plasticity in the form of phonetic convergence (§2.2.3). Convergence occurs as the consequence of exemplar storage of phonetic details of every heard instance (Principle I). When a phonetically novel instance of a word is encountered in the speech of another talker, *all* perceived phonetic properties of the novel form are encoded, and *a priori* have an equal potential to influence subsequent experiences of producing (and perceiving) the word. Yet studies of phonetic convergence show non-uniform effects, with convergence-related shifts in production observed for some, but not all, phonetic properties of word forms. For example, Pardo et al. (2017) examined acoustic evidence of phonetic convergence in a large single-word shadowing experiment with American English speakers, and report convergence in word duration, two-dimensional vowel space (F1xF2), and F2 alone, but not in F1 alone or F0. Clopper and Dossey (2020) performed a similar experiment, using single-word shadowing to test convergence on several distinguishing features of Southern American English with native speakers of non-Southern varieties. Acoustic measures from that study show convergence in word duration and vowel backness (F2), but not on the duration of vowel formant trajectories as correlates of monophthongization. Disparities in phonetic convergence across measures are also reported in Gessinger et al. (2021), which used a sentence-shadowing task with German speakers to compare phonetic evidence for convergence to regional dialectal variants in local (phone level) and global (phrasal prosody) measures, and in naturally produced and synthesized speech. Phonetic convergence to natural speech stimuli was observed for one phonetic variable (whether a word-final ⟨-ig⟩ is realized as [ɪç] or [ɪk], based on narrow phonetic transcription), but not in another (whether word-final <-en> is realized as a syllabic nasal [n̩] or with an epenthetic schwa [ən] based on acoustic duration of phonetically transcribed vocalic intervals). Examining convergence both at the group level and for individual participants, Gessinger et al. explicitly remark that convergence in one phonetic feature does not predict convergence in another feature (see also Cohen Priva and Sanker, 2020, for similar findings of variable convergence on acoustic prosodic measures in spontaneous conversation). Further evidence for variation in convergence comes from Ostrand and Chodroff (2021), who measure seven acoustic-phonetic and temporal measures from lexically unrestricted and variable speech elicited in a dyadic interactive game with American English speaking participants. They find evidence for partner-specific convergence in two temporal features (pause duration, speech rate), but not in five spectral-phonetic features (F1 for four vowel phones and A-I-U dispersion).

To summarize, studies testing convergence in shadowing tasks and in interactive speech report evidence of phonetic convergence for some but not all of the measured acoustic and phonological parameters, which differ across studies, and also point to non-uniformity of convergence across individual speakers. Understanding the nature and limits of this non-uniformity requires additional research. There is substantial variation across studies in the elicitation methods and in the measured acoustic and phonological variables. Furthermore, as

discussed by Cohen Priva & Sanker (2019), some studies fail to properly control for intra-speaker variability and the baseline similarity of the talker to the target of convergence, which can lead to spurious results. Although this is unlikely to account for the full range of results above (see, e.g., Gessinger et al. 2021 and Ostrand & Chodroff, 2019, for discussion), it increases our uncertainty about the precise extent of (non-)uniformity in convergence across acoustic parameters. What is clear from the available evidence is that there are limits on convergence – limits that are not predicted under core exemplar theory.

5. Twenty first century theoretical advances: Hybrid exemplar models

5.1 Principles of hybrid models

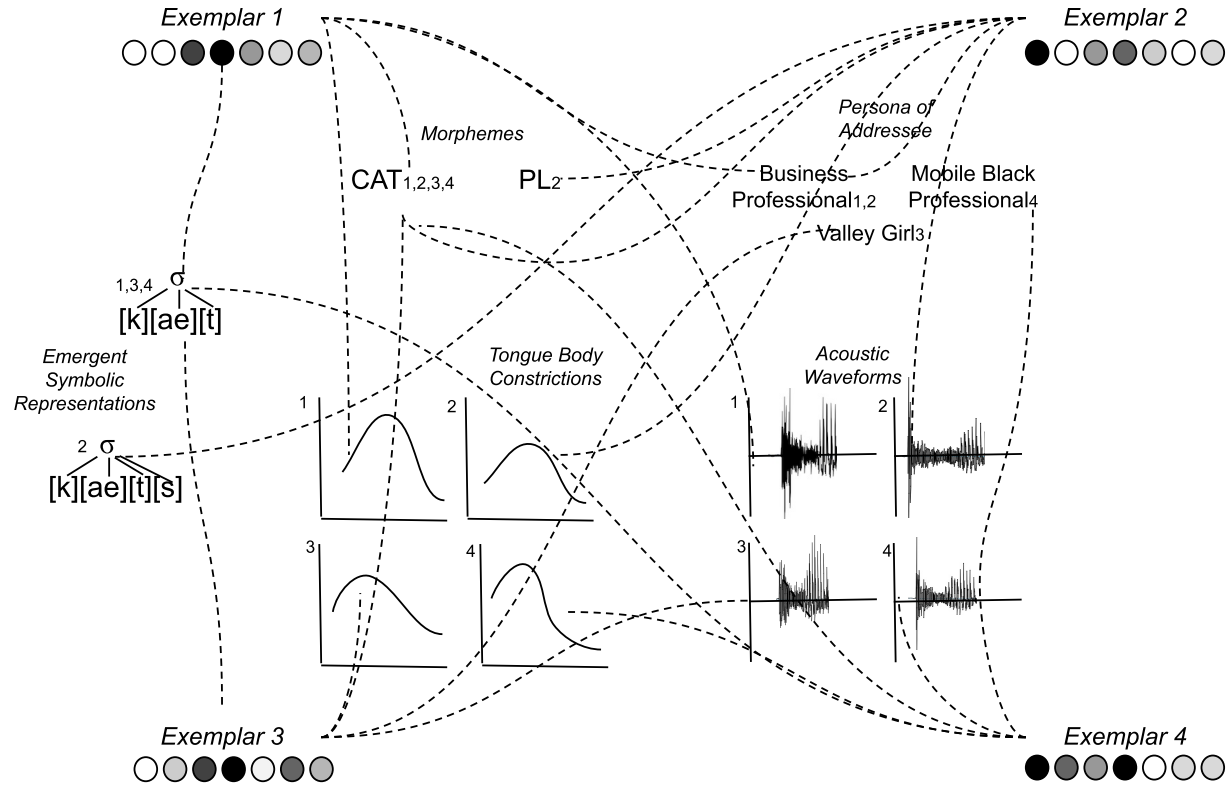


Figure 4. A simplified illustration of the hybrid exemplar architecture showing four exemplars: three instances of the word *cat,* and one of the word *cats*. Memory representations (here, a seven-dimensional vector, analogous to representations in eliminativist connectionist or deep learning systems) are particular speech experiences or *exemplars*. These are associated (as shown by dotted lines and subscripts) with various dimensions of information: morphological, phonetic, and social information (here, personae, see D'Onofrio, 2021). Critically, in a hybrid architecture, exemplars are associated to symbolically-structured representations (here, shown as phones associated with a syllable) that have emerged based on exemplar processing during learning.

Hybrid exemplar models of production (Pierrehumbert, 2002; anticipated in Goldinger, 1998:265-266; for reviews of a diverse array of specific proposals[3], see: Davis & Redford, 2019; Pierrehumbert, 2006; Morley, 2019; Shattuck-Hufnagel, 2014; Todd et al., 2019; Walsh et al., 2010; Wedel & Fatkullin, 2017) modify a core representational principle of eliminativist exemplar models §3.2 (III) (new text in *italics*):

[3] The Analogical Modeling of Language framework (e.g., Eddington, 2000) can also be viewed as a type of hybrid model. In this framework, exemplars are coded with respect to a set of 'variables' which can represent quite abstract aspects of phonological and morphological structure.

Principle III'. Emergent *symbolic and non-symbolic* phonological structures. Effects attributed to symbolic phonological structure emerge from non-symbolic representations and processes during learning and processing. *The resulting representations will, under certain circumstances, share key properties with symbolic phonological representations.*

Hybrid models stake a claim as to the products of emergence. For example, in Figure 4, emergent representations include a compositional syllable representation and discretely represented phones. This allows them to capitalize on the strength of phonological abstractions while also respecting the explanatory power of exemplars. Interestingly, there are parallels in the development of psycholinguistic models of speech production. Although some connectionist models of speech production pursued an eliminativist perspective (reviewed in §3.1.2), hybrid frameworks have been widely used. For example, the highly influential production model of Dell (1986) assumed that retrieval of word forms from long term memory is a dynamical, gradient spreading activation process. Out of this graded activation of multiple elements, single form elements (e.g., a consonant or a vowel) are then selected to fill particular positions in a discrete symbolic structure (e.g., a prosodic frame) that serves to guide subsequent production. This approach continues to guide theoretical development in the field (see, e.g., Levelt et al., 1999; O'Seaghdha et al., 2010). This convergence is not accidental; Pierrehumbert's (2002) hybrid proposal draws on work in speech perception that makes use of a hybrid connectionist architecture (Norris et al., 2000).

It is important to emphasize that (III') does not claim that *all* emergent representations are similar to symbolic phonological structures. Abstractions can also emerge in the context of learning and processing. For example, Cole (2009) examines harmony systems from an exemplar perspective. In this proposal, such systems emerge from patterns of association between sub-lexical and lexical units, without any recourse to explicit feature structure representations. It's also important to note that in contrast to generative learning theories, exemplar theories of emergence have typically assumed that learning is not subject to strong biases. The lack of strong constraints on learning yields a necessary trade-off with between-learner variability (Geman et al., 1992), leading researchers to seek other explanations for the convergence of linguistic communities to common patterns (see, e.g., Pierrehumbert, 2003, for discussion).

5.2 Successes of hybrid models

Pierrehumbert (2002) uses generalization over sub-lexical structures (§4.1) to motivate a hybrid exemplar architecture (see also Pierrehumbert 2006, 2016, for discussion). Experiences with these compositionally-structured sublexical phonological units can modulate processes applying across all word contexts (e.g., lenition; §4.1.1). Learning can target these compositional units, allowing for rapid, systematic generalization across all contexts (§4.1.3). The productivity of type frequency effects (§4.1.2) can be attributed to multiple levels of abstraction within hybrid models. Generalizations are formed based on the distribution of these compositional sub-lexical units across the distinct lexical entries that they occur within (Pierrehumbert 2003, 2006, 2016).

5.3 The limitations of hybrid models

5.3.1. The emergence of structure

While learning in eliminativist exemplar models can be quite simple (reducing simply to storage and activation of exemplars), hybrid exemplar learning requires a process for identifying abstract analyses out of a very large space of possibilities – an extremely challenging problem (see Baayen & Ramscar, 2019, for discussion[4]). While there has been extensive computational exploration of the dynamics[5] of hybrid models (e.g., Morley, 2019; Todd et al., 2019; Wedel & Fatkullin, 2017), assessing their ability to account for a variety of phenomena in language change, such work has not tackled the emergence of novel abstract representations. Pierrehumbert (2003; see also Pierrehumbert, 2006, 2016) proposes that phonetic category formation is initiated by distributional learning (i.e., using modes to infer the number of categories; Maye, Werker, & Gerken, 2002). However, recent work has cast significant doubt on the ability of distributional learning to model phonetic category learning, particularly when examining learning of naturalistic (as opposed to lab-based) speech (e.g., Hitczenko et al., 2020; see Feldman et al., 2021, for a review). It is important to note that Pierrehumbert (2003 et seq.) acknowledges the limitations of distributional learning, proposing that these initial categories are further refined by abstractions over the lexicon (as discussed above). However, the precise mechanisms underlying this complex set of clustering and abstraction processes have not been fully articulated. Current theory therefore leaves Principle III' only half-realized.

5.3.2. The complexity of word-specific phonetics and convergence

The diverse set of empirical effects reviewed in §4.2-3 has not been addressed in any detailed way by new theoretical proposals. There has been discussion of directions in which theories need to be extended. For example, Clopper and Pierrehumbert (2008) discuss how exemplars associated with dialect-specific variants may be accessed more rapidly, resulting in interactions between neighborhood density and social variables. Clopper and Dossey (2020), in discussing the non-uniformity of phonetic convergence, note that convergence of specific phonological and phonetic variables may depend on linguistic factors like the baseline distance between talkers, or on social factors like the stereotyped prestige of a particular variant. The implications of these findings for exemplar theory is that distinct phonetic parameters in exemplar encoding may be differently weighted, and therefore behave differently in effects of lexical relatedness (§4.2) or convergence (§4.3). The differential weighting of acoustic cues has also been proposed to explain individual differences among listeners in the mapping from acoustic cues to phone categories in speech perception, and ultimately, as a path to sound change (Schertz & Clare, 2020). Moreover, Gessinger et al. (2021) note a role for speaker traits

[4] Connectionist advances in this area have been limited as well. With the exception of Dell and colleagues' work on the emergence of representations encoding phonotactic regularities (reviewed in Dell et al., 2021), there has been no progress on this general issue in connectionist models of speech production.

[5] Eliminativist exemplar evolutionary models have examined the emergence over time of speech signals with apparent (but not explicit) compositional structure (e.g., Zuidema & deBoer, 2009), with success limited to highly restricted artificial domains (see Little et al., 2017, for a recent critical review).

related to innate phonetic talent, cognitive function, and personality (Yu et al., 2013) in determining the degree of phonetic convergence, suggesting a tighter integration between linguistic knowledge and the cognitive or neural systems governing other behaviors. However, in general, more systematic and detailed theoretical extensions have not been proposed. (To be clear, this is not due to intrinsic issues with the exemplar framework; see Todd et al., 2019, for the use of detailed exemplar models to examine complex interactions between word frequency and sound change arising in findings that both support and contradict previous claims about Neogrammarian (regular) sound change.)

6. Rising to the challenge: Advancing hybrid exemplar theory

6.1. Situating exemplar processing in a broader theory of speech production

Exemplar modeling has typically incorporated influences on speech production external to the core exemplar mechanisms, from random noise reflecting imprecision in executing motor targets (e.g., Wedel, 2006) or imprecision combined with systematic biases (e.g., lenition pressures; Pierrehumbert, 2001; see Morley, 2019; Todd et al., 2019, for two recent implementations). As these papers and other work shows, modeling the entirety of the rest of the production system through noise and biasing terms has been very productive from a modeling standpoint; in fact, it's likely that such simplifications are what have made insight into the model's dynamics possible. However, in order to confront the much wider set of production data reviewed above, as discussed by Ernestus (2014) and Fink and Goldrick (2015) it is imperative that exemplar processing be situated within a broader model of speech production. By explicitly modeling the influence of processes preceding exemplar retrieval (e.g., selection of lexical items to convey an intended message) and those following retrieval (e.g., articulatory planning and execution), exemplar theories may find a means to generate a more complex set of predictions for production phenomena. Here, we see promise in renewing and deepening links to hybrid connectionist models of speech production as well as exploring links with dynamical models of speech planning and articulation.

Within the context of a hybrid model, it is critical to clarify how structure plays a role in planning or exemplar updating. There is a diverse array of approaches in current use. At one extreme, Walsh et al. (2010) treat these as two separate processing "routes" for production (such that plans are either assembled compositionally via abstract representations, or via holistic exemplars). In a more integrated architecture, it will be important to consider how exemplar processing is influenced by, and influences, structure-sensitive processes. Shattuck-Hufnagel (2014) sketches one such approach, where articulatory planning of an utterance can either draw on stored exemplars at multiple levels of structural granularity (from adjacent segments to whole fixed phrases) or utilize more general processes that translate this plan into a structured phonetic representation. While they do not make use of structure-based utterance plans, other hybrid models (e.g., Davis & Redford, 2019; Pierrehumbert, 2002; Morley, 2019; Todd et al., 2019; Wedel & Fatkulin, 2017) also assume general phonetic processes that apply across all productions (as discussed above). In such models, abstract representations are integrated into exemplar processing during similarity calculations, defining the relevant set of exemplars to be utilized. For example, Wedel (2012:332) proposes that productions are "biased

towards previously heard exemplars at both the word *and sound* levels. [emphasis added]" This influence of sub-lexical similarity creates a pathway for phonetic effects driven by competition between specific words (e.g., minimal pairs) to spread throughout the lexicon (accounting for effects of functional load on probability of phone category mergers; see §4.2.3 for discussion).

The possibility of a division of labor between structure sensitive vs. exemplar processes raises an important analytical challenge for exemplar theories: whether to attribute some or all of a particular empirical effect to one (or both) processes (Ernestus, 2014). This is a pervasive issue in cognitive science more broadly. Onnis and Huettig (2021) discuss whether frequency effects for multi-word sequences (e.g., Arnon & Cohen Priva, 2014) necessarily reflect storage or are instead due to active processing (i.e., pre-activation of upcoming material and ease of integration for previous planned material). Importantly, under certain computational frameworks, this question might be ill-posed. For example, in connectionist models of memory it's unclear whether it's possible to draw a line between the mechanisms implementing retrieval of a known, 'stored' item and those that generate novel forms (see Hinton et al.,1986, for discussion).

This division of storage and computation is highly salient when considering how exemplar models – and highly lexicalist psycholinguistic models – might move beyond isolated words to consider how prosodically-conditioned phonetic variation should be incorporated into a hybrid speech production model. A range of models have been proposed in recent work. In some, there is an emphasis on computation. As noted above, Shattuck-Hufnagel (2014) has proposed a multiple-route production model. One set of processes are sensitive to explicit rules relating prosodic structure to phonetic variation. These processes are complemented by those that access a "prosodicon of constituents, *separate* from form-meaning pairings in the lexicon (p. 269; emphasis added)." In a related vein, Cho (2011; see also Cho 2022) proposes a computational process by which abstract categories are drawn from the lexicon after which post-lexical prosodic structure is computed, taking into account syntactic and pragmatic context. Phonetic parameter values for abstract phonological categories are then determined through selection of an appropriate token from the stored exemplar cloud, in which each stored exemplar is associated with information about its prosodic context. Subsequent computation is responsible for further fine-tuning of the phonetic parameter values before motor implementation. Thus, even in proposals with a processing route emphasizing storage, prosodic variation requires computations that integrate stored constituents with stored segmental and sub-segmental representation. In contrast, on the basis of lexical frequency effects on prosodic variation, Schweitzer et al. (2015) and Tang and Shaw (2021) argue that prosodic variation is stored within lexical exemplars. Further elaboration of these proposals within a broader model of speech production is a critical avenue for development of hybrid exemplar theories.

6.2 The seeds of emergence

Understanding the emergence of complex structure or behavior from the interaction of simpler primitives is a highly challenging problem. In exemplar theories, the emergence of patterns has frequently been explored through model-/simulation-based methods. In such an approach, a simplified computational model of the exemplar account is constructed. As these models include randomness (e.g., imprecision in realizing articulatory targets), multiple simulations will produce a range of results. The behavior of the models can then be studied to

examine if the desired patterns emerge (see Morley, 2019; Todd et al., 2019; Wedel & Fatkulin, 2017, for recent examples). Incorporating an analytic perspective as well may help advance the development of hybrid exemplar accounts. For example, Iskarous (2017, 2019) examines how a small set of principles can give rise to complex dynamical behavior. This relies not on simulation but on an understanding of the core mathematical principles underlying the (relatively) complex mechanisms explored in simulation studies. (See Plaut et al., 1996, for an illustration of how both simulation and analytical approaches inform the development and testing of a connectionist account of emergence.)

The grist for the mill of emergence is the set of exemplars that are stored by the learner. Another key area for both theoretical and empirical development is developing a better understanding of the processes underlying storage. At the point of encoding, current hybrid models incorporate mechanisms that filter out ambiguous exemplars (e.g., Pierrehumbert, 2002; see Morley, 2019, for a review). Exclusion of these tokens alters the phonetic properties of sound categories that emerge over time. These mechanisms could be elaborated and extended to account for the weak encoding (or complete failure to encode) of exemplars along a variety of dimensions (as discussed in §4.3) , or by including a mechanism for differential (cue) weighting of phonetic parameters in exemplar encoding (§5.3.2).

There is also evidence that the content of exemplars can be impacted by processing during encoding. Goldinger (1996) found that the strength of indexical effects on memory performance was modulated by which stimulus dimensions a listener's attention was directed to. Listeners who made gender classifications at encoding (e.g., is the word you heard spoken by a male or female talker?) showed stronger indexical effects than listeners who made syntactic category classifications (e.g., is the word you heard a noun or a verb?). Clopper and Dossey (2021) discuss a related finding, where convergence effects in word shadowing are stronger when shadowers are explicitly instructed to imitate, implicating a role for attention in modulating exemplar effects on speech production (see also Schertz & Paquette-Smith, 2023, for differential patterns of convergence in an explicit imitation task). Although we're not aware of any work explicitly examining this question, the overall framework predicts that such modulations of exemplar content should have downstream consequences for speech production. Results from such studies could inform the development of more detailed accounts of the role of attention in the mechanisms of exemplar encoding.

7. Conclusions

At the dawn of the twenty first century, the core principles of exemplar models of lexical encoding had emerged, providing a novel perspective for understanding phonological and phonetic aspects of speech production. Phonologists and phoneticians rapidly realized the shortcomings of an eliminativist stance within this framework, leading to the introduction of hybrid exemplar models that integrate compositional representations of phonological structure with exemplar storage.

As we enter the third decade of the twenty-first century, the empirical challenges faced by current exemplar theories suggest it is time for exemplar accounts to move beyond splendid isolation – to consider, in detail, how exemplar processing is integrated with other components of the speech processing system. Integration with speech production mechanisms that trigger

exemplar retrieval, connect retrieved information with articulation, and execute speech in real time may help exemplar accounts capture the complexity of speech production data. A deeper specification of memory encoding mechanisms will help serve the development of more explicit theories of the emergence of linguistic structure. More broadly, a re-balancing of our collective effort is in order; greater prioritization of theory development is necessary to catch up to the tremendous volume of empirical data inspired by exemplar principles.

Acknowledgments
Supported in part by National Science Foundation grants to JC (BCS-1944773) and MG (EHR 2219843). Thanks to Ann Bradlow and José I. Hualde for helpful comments and suggestions.

Reference List

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition, 15*, 517–530.

Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: Evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Frontiers in Psychology, 7*, 617.

Alderete, J., & Tupper, P. (2018). Phonological regularity, perceptual biases, and the role of phonotactics in speech error analysis. *Wiley Interdisciplinary Reviews: Cognitive Science, 9*, e1466.

Anderson, K., Milostan, J., & Cottrell, G. W. (1998). Assessing the contribution of representation to results. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the Cognitive Science Society* (pp. 48- 53). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Arnon, I., & Priva, U. C. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon, 9*, 377-400.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*, 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40,* 177–189. https://doi.org/10.1016/j.wocn.2011.09.001.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes, 24*, 527-554.

Baayen, R. H., & Ramscar, M. (2019). Abstraction, storage and naive discriminative learning. In E. Dąbrowska & D. Divjak (Eds.) *Cognitive linguistics-foundations of language* (pp. 115-139). De Gruyter Mouton.

Braver, A. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology, 36*, 1-36.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica, 49*, 155-180.

Burdin, R. S., Turnbull, R., & Clopper, C. G. (2015). Interactions among lexical and discourse characteristics in vowel production. *Proceedings of Meetings on Acoustics, 22*, 060005.

Bybee, J. (1999). Usage-based phonology. In M. Darnell, E. Moravcsik, F. Newmeyer, M. Noonan, & K. Wheatley (Eds.) *Functionalism and formalism in linguistics volume I: General papers* (pp. 211-242). Amsterdam: John Benjamins.

Bybee (2001). *Phonology and language use*. Cambridge: Cambridge University Press

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*, 711-733.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*, 381-410.

Carr, P. (1999). Sociophonetic variation and generative phonology: the case of Tyneside English. *Cahiers de grammaire, 24*, 7-15.

Chang, C. B. (2019). Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics, 74*, 96-113.

Cho, T. (2011). Laboratory phonology. In N.C. Kula, B. Botma, & K. Nasukawa (Eds.), *The Continuum companion to phonology* (pp.343-368). London/New York: Continuum.

Cho, T. (2022). Linguistic functions of prosody and its phonetic encoding with special reference to Korean. In K. Horie, K. Akita, Y. Kubota, D. Y. Oshima and A. Utsugi (Eds.), *Japanese/Korean linguistics* (vol. 29, pp. 1-24). Palo Alto, CA: CSLI Publications.

Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*, 205-235.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America, 124*, 1682-1688.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology, 6*, 243-278.

Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language, 93*, 569-597.

Cohen Priva, U., & Sanker, C. (2019). Limitations of difference-in-difference for measuring convergence. *Laboratory Phonology, 10*, 15. https://doi.org/10.5334/labphon.200

Cohen Priva, R. & Sanker, C. (2020). Natural leaders: Some interlocutors elicit greater convergence across conversations and across characteristics. *Cognitive Science, 44*, e12897.

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology, 10*, 43-81.

Cole, J. (2009). Emergent feature structures: Harmony systems in exemplar models of phonology. *Language Sciences, 31*, 144-160.

Cole, J., & Hualde, J. I. (2011). Underlying representations. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (eds.), *The Blackwell companion to phonology* (vol 1., pp. 1-26). Malden, MA: Wiley-Blackwell.

Clopper, C. G., Mitsch, J. F., & Tamati, T. N. (2017). Effects of phonetic reduction and regional dialect on vowel production. *Journal of Phonetics, 60*, 38-59.

Clopper, C. G., & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler & M. Zellers (Eds.) *Rethinking reduction* (pp. 25-72). De Gruyter Mouton.

Clopper, C. G., & Dossey, E. (2020). Phonetic convergence to Southern American English: Acoustics and perception. *The Journal of the Acoustical Society of America, 147,* 671-683.
Davis, M., & Redford, M. A. (2019). The emergence of discrete perceptual-motor units in a production model that assumes holistic phonological representations. *Frontiers in Psychology, 10*, 2121.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283-321.
Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 4*, 313-349.
Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*, 149–195.
Dell, G. S., & Kim, A. E. (2005). Speech errors and word form encoding. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.) *Phonological encoding and monitoring in normal and pathological speech* (pp. 29-53). London: Psychology Press.
Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1355-1367.
Dell, G. S., Kelley, A. C., Hwang, S., & Bian, Y. (2021). The adaptable speaker: A theory of implicit learning in language production. *Psychological Review, 128*, 446-487.
Denby, T., Schecter, J., Arn, S., Dimov, S., & Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 44, 280-294.
D'Onofrio, A. (2021). Sociolinguistic signs as cognitive representations. In L. Hall-Lew, E. Moore, & R. J. Podesva (Eds.), *Social meaning in linguistic variation: Theorizing the third wave* (pp. 153-175). Cambridge University Press.
Drager, K., & Kirtley, M. J. (2016). Awareness, salience, and stereotypes in exemplar-based models of speech production and perception. In A. M. Babel (Ed.) *Awareness and control in sociolinguistic research* (pp. 1-24). Cambridge: Cambridge University Press.
Dresher, B. E. (1999). Charting the learning path: Cues to parameter setting. *Linguistic Inquiry, 30*, 27-67.
Eddington, D. (2000). Spanish stress assignment within the analogical modeling of language. *Language, 76*, 92-109.
Edwards, J., Beckman, M. E., & Munson, B. (2015). Frequency effects in phonological acquisition. *Journal of Child Language, 42*, 306-311.
Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua, 142*, 27-41.
Ernestus, M., & Baayen, R. H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In L. M. Goldstein, D. H. Whalen, and C. T. Best, *Laboratory phonology 8* (pp.27-49). Berlin: Mouton de Gruyter.
Feldman, N. H., Goldwater, S., Dupoux, E., & Schatz, T. (2021). Do infants really learn phonetic categories? *Open Mind, 5*, 113-131.
Fink, A., & Goldrick, M. (2015). The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard, 1*, 215-225.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409-438.
Fricke, M., Baese-Berk, M. M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition and Neuroscience, 31*, 639-645.
Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language, 84*, 474-496.
Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics, 49*, 96-116.
Gahl, S., & Strand, J.F. (2016). Many neighborhoods: Phonological and perceptual neighborhood density in lexical production and perception. *Journal of Memory and Language, 89*, 162-178
Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789-806.
Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4*, 1-58.
German, J. S., Carlson, K., & Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics, 41*, 228-248.
Gessinger, I., Raveh, E., Steiner, I., & Möbius, B. (2021). Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing. *Speech Communication, 127*, 43–63. http s://doi.org/10.1016/j.specom.2020.12.004.
Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166-1183.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251-279.
Goldinger, S. D. (2007). A complementary-systems approach to abstract and episodic speech perception. In J. Trouvain and W. J. Barry (Eds.) *Proceedings of the 16th international congress of phonetic sciences* (pp. 49-54).
Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*, 716-722.
Goldrick, M. (2011). Linking speech errors and generative phonological theory. *Language and Linguistics Compass, 5*, 397-412.
Goldrick, M. (2017). Encoding of distributional regularities independent of markedness: Evidence from unimpaired speakers. *Cognitive Neuropsychology, 34*, 476-481.
Hale, M. (2003). Neogrammarian sound change. In B.D. Joseph and R.D. Janda (Eds.), *The handbook of historical linguistics* (pp. 343-368). Oxford: Blackwell.
Hay, J., Jannedy, S., & Mendoza-Denton, N. (1999). Oprah and/ay: Lexical frequency, referee design and style. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1389-1392). Berkeley, CA: University of California.
Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In Rumelhart, D. E., McClelland, J. L., & the PDP research group. *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I* (pp. 77-109). Cambridge, MA: MIT Press.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.
Hitczenko, K., Mazuka, R., Elsner, M., & Feldman, N. H. (2020). When context is and isn't helpful: A corpus study of naturalistic speech. *Psychonomic Bulletin and Review, 27*, 640-676.
Iskarous, K. (2017). The relation between the continuous and the discrete: A note on the first principles of speech dynamics. *Journal of Phonetics, 64*, 8-20.
Iskarous, K. (2019). The morphogenesis of speech gestures: From local computations to global patterns. *Frontiers in Psychology, 10*, 2395.
Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Language Learning, 66*, 324-353.
Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in spoken production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824-843.
Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. In Solé, M.J., Beddor, P. & Ohala, M. (eds) *Experimental approaches to phonology: In honor of John Ohala* (pp. 25-40). Oxford: Oxford University Press.
Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*. Institute for Cognitive Science Technical Report 8604. La Jolla, CA: University of California at San Diego. [Reprinted in J. W. Donahoe & V. P. Dorsel (Eds.), (1997). *Neural-network models of cognition: Biobehavioral foundations* (pp. 221- 277). Science Press.]
Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology, 2*, 125–156. https://doi.org/10.1515/labpho n.2011.004.
Kirchner, R., Moore, R. K., & Chen, T. Y. (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics, 38*, 540-547.
Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of lexical frequency in word production? Insights from aphasic picture naming errors. *Cognitive Neuropsychology, 25*, 463–492.
Krämer, M. (2012). *Underlying representations*. Cambridge University Press.
Kruschke, J. K (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.
Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.) *The Cambridge handbook of computational psychology* (pp. 267-301). New York: Cambridge University Press.
Labov, W. (1981). Resolving the Neogrammarian controversy. *Language*, *57*, 267-308.
Labov, W. (2006). A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics, 34*, 500-515.
Laganaro, M. (2019). Phonetic encoding in utterance production: A review of open issues from 1989 to 2018. *Language, Cognition and Neuroscience, 34*, 1193-1201.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-38.
Levelt, W.J.M. & Wheeldon, K. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239-269.

Levi, S. V. (2015). Generalization of phonetic detail: Cross-segmental, within-category priming of VOT. *Language and Speech, 58*, 549-562.
Lindsay, S., Clayards, M., Gennari, S., & Gaskell, M. G. (2022) Plasticity of categories in speech perception and production, *Language, Cognition and Neuroscience, 37*, 707-731.
Little, H., Rasilo, H., Van Der Ham, S., & Eryılmaz, K. (2017). Empirical approaches for investigating the origins of structure in speech. *Interaction Studies, 18*, 330-351.
Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The Weckud Wetch of the Wast: Lexical adaptation to a novel accent. *Cognitive Science, 32,* 543–562. https://doi.org/10.1080/03640210802035357.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.
McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science, 2*, 751-770.
McClelland, J. L., Rumelhart, D. E., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume II.* Cambridge, MA: MIT Press.
MacLeod, B. (2021). Problems in the Difference-in-Distance measure of phonetic imitation. Journal of Phonetics, 87, 101058. https://doi.org/10.1016/J.WOCN.2021.101058
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.
Morley, R. L. (2019). *Sound structure and sound change: A modeling approach*. Language Science Press. DOI: 10.5281/zenodo.3264909
Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. S. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 201-228). New York: Mouton de Gruyter.
Nelson, N. R., & Wedel, A. (2017). The phonetic specificity of competition: Contrastive hyperarticulation of voice onset time in conversational English. *Journal of Phonetics, 64*, 51-70.
Nicenboim, B., Roettger, T. B., & Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in German. *Journal of Phonetics, 70*, 39-55.
Nielsen, K. (2011). Specificity and abstractness of VOT imitation. Journal of Phonetics, 39(2), 132-142.
Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-325.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
Onnis, L., Huettig, F. (2021). Can prediction and retrodiction explain whether frequent multi-word phrases are accessed ‘precompiled’ from memory or compositionally constructed on the fly? *Brain Research, 1772*, 147674.
O’Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282-302.

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, *88*, 101074.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. The Journal of the Acoustical Society of America, 119(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology, 4*, 1–5. h ttps://doi.org/10.3389/fpsyg.2013.00559.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. Journal of Phonetics, 40(1), 190–197. https://doi.org/10.1016/J.WOCN.2011.10.001

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics, 79*, 637–659. https://doi.org/10.3758/s13414-016 -1226-0.

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. Journal of Phonetics, 69, 1–11. https://doi.org/10.1016/j.wocn.2018.04.001

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Tübingen: Max Niemeyer.

Pierrehumbert, J. B. (1994). Knowledge of variation. In *CLS-30: Papers from the 30th regional meeting of the Chicago Linguistic Society, Vol 2: Parasession on Variation and Linguistic Theory* (pp. 232-256). Chicago: Chicago Linguistics Society.

Pierrehumbert, J. B. (2001a). Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee and P. Hopper (Eds.) *Frequency and the emergence of linguistic structure* (pp. 137 - 158). Amsterdam: John Benjamins.

Pierrehumbert, J. (2001b). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes, 16*, 691-698.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In: Gussenhoven, C., Warner, N., Rietveld, T. (Eds.), *Phonology & phonetics [Laboratory Phonology 7]* (pp. 101-140). Berlin: Mouton.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.

Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics, 4*, 516-530.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2,* 33-52.

Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech perception and comprehension: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381- 415). Mahweh, NJ: Erlbaum.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics, 13*, 455-471.

Richtsmeier, P., Gerken, L., & Ohala, D. (2011a). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language, 38*, 951-978.

Richtsmeier, P. T. (2011b). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology, 2*, 157–183.

Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I.* Cambridge, MA: MIT Press.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Scarborough, R., & Zellou, G. (2013). Clarity in communication:"Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *Journal of the Acoustical Society of America, 134*, 3793-3807.

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2), e1521.

Schertz, J., & Paquette-Smith, M. (2023). Convergence to shortened and lengthened voice onset time in an imitation task. *JASA Express Letters*, *3*(2), 025201.

Schweitzer, K., Walsh, M., Callhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication, 66*, 65-81.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*, 140-155.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial order mechanism in sentence production. In W.E. Cooper & E.C.T. Walker (Eds.), *Sentence processing* (pp. 295-342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (2014). Phrase-level phonological and phonetic phenomena. In Goldrick, M., Ferreira, V., & Miozzo, M. (Eds.) *The Oxford handbook of language production* (pp. 259-274). Oxford: Oxford University Press.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*, 422-429.

Smalle, E. H. M., & Szmalec, A. (2022). Quick learning of novel vowel-consonant conjunctions within the mature speech production system – a commentary on Dell et al. (2019). *Language, Cognition, and Neuroscience, 37*, 532-536.

Smolensky, P., McCoy, R. T., Fernandez, R., Goldrick, M., & Gao, J. (2022). Neurocompositional computing: From the central paradox of cognition to a new generation of AI systems. *AI Magazine, 43,* 308-322.

Sonderegger, M., Bane, M., & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language, 93*, 598-640.

Strycharczuk, P. (2019). Phonetic detail and phonetic gradience in morphological processes. *Oxford Research Encyclopedia of Linguistics*.

Tang, K., & Bennett, R. (2018). Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan). *Journal of the Acoustical Society of America, 144*, 997-1017.

Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition, 210*, 104601.

Tenpenny, P. L. (1995). Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin & Review, 2*, 339-363.

Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry, 29*, 229-268.

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition, 185*, 1-20.

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2021). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics, 57*, 123-161.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science, 34*, 537-582.

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review, 23*, 247-274.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition, 4*, 319-355.

Wedel, A., & Fatkullin, I. (2017). Category competition as a driver of category contrast. *Journal of Language Evolution, 2*, 77-93.

Wedel, A., Nelson, N., & Sharp, R. (2018). The phonetic specificity of contrastive hyperarticulation in natural speech. *Journal of Memory and Language, 100*, 61-88.

Wright, R. A. (1997). Lexical competition and reduction in speech: A preliminary report. In D. Pisoni (Ed.) *Research on spoken language processing progress report 21* (pp. 471-485). Bloomington, IN: Speech Research Lab, Psychology Department, Indiana University.

Wright, R. A. (2004). Factors of lexical competition in vowel articulation. In J. J. Local, R. Ogden, & R. Temple (Eds.), *Laboratory phonology* (Vol. 6, pp. 26-50). Cambridge, UK: Cambridge University Press.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PLoS ONE, 8*, e74746. https://doi.org/10.137 1/journal.pone.0074746.

Zuidema, W., & de Boer, B (2009). The evolution of combinatorial phonology. *Journal of Phonetics, 37*, 125-144.