CrossMark

Variation in the speech signal as a window into the cognitive architecture of language production

Audrey Bürki [1,2]

Abstract
The pronunciation of words is highly variable. This variation provides crucial information about the cognitive architecture of the language production system. This review summarizes key empirical findings about variation phenomena, integrating corpus, acoustic, articulatory, and chronometric data from phonetic and psycholinguistic studies. It examines how these data constrain our current understanding of word production processes and highlights major challenges and open issues that should be addressed in future research.

Keywords Language production · Variation · Psycholinguistic models · Phonology · Phonetics

Introduction

Standard dictionaries list the canonical forms of words, yet in connected speech the pronunciation of these words may take various forms. Pronunciations with fewer segments or syllables, for instance, are frequent in many languages. Johnson (2004) reported that in American English, 25% of words have lost one segment, and 6% have lost a full syllable, with respect to their described canonical forms. In French, many words have a schwa, and this schwa is often not produced (e.g., Racine & Grosjean, 2002). Reductions can be even more drastic. Hawkins (2003) discussed the pronunciation of the English sentence *I do not know*, often realized as *dunno*. Ernestus (2014) provided Dutch examples such as the words *eigenlijk* /ɛixələk/ "actually" and *natuurlijk* /natyrlək/ "of course" realized as, respectively, ['ɛik], and [tyk]. The pronunciation of many words also often depends on contextual constraints. For instance, in many dialects of English, the phoneme /r/ is not realized in final position (e.g., [kɑ:] for *car*), but can be realized in connected speech when the word is resyllabified with the following word (e.g., *the car is*

[kɑ:rɪz]). Similarly, in many languages, word-final sounds can be produced differently depending on the phonological properties of the following word—for instance, English place assimilation, as in *plain* /pleɪn/ being realized as [pleɪm] in *plain bun*; French voice assimilation, as in *sac* /sak/ "bag" being realized as [sag] in *sac de graines* "bag of seeds"; or postnasal devoicing in Tswana, where /b/ is optionally devoiced after nasals.

Variation is an inherent property of speech. It occurs all the time in everyday speech events. Any theory of speech production whose ambition is to describe speech processing in everyday communication must thus account for and explain the mechanisms and representations underlying the production of pronunciation variants.[1] In turn, the study of variation phenomena is likely to provide essential insights into the architecture of the language production system. In his seminal book *Speaking*—which in many ways set the basis of the next 30 years of experimental research in language production—Levelt (1989) describes several types of variation and discusses possible mechanisms that could account for how this variation is generated by the language production system. Surprisingly, however, the cognitive processes and representations underlying the production of pronunciation variants have long been neglected in the psycholinguistic community.

✉ Audrey Bürki
buerki@uni-potsdam.de

1 Department of Linguistics, Cognitive Sciences, University of Potsdam, Potsdam, Germany

2 Faculty of Psychology and Educational Sciences, University of Geneva, Genève, Switzerland

---

[1] In this review, the focus is on the *cognitive* mechanisms and representations underlying the generation of speech. The review therefore focuses on within-speaker variation and does not cover differences across individuals or groups (e.g., gender, social class; see, e.g., Babel & Munson, 2014; Docherty & Mendoza-Denton, 2012).

Springer

Despite recent efforts that specifically target aspects of phonetic variability (e.g., Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Buz & Jaeger, 2016; Fink & Goldrick, 2015; Gahl, Yao, & Johnson, 2012; Kahn & Arnold, 2012; Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2010), the relevance of variation data for our understanding of the language production system is clearly not sufficiently recognized. Too few studies have applied the methods of experimental psychology to study the encoding of pronunciation variants; in particular, there is an important gap between the range of variation phenomena reported in other fields of language sciences, and the restricted number of variation phenomena examined in psycholinguistic studies. This contrasts sharply with the large body of experimental studies on the *recognition* of pronunciation variants that have been conducted since the late nineties (e.g., Ernestus, Baayen, & Schreuder, 2002; Gaskell, 2001; Gaskell & Marslen-Wilson, 1996; Kuijpers, van Donselaar, & Cutler, 1996; McLennan, Luce, & Charles-Luce, 2003; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009; Ranbom & Connine, 2007; Spinelli & Gros-Balthazard, 2007; Sumner & Samuel, 2005). Data from these studies have been used to constrain models of spoken word recognition (e.g., Bybee, 2000; Goldinger, 1998; Hawkins, 2003; Lahiri & Reetz, 2002; Nguyen, 2012), sparking fruitful debates about the nature of lexical representations. These data have challenged the long-standing view that word form representations are mere collections of abstract units, and suggested that more detailed information may also be stored in memory (e.g., McQueen, Dahan, & Cutler, 2003).

This review summarizes the state of the art in variation phenomena, integrating findings from different research fields and research lines. Pronunciation variation has been studied in several domains other than psycholinguistics, including phonetics, phonology, and sociolinguistics. The empirical study of variation phenomena has a long tradition in phonetic sciences. In the last two decades, the amount and diversity of empirical data on variation phenomena has further drastically increased in some fields. Part of this increase has been due to the emergence of laboratory phonology (e.g., Cohn, Fougeron, & Huffman, 2012; Pierrehumbert, Beckman, & Ladd, 2000; Pierrehumbert & Clopper, 2010). Whereas "traditional" phonologists mostly rely on intuitions or very restricted corpora to elaborate theories of sound structure, laboratory phonologists consider the collection of empirical data to be essential to this process. Within this community, the study of the acoustic (or articulatory) signal has become a privileged research tool. Empirical studies on variation processes have also benefited from advances in automatic speech processing research and from the greater availability of its tools. Large annotated corpora have been made available and used to conduct large-scale corpus studies (e.g., Adda-Decker & Snoeren, 2011; Bell et al., 2009; Torreira & Ernestus, 2011).

The different threads of research on variation phenomena are usually discussed separately, and a general overview is clearly lacking. Moreover, variation phenomena that bear strong similarities with or are related to one another have been studied under different names and research lines (e.g., reduction, acoustic prominence, and predictability effects). The present review provides a general and integrative view of these phenomena.

This review has three major aims. The first is to provide an overview of the available empirical findings and to discuss how these findings constrain the modeling of the language production system. The second is to highlight the urgent need for further empirical studies on variation phenomena and to pinpoint important open issues that must be addressed in future research. The third aim is to highlight some of the methodological challenges in the study of variation and to sketch potential solutions for future research.

This review complements and builds on previous attempts to question models of spoken word production in light of variation data. Pierrehumbert (2002), for instance, reported on acoustic and articulatory data on word-specific variation and presents a model that can account for these data. Ernestus (2014) reviewed data on acoustic reduction and discusses their implications for both recognition and production. Fink and Goldrick (2015) discussed recent findings on phonetic variation and their implications for the mental lexicon. The present review further builds on these previous efforts to integrate views from different fields in order to get a better view of the language-processing system (see, e.g., Hickok, 2014).

The standpoint taken here is clearly that of a cognitive psychologist, and the discussion is framed in terms of the dominant cognitive models of language production (e.g., G. S. Dell, 1986, 1988; Levelt, Roelofs, & Meyer, 1999). In the remainder of this introduction, I summarize the architecture of the word production system that is commonly and quite consensually adopted in the psycholinguistic literature, and discuss some open and debated issues regarding more specifically the processes underlying the encoding of the sound structure of words (i.e., word forms). Note that the review mainly focuses on variation as it can be observed at the word level (i.e., different pronunciations of the same word) and in error-free speech. Speech errors or variation in sublexical units (e.g., morphemes) are beyond the scope of this review and are only mentioned in passing.

Cognitive processes involved in language production

Data from errors, response times, and neurophysiological measures collected in the last 30 years have allowed us to build precise models of the cognitive processes involved in language production. Several models concur in describing the

cognitive representations and processes involved (e.g., Caramazza, 1997; G. S. Dell, 1986, 1988; Levelt et al., 1999; see Roelofs, 1997, for a computational implementation of the word form encoding process). These models do not agree on all aspects of the word production process, but they do agree on the following view. The generation of utterances involves three major processes: conceptualization, formulation, and articulation. During the *conceptualization* process, the speaker activates the concepts associated with the meaning (s)he wants to convey. The *formulation* process is itself divided into three encoding processes: the grammatical, the phonological, and the phonetic encoding processes. During *grammatical* encoding, the syntactic and semantic properties of the words (often called *lemmas*) are retrieved, syntactic functions are assigned to these representations, and a syntactic frame (an ordered set of word and morpheme slots; see also Bock & Levelt, 1994, or Garrett, 1980) is generated in which the representations are to be inserted (note that in Caramazza, 1997, direct links are assumed between lexical–semantic representations—i.e., concepts—and word form representations). The generation of the sound structure of words, in which I will be particularly interested here, starts with the *phonological* encoding process. Data on speech errors (G. S. Dell, 1986, 1988; Fromkin, 1971; Garrett, 1975; Shattuck-Hufnagel, 1979) pointed early on to the need to posit "frames" and "fillers" during phonological encoding. That is, word forms (i.e., lexemes) are not represented in a holistic way but come in two parts, an unordered collection of sublexical units (fillers) and a corresponding "metrical" structure (or frame). During the phonological encoding process, the frames and fillers are retrieved from the lexicon, and the latter are inserted into the former to form a phonological word or phrase (i.e., *segmental spell-out*; see also Levelt et al., 1999, or Meyer, 1996, for a review of the speech error evidence).

The last stage of the formulation process is that of *phonetic* encoding. At this stage, the speaker must map the abstract phonological word or phrase onto concrete motor programs. The phonetic encoding process has received little attention in the psycholinguistic literature. In Levelt et al.'s (1999) model, for instance, speakers have access to "a repository of gestural scores for the frequently used syllables of the language" (p. 5), or a *syllabary*. These syllable scores (which specify the articulatory gestures and their temporal relationships; Roelofs, 1997) are activated by the phonological syllables in the phonological word. Once the phonetic syllables have been retrieved or computed (for novel or less frequent syllables), free parameters have to be set to define the syllable loudness, pitch, and duration and to specify how the articulatory tasks will be aligned in time. Finally, the production process ends with the execution of the motor programs.

Beyond the general agreement about this architecture, many issues are still debated, and others await empirical support. A first debate concerns the dynamics of the phonological

and phonetic encoding processes—that is, how these processes relate to one another as well as to previous and subsequent processes. Examples of the questions here are, to what extent a given process is fully completed before the onset of the next process or to what extent subsequent processes can influence previous ones (see, e.g., Goldrick, 2006). A second debate concerns the generation/storage of syllabic structure. Whereas some models assume that this structure is stored in the mental lexicon (G. S. Dell, 1988), others (Levelt et al., 1999; Roelofs, 1997) assume that it is computed on the fly, during segmental spell-out. Other open issues concern the phonetic encoding process more specifically. As was mentioned above, few studies in psycholinguistics have examined this process empirically. Levelt's view that the phonetic encoding process consists in the retrieval of stored syllable-sized gestural scores has been little challenged, but it has also received little empirical support. Most of the evidence so far is also consistent with an online assembly of gestural scores that proceeds syllable by syllable. A related issue concerns the exact content of phonetic encoding units (i.e., articulatory targets, perceptual targets, and gestures). Another open issue relates to the interface between the production and perception processes, and in particular to whether individuals' long-term knowledge about how words sound and are pronounced/articulated is shared across modalities. Another fundamental open issue relates to how words are combined to build utterances. Most studies so far in psycholinguistic research have focused on modeling isolated word production, and there is an urgent need to extend these models to larger utterances. Finally, note that the traditional "psycholinguistic" view of the language production system has detractors. Exemplarist accounts (Bybee, 2001; Goldinger, 1998; Port, 2007; see Fink & Goldrick, 2015, for a discussion) question the abstract nature of word form representations. More recently, Hickok (2014) questioned the relevance of the phonological encoding process and proposed a direct mapping between syntactico-semantic representations (lemmas) and syllabic motor programs. The study of variation phenomena has the potential to address many of these open issues and has, in fact, already provided crucial answers to many of them.

The present review is structured as follows. The next section provides an overview of variation phenomena, with two aims: The first is to describe the types of output that models of language production must be able to account for. The second is to draw attention to a wide range of phenomena whose investigation with the methods of experimental psychology has not yet been undertaken, but that have the potential to provide crucial insights into the language production system. Section 3 describes the variables that influence how words are pronounced. Section 4 concerns how the available evidence reported in the two previous sections constrains our understanding of the word production system, as well as the challenges that this evidence raises for current models. It further

highlights important open issues to be examined in future research.

Description and classification of variation phenomena

The terms *variability* and *variation* in pronunciation merely characterize the observation that words are pronounced with variable acoustic and/or articulatory properties. Defined in this way, variability is ubiquitous in language production, as two utterances of the same word, even produced by the same speaker in a similar situation, are never pronounced in exactly the same way if the utterances are examined with sufficiently fine-grained measures. All researchers familiar with variation phenomena will agree that there are different types of variation. For instance, most researchers probably consider that variability in the duration of the vowel in two successive productions of the word *bat* by the same speaker is not exactly the same type of variation as, for instance, the variability in British English between the word *car* produced in isolation ([kɑ:]) versus followed by a vowel-initial word ([kɑ:r]). But this is where the agreement ends. How best to categorize the different types of variability/variation phenomena remains an open and difficult issue.

A mere listing of variation phenomena is hardly informative. Classifications are crucial to finding patterns of consistency in variation phenomena, and these patterns can in turn inform our understanding of the architecture of the language production system. For instance, if it can be shown that some variation phenomena are characterized by categorical changes across variants, whereas others are characterized by gradient changes, the description of word production processes must entail mechanisms that are able to generate both gradient and categorical changes in pronunciation.

In an ideal world, there would be a descriptive classification of variation phenomena, and this could be used to better understand the language production system that generates them. In the real world, however, classifications are not merely descriptive: they tend to depend on the theoretical framework, so that phenomena are classified according to assumed cognitive mechanisms. For instance, the distinction between phonetic and phonological variation phenomena (see, e.g., Spinelli & Ferrand, 2005) assumes that the production system has distinct phonological and phonetic components.

In this review, I adopt the following criteria to classify variation phenomena, in an attempt to be as descriptive as possible. These criteria are not exempt from criticisms, however. They were selected because they build on the properties that are generally used to describe specific variation phenomena in the literature. I first categorize variation phenomena according to whether the phenomenon is *specific* versus *general*. Specific processes only apply to a subset of words (i.e., all words with a schwa, a liaison consonant, a final "r," etc., in a given language). By contrast, general processes potentially apply to any given word in a given language or across languages (e.g., coarticulation), even though they may affect different words to different extents.

Within these two categories, I further distinguish variation phenomena in terms of the *type* of acoustic/articulatory change(s) that can be observed between the pronunciation generally assumed to be canonical and the noncanonical forms (or variants), and in terms of the *nature* of these changes. Types of changes between a canonical form and a variant include insertions, deletions, and substitutions of acoustic/articulatory material. The nature of these changes describes whether the changes are gradient or categorical. *Categorical* in this context means that the pronunciations of a given word, for a given type of change, can be classified in two groups (or possibly more) according to whether or not the change has occurred. By contrast, *gradient* refers to those changes that cannot be classified in this way, because the articulatory/acoustic changes occur along a continuum—that is, to different extents across occurrences. Note that categorical and gradient processes may bear strong similarities with one another in their outputs, and therefore may be hard to disentangle (see also Cohn, 2007, or Flemming, 2001). For instance, most accounts and empirical studies describe the alternation between schwa and nonschwa variants in French as a categorical process (e.g., Côté, 2011; F. Dell, 1985). In addition, all segments, and especially unaccented ones, undergo gradient reduction processes in connected speech. On the basis of a large corpus of French interviews, Adda-Decker, Boula de Mareüil, Adda, and Lamel (2005) report that French vowels were completely deleted in 6% of the tokens. These extreme phonetic reductions have been shown to affect the schwa vowel as well (Bürki, Fougeron, Gendrot, & Frauenfelder, 2011b). Consequently, some of the word tokens in which no schwa can be heard necessarily result from a gradient reduction, irrespective of whether the schwa is also deleted categorically in other word tokens. Similarly, assimilation processes co-occur with coarticulation processes. For instance, as is reported in Cohn (2007), in Sundanese, vowels following nasal consonants are nasalized. This assimilation is considered a categorical process. In addition to this categorical process, gradient coarticulatory nasalization is observed in the same context. Consequently, acoustic traces of the nonassimilated vowels in a subset of assimilated variants may result from extreme coarticulation.

In addition, the empirical classification of variation phenomena as gradient versus categorical involves important methodological challenges. Any two occurrences of the same word differ if they are examined with sufficiently fine-grained measurements, and distinguishing between gradient and

categorical phenomena requires defining, a priori, what counts as meaningful gradient reduction. Moreover, such classification requires that invariance (i.e., the absence of difference) be assessed. Despite these challenges, such classification is useful, as it may inform on the nature of the underlying production processes. I will come back to such methodological issues in section 4.

In what follows, I describe a selected set of studies on a selected range of variation phenomena. The aim of this section is to illustrate the different types of variability that have been reported, as well as the challenges that their description may entail. This review is by no means exhaustive and tends to focus on the most-studied phenomena and languages.

Specific variation processes

As I defined above, specific processes only apply to a subset of words. In this section I provide examples of specific variation phenomena. These examples are further classified according to whether the variants have more acoustic content than the assumed canonical pronunciation (acoustic insertions), less acoustic content (reduction or deletions), or different acoustic content (substitutions).

Acoustic insertions

So-called *linking /r/* in British English is typically described as a specific variation phenomenon involving the insertion of acoustic content (note that this phenomenon is also sometimes described as a deletion process; e.g., Cohen-Goldberg, 2015). The word-final orthographic "r" is not pronounced when the word is used in isolation but tends to be pronounced when the word is resyllabified with the following word (e.g., Durand, 1997). This process is specific, in that it only concerns words with a final "r." The pronunciations with and without [r] are quite consensually assumed to differ in a categorical way—that is, the [r] is either present (i.e., "planned" by the speaker) or absent.

French liaison is another example. In French, some words end in a vowel when they are produced in isolation (e.g., [pə ti] *petit* "small") or precede a consonant-initial word (e.g., [pə tiʃa] *petit chat* "small cat"), but can (or must) be produced with an additional final consonant when they precede a vowel-initial word in certain contexts (e.g., [pətit] as in [pətitan] *petit âne* "small donkey"; see, e.g., Côté, 2011). The liaison consonant is represented in the word's spelling. Here again, there is little discussion in the literature about the fact that the distinction between the variants with and without the liaison consonant is a categorical one (see Nguyen, Wauquier, Lancia, & Tuller, 2007, for empirical acoustic data on liaison consonants and reviews of similar work). French liaison is a specific process in that it only concerns a subset of French words that in

isolation end with a vowel and have a mute final consonant letter.

In Dutch, a schwa can be inserted in some consonant clusters, including /l + noncoronal consonant/ clusters. The word *melk* "milk," for instance, can be pronounced [mɛlk] or [mɛlək]. The nature of the distinction between the two variants has been examined empirically. Warner, Jongman, Cutler, and Mücke (2001) compared the articulatory properties of /l/ in words produced with an inserted schwa, in words with an underlying schwa, and in words with /l/ in coda position and no schwa. They found that the articulation of /l/ before a schwa, whether epenthetic or underlying, was similar (light) and differed from the articulation of this same phoneme in coda position (dark [l]). This suggests that in the epenthetic-schwa and phonological-vowel conditions the preceding [l] is articulated in the same way. The authors took this finding to suggest that the schwa is inserted during the processing of abstract phonological units, and that the two variants of the words thus differ in a categorical way.

As Gick and Wilson (2006) summarized, in many dialects of English, sequences involving a high tense vowel followed by a liquid may be realized with what sounds like an inserted schwa (e.g., *feel* realized as [fi:əl]). The authors discuss findings from other studies, showing that this additional element does not increase the acoustic duration of the syllable (Lavoie & Cohn, 1999) or the articulatory timing within the syllable (Gick & Wilson, 2001). According to Gick and Wilson, these findings suggest that the schwa-like element does not result from the insertion of a phonological vowel. They argue that the insertion of the schwa-like element results from a gestural conflict between opposing tongue root targets. They report ultrasound and acoustic data from two participants that support the gradient, articulatory nature of the insertion.

In all four examples described above, the words that are concerned in the variation phenomenon share common properties. By contrast, in some cases an insertion can be word-specific—that is, can concern one word only. For instance, in English the indefinite singular determiner is realized [ə] in isolation or before consonant-initial words, and is realized [ən] in front of vowels (e.g., Raymond, Fisher, & Healy, 2002). The change between [ə] and [ən] is quite consensually assumed to be a categorical one.

Acoustic deletions/reductions

Specific changes may also lead to the pronunciation of variants with less acoustic material than the assumed canonical form of the word. In French, for instance, words with a schwa in their first syllable (e.g., fenêtre "window") can be produced without the schwa in connected speech ([fnɛtR]). The same applies for monosyllabic clitics (here, the deletion is also present in the spelling; e.g., *le chien* "the dog" vs. *l'âne* "the

donkey," *il me donne* "he gives me" vs. *il m'offre* "he offers me"). Such deletions are traditionally described in categorical terms (i.e., the schwa is either planned or not planned by the speaker; e.g., F. Dell, 1985). Several studies have examined the acoustic properties of schwa words. Bürki, Fougeron, and Gendrot, (2007; see Torreira & Ernestus, 2011, for similar conclusions on the French utterance *c'était* "it was"), for instance, examined the distribution of the duration of the French schwa vowel in a large corpus of broadcasted news, focusing on polysyllabic words with a schwa in their initial syllable. Bürki et al. (2007) found a bimodal distribution of schwa duration with a first mode at 0 ms (a complete absence of voicing and formant[2] structure) and a second mode at 49 ms. They further reported a gap between the first and second modes and concluded that schwa deletion in this language is a categorical process. Several studies, however, also reported evidence that French nonschwa variants retain acoustic and/ or articulatory traces of the schwa vowel (Barnes & Kavitskaya, 2002; Rialland, 1986; but see Côté & Morrison, 2007). Bürki, Ernestus, Gendrot, Fougeron, and Frauenfelder (2011a) suggested that occurrences of nonschwa variants may result from two distinct processes, a specific categorical deletion process and a general gradient reduction process.

Schwa deletion also occurs in other languages, such as Dutch (e.g., Booij, 1995; Kuijpers & van Donselaar, 1998) or English. In Dutch the deletion is generally assumed to be a categorical one; in English, it has been suggested that the nature of the change might depend on the position of the schwa. According to Kager (1997; see also Patterson, LoCasto, & Connine, 2003), for instance, the two variants of post-stress schwa words (e.g., [ˈkam(ə)rə]) are categorically distinct from one another, whereas nonschwa variants of pre-stress schwa words (e.g., [p(ə)ˈteɪtəʊ]) are assumed to result from a gradient reduction process (see also Davidson, 2006, and Bürki & Gaskell, 2012, for psycholinguistic evidence that the two word types are processed differently).

Specific consonants can also optionally be deleted. French words with final consonant clusters in which the second consonant is a liquid (/r/ or /l/) can be produced both with and without the last consonant (Chevrot, Beaud, & Varga, 2000; Kemp, Pupier, & Yaeger, 1980; Laks, 1977). In English, the phonemes /t/ and /d/ are variably deleted in some contexts (Raymond, Dautricourt, & Hume, 2006). Word-final clusters with a /t/–/d/ in final position can be pronounced either with or without the final phoneme (e.g., Labov, 1968; Wolfram, 1969). Although this deletion has been traditionally described as a categorical process, there is some evidence that this might not always be the case. Browman and Goldstein (1992) observed, for instance, that /t/ in sequences such as *perfect*

memory may be inaudible yet are not completely absent in the articulation.

Acoustic substitutions

Several specific variation phenomena involve qualitative rather than quantitative changes. That is, a variant and its canonical form have the same number of segments, but these differ. Assimilation is an example of specific substitution. In assimilation, the word-final or word-initial segment adapts to a phonological feature of the next or of the previous sound (e.g., the English word *run* /rʌn/ can be realized as [rʌm] when followed by a bilabial consonant). As a result, the words have two pronunciation variants that depend on the phonological context in which they occur. Assimilation is present in many languages, but the context that triggers the change and the segments that are concerned in this change differ across languages. In English, for instance, word-final coronals can adopt the place of articulation of the following sound when this sound is labial, dental, or velar (place assimilation, as in the *run* example above). In French, when two consonants of different voicing are produced in a row, the first may become voiced or voiceless, depending on whether the second is voiced or voiceless (e.g., *robe sale* "dirty dress" realized as [Rɔpsal] e.g., Hallé & Adda-Decker, 2011; Snoeren, Hallé, & Segui, 2006). There are many other examples of assimilation across languages—for instance, regressive voice assimilation in Dutch (e.g., Ernestus, Lahey, Verhees, & Baayen, 2006), voice assimilation in Hungarian (Gow & Im, 2004), palatalization of alveolar plosives in the context of front vowels in Bulgarian (Wood, 1996), and place assimilation in Korean (Kochetov & Pouplier, 2008; see Gafos & Goldstein, 2012, for other examples).

Assimilation has been the focus of many acoustic and articulatory studies. Many of these studies have examined whether changes are categorical or gradient (see the review in Ernestus, 2011). In some of the phenomena examined, the assimilation was usually complete (no trace of the original segment in the actual pronunciation; e.g., Kochetov & Pouplier, 2008); in others it was mostly gradual, such as English place assimilation (Nolan, 1992; see also Wright & Kerswill, 1989); both French and English place assimilation in sequences of alveolar (e.g., /s/) and postalveolar (e.g., /ʃ/) sibilants (Niebuhr, Clayards, Meunier, & Lancia, 2011); and French voice assimilation (Snoeren et al., 2006) or regressive voice assimilation in Dutch (Ernestus et al., 2006). In some cases, speakers were found to vary in whether or not they assimilated, and if so, in whether the assimilation was complete or gradual (e.g., Ellis & Hardcastle, 2002).

Assimilations are not the only examples of specific substitutions. In Welsh, as a result of a process called "soft mutation," the pronunciation of many words differs depending on the phonological context. For instance, after the

---

[2] A formant corresponds to a concentration of acoustic energy in a particular frequency band. Vowels are characterized by the values (i.e., frequencies) of their first two (or three) formants.

definite determiner *y*, the words *mam* "mother" and *basged* "basket" are realized (and written), respectively as *fam* and *fasged*. Other substitutions of phonemes occur, for instance, in the determiners of many languages. These determiners have two pronunciation variants (e.g., *a/an* in English, *lo/il* "the" in Italian, *ce/cet* "this" in French). In Welsh, the definite determiner has three variants—*y* /ə/, *yr* /ər/, and *'r* /r/—and its pronunciation depends strictly on the phonological properties of the following word (Hannahs & Tallerman, 2006). The alternation can be reflected in the word's spelling, as in the examples above, but this is not necessarily the case. For instance, in French, the masculine definite determiner is written *le* before consonants, but can be produced as [le] or [l] in this context (e.g., *c'est l'mien* "this is mine").

Flapping is another specific variation phenomenon. It has been described, for instance, in some dialects of English, such as American English, and only concerns a subset of phonemes in a subset of contexts. Flaps are allophones of /d/ and /t/. For instance, when these phonemes occur between vowels word-internally and when the following vowel is unstressed, these phonemes may become flapped; that is, the tongue makes a brief contact with the alveolar ridge (e.g., Turk, 1992; Zue & Laferriere, 1979; for a short review of experimental studies on the phonetic properties of flapped and nonflapped pronunciations, see Herd, Jongman, & Sereno, 2010). Flapped variants are more frequent than [t] or [d] variants (Patterson & Connine, 2001), and although the process was long considered categorical, there are data suggesting an incomplete process (see Braver, 2014, for a recent review of this issue).

Notably, all the variation phenomena described in this section were first described as fully categorical phenomena. For most of them, there is indeed empirical evidence suggesting that the variants differ in categorical ways. As we have seen, however, some of them show evidence in favor of both categoriality *and* gradience.

Nonspecific (or general) variation phenomena

In contrast to specific phenomena, general variation phenomena are not restricted to subsets of words or phonological contexts. This is not to say that all words and phonemes are affected to a similar extent. The actual changes may vary depending on the sound structure of the words, and how often these apply may depend on word-specific properties such as lexical frequency. General, like specific phenomena may also involve deletion-like, addition-like, and substitution-like processes.

Acoustic deletion/reduction

In spontaneous speech, words are often produced with a reduced content. Unlike the specific schwa deletion phenomena discussed above, these acoustic reductions are not restricted to a subset of words or phonological contexts. A review of acoustic reduction phenomena can be found in Ernestus (2014; note that the term *acoustic reduction* in Ernestus encompasses both general and specific deletion phenomena). Acoustic reduction may lead to the complete deletion of segments or syllables (e.g., Johnson, 2004). Acoustic reduction may also refer to more subtle changes, such as a shorter duration of segments (e.g., Bell et al., 2009; Kahn & Arnold, 2012), centralization of vowels (i.e., the first [F1] and/or second [F2] formant values get closer to the center of the two-dimensional space formed by the F1/F2 values of all the vowels; e.g., Buz & Jaeger, 2016; van Bergem, 1993; van Son, Bolotova, Lennes, & Pols, 2004), or weakening of consonants (i.e., consonants become more vowel-like/less consonant-like; Ashby & Maidment, 2005; Trask, 2000; see also van Son & Pols, 1999).

Many empirical studies have examined reduction phenomena, and in particular the variables that influence reduction. These will be discussed in section 3. Acoustic reduction is generally thought of as a gradient process (e.g., Ernestus, 2014). There is some evidence, however, that some reductions in some words may take a categorical form (e.g., Torreira & Ernestus, 2011).

Acoustic prominence/enhancement

Reduction, as defined in the previous paragraphs, is a relative concept. The definition of acoustic reduction implicitly suggests that there is an unreduced pronunciation of the word that reduced forms can be compared with. But consider, for instance, segment duration. What would be a canonical duration for a given word? This issue becomes particularly relevant when looking at the literature on acoustic prominence. According to Arnold and Watson (2015), variation in acoustic prominence, at least in English, involves variation in duration, pitch, and amplitude. Studies on acoustic prominence often examine, among other cues, segment or word duration (e.g., Lam & Watson, 2010; Watson, Arnold, & Tanenhaus, 2008). In this context, a prominent word is a word with particularly long segments. Implicitly, prominent words are thus more prominent than a canonical word form. Acoustic prominence, at least when defined with segment or word duration, can therefore been seen as the other end of the reduction continuum (see Jaeger & Buz, 2016, for a similar discussion). Like acoustic reduction, acoustic prominence is usually described as a gradient phenomenon.

Acoustic changes

In connected speech, the articulation of all segments is influenced by the articulatory properties of neighboring sounds. This "local articulatory adjustment of all phoneme instantiations to their current neighbors" (Wood, 1996, p. 139) is

known as *coarticulation*. Coarticulation occurs both within and between words and is generally viewed as a consequence of the biomechanical properties of the articulators (e.g., Niebuhr et al., 2011). As a consequence, coarticulation is not optional, and the fine acoustic and articulatory details of any given word will depend on the context in which the word is produced. The degree of coarticulation is known to depend on speech rate (e.g., Gay, 1978). In practice, it may be difficult to disentangle coarticulation from assimilation phenomena on the basis of the sole acoustic signal or to find clear criteria to distinguish between the two phenomena. On this ground, some authors consider that the two should not be conceptually distinguished (Flemming, 1995; Ohala, 1993). It is generally agreed upon, however, that the two phenomena differ in their domains of application, with coarticulation being obligatory in any context, but assimilation only applying in restricted contexts (see also Hoole, Nguyen-Trong, & Hardcastle, 1993, for evidence suggesting speaker-specific coarticulation patterns).

Variable variation

The variation phenomena reviewed in this section reveal that variation can take many different forms. Obviously, models of language production must be able to account for all kinds of variability, not just a subset of them, and it is unlikely that a single mechanism can account for all of them at once. The understanding of the language production system requires that the existing evidence be taken into account, but also that additional data be collected. Many variation phenomena, especially in less studied languages, remain to be documented. Moreover, additional data are needed on those phenomena that have already been examined. Some phenomena have only been described in one or two studies, and the methodology or analyses used in these study have not always necessarily been sound (i.e., very few participants or items, lack of statistical analysis). Finally, and as was mentioned already, specific efforts should be undertaken to examine variation phenomena with the methods of experimental psychology. Information about processing times provide crucial complements to articulatory and acoustic data, but so far have been obtained for only a very restricted subset of variation phenomena.

In the next section I describe another aspect of variation—that is, the variables that influence how a given word is pronounced. As we will see, some of these variables influence specific rather than—or more than—general phenomena, or vice versa. Knowing which variables influence the pronunciation of words, and in what ways, is just informative about the architecture of the language production system as the descriptions of variation phenomena and their properties.

Factors that influence variability in pronunciation

Knowledge about the determinants of variation is a crucial source of information to model the language production system, including the nature and organization of linguistic knowledge in memory, or the encoding processes by which a word is encoded, and their time courses. Interestingly, many of the determinants of variability also influence the speed with which a word or word variant is encoded, as measured in carefully controlled psycholinguistic experiments. This joint influence provides particularly interesting insights into the mechanisms underlying the production of these words or variants.

Influence of lexical information

Information load and phonological neighborhood

There is abundant evidence that word-specific properties influence articulation. Baese-Berk and Goldrick (2009) refer to this phenomenon as "lexically-conditioned phonetic variation" (see also Fox, Reilly, & Blumstein, 2015). For instance, studies converge to suggest that the detailed acoustic realization of a given word is influenced by the relationship between this word and other words in the speaker's lexicon. Van Son and Pols (2003) found that the information load of a given segment in a given word token—that is, the amount to which this segment contributes to reducing the set of possible candidates—influences the actual acoustic realization of the segment. Segments with higher information load tend to be longer and hyperarticulated. In a similar vein—and as reviewed, for instance, in Buz and Jaeger (2016) or Gahl et al. (2012)—the pronunciation of a given word is influenced by the number of similar-sounding words in the mental lexicon (i.e., its phonological neighborhood; e.g., Munson & Solomon, 2004; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013; Scarborough, 2010; Wright, 2004; see Caselli, Caselli, & Cohen-Goldberg, 2016, for evidence that the phonological neighborhood of inflected words [e.g., *eating*] influences the duration of monomorphemic words). Many authors operationalize this measure by counting the number of words that differ by one phoneme from the target word (i.e., neighborhood density; Luce, 1986). Other studies have observed acoustic differences between words with and words without a minimal-pair phonological neighbor (i.e., words that differ from the target word by one phoneme—e.g., *pox–box*; Baese-Berk & Goldrick, 2009; see also Clopper & Tamati, 2014).

Findings have diverged as concerns the exact direction of this influence. On the one hand, several studies have reported that words with more phonological neighbors or a minimal-pair neighbor are hyperarticulated when compared to similar words in sparser neighborhoods or without a minimal-pair

neighbor (e.g., greater vowel dispersion: Munson & Solomon, 2004; Scarborough, 2010; Scarborough & Zellou, 2013; Wright, 2004; longer voice onset times (VOT) for some stop consonants: Baese-Berk & Goldrick, 2009; Peramunage et al., 2010; or longer word durations: Buz & Jaeger, 2016; Scarborough & Zellou, 2013). Other studies have found the reverse—that is, a reduced pronunciation for words in denser neighborhoods (e.g., Gahl & Strand, 2016; Gahl et al., 2012; see also Clopper, Mitsch, & Tamati, 2017, for null effects). It has further been suggested that the influence of phonological neighborhood might depend on the phone being measured and on its position in the word. Goldrick, Vaughn, and Murphy (2013) observed no difference in mean VOTs or mean rates of prevoicing for words with voiced initial stops with and without a minimal-pair neighbor. By contrast, for word-final stops, they reported a shorter duration for words with a minimal-pair neighbor than for words without, but this was true only for voiced final stops, not for voiceless stops. In their analysis of VOT duration in a corpus of conversational English, Nelson and Wedel (2017) observed shorter VOT for voiced word-initial stops and longer VOT for voiceless word-initial stops for words with a minimal-pair neighbor.

The discrepant findings regarding the influence of phonological neighborhood are likely to be related to a major methodological challenge in the study of lexical influences on word production. The way that words are articulated is influenced by many variables at once. When attempting to determine the influence of a single variable, there is a major risk of overlooking potential confound variables. Some confound variables in studies of phonological neighborhood have already been pointed out, including lexical frequency, segmental context (see Gahl, 2015; Gahl et al., 2012), and stimulus set (Buz & Jaeger, 2016). Other variables known to influence articulation, however, have not been considered yet; these include the word's age of acquisition and phonotactic variables such as digram or trigram frequency. Differences in the direction of neighborhood effects might also be related to the materials used for the analyses (e.g., isolated words vs. sentences, laboratory speech vs. corpora).

Notably, phonological neighborhood not only influences articulation, it also affects the duration of the encoding process (see Sadat, Martin, Costa, & Alario, 2014, or Buz & Jaeger, 2016, for recent reviews). As is discussed at length in Sadat et al. (2014), however, the results have (again) been inconsistent across studies. Some have shown longer speech onset latencies (i.e., the time interval between the onset of picture presentation and the onset of articulation) in picture-naming tasks for words with many phonological neighbors (Buz & Jaeger, 2016; Vitevitch & Stamer, 2006), whereas others have reported shorter latencies for these words (Vitevitch & Sommers, 2003) or null effects (e.g., Gordon & Kurczek, 2013). Here again, the lack of appropriate control of some potential confound variables may explain the discrepancies.

In an effort to document the role of phonological neighborhood on speech production latencies further, and potentially to resolve some of the existing discrepancies, Sadat et al. reanalyzed the datasets from previously published experiments using statistical models that would allow for better control of the covariates, and analyzed additional data from a large-scale picture-naming experiment. The findings converged to suggest that denser phonological neighborhoods tend to result in longer speech onset latencies.

Lexical frequency

Another lexical property found to affect the acoustic realization of words is lexical frequency (e.g., Gahl, 2008, on comparing homophones with different frequencies; a comprehensive review can be found in Bell et al., 2009). *Lexical frequency* refers to how often a word is used in the language and is often operationalized as the number of occurrences per million. Frequent words are produced with shorter durations (e.g., Bell et al., 2009; see also Tremblay & Tucker, 2011, on utterance duration) and tend to be produced with more contracted vowels (e.g., Munson & Solomon, 2004) and more segment deletions (e.g., French schwa deletion: Hansen, 1994; Racine & Grosjean, 2002; English schwa deletion: Patterson et al., 2003; English t/d deletion: Jurafsky, Bell, Gregory, & Raymond, 2000; Dutch schwa deletion: Hinskens, 2011; /r/ deletion in nonrhotic varieties of English: Cohen-Goldberg, 2015; see also Bybee, 2001). Lexical frequency has also been shown to influence the realization of morphemes. For instance, in a study on Dutch, Pluymaekers, Ernestus, and Baayen (2005) reported a negative correlation between lexical frequency and the duration of three of the four affixes (in terms of both total duration and duration of the individual phones) examined in the study. Caselli et al. (2016) examined the durations of the -ing, -s, and -ed suffixes in English verbs and observed that the frequency of the whole word and that of the root independently predicted the duration of the suffix. As was the case for phonological neighborhood, lexical frequency influences the speed of the word encoding process, with shorter latencies for more-frequent words (e.g., Alario et al., 2004; Barry, Morrison, & Ellis, 1997; Ellis & Morrison, 1998; Jescheniak & Levelt, 1994; Mousikou & Rastle, 2015).

Notably, several studies have also reported evidence that the frequency of pronunciation variants influences both the acoustic realization of these variants and the length of the encoding process. Racine and Grosjean (2002) found that in French, schwas in words in which the schwa is often deleted have a shorter duration than schwas in words in which deletion is less frequent. Bürki, Ernestus, and Frauenfelder (2010) found that picture-naming latencies for French schwa words correlate with variant frequency. The influence of variant frequency on encoding times suggests that information about the

frequency of these variants (and therefore about the variants themselves) is encoded in the mental lexicon (see Bürki & Gaskell, 2012, for converging evidence).

The studies reviewed above converge to reveal that variation in word production emerges at least partly as a consequence of the organization of the lexicon. This further suggests that during the word production process, information at the lexical level influences the articulation. To account for the influence of lexical information on word pronunciation, the architecture of the word production system must allow information at higher processing levels to be passed on to lower levels, and to do so in a gradient way.

Sublexical frequency

There is some evidence that syllable frequency influences how words are actually pronounced. Herrmann, Whiteside, and Cunningham (2009) reported, for instance, that monosyllabic words that consist of highly frequent syllables are produced with greater coarticulation and a shorter duration than are words that consist of low-frequency syllables. Schweitzer and Möbius (2004) measured the durations of syllables and of their constituents in a large corpus of read speech. They reported that the relationship between the duration of syllables and their constituent phones is influenced by syllable frequency, and that this relationship is stronger for low-frequency syllables. Interestingly, several studies have reported that words with high-frequency syllables are produced with shorter response latencies (e.g., Cholin, Levelt, & Schiller, 2006; Laganaro & Alario, 2006; Levelt & Wheeldon, 1994).

Influence of segmental context

The pronunciation of words is influenced by the properties of the surrounding words, and in particular by the phonological properties of these words. A first important source of variability in how many words are pronounced is the resyllabification process. In connected speech, the actual syllabic structure of many words depends on the phonological properties of the preceding and/or following word(s). In many languages, word-final consonants produced before a vowel-initial word are very likely to attach to the next word and to be realized as a syllable onset. For instance, in English, the words *an* and *idiot* are produced, respectively, with the syllable structures *an* and *i-diot* when uttered in isolation. When produced together, they are resyllabified as *a-ni-diot*. It has been shown repeatedly that the phonetic and articulatory properties of segments depend on their position in the syllable (see Fougeron, 1999, for a review). As a consequence, the pronunciation of many words varies depending on whether the next word starts with a vowel or a consonant.

The segmental context may also call for variants that differ from the canonical form in more than fine phonetic details.

For instance, in English the indefinite determiner is realized (and spelled) *a* before consonants and *an* before vowels (e.g., Gaskell, Cox, Foley, Grieve, & O'Brien, 2003, or Raymond et al., 2002, for empirical data; see also Skousen, 1989). In Italian, the masculine definite singular determiner is pronounced (and spelled) *l* before vowels; *lo* before /ʃ/, /s + consonant/, /gn/, /ts/, and /dz/; and *il* in all other cases. Similarly, in Welsh the definite determiner has three variants—*y* /ə/, *yr* /ər/, and *'r* /r/—and its pronunciation depends strictly on the phonological properties of the following word (Hannahs & Tallerman, 2006; see also Caramazza, Miozzo, Costa, Schiller, & Alario, 2001, for similar examples in French or Spanish). In the same language, as a result of the "soft mutation" process described above, the pronunciation of many words differs depending on the phonological context.

In all the examples cited above, the constraints operate in a systematic way. For other variation phenomena, the phonological constraints are less systematic. For instance, the English definite determiner tends to be pronounced [ðɪ] when followed by a vowel-initial word, and [ðə] when followed by a consonant-initial word. The two pronunciations are, however, found before both vowels and consonants (e.g., Raymond et al., 2002). Similarly, although French schwa and nonschwa variants can occur in any context, schwa variants are preferred when the previous word ends in a consonant or consonant cluster (e.g., F. Dell, 1985; Grammont, 1914). Although the contexts of the application of assimilation phenomena are highly constrained (e.g., in English, only coronal consonants assimilate to the place of articulation of the next sound), the assimilation is not systematic in these contexts (e.g., a coronal consonant does not systematically assimilate before a labial or velar consonant).

Psycholinguistic studies have investigated the production of determiners whose realization is constrained by their phonological context in several languages. They have found that the phonological properties of the surrounding words influence both the length of the encoding process and the fine phonetic realization of the determiners. Alario and Caramazza (2002) compared the speech onset latencies when participants named determiner + adjective + noun phrases in which the adjective and noun called for the same determiner form (i.e., consistent utterances; e.g., *ma nouvelle fenêtre* "my new window") to speech onset latencies for utterances in which the two words called for different determiner forms (inconsistent utterances; e.g., *ma nouvelle amie* "my new friend"). With the French determiners *ma* "my" (*mon* before vowels) and *ce* "this" (*cet* before vowels), they found shorter naming latencies for consistent noun phrases (i.e., a phonological consistency effect; see Miozzo & Caramazza, 1999, Exp. 5, for similar evidence with *lo/il* in Italian; Spalek, Bock, & Schriefers, 2010, for English *a/an* and *the/thee*; or Bürki, Laganaro, & Alario, 2014, for additional evidence in French). Bürki, Frauenfelder, and Alario (2015)

replicated this finding for the French determiner *un* (produced with a liaison consonant before vowel-initial words). In addition, they showed that the determiner has a longer acoustic duration in inconsistent utterances. According to most authors, phonological consistency effects suggest that the two variants of the determiner have a corresponding underlying phonological representation in memory. Differences in speech onset latencies between consistent and inconsistent sequences arise because the adjective and noun send activation to the same representation in the former case, and to two different representations in the latter (but see Spalek et al., 2010, for an alternative account, and Bürki et al., 2014, for an extensive discussion). This implies that these determiners have (at least) two representations in memory.

The observation that the segmental context influences the pronunciation of words imposes constraints on the time course of utterance production. Several studies have examined the production of noun phrases with the picture–word interference paradigm. In some versions of this paradigm, participants are asked to name pictures and to ignore distractor words superimposed on the picture. Schriefers (1993) asked Dutch participants to perform such a task using determiner–adjective–noun utterances or adjective–noun utterances. The distractor had either the same gender as the target word or a different gender. Gender-incongruent picture–word pairs were responded to more slowly (i.e., a gender congruency effect). This effect is robust in Dutch and German (La Heij, Mak, Sander, & Willeboordse, 1998; Schiller & Caramazza, 2003; van Berkum, 1997) but has not been found with the same experimental design in Romance languages (Alario & Caramazza, 2002; Miozzo & Caramazza, 1999; Miozzo, Costa, & Caramazza, 2002). According to Caramazza et al. (2001; see also Costa, Alario, & Sebastián-Gallés, 2007), this discrepancy results from different time courses of word selection within noun phrases across languages. In Romance languages, the phonological content of the noun must be accessed in order to select the appropriate determiner form. Determiner phonological form selection is therefore delayed. When the determiner is actually selected, the activation of alternative forms driven by the processing of the distractor has become too weak to influence the selection of the to-be-produced determiner. Note, again, that this explanation assumes that determiners with two pronunciations have two corresponding representations in long-term memory.

Influence of repetition

Repeated words tend to be acoustically reduced. Fowler and Housum (1987) reported, for instance, that words that have been produced before by the same speaker (*old* words) are produced with reduced duration and amplitude and are harder to identify by a group of independent listeners than words that have not yet been encountered (*new* words). Many follow-up studies have examined the conditions under which this repetition reduction occurs, so as to understand the mechanism(s) underlying this effect (e.g., Bard et al., 2000; Fowler, 1988). For instance, important issues are whether speakers only reduce words that they have produced themselves or whether they reduce words that have been mentioned by others, and relatedly, whether reduction is influenced by what they think their interlocutor knows about which words have been produced previously. Many studies have addressed these issues in an attempt to disentangle two concurring views about repetition reduction effects. According to the "audience-design" account, repetition reduction effects are driven by the speaker's adjustments to the listener's needs. This account is along the lines of Lindblom's (1990) H&H theory, which explains phonetic variation by adaptive processes. Accordingly, speakers would like to reduce the effort they put into articulating as much as possible but are careful not to reduce too much when the reduction would impair intelligibility (see also Ernestus, 2014). Variation thus arises because speakers adapt to their interlocutors' needs. Alternatively, according to the "speaker-processing" account (or production facilitation account), reduction is driven by speaker-internal production processes. Importantly, the two mechanisms are not mutually exclusive. A review of the evidence in favor of each view is found in Kahn and Arnold (2015). For instance, Galati and Brennan (2010) compared words in a story told for the second time either to the same addressee or to a new addressee and found that the speech was more reduced in the first than in the second condition. This finding suggests that reduction is at least partly driven by adjustments of the speaker to the listener's needs (but see Bard & Aylett, 2004). Several studies have further reported evidence supporting the idea that reduction occurs because processing is facilitated. For instance, Bard et al. (2000) reported reduction for repeated words even when the speaker was aware that the listener had not heard the first mention. Kahn and Arnold found that instructions to listeners showed reduction after an auditory prime, irrespective of whether the speaker thought the listener had heard the prime.

Studies have also examined the locus of the speaker-driven repetition effect. Facilitation could in principle originate at any level of the word production process—that is, at the conceptual level, the lemma level, the phonological level (either lexeme access or metrical spell-out), the phonetic encoding level, or during articulation (see also Jacobs, Yiu, Watson, & Dell, 2015; Kahn & Arnold, 2015). Kahn and Arnold (2015) further found that saying the word or having heard the word previously led to similar amounts of reduction, whereas simply thinking about the word did not result in as much reduction as having previously articulated the word. This suggests that the repetition reduction effect does not merely originate in articulatory processes. Lam and Watson (2014) reported evidence that the repetition of a given word led to reduced

duration and intensity when compared with the first mention of the word, whereas repetition of the referent without repetition of the word led to a reduction in intensity but not in duration, suggesting that part of the repetition effect originates at the lexical (lemma or lexeme) level. Jacobs et al. (2015) further addressed this issue by comparing the reduction of second mentions of words, following a first mention that was spoken either aloud or silently, and within the latter condition, they compared first mentions that involved mouthed versus unmouthed inner speech. Reduction in acoustic duration was only found in the speaking-aloud condition. In another experiment, they found reduction after previous naming of homophones. In an attempt to interpret their findings in combination with the finding that previously hearing a phonological form generates reduction, they suggested that repetition reduction might be at least partially driven by auditory feedback: The more efficient the feedback, the faster the production process and the shorter the acoustic duration. The efficiency of feedback might be improved by previous auditory experience with the phonological form of a word.

If repetition occurs because processing is facilitated during a given encoding process, reduced productions should also be produced with shorter encoding times. Repeated words tend to have shorter encoding latencies (e.g., Cave, 1997; Mitchell & Brown, 1988). A review of the empirical evidence on repetition priming effects in chronometric and neuroimaging studies may be found, for instance, in Francis (2014), who concluded that repetition priming effects do not arise in articulation processes. Similar repetition effects are indeed found when the prime is not overtly articulated (see, e.g., Ferrand, Humphreys, & Segui, 1998). According to Francis, however, the extent to which repetition priming arises as a result of facilitated lemma or lexeme access remains to be determined.

Influence of the word's predictability

The pronunciation of a given word is influenced by how predictable the word is, given the context[3] (see Bell et al., 2009, and Jaeger & Buz, 2016, for reviews). Various measures of predictability have been found to correlate with acoustic reduction (the probability of a syllable given the previous syllable and given the previous two syllables: Aylett & Turk, 2006; the joint probability of a word given previous and/or following words: Jurafsky et al., 2000; the conditional probability given the previous or following word: Bell et al., 2009, and Jurafsky

---

[3] Note that lexical frequency effects and repetition effects are sometimes considered measures of predictability. This makes sense, given that more frequent words and words that have already been used in the preceding utterances are more probable. In this review I discuss lexical frequency, repetition, and predictability influences on pronunciation separately, since it is not yet clear whether these effects can fully be accounted for by the same mechanisms. Moreover, there is empirical evidence that these effects may be independent of one another. Bell et al. (2009) found, for instance, that the effect of repetition is independent of that of word frequency or predictability.

et al., 2000; and mutual information: Bell et al., 2009, and et al., 2011a). It has further been shown that word duration is influenced by how predictable a word is in a given context. Seyfarth (2014) examined the influence of word informativity, defined as "the average of a word's bigram probability across the contexts that it occurs in, weighted by how frequently it occurs in each of those contexts" (p. 144), with the context being either the following or the previous word. He found that the pronunciation of words is influenced by whether the words are predictable versus unpredictable in a given context. Importantly, this effect is independent of local contextual predictability (e.g., the conditional probability of a word given the previous or following word), word frequency, or number of phonemes. In short, words that occur more often in predictable contexts tend to be reduced, whether they occur in predictable or unpredictable contexts.

A word can also be more or less predictable given the larger semantic or syntactic context. Lieberman (1963), for instance, found that words in sentences were produced with less stress and were less intelligible when they were more predictable given grammatical or semantic information (see also Clopper & Pierrehumbert, 2008).

As was reviewed in Jaeger and Buz (2016), more-predictable words or syllables tend, for instance, to be produced with shorter acoustic durations (but see Kuperman, Pluymaekers, Ernestus, & Baayen, 2007), more centralized vowels, more weakening of consonants, segment or word deletions, and more morphological contractions. The consequences of predictability on word pronunciation are not restricted to reduction, however. Bybee (2001) discussed, for instance, optional liaison in French and reported evidence showing that frequent utterances tend to be produced more often with a liaison consonant than are less-frequent word combinations.

Three accounts are regularly called on to explain predictability effects: the production facilitation account and the audience design account reviewed above, and the exemplarist (or representational) account. As we will see in detail in section 4, the exemplarist account assumes that speakers store detailed phonetic exemplars of the speech of their interlocutors and that these exemplars contain information about the contexts in which the words were previously produced. Which of the three accounts best explains the available data is still under debate. I refer the reader to Jaeger and Buz (2016) for an extensive discussion, and will come back to this issue in section 4.

Influence of phrase frequency

A few studies have set out to examine whether phrase frequency influences the acoustic properties of the words within the

phrase or of the whole phrase (see Arnon & Cohen Priva, 2013, for a review). Tremblay and Tucker (2011), for instance, examined the contributions of word and sequence frequencies to utterance duration. They found that *N*-gram frequencies (two, three, or four words) were a poor predictor of utterance duration, whereas lexical frequency accounted for a large part of the variance in utterance duration. By contrast, Arnon and Cohen Priva (2013) observed that the higher the phrase frequency of three-word sequences, the shorter the duration of the sequence. This was true in both controlled elicited productions and corpus data, after controlling for lexical frequency. Moreover, the same effect was found whether the phrase consisted of one constituent or crossed constituents. In a subsequent study, Arnon and Cohen Priva (2014) showed that the effect of phrase frequency was not restricted to the final word of the utterance and was present over and beyond the influence of word predictability. These authors further observed that whereas the effects of word and phrase frequency were present for high-frequency phrases, the effect of lexical frequency decreased for these phrases (as compared with less frequent phrases), while the effect of phrase frequency increased.

Notably, phrase frequency has also recently been found to influence production latencies. For instance, Janssen and Barber (2012) found that frequent noun phrases are encoded faster than less-frequent noun phrases. In Henrix, Bolger, and Baayen (2017), phrase frequency was not a significant predictor of the naming times but did influence the event-related potentials. The authors reported a long-lasting, nonlinear effect, with high-frequency phrases being overall more negative than low-frequency phrases.

Influence of the speech of interlocutors

In recent years, an increasing number of studies have examined the extent to which a speaker's speech is influenced by previous utterances from his or her interlocutors. This issue is particularly interesting because it can provide information on the relationship between the production and perception systems, as well as on the nature of the representations used in production and perception.

Several studies have reported that speakers tend to adapt their pronunciation to that of their interlocutor. This phenomenon is termed *phonetic imitation* or *phonetic convergence* and occurs in natural conversations (e.g., Kim, Horton, & Bradlow, 2011; Pardo, 2006; Pardo, Jay, & Krauss, 2010), as well as in repetition (or shadowing) tasks. Goldinger (1998), for instance, asked participants to produce isolated words and nonwords and then to repeat the same stimuli immediately upon hearing them (immediate shadowing task) or after a short delay (delayed shadowing task). He then asked a second group of participants to judge the degree of imitation between the stimuli that the first group heard during the shadowing task (i.e., the model) and tokens of the same stimuli they had produced before and during the shadowing task. He found that tokens registered during the immediate shadowing task were judged to be more similar to the models than were tokens registered before the shadowing task. Moreover, the imitation was stronger for low-frequency words in both the immediate and delayed naming tasks. Goldinger and Azuma (2004) recorded baseline productions of isolated words (reading task) by a group of participants. They then exposed these participants to tokens of the same words produced by four different speakers. One week after exposure, they asked their participants to read the words again. They observed a strong imitation effect (as judged by independent speakers). Imitation was again stronger for low-frequency words and increased with repetition (but see Pardo et al., 2013; Shockley, Sabadini, & Fowler, 2004).

Several studies have attempted to determine the acoustic correlates of perceptually attested imitation (e.g., Babel & Bulatov, 2012; Shockley et al., 2004), but no clear picture has yet emerged (see also Gambi & Pickering, 2013). Pardo et al. (2010) found that tokens judged to be perceptually similar did not necessarily show similarity in articulation rates or vowel spectra. Pardo et al. (2013) reviewed the acoustic features examined in several imitation studies and found "a great deal of variability and inconsistencies across measures" (p. 184). In the same article, the authors reported that none of the acoustic features they examined (duration, vowel formants, and fundamental frequency) was a significant predictor of perceptual judgments of imitation, but that the combination of all three features did predict these perceptual judgments.

Several studies have also reported acoustic evidence suggesting that speakers imitate specific phonetic features of their interlocutor's speech. In these studies, and unlike in the phonetic imitation studies described above, specific acoustic features in the interlocutor's speech differed consistently from the speaker's own productions. Nielsen (2011) manipulated the duration of the VOT of /p/-initial stimuli and found that speakers imitated this duration when producing the same words, novel /p/-initial words and—to a lesser extent—novel /k/-initial words (see also Fowler, Brown, Sabadini, & Weihing, 2003). Dufour and Nguyen (2013) asked Southern French speakers to repeat or imitate the productions of a Standard French speaker. Whereas in Standard French there is a distinction between open and closed /e/ word-finally, in Southern French, words all end with a closed /e/. Acoustic measurements revealed that when shadowing words with a final /e/ produced by a Standard-French speaker, participants converged toward the vowels of the model. Mitterer and Ernestus (2008) manipulated the amount of prevoicing in Dutch stimuli with voiced stops (i.e., no prevoicing and 6 or 12 cycles of prevoicing). They observed that the presence versus absence of voicing was imitated, but that the length of the prevoicing was not. Tilsen (2009) asked participants

to produce isolated vowels, which were sometimes preceded by vowels whose formants had been centralized. Speakers tended to produce more centralized vowels following those primes. Note that other studies have suggested a more nuanced picture. For instance, Kraljic, Brennan, and Samuel (2008) found that whereas listeners change their perceptual representations to accommodate the specificities of their interlocutor's speech (see also Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003), these perceptual shifts do not necessarily become apparent in the listeners' own later productions.

Finally, at least one study has reported that participants tended to reproduce the pronunciation variants heard in the speech of others. Brouwer, Mitterer, and Huettig (2010) had participants shadow sentences from a corpus of casual speech containing both canonical and noncanonical (reduced) pronunciations. They found that participants tended to produce a canonical (unreduced) pronunciation more often (88% of the time) than a reduced one. However, the probability of a reduced variant was higher when the model was a reduced as opposed to a full pronunciation. In addition, the acoustic duration was longer after canonical than after noncanonical models.

Influence of communication/conversational context

The way a given speaker pronounces words is further influenced by contextual factors such as social variables (i.e., status of interlocutor, degree of formality, etc.). Several studies have reported variation in pronunciation depending on speech style. Labov (1966; see Babel & Munson, 2014, for a discussion and further studies), for one, studied the presence of postvocalic /ɹ/ in New York English, varying in different types of speech (casual speech, careful speech) and tasks (reading, lists of words, etc.), and reported different rates of /ɹ/ presence depending on these dimensions. Studies on phonological variation phenomena also highlighted from early on the role of stylistic variables. In French, for instance, the rate of schwa presence depends on speech style (Léon, 1971), register (Verluyten, 1988), conversation topic, or degree of formality (Malécot, 1976).

Moreover, studies have shown that the degree of imitation (convergence) is influenced by social variables such as the desire to be socially accepted (e.g., Natale, 1975), the social status of the interlocutor (Gregory & Webster, 1996), or the sex of the interlocutor (Pardo, 2006). Babel (2010) further found more imitation when speakers have a greater affinity with their interlocutor.

Another line of research has reported that speakers modify the way they articulate words when words or word parts are misunderstood (or likely to be misunderstood) by their interlocutors (e.g., Oviatt, Levow, Moreton, & MacEachern, 1998; Schertz, 2013; Seyfarth, Buz, & Jaeger, 2016; Stent, Huffman,

& Brennan, 2008). Finally, the influence of word-specific properties discussed above, such as phonological neighborhood, further depends on task or stimulus set (Baese-Berk & Goldrick, 2009; see Fink & Goldrick, 2015, for discussion).

Integrating variation phenomena in modeling the language production system

The aim of this section is to bring together the empirical findings discussed in the previous sections and to discuss how these findings constrain our understanding of the word production system. Existing studies on variation phenomena usually aim to account for a given phenomenon or set of phenomena (e.g., repetition reduction effects, word-specific effects, imitation) within a given model/proposal. This section goes a step further by considering a wide range of empirical observations jointly. It further considers different issues, including the nature and content of word form representations, how the information flows in the system, the interface between production and perception, and the production of multiword utterances. Methodological concerns and challenges, as well as open issues, are further discussed.

The review shows that variation is not a homogeneous phenomenon. Some word variants differ from one another in a mostly categorical way, others differ from one another in a mostly gradient way. Moreover, some variants involve the deletion or insertion of acoustic materials (when taking as a baseline a neutral pronunciation in isolation), others involve substitution of acoustic materials. Some of the changes are general, others are word-specific. Changes in pronunciation for a given word may furthermore be strongly or less constrained by the context. Finally, many variables influence how a word is pronounced. In considering models and theories of word production, this heterogeneity must be accounted for. This will necessarily result in a complex model in which different mechanisms may be necessary to explain different types of variation, and in which the pronunciation of a given word will not necessarily result from only one of the mechanisms.

Nature and content of representations

To produce words, speakers must retrieve information in their long-term memory about the properties of the words. In this discussion I focus on *word form* representations and assume that during the word production process these representations receive activation from higher-level representations (lemmas; see, e.g., Levelt et al., 1999).

Questions about the nature of word form representations are often framed in the abstractionist-versus-exemplarist debate. According to the abstractionist view, implemented in traditional psycholinguistic models of word production

(Caramazza, 1997; G. S. Dell, 1986, 1988; Levelt et al., 1999), representations of the sound structure of words are made of abstract units. Moreover, word form information is stored at two representational levels, the phonological and phonetic levels. As was described in the introduction, information at the phonological level is stored in two parts, the *fillers* and the *frame*. The fillers are most often assumed to be phonemes. At the phonetic level, abstract motor commands are stored, possibly in a syllabic format (e.g., Levelt & Wheeldon, 1994). The same abstract programs are retrieved, for instance, when articulating the first syllable shared by the words *carrot* and *café*. Abstractionist models are inherited from the generative framework (Chomsky & Halle, 1968) and have long dominated the field of phonology. In this context, many variation phenomena are explained with so-called phonological rules (e.g., F. Dell, 1985). Accordingly, words with two variants have one phonological form stored in memory (the underlying form), and the other variant (the surface form) is derived via a rule that adds, deletes, or substitutes a phoneme from the underlying form.

According to the exemplarist account (e.g., Bybee, 2007; Goldinger, 1998; Port, 2007; see Kirchner, Moore, & Chen, 2010, for a computational implementation), speakers store detailed phonetic exemplars of each known word. As a consequence, each word is represented by a collection of representations (or exemplars). In such models, there is no need for a separation between phonological and phonetic levels of representations. "Abstraction" may exist in exemplarist accounts as a consequence of the organization of the exemplars. These exemplars are grouped under category labels, which can be specific words (e.g., all exemplars of the word *car* are stored under the label CAR) but could in principle also be different variants of the same word (i.e., a label for the variant with R and another label for the variant without). Moreover, the exemplars are shared between the production and recognition systems. Some models assume the storage not only of words but also of utterances, especially the frequent ones (e.g., Bybee, 2001). Exemplar-based models are most often discussed in the context of spoken-word recognition tasks (e.g., Goldinger, 1998; Johnson, 2004), and they face specific challenges when it comes to explaining the production process. Unlike for recognition tasks, in which a simple mapping can be assumed between the speech input and an exemplar, an exemplarist model of word production is necessarily more complex and must specify how the exemplar is transformed into an articulated signal. Only few attempts to specify how this is actually done can be found in the literature (see Kirchner et al., 2010, for an example). Moreover, as was highlighted by Pierrehumbert (2002), the data from psycholinguistic studies on both speech errors and production latencies clearly show that words are not retrieved as wholes during production, but that their phonological content must be assembled online. If exemplar-based models are to be considered a viable alternative to traditional psycholinguistic models, they must explain how the system can generalize to novel forms, produce errors following the observed patterns (see G. S. Dell, 2014), and explain a wide range of reaction time results, usually interpreted within the abstractionist framework (e.g., phonological facilitation effects in picture–word interference studies [see, e.g., Meyer & Schriefers, 1991] or sequentiality in phonological encoding [Meyer, 1990], to cite just two empirical phenomena). As Baayen, Hendrix, and Ramscar (2013) discussed in the context of multiword utterances, another issue with exemplar-based models is the plausibility of a system with a huge number of lexical representations. If all encounters with a given word result in a memory trace, the number of traces would soon become enormous. Although such storage may be argued to be neurobiologically possible, processing exemplars with such a high number of possibilities would be time-consuming (Baayen et al., 2013; Hendrix, Bolger, & Baayen, 2017).

Pierrehumbert (2002) describes a hybrid model, with both abstract representations and detailed phonetic exemplars. This model avoids many of the limitations of the exemplarist approach described above. This model assumes exemplars at the segmental level. The phonological encoding process and its representations are as in standard psycholinguistic models. The output of the phonological encoding process does not map onto abstract motor programs, however, but "probabilistically evokes regions of the exemplar space as production goals" (p. 121). Each phoneme is associated with a cloud of exemplars, each exemplar being a particular phonetic realization of that phoneme. The most activated exemplar is chosen for production. This model further assumes shared phonetic exemplars for recognition and production tasks. Building on this model, Schweitzer and Möbius (2004) suggested that exemplars can also be syllable-sized. More specifically, the researchers argued that the frequent syllables of a language are represented as phonetic exemplars, whereas infrequent syllables likely have to be computed from exemplar phones (see also Walsh, Schütze, Möbius, & Schweitzer, 2007).

A related issue concerns the status of phonological representations/the phonological encoding process. In traditional psycholinguistic models, word form representations are made of sets of abstract phoneme representations that are assembled during segmental spell-out. In a recent proposal, Hickok (2014) described a model in which there is no intermediate level of encoding at which abstract phonemes are assembled to form phonological syllables. In this model, the "phonological component" actually consists of two components, an auditory-based and a motor-based component. Syntactico-semantic representations (lemmas) send activation directly to (syllabic) motor programs and (syllabic) auditory targets. The syllabic motor programs can be seen as equivalent to the syllable-sized phonetic representations described by Levelt and Wheeldon (1994). Motor programs and auditory

targets are gradient representations (see, e.g., Rapp, Buchwald, & Goldrick, 2014). This model has been criticized in the psycholinguistic literature on the grounds that it cannot account for several empirical findings, including speech error patterns in pathological and nonpathological speech (see Pierrehumbert, 2002, for a similar discussion, and Buchwald and Miozzo, 2011, for empirical evidence). Others have argued that because this model does not allow for the reorganization of syllables as a function of the phonological context, it can hardly explain resyllabification (see Indefrey, 2014, or Roelofs, 2014).

In what follows, I examine these different issues in light of the empirical data reviewed in sections 2 and 3.

Abstraction, categories, and phonemes in mental representations

Variability in the speech signal, such as the presence of reduced forms (e.g., Arnold, Tomaschek, Sering, Lopez, & Baayen, 2017), is often cited as a problem for models in which word form representations are made of abstract phoneme-sized units in speech comprehension. Do the available data on variation challenge the view that the *production* of language involves accessing and ordering phonemes? Several of the empirical findings reported in the present review instead support the alternative hypothesis, that the sound structure of words is represented, at some level of processing, as units that we can describe in terms of an "abstract collection of segments" (e.g., Levelt et al., 1999). A first line of such evidence comes from phonetic studies showing that some words have two categorically different variants (e.g., French liaison, French schwa deletion, and Welsh soft mutation), which differ in one phoneme. A second line of evidence comes from the finding that speakers can rate the relative frequencies of the two pronunciation variants of some words that differ in one phoneme, and that these relative frequencies influence naming times (e.g., Bürki et al., 2010). This finding clearly suggests that the two variants of these words are two different entities in the mental lexicon of the speaker. Thirdly, psycholinguistic evidence on determiners with varying pronunciations (e.g., Alario & Caramazza, 2002; Miozzo & Caramazza, 1999; Spalek, Bock, & Schriefers, 2010) also points to the representation of categorically distinct word forms for these determiners.

These findings suggest that mental representations involve categories that differ from one another in one (or more) phonemes. Note that such categories can be obtained in different ways. In the exemplarist view, one would assume that phonetically detailed exemplars are grouped under category labels that are formed with experience/usage (see, e.g., Wedel, 2007). Similar categories could emerge in the lexicon on the basis of acoustic similarity, without positing the storage of multiple exemplars or the phoneme as a processing unit.

Notably, the relevance of such categories at the word form level confirms the psychological reality of the distinction between phonological and phonetic processes. Data on variation thus complement studies on speech errors or patients (e.g., Buchwald & Miozzo, 2011; see also Buchwald, Rapp, & Stone, 2007; Rapp et al., 2014) in showing that the word production system encodes segments in an abstract form—or in other words, that the contrastive segments of a given language are stored under different category labels.

At least two categorically distinct word form representations for some words

The data on variation further suggest that words with categorical variants may have (at least) two corresponding word form representations in the lexicon. This observation is at odds with the classical psycholinguistic model, since these models assume a single representation per word at the phonological encoding level (Caramazza, 1997; Levelt et al., 1999). These models would need to be extended to allow for duplicate word form representations. Note that Levelt (1989) discussed the possibility that extremely frequent variants (e.g., *don't*) are stored as lexemes. The integration of abstract variants in the lexicon has been considered in related fields. Some phonologists have accounted for variation phenomena by assuming representations for two abstract categorical variants in memory (i.e., allomorphs, in linguistic terms; e.g., Nevins, 2011; Zwicky, 1986). A similar proposal is found in the context of psycholinguistic spoken-word recognition models (e.g., Connine, Ranbom, & Patterson, 2008; Pinnow & Connine, 2014).

Although the inclusion in the lexicon of a second lexeme for some words may appear a relatively small conceptual move, it has nontrivial consequences. Firstly, the inclusion of noncanonical variant representations may ask for a reconsideration of the notion of "phonotactic legality." A given syllable is traditionally considered legal in a given language if it occurs in this language. For example, in French, the syllable /ʃmi/ is illegal because no French word contains that syllable. If the word variant /ʃmine/ (nonschwa variant of the word *cheminée* "chimney") is stored in the lexicon, /ʃmi/ would become a legal syllable in this language. Secondly, adding words in the mental lexicon is likely to impact the processing of other related words by modifying, for instance, the phonological neighborhoods of many words. For instance, if the French nonschwa variant /plɔt/ is represented in the mental lexicon, it becomes a phonological neighbor of the word *plate* "flat" and will in turn influence the production/pronunciation of this word.

Importantly, the psycholinguistic evidence supporting the representation of categorically distinct pronunciation variants in the speaker's lexicon so far has only concerned a restricted number of variation phenomena or words. The scope of this phenomenon remains to be defined. Psycholinguistic studies

have shown the relevance of chronometric paradigms to examining representation issues, and further studies should build on this knowledge to study different phenomena and languages. For instance, the phonological consistency effect described in section 3.2 could be further used to examine other types of variation phenomena (e.g., the reduction of consonant clusters involving liquids in front of consonants; e.g., *notre chat* "our cat" realized as, [nɔtʃa] or assimilation phenomena). Studies would further benefit from a joint examination of online measures (reaction times, electroencephalography, or eye movements) and acoustic or articulatory evidence. A combination of paradigms and measures would allow, for instance, comparing the range of variability/similarity across variants for which online measures suggest a categorical abstract distinction with the range of variability/similarity across variants for which psycholinguistic measures do not show evidence for two representations. A major obstacle in making sense of phonetic variability resides in the difficulty of assessing where meaningful variation ends. Any two occurrences of the same word are likely to vary, at least slightly, on a phonetic dimension. Quantifying the range of phonetic variation across pronunciations that are issued from the same versus from two different phonological representations could help set phonetic criteria to distinguish between categorical and gradient phenomena.

The need to specify the limits of meaningful variability is a pervasive problem in phonetic/articulatory studies. Because the absence of a difference cannot be taken as evidence for similarity, most efforts in language sciences have focused on describing *differences* across variants, resulting in a bias toward variability. Most phenomena once thought to be categorical have at some point shown evidence for gradience. Importantly, however, any variability is not necessarily meaningful. Variability in pronunciation is a useful source of information about the architecture of the word production system, but invariants are just as important to understanding this system. As Dienes (2008) highlighted, invariants are just as crucial and meaningful to understanding the world and testing our theories about it as variance is. The study of variation is difficult because it requires disentangling random from "meaningful" (i.e., from the researcher's perspective) variation.

Empirical phonetics, laboratory phonology, and psycholinguistics, like many fields in the cognitive sciences, are anchored in the frequentist framework. In this framework, it is only possible to test hypotheses about differences. The null hypothesis—that is, the no-difference hypothesis—can only be rejected, it can never be accepted. The scientific literature is therefore biased toward reporting changes and lacks information about invariants. Tools from the Bayesian framework, such as Bayes factors, have recently been made available in regular software. Bayes factors allow for assessing which, of the alternative and the null hypothesis, is more likely given the data (see, e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009, for Bayes factors applied to comparisons between means). Bayesian analyses could provide a useful complement to standard frequentist analyses in the study of variation phenomena.

Beyond monomorphemic words

So far, the discussion of lexical representations has been limited to monomorphemic words. Multimorphemic words (e.g., *eats* and *eating* have two morphemes, the stem "eat" and a suffix; compound words are another example) are beyond the scope of this article, but here note that data on variation may also inform as to the representations and processes underlying the production of these words. As Caselli et al. (2016), for instance, have reviewed, there is ample evidence that listeners store and use both whole-word and morpheme representations in spoken recognition/comprehension tasks involving multimorphemic words, but the evidence from production tasks, especially from chronometric studies, points instead to morphemic representations. This is especially true for inflected forms. The observation (see section 3.1) that lexical and root frequency influence the duration of affixes (e.g., Caselli et al., 2016) or that the number of phonological neighbors, computed from inflected words, influences the monomorphemic word duration (Caselli et al., 2016) is as predicted under the hypothesis that the production of inflected forms involves both whole-word and monomorphemic representations.

Storage of detailed phonetic information?

Evidence from variation data, and in particular from imitation studies, is often taken to suggest that speakers store detailed phonetic features of the speech of their interlocutor, at least temporarily. Goldinger (1998), for instance, accounted for his finding that speakers imitate in shadowing tasks by assuming that speakers store *detailed representations* of the speech of their interlocutor and reuse these word representations during the word production process.

Whereas many aspects of the variation observed in the output signal are consistent with the idea that speakers store phonetically detailed exemplars, the data available to date on variation offer little direct support in favor of this hypothesis. As was noted by a reviewer, although imitation data are as predicted by models in which phonetic details are stored in memory, they do not provide evidence on whether phonetic detail is or is not actually represented (see Linke, Bröker, Ramscar, & Baayen, 2017, for a discussion of the link between behavior and representations). Moreover, under the hypothesis that speakers store detailed phonetic exemplars of the speech of others and reuse these exemplars to plan their own productions, acoustic similarity is to be expected between

the speech of interlocutors. As is evident from the literature review on phonetic convergence (see the further discussion in section 4.3), imitation effects, as quantified with acoustic measures, are subtle and inconsistent. They also tend to differ across speakers (Pardo, Urmanche, Wilman, & Wiener, 2017; see also Yu, Abrego-Collier, & Sonderegger, 2013).

Syllables in word production

Data on variation also inform as to the role of syllables in the representation of word forms. As was mentioned already, an important source of variation in how words are actually pronounced comes from the need to resyllabify in connected speech. The phonetic properties of a given segment will differ depending on its position in the syllable (see Fougeron, 1999, for a review). This suggests, first, that words are not stored as sequences of syllabic chunks (e.g., G. S. Dell, 1988) at the phonological level, but that syllables are computed on the fly in phonological units that may encompass several words (Roelofs, 1997). As was discussed above, the observation that the syllable structure of words depends on the phonological properties of the neighboring words is further taken as evidence in favor of a distinct phonological component.

An interesting goal for further studies would be to investigate whether resyllabification can be triggered by the articulatory properties of the next word onset rather than by the phonological properties of its first phoneme. In the spoken-word recognition literature, Arnold et al. (2017) recently demonstrated that an error-driven learning algorithm, which learns to map acoustic features directly to semantics, can recognize spoken words extracted from a corpus of spontaneous speech. This shows that it is possible to approach spoken-word *recognition* without phonemic representations. One could imagine that a learning algorithm of the kind described in Arnold et al. could learn to associate a grammatical frame to articulatory programs, the articulatory programs for a given word being selected depending on the articulatory programs attached to the next word.

Findings from several studies, mainly by Levelt and colleagues, have suggested that speakers make use of syllable-sized phonetic representations, at least for frequent syllables (Cholin & Levelt, 2009; Cholin et al., 2006; Laganaro & Alario, 2006; Levelt & Wheeldon, 1994; see Carreiras, Mechelli, & Price, 2006, or Riecker, Brendel, Ziegler, Erb, & Ackermann, 2008, for neuroimaging data; see also Hickok, 2014). The observations that the phonetic properties of a given segment differ depending on the position of the segment in the syllable and that coarticulation is greater in frequent than in infrequent syllables provide further support for the idea that syllables play a role during the phonetic encoding process (see also Cholin et al., 2006).

Rules

As was described above, in the generative framework that inspired dominant psycholinguistic models of language production, the production of a subset of variants is assumed to be derived by phonological rules (e.g., F. Dell, 1985; see also Levelt, 1989). Support for the psychological reality of rules is found, among other places, in the literature on speech errors. When segments are exchanged between words, they tend to be pronounced as appropriate not for the context in which they were intended but for the context in which they are actually produced (e.g., *run[z] out* realized as *run out[s]*; e.g., Garrett 1980; see also Stemberger & Lewis, 1986, or Stemberger, 1983, for similar findings at the submorphemic level; see Goldrick, 2011, for a review). In error-free speech, the evidence supporting the psychological reality of rules is scarcer. Before we speculate on this lack of empirical support, we need to ask what would count as evidence that a variant was generated via a phonological rule. A likely prediction of rule-based accounts is that the production of the nonrepresented variant takes more time than the production of the canonical form. This hypothesis may not be easy to test for many variation phenomena, because the variants of a given word differ in many respects, including the contexts in which they are produced or, often, whether or not the variant is resyllabified with the neighboring word. For instance, in Bürki et al. (2010, Exps. 2 and 3), participants took more time to name the nonschwa than the schwa variant. This result would be as expected if nonschwa variants result from the application of a phonological rule (assuming that the application of the rule is costly). However, this result could just as well be due to the fact that no-schwa variants are used less frequently in formal contexts.

The observation that the selection of a given variant is highly constrained by the phonological context (*a/an* in English, *le/l* in French, soft mutation in Welsh, French liaison, etc.) is sometimes taken as evidence in favor of phonological rules (e.g., Raymond et al., 2002). Similar regularity could, however, be expected if speakers store whole utterances, or keep track of the associations between variants and neighboring words (e.g., *eagle–an, dog–a*) and produce novel utterances based on their analogy with previously heard sequences (see Skousen, 1989, for such a proposal; see also Baayen, 2011). I have already discussed the limits of accounts in which all utterances are stored. An interesting goal for future work would be to determine the extent to which rule-like outputs can be obtained with algorithms that do not implement rules but learn to associate variants with neighboring words or word properties (e.g., Baayen, 2011; Baayen et al., 2013).

To summarize, taken together, the empirical data on variation from phonetic and psycholinguistic studies are in line with the ideas that the generation of word forms is a two-step process involving a phonological and a phonetic

component, that some words are represented in memory with more than one representation, and that phonemes and syllables are functional units in speech production. The available evidence provides only little support for the idea that speakers store detailed phonetic information or that some variants result from the application of phonological rules during the phonological encoding process, but it does not provide evidence against these proposals, either. Further insights into the nature of mental representations will be gained by combining phonetic and psycholinguistic investigations and extending the psycholinguistic investigation to a range of variation phenomena. Building on recent proposals in the spoken-word recognition literature (e.g., Arnold et al., 2017; Baayen et al., 2013), interesting insights might further be gained by examining to what extent linguistic behavior (including the reaction times in production experiments) can be accounted for by algorithms that learn to associate pronunciations with specific contexts.

Information flow in the model

On top of defining the nature of representations, models of word production must specify how the information flows in the system. Several important issues must be addressed. A first issue concerns the type/source of information that becomes available, at a given encoding level, to select the relevant representations. For instance, are word form representations selected solely on the basis of information from the semantic–syntactic encoding level, or can contextual information also influence which representation is selected? A second issue relates to whether the information circulates in the system in a modular manner or whether information available at higher encoding levels cascades to all subsequent levels. Another issue concerns the direction of the information flow—that is, whether it only goes from earlier to subsequent encoding levels, or whether the information activated at some later level can feed back to previous encoding levels. A fourth and related issue concerns the degree of parallelism/ sequentiality in encoding processes—that is, whether a given encoding process must be fully accomplished before the onset of the subsequent level, or whether several processes can run in parallel. Several of these issues are for instance discussed in detail in Goldrick (2006) and are at the center of intense debates in the psycholinguistic word production literature (Indefrey & Levelt, 2004; Munding, Dubarry, & Alario, 2015; Strijkers & Costa, 2016; Strijkers, Costa, & Pulvermüller, 2017).

Data on variation provide interesting insights into some of these issues. The observation that word-specific properties influence acoustic details, in particular, shows that information activated/accessed at the word form level must be passed on to the articulators. That contextual variables (e.g., conversational context, word predictability, or repetition) influence articulation reveals that both linguistic and extralinguistic

contextual information participate to the generation of pronunciation.

Several authors have accounted for the influence of higher-level information on word pronunciation with a coordination mechanism. According to Pierrehumbert (2002), for instance, information regarding the ease of retrieval for a given lexeme could be computed and passed on to the phonetic encoding process, where it would be used to control the amount of effort and clarity put into the articulation. Words in low-density neighborhoods or with a high frequency are retrieved more easily (as shown by their shorter onset production latencies in naming tasks), and would thus be produced with less effort, resulting in higher degrees of reduction. A similar proposal is found in Pluymaekers et al. (2005). Here the articulation process continuously adapts its efforts such that less-informative words (more frequent or more predictable) are produced with less articulatory effort. Finally, in Bell et al. (2009) a coordination mechanism regulates the pace of articulation as a function of the pace of the phonological encoding process. In connected speech, the pace of this process is determined in part by how quickly the lexemes can be accessed and selected, whereas the rate of articulation is mainly determined by the syllabic structure (Bell et al., 2009, p. 106). The coordination mechanism helps synchronize the two processes. Accordingly, when lexical access takes longer, the execution of the articulatory plan is slowed, leading to longer acoustic durations. An interesting aspect of these mechanisms is that they account for both word-specific effects (e.g., lexical frequency and phonological neighborhood) and predictability/repetition effects within the same mechanism. Words that tend to be predictable or are repeated are accessed faster, and this speed of access further influences later encoding stages. Interestingly, similar mechanisms could account for phrase frequency effects on acoustic duration. According to Arnon and Cohen Priva (2013), phrase frequency effects are hard to reconcile with traditional models of language processing, in which words and phrases are processed via different mechanisms, with words being directly accessed and phrases requiring in addition a combinatorial process. It could be assumed, however, that lemmas or lexemes that are produced together more often would require less effort to be inserted in the syntactic structure at the grammatical (lemma) level or in the metrical structure at the phonological level. This could, in turn, be reflected in the amount of articulatory effort. A coordination mechanism of this kind would predict both shorter response times and less effort in the articulation of more-frequent phrases. Note that this would require that the system keep track of how often specific lemmas or lexemes are combined.

A second option to allow for lexical–phonological, or more generally for higher-order, information to impact processing at the phonetic encoding and articulation levels is found in cascading-activation models (e.g., Goldrick, 2006). Discrete

models (e.g., Levelt et al., 1999) assume that encoding processes take place sequentially, with a given process only being initiated once the previous one has been completed. Moreover, in such models, only selected representations can send activation to the units at the following processing level (i.e., a representation that has been activated but not selected cannot influence subsequent processing), and subsequent processes do not influence previous ones. In Levelt et al.'s model, for instance, the phonological syllables that result from the phonological encoding process are shared across words. For instance, the words *cupcake* and *cupboard* both contain the syllable /kʌp/. The output of phonological encoding contains the same abstract phonological syllable for both words, and this abstract phonological syllable does not contain information about the word in which the syllable is embedded. This information thus can not influence subsequent encoding processes. Moreover, at the phonetic encoding level, motor commands are again abstract and shared across words. This model cannot explain the influence of word-specific properties on pronunciation without a coordination mechanism of the kind described above.[4]

In cascading-activation models, by contrast, activated representations can pass on activation to the next processing level, irrespective of whether the representations end up being selected. In these models, processing at a given level is initiated before the selection at the previous level is completed. Lexically conditioned variation finds a ready explanation within such systems. Baese-Berk and Goldrick (2009) discussed the effects of phonological neighborhood. During the production of a word from a dense phonological neighborhood, many candidates (the neighbors) are activated and compete with the target representation for production. To be selected, the target representation thus requires more activation than would be necessary for a word from a sparser neighborhood. Because of cascading activation from the phonological to the phonetic encoding level, more activated phonological representations lead to more activated phonetic representations, resulting in more extreme articulation. Incomplete "phonological" variation phenomena or acoustic remnants of the canonical form in the articulation of noncanonical variants also find a ready explanation in such models, since segments of the nonintended variant are partially activated during the encoding process. This partial activation cascades to later processing levels and influences the actual realization of the intended phonemes. For instance, in *run past . . .* , even though the *n* can be assimilated (and thus produced as an [m]), its articulation may be influenced by the fact that /n/ has received some activation and has passed it on to the articulators. Cascaded activation can also readily explain why speech

errors show acoustic features of the intended production (Goldrick & Chu, 2014; see also McMillan & Corley, 2010). These accounts assume, as in Fink and Goldrick (2015), that although representations are abstract, their activation is gradient (in line with proposals in articulatory phonology; e.g., Browman & Goldstein, 1990). Cascading-activation accounts can also explain predictability/repetition effects. A repeated or predictable word would be accessed more easily at one or several processing levels, and this boost of activation can in turn cascade to the articulation and influence the amount of effort there.

Whereas cascading activation allows for units that are activated at a given level to influence any subsequent encoding process, the coordination mechanism described above assumes that information about the ease of processing only influences the articulation. Accordingly, the ease with which a given lexeme is retrieved from the mental lexicon will not influence metrical spell-out or the phonetic encoding process, only the actual articulation of the word. By contrast, in cascading-activation models, the ease with which a given lexeme is retrieved will influence all subsequent processes. The use of online measures such as electroencephalography (EEG) could help further disentangle the two accounts (although in principle cascading activation and coordination mechanisms could coexist). EEG allows for monitoring ongoing processes, from the onset of the stimulus to the onset of articulation. As such, it can target specific encoding processes and could potentially answer the issue of whether word production is achieved through cascading activation, coordination mechanisms, or possibly both. Note, however, that studies in which the EEG is recorded while the participants perform an overt production task have usually avoided exploiting the signal during the articulation or shortly before. The electromyographic activity resulting from movement of the articulators contaminates the EEG signal during and before the articulation (e.g., De Vos et al., 2010; Ganushchak & Schiller, 2008; Goncharova, McFarland, Vaughan, & Wolpaw, 2003; Ouyang et al., 2016; Porcaro, Medaglia, & Krott, 2015). How long before the articulation the signal is contaminated is a matter of discussion (e.g., Fargier, Bürki, Pinet, Alario, & Laganaro, 2017; Ouyang et al., 2016; Porcaro et al., 2015), but several studies have suggested that the phonological encoding process and at least part of the phonetic encoding processes can be investigated with EEG (see Bürki, Pellet-Cheneval, & Laganaro, 2015, or Laganaro, Python, & Toepel, 2013). Important efforts have also been put into developing and testing ways of removing artifacts from the EEG signal of interest (James & Hesse, 2005; Ouyang et al., 2016; Pham, Fine, & Semrud-Clikeman, 2011; Urigüen & Garcia-Zapirain, 2015; Vigario & Oja, 2008).

Note that both the coordination mechanism and cascading-/gradient-activation accounts are specific implementations of the production facilitation (or speaker-processing) account

---

[4] The question of whether a modular and sequential system with an added coordination mechanism still qualifies as a modular and sequential system is legitimate here, but beyond the scope of this article.

discussed in section 3, which assumes that variation is driven by speaker-internal production processes. Speaker-processing accounts assume that the effects of some variables both on encoding processes (and their speed) and on the articulation have the same source. Accordingly, when processing is made easier (i.e., because the word is more frequent or more predictable), this ease of processing directly impacts the way the word is articulated. Buz and Jaeger (2016) dubbed this view the "planning-drives-articulation hypothesis." As these authors highlighted, the parallel influences of a given variable on pronunciation and processing speed are not sufficient to link the two effects to the same source. To date, there has been little empirical evidence on whether processing speed and variation in articulation have the same sources, because in many cases the effects of a given variable on articulation and processing speed are not even documented within the same study. An important aim for further studies will be to examine more systematically the relationship between processing speed and articulation. Here again, EEG data could be used to investigate the origin of the speeding of latencies/ shortening of acoustic duration (i.e., whether these effects arise in the grammatical encoding process, in the phonological encoding process, in the phonetic encoding process, or in several of these processes).

Cascading activation and coordination mechanisms can be implemented in abstractionist models, exemplarist models, or Pierrehumbert's (2002) "hybrid" model. Note that exemplarist models do not need such mechanisms to account for the influence of lexical frequency on how words are actually pronounced. In these models, variation arises not only because words are stored as detailed exemplars, but also because articulatory routines may become more or less automatized (e.g., Bybee, 2001; Pierrehumbert, 2001, 2002). Frequent words will therefore tend to be more reduced, and these reduced exemplars will in turn be stored, influencing subsequent productions. Exemplarist accounts can also account for some contextual effects, including social factors, by storing knowledge about the context in which the words are produced together with the exemplars (see also Fink & Goldrick, 2015), or by assuming that the speaker's representation of the state of the listener influences the selection of exemplars (e.g., more reduced exemplars selected when the word is predictable). These models can also account for phrase frequency effects by assuming that not only words but also whole utterances are represented in long-term memory (see Arnon & Cohen Priva, 2013; but see Baayen et al., 2013; Hendrix et al., 2017). Exemplar-based models cannot account for all observations, however (see Fink & Goldrick, 2015, for an extensive discussion). For instance, whereas exemplar-based models could account for phonological neighborhood effects by storing this lexical property with the exemplars, they cannot readily explain why these effects vary across tasks. In addition, these models

do not provide an account of why phonological neighborhood density also influences encoding speed.

In discussing predictability/repetition reduction effects in section 3, I mentioned audience-design (or intelligibility-based) accounts. These assume that speakers adjust their speech in order to be understood. The models assume that variation driven by contextual factors originates in late articulatory processes and results from adaptation to the listener (e.g., Fox Tree & Clark, 1997). Such accounts are independent of the kind of information stored in memory, and could thus be implemented in exemplarist models, abstractionist models, or the hybrid model discussed above. The influences of predictability, repetition, communicative goals, lexical frequency, and social factors can, in principle, all be interpreted in this framework. In the latter case, for instance, speakers are careful to hyperarticulate in contexts in which reduction is not expected or would contribute to lowering their social image. Speakers take into account all relevant variables when "deciding" to what extent they should hyperarticulate or can afford to hypoarticulate. Note that this decision is assumed to influence the amount of effort put into the articulation process, not into other encoding stages. An interesting goal for further studies in this context would be to determine how the different factors are weighted to estimate the degree of effort to be put into the articulation. Moreover, and as highlighted by studies on repetition or predictability, patterns of variation are not entirely captured by listener-oriented mechanisms. It is likely that both production-based and intelligibility-based mechanisms work in tandem during speech production, and an important goal for future studies will be to determine how speakers integrate information from their listener's needs with the requirements of their speech processing system during speech production.

Information about how information flows in the language production system is provided by studying the variables that influence pronunciation (and the speed of encoding). The study of these variables is not exempt from methodological concerns. In section 3, the influences of about ten types of variables were discussed. This has important methodological implications. If some of these variables are not taken into account in a given study, they may act as confounding variables. The effects of a given manipulation may in fact be a byproduct of differences in other dimensions across groups. This is particularly problematic when different words are compared, because it is almost impossible to equate word lists across conditions in all relevant dimensions. Some studies avoid this concern by using clever manipulations in the design (such as use of the same materials with different primes or distractors, or manipulation of the stimulus set; e.g., Baese-Berk & Goldrick, 2009; Buz, Tanenhaus, & Jaeger, 2016; Kirov & Wilson, 2012; Roon & Gafos, 2015; Yuen, Davis, Brysbaert, & Rastle, 2010). This is not always possible, and some research questions require the use of different linguistic materials across conditions. There are different approaches to

controlling for confounding variables. One is to balance the items in the two conditions of interest (e.g., low and high neighborhood density) with respect to these variables. Balance is "achieved" by ensuring that the *p* values when comparing the two sets are not significant. Importantly, however, differences that are not significant across sets may still have significant influence on how the words are actually pronounced. An alternative approach is to account for other variables in the statistical analysis (see, e.g., Fricke, Baese-Berk, & Goldrick, 2016, for the reanalysis of previous datasets with an analysis that takes into account the influences of additional variables). As was noted by Buz and Jaeger (2016), controlling for all potential confounding variables is often not possible. However, unless this is done, the extent to which a given variable accounts for an individual amount of phonetic variation and how it does so will remain open issues. Another option to avoid potential confounds is to design novel-word studies. Novel-word studies are regularly used in psycholinguistics, especially in spoken-word recognition research (e.g., Gaskell & Dumay, 2003; Magnuson, Tanenhaus, Aslin, & Dahan, 2003) and could provide interesting insights into the role of contextual factors on articulation.

To summarize, the evidence suggests that variation in pronunciation involves both speaker-internal and listener-oriented mechanisms. Further research should try to determine whether cascading activation or a coordination mechanism provides a better account of the observed patterns of variation. Another important issue for future research will be to determine how information about the listener's need and other variables is integrated during the word production process.

Interface between production and perception systems

For many years, production and perception processes were studied separately, by different groups of researchers. Understanding and modeling the architecture of the word production system requires specifications about how this system relates to the spoken-word recognition system (and vice versa; see the special issue "Speaking and Listening: Relationships Between Language Production and Comprehension" by Meyer, Huettig, & Levelt, 2016). Important issues, for instance, are whether or not phonological and phonetic representations are shared between the two systems, and if they are not shared, to what extent and how they influence one another. Variation phenomena are likely to offer a unique window into this interface. For instance, the speech of a given speaker is necessarily much less variable than the speech that this speaker is exposed to. An important goal for future research should be to exploit this gap to study the interface between the two systems. For instance, studies have shown that the frequency with which a given variant is encountered influences recognition times (see, e.g., Connine et al., 2008, for English schwa

words; Ranbom & Connine, 2007, for nasal flap variants in American English; or Bürki & Frauenfelder, 2012, for French schwa words). Other studies have found an influence of the speaker's own production frequency on the time taken to encode a variant in speech production tasks (e.g., Bürki et al., 2010). The question arises of whether production and recognition tasks are influenced by the frequency of the variants in their respective modalities or whether they are influenced by exactly the same frequency measures, and whether these frequencies come mainly from a speaker's own production, from the speech (s)he is exposed to, or from a combination of both. Answering these questions may inform as to whether word form representations are shared or separate across modalities.

Studies on phonetic imitation also have the potential to provide information on the interface between comprehension and production. The observation that speakers imitate fine phonetic features of the speech of their interlocutor reveals that in conversations, perception and production are tightly linked (see also Mitterer & Ernestus, 2008), and more specifically, that speakers reuse detailed phonetic information from their interlocutor's speech in their own productions. Models that assume shared or closely related phonetic representations between the production and comprehension systems (e.g., Goldinger, 1998; Hickok, 2014; Pierrhumbert, 2002) are in a better position to account for these effects. Models in which phonetic representations are abstract and static (e.g., Levelt & Wheeldon, 1994) cannot readily account for such data.

Several important issues remain to be resolved, however, in order to fully understand how the speech production system makes use of the details of the speech of others in phonetic imitation studies and to test models of language production against these data. In many studies in which imitation is assessed by independent judges, it is not clear which phonetic parameter was imitated (see Pardo et al., 2017, for a recent review). Understanding the nature of the phonetic input that speakers reuse in these studies will be crucial to understanding the nature of the relationship between perception and production processes. The question arises of whether participants imitate anything beyond general prosodic patterns (i.e., intensity, prosodic contour). Another interesting issue is whether speakers only imitate when presented with the same words to repeat (as in Goldinger's, 1998, study and similar ones) or whether imitation can be found as well when the target word is preceded by a partially overlapping word (e.g., *carrot–cabin*). Whereas Goldinger's account of imitation effects predicts imitation only when words fully overlap with one another, Pierrehumbert's model predicts imitation when a subset of segments are shared between the speech of a given speaker and that of his or her interlocutor. Moreover, if the syllable is the unit of phonetic encoding, at least for frequent syllables, imitation should be found when the overlap concerns a syllable, not necessarily when it concerns sequences of segments that cross syllable boundaries.

Beyond isolated words

Speakers rarely utter words in isolation, and the ultimate aim of language production models is to describe not only how the information about words is stored and encoded, but also how words are combined to build utterances. The study of utterance production must address three issues in particular. The first concerns the scope of advance planning—that is, how many words speakers plan ahead before the onset of articulation. For instance, when speakers produce a sentence such as *the big dog barked at him and left*, how many words do they prepare in advance at each processing level? Do they only plan one word (*the*), a phonological word (*the big*), a simple noun phrase (*the big dog*), a complex noun phrase (*the dog and the boy*), a whole clause (*the dog barked at him*), or the whole sentence (*the big dog barked at him and left*) in advance? This issue has been addressed in several studies, but the conclusions have been highly heterogeneous and have greatly depended on which dependent measure was used. Most studies relying on eyetracking data have found a restricted scope of advance planning (e.g., Meyer, Sleiderink, & Levelt, 1998; Meyer & van der Meulen, 2000; Spieler & Griffin, 2006; see Brown-Schmidt & Konopka, 2008, for a restricted scope of planning at the message encoding level). The scope of advance planning has also been investigated with indirect measures, such as errors, pauses, or speech onset latencies. The findings are roughly consistent with the idea that speakers plan more at the grammatical than at the phonological level, but exactly how much they plan at each level is far from consensus (e.g., Costa & Caramazza, 2002; Damian & Dumay, 2007; Martin, Crowther, Knight, Tamborello, & Yang, 2010; Martin, Miller, & Vu, 2004; Meyer, 1996; Michel Lange & Laganaro, 2014; Oppermann, Jescheniak, & Schriefers, 2010; Schnur, 2011; Smith & Wheeldon, 1999; Wheeldon & Lahiri, 1997). Disparate conclusions across studies are sometimes interpreted in terms of flexibility (i.e., speakers use flexible planning units; e.g., Norcliffe, Konopka, Brown, & Levinson, 2015). An important goal for future studies will be to determine the variables that influence and constrain this flexibility.

The second issue concerns the time course (i.e., alignment) of the encoding processes for the successive words within planning units. Do speakers select and encode all representations at a given level at once, or do they select them sequentially, and if so, in which order? If they activate several words at once, how do they manage to still produce the words in the correct order? Moreover, do they fully encode the utterance at a given level (e.g., grammatical encoding) before the initiation of the next level (e.g., phonological encoding)? This issue has largely been overlooked. Possible accounts have been discussed, but the available evidence is far from constraining (G. S. Dell, 1986; Jescheniak, Schriefers, & Hantsch, 2003; Kempen & Huijbers, 1983; Meyer, 1996; Schriefers, 1992).

Another important issue to be addressed relates to the size of the representational units in long-term memory. Traditional psycholinguistic models assume that words are stored and accessed as separate entities. Exemplarist accounts assume that longer utterances could be stored (e.g., Bybee, 2001; see Tremblay & Baayen, 2010, for a discussion).

The study of variation phenomena can provide (and has to some extent already provided) crucial information on these issues. For instance, the fact that the choice of the pronunciation variant for a given word is influenced by the phonological content of the next word clearly indicates that the two words are part of the same phonological encoding unit, and that the phonology of the second word is accessed before the first word is phonologically encoded. Speakers of French can, for example, only produce utterances such as *le grand chat* "the big cat" without errors if they encode the whole utterance at the phonological level before they initiate the phonetic encoding of the determiner and if they encode the noun before the adjective and the adjective before the determiner. The study of contextually constrained phonological variation may help further define the scope of advance planning at the phonological level, its flexibility, and the constraints that limit this flexibility (see Kilbourn-Ceron, 2017, for an example concerning French liaison). An interesting issue to be addressed would be whether the scope of phonological encoding and the order in which words are encoded, within a given language and for similar types of utterances, strictly depend on the presence of phonological contextual constraints, or whether they are similar, within a language, for utterances with and without such constraints, but are large enough to accommodate phonological constraints when present.

Along the same lines, the phonological-consistency effect reported in psycholinguistic studies and discussed in section 3.2 (longer onset latencies to produce determiner–adjective–noun utterances when the onset of the noun and the onset of the adjective call for different determiner forms) provides evidence that the whole noun phrase is planned at the grammatical and phonological levels before the onset of the next stage, and that the noun is accessed first during phonological encoding. This effect also clearly suggests that words are accessed as separate entities and combined online. Phonological consistency effects are hard to explain within a system in which utterances are stored as single units in the speaker's lexicon. Further investigation into the phonological-consistency effect could help resolve the apparent conflict between the effects of utterance frequency on speech onset latencies and articulation (Arnon & Cohen Priva, 2013, 2014; Janssen & Barber, 2012) and the evidence supporting access to individual words. Evidence that phrase frequency influences naming latencies and utterance duration is often taken as evidence that the whole utterance is stored in the mental lexicon (Arnon & Cohen Priva, 2013; Janssen & Barber, 2012). A first alternative hypothesis, to be tested in

further studies, assumes that frequency effects (on reaction times and articulation) do not reflect the storage of the noun phrase, but rather the fact that the lexicon stores information about how frequently words are produced together. A second alternative hypothesis assumes that very frequent noun phrases are stored, whereas less-frequent phrases result from combinatorial processes. Studies in which both phonological consistency and utterance frequency are manipulated would allow for addressing this issue. The second alternative hypothesis, for instance, predicts an interaction between phonological consistency and noun phrase frequency.

Finally, as was discussed in section 2, many variation phenomena are constrained by the context in which they are produced. Some are highly constrained, others are constrained by more than just one factor (e.g., the production of schwa words in French depends on the number of consonants in the preceding word, but also on the formality of the situation, the position of the word in the utterance, or the speech rate). The comparison of utterances whose pronunciation is constrained to different degrees and by various types of constraints would inform theories of how the building of utterances (i.e., the scope of advance planning and the order of word access) is influenced by contextual linguistic and nonlinguistic information, and as such, would provide further insight into the scope of advance planning, its flexibility, and the constraints operating on this flexibility.

Conclusion

An increasing number of researchers in psycholinguistic research have pointed toward the study of language processes in more naturalistic contexts (Hasson & Honey, 2012; Healey, Purver, & Howes, 2014; Jaeger, 2013). Variability is part and parcel of natural forms of speech, and its study offers unique insights into the architecture of the word production system. Phonologists, for instance, have long realized the importance of variation and have often evaluated theories of sound structure on the basis of their ability to handle variation phenomena (e.g., Coetzee, 2012; Durand & Lyche, 2008). By contrast, in psycholinguistic research, the relevance of variation phenomena is often underestimated. The dominant models in the field do not explicitly model the production of pronunciation variants. This review has highlighted the relevance of the available data on variation phenomena to understanding and modeling the cognitive processes and representations underlying language production, and has pinpointed the many open issues that the study of variability has the power to address in future research. As is evident from the review, there are different ways of accounting for the observed patterns of variation. Advances in understanding the cognitive architecture of the word production system will benefit from the design of empirical studies that specifically contrast the predictions of these accounts, from the implementation of these proposals in computational models, and from the evaluation of the output of these models against both existing and novel data.

References

Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, *46*, 119–139. https://doi.org/10.1016/j.specom.2005.03.006

Adda-Decker, M., & Snoeren, N. D. (2011). Quantifying temporal speech reduction in French using forced speech alignment. *Journal of Phonetics*, *39*, 261–270. https://doi.org/10.1016/j.wocn.2010.11.011

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

Alario, F. X., Ferrand, L., Laganaro, M., New, B., Frauenfelder, U. H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, Instruments, & Computers*, *36*, 140–155. https://doi.org/10.3758/BF03195559

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., & Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PLoS ONE*, *12*, e0174623:1–16. https://doi.org/10.1371/journal.pone.0174623

Arnold, J. E., & Watson, D. G. (2015). Synthesizing meaning and processing approaches to prosody: Performance matters. *Language, Cognition and Neuroscience*, *30*, 88–102. https://doi.org/10.1080/01690965.2013.840733

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*, 349–371. https://doi.org/10.1177/0023830913484891

Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *Mental Lexicon*, *9*, 377–400. https://doi.org/10.1075/ml.9.3.01arn

Ashby, M., & Maidment, J. (2005). Introducing phonetic science. Cambridge, UK: Cambridge University Press.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*, 3048–3058. https://doi.org/10.1121/1.2188331

Baayen, R. H. (2011). Corpus linguistics and naive discriminative learning. *Revista Brasileira de Linguística Aplicada*, *11*, 295–328. https://doi.org/10.1590/S1984-63982011000200003

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of *n*-gram frequency effects based on Naive Discriminative Learning. *Language and Speech*, 56, 329–347. https://doi.org/10.1177/0023830913484896

Babel, M. (2010). Dialect convergence and divergence in New Zealand English. *Language in Society*, 39, 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, 55, 231–248. https://doi.org/10.1177/0023830911417695

Babel, M., & Munson, B. (2014). Producing socially meaningful linguistic variation. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds). *Oxford handbook of language production* (pp. 308–325). Oxford, UK: Oxford University Press.

Baese-Berk, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24, 527–554. https://doi.org/10.1080/01690960802299378

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42, 1–22. https://doi.org/10.1006/jmla.1999.2667

Bard, E. G., & Aylett, M. (2004). Referential form, duration, and modeling the listener in spoken dialogue. In J. Trueswell & M. Tanenhaus (Eds.), Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions (pp. 173–191). Cambridge, MA: MIT Press.

Barnes, J., & Kavitskaya, D. (2002). Phonetic analogy and schwa deletion in French. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* (Vol. 28, pp. 39–50). Berkeley, CA: University of California, Berkeley Linguistics Society.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, 50, 560–585. https://doi.org/10.1080/783663595

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

Bock, K., & Levelt, W. J. M. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), Handbook of psycholinguistics (pp. 945–984). Orlando, FL: Academic Press.

Booij, G. E. (1995). Phonology of Dutch. Oxford, UK: Oxford University Press.

Braver, A. (2014). Imperceptible incomplete neutralization: Production, non-identifiability, and non-discriminability in American English flapping. *Lingua*, 152, 24–44. https://doi.org/10.1016/j.lingua.2014.09.004

Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *Journal of the Acoustical Society of America*, 128, EL32–EL37. https://doi.org/10.1121/1.3448022

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology. *An Overview Phonetica, 49*, 155–180

Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, 109, 274–280. https://doi.org/10.1016/j.cognition.2008.07.011

Buchwald, A., & Miozzo, M. (2011). Finding levels of abstraction in speech production: Evidence from sound-production impairment. *Psychological Science*, 22, 1113–1119. https://doi.org/10.1177/0956797611417723

Buchwald, A. B., Rapp, B., & Stone, M. (2007). Insertion of discrete phonological units: An articulatory and acoustic investigation of aphasic speech. *Language and Cognitive Processes*, 22, 910–948. https://doi.org/10.1080/01690960701273532

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, 62, 421–437. https://doi.org/10.1016/j.jml.2010.01.002

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011a). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *Journal of the Acoustical Society of America*, 130, 3980–3991. https://doi.org/10.1121/1.3658386

Bürki, A., Fougeron, C., & Gendrot, C. (2007). On the categorical nature of the process involved in schwa elision in French. In *Proceedings of Interspeech* (pp. 1026–1029). New York, NY: ISCA Archive.

Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. (2011b). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, 39, 279–288. https://doi.org/10.1016/j.wocn.2010.07.003

Bürki, A., & Frauenfelder, U. H. (2012). Producing and recognizing words with two pronunciation variants: Evidence from novel schwa words. *Quarterly Journal of Experimental Psychology*, 65, 796–824. https://doi.org/10.1080/17470218.2011.634915

Bürki, A., Frauenfelder, U. H, & Alario, F.-X. (2015). On the resolution of phonological constraints during utterance production. *Journal of the Acoustical Society of America*, 138, EL429. https://doi.org/10.1121/1.4934179

Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 617–631. https://doi.org/10.1037/a0026167

Bürki, A., Laganaro, M., & Alario, F.-X. (2014). Phonologically driven variability: The case of determiners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 1348–1362. https://doi.org/10.1037/a0036351

Bürki, A., Pellet-Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language*, 150, 90–102. https://doi.org/10.1016/j.bandl.2015.08.006

Buz, E., & Jaeger, T. F. (2016). The (in)dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 31, 404–424. https://doi.org/10.1080/23273798.2015.1105984

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, 89, 68–86. https://doi.org/10.1016/j.jml.2015.12.009

Bybee, J. (2000). Lexicalization of sound change and alternating environment. In M. B. Broe & J. B. Pierrehumbert (Eds.), Papers in laboratory phonology V: Acquisition and the lexicon (pp. 250–268). Cambridge, UK: Cambridge University Press.

Bybee, J. (2001). Phonology and language use. Cambridge, UK: Cambridge University Press.

Bybee, J. (2007). Frequency of use and the organization of language. Oxford, UK: Oxford University Press.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208. https://doi.org/10.1080/026432997381664

Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X. (2001). A crosslinguistic investigation of determiner production. In E. Dupoux (Ed.), Language, brain and cognitive development: Essays in honor of Jacques Mehler (pp. 209–226). Cambridge, MA: MIT Press.

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, 27, 963–972. https://doi.org/10.1002/hbm.20236

Caselli, N. K., Caselli, M. K., & Cohen-Goldberg, A. M. (2016). Inflected words in production: Evidence for a morphologically rich lexicon.

Quarterly Journal of Experimental Psychology, 69, 432–454. https://doi.org/10.1080/17470218.2015.1054847

Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science, 8*, 322–325. https://doi.org/10.1111/j.1467-9280.1997.tb00446.x

Chevrot, J.-P., Beaud, L., & Varga, R. (2000). Developmental data on a French sociolinguistic variable: Post-consonantal word-final /R/. *Language Variation and Change, 12*, 295–319.

Cholin, J., & Levelt, W. J. M. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes, 24*, 662–684. https://doi.org/10.1080/01690960802348852

Cholin, J., Levelt, W. J. M., & Schiller, N. (2006). Effects of syllable frequency in speech production. *Cognition, 99*, 205–235. https://doi.org/10.1016/j.cognition.2005.01.009

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Cambridge, MA: MIT Press.

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. https://doi.org/10.1016/j.wocn.2016.11.002

Clopper, C., & Pierrehumbert, J. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *Journal of the Acoustical Society of America, 124*, 1682–1688. https://doi.org/10.1121/1.2953322

Clopper, C. G., & Tamati, T. N. (2014). Effects of local lexical competition and regional dialect on vowel production. *Journal of the Acoustical Society of America, 136*, 1–4. https://doi.org/10.1121/1.4883478

Coetzee, A. W. (2012). Variation: Where laboratory and theoretical phonology meet. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *Oxford handbook of laboratory phonology* (pp. 62–75). Oxford, UK: Oxford University Press.

Cohen-Goldberg, A. M. (2015). Abstract and lexically specific information in sound patterns: Evidence from /r/-sandhi in rhotic and non-rhotic varieties of English. *Language and Speech, 58*, 522–548. https://doi.org/10.1177/0023830914567168

Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory, 16*, 1–31.

Cohn, A. C., Fougeron, C., & Huffman, M. K. (2012). Introduction. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), Oxford handbook of laboratory phonology (pp. 1–9). Oxford, UK: Oxford University Press.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics, 70*, 403–411. https://doi.org/10.3758/PP.70.3.403

Costa, A., Alario, F. X., & Sebastián-Gallés, N. (2007). Cross-linguistic research on language production. In M. G. Gaskell (Ed.), Oxford handbook of psycholinguistics (pp. 531–546). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198568971.013.0032

Costa, A., & Caramazza, A. (2002). The production of noun phrases in English and Spanish: Implications for the scope of phonological encoding in speech production. *Journal of Memory and Language, 46*, 178–198. https://doi.org/10.1006/jmla.2001.2804

Côté, M.-H. (2011). French liaison. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *Companion to phonology* (pp. 2685–2710). Malden, MA: Wiley-Blackwell.

Côté, M., & Morrison, G. S. (2007). The nature of the schwa/zero alternation in French clitics: Experimental and non-experimental evidence. *Journal of French Language Studies, 17*, 159–186. https://doi.org/10.1017/S0959269507002827

Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language, 57*, 195–209. https://doi.org/10.1016/j.jml.2006.11.001

Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica, 63*, 79–112. https://doi.org/10.1159/000095304

Dell, F. (1985). Les règles et les sons. Paris, France: Hermann.

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

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124–142. https://doi.org/10.1016/0749-596X(88)90070-8

Dell, G. S. (2014). Phonemes and production. *Language, Cognition and Neuroscience, 29*, 30–32. https://doi.org/10.1080/01690965.2013.851795

De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., . . . Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics, 8*, 135–150. https://doi.org/10.1007/s12021-010-9071-0

Dienes, Z. (2008). Understanding psychology as a science: An introduction to scientific and statistical inference. London, UK: Palgrave Macmillan.

Docherty, G., & Mendoza-Denton, N. (2012). Speaker-related variation—Sociophonetic factors. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), Oxford handbook of laboratory phonology (pp. 43–60). Oxford, UK: Oxford University Press.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Cognitive Science, 4*, 346. https://doi.org/10.3389/fpsyg.2013.00346

Durand, J. (1997). Linking "r" in English : Constraints, principles and parameters, or rules ? *Histoire Epistémologie Langage, 19*, 43–72. https://doi.org/10.3406/hel.1997.2572

Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language Studies, 18*, 33–66. https://doi.org/10.1017/S0959269507003158

Ellis, A. W., & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 515–523. https://doi.org/10.1037/0278-7393.24.2.515

Ellis, L., & Hardcastle, W. J. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *Journal of Phonetics, 30*, 373–396. https://doi.org/10.1006/jpho.2001.0162

Ernestus, M. (2011). Gradience and categoricality in phonological theory. In M. Van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), Blackwell companion to phonology (pp. 2115–2136). Wiley-Blackwell.

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

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162–173. https://doi.org/10.1006/brln.2001.2514

Ernestus, M., Lahey, M., Verhees, F., & Baayen, R. H. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America, 120*, 1040–1051. https://doi.org/10.1121/1.2211548

Fargier, R., Bürki, A., Pinet, S., Alario, F.-X., & Laganaro, M. (2017). Word onset phonetic properties and motor artefacts in speech production EEG recordings. *Psychophysiology.* Advance online publication. https://doi.org/10.1111/psyp.12982

Ferrand, L., Humphreys, G. W., & Segui, J. (1998). Masked repetition and phonological priming in picture naming. *Perception & Psychophysics, 60*, 263–274. https://doi.org/10.3758/BF03206035

Fink, A., & Goldrick, M. (2015). The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard, 1*, 215–225. https://doi.org/10.1515/lingvan-2015-1003

Flemming, E. (1995). Phonetic detail in phonology: Evidence from assimilation and coarticulation. In K. Suzuki & D. Elzinga (Eds.), Proceedings of the Southwest Workshop on Optimality Theory: Features in OT (pp. 39–50). Tucson, AZ: University of Arizona, Coyote Working Papers.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology, 18*, 7–44.

Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *U.C.L.A. Working Papers in Phonetics, 97*, 1–74.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31*, 307–319. https://doi.org/10.1177/002383098803100401

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*, 396–413. https://doi.org/10.1016/S0749-596X(03)00072-X

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504. https://doi.org/10.1016/0749-596X(87)90136-7

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighbourhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language, 83*, 97–117. https://doi.org/10.1016/j.jml.2015.04.002

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151–167.

Francis, W. S. (2014). Repetition priming in picture naming: Sustained learning through the speeding of multiple processes. *Psychonomic Bulletin & Review, 21*, 1301–1308. https://doi.org/10.3758/s13423-014-0610-9

Fricke, M., Baese-Berk, M. M., & Goldrick, M. (2016). Dimensions of similarity in the mental lexicon. *Language, Cognition, and Neuroscience, 31*, 639–645. https://doi.org/10.1080/23273798.2015.1130234

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27–52.

Gafos, D., & Goldstein, L. (2012). Articulatory representation and phonological organization. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), Handbook of laboratory phonology (pp. 220–231). Oxford, UK: Oxford University Press.

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

Gahl, S. (2015). Lexical competition in vowel articulation revisited: Vowel dispersion in the Easy/Hard database. *Journal of Phonetics, 49*, 96–116. https://doi.org/10.1016/j.wocn.2014.12.002

Gahl, S., & Strand, J. F. (2016). Many neighbourhoods: Phonological and perceptual neighbourhood density in lexical production and perception. *Journal of Memory and Language, 89*(Suppl. C), 162–178. https://doi.org/10.1016/j.jml.2015.12.006

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighbourhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789–806. https://doi.org/10.1016/j.jml.2011.11.006

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51. https://doi.org/10.1016/j.jml.2009.09.002

Gambi, C., & Pickering, M. J. (2013). Prediction and imitation in speech. *Cognitive Science, 4*, 340. https://doi.org/10.3389/fpsyg.2013.00340

Ganushchak, L. Y., & Schiller, N. O. (2008). Brain error-monitoring activity is affected by semantic relatedness: An event-related brain potentials study. *Journal of Cognitive Neuroscience, 20*, 927–940. https://doi.org/10.1162/jocn.2008.20514

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of learning and motivation* (pp. 133–177). New York, NY: Academic Press.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production* (Vol. 1, pp. 177–220). London, UK: Academic Press.

Gaskell, M. G. (2001). Phonological variation and its consequences for the word recognition system. *Language and Cognitive Processes, 16*, 723–729. https://doi.org/10.1080/0169096014300128

Gaskell, M. G., Cox, H., Foley, K., Grieve, H., & O'Brien, R. (2003). Constraints on definite article alternation in speech production: To "thee" or not to "thee"? *Memory & Cognition, 31*, 715–727. https://doi.org/10.3758/BF03196110

Gaskell, M. G, & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132. https://doi.org/10.1016/S0010-02770300070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158. https://doi.org/10.1080/01690969508407097

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63*, 223–230.

Gick, B., & Wilson, I. (2001) Pre-liquid excrescent schwa: What happens when vocalic targets conflict. In P. Dalsgaard, B. Lindberg, & H. Benner (Eds.), Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001) (pp. 273–276). Aalborg, Denmark.

Gick, B., & Wilson, I. (2006). Excrescent schwa and vowel laxing: Cross-linguistic responses to conflicting articulatory targets. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), Papers in Laboratory Phonology VIII (pp. 635–659). Berlin, Germany: Mouton de Gruyter.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279. 10.1.1.380.9874

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*, 716–722. https://doi.org/10.3758/BF03196625

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language and Cognitive Processes, 21*, 817–855. https://doi.org/10.1080/01690960600824112

Goldrick, M. (2011). Linking speech errors and generative phonological theory. *Linguistics and Language Compass, 5*, 397–412. https://doi.org/10.1111/j.1749-818X.2011.00282.x

Goldrick, M., & Chu, K. (2014). Gradient co-activation and speech error articulation: Comment on Pouplier and Goldstein (2010). *Language, Cognition and Neuroscience, 29*, 452–458. https://doi.org/10.1080/01690965.2013.807347

Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbours on stop consonant articulation. *Journal of the Acoustical Society of America, 134*, 172–177. https://doi.org/10.1121/1.4812821

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: Spectral and topographical characteristics. *Clinical Neurophysiology, 114*, 1580–1593. https://doi.org/10.1016/S1388-2457(03)00093-2

Gordon, J. K., & Kurczek, J. C. (2013). The ageing neighbourhood: Phonological density in naming. *Language, Cognition and Neuroscience, 29*, 326–344. https://doi.org/10.1080/01690965.2013.837495

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279–296. https://doi.org/10.1016/j.jml.2004.05.004

Grammont, M. (1914). *Traité pratique de prononciation française*. Paris, France: Delagrave.

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*, 1231–1240.

Hallé, P. A., & Adda-Decker, M. (2011). Voice assimilation in French obstruents: Categorical or gradient? In J. Goldsmith, E. Hume, & L. Wetzels (Eds.), Tones and features: A festschrift for Nick Clements (pp. 149–175). Berlin, Germany: Mouton de Gruyter.

Hannahs, S. J., & Tallerman, M. (2006). At the interface: Selection of the Welsh definite article. *Linguistics*, *44*, 781–816. https://doi.org/10.1515/LING.2006.025

Hansen, A. B. (1994). Etude du E caduc—Stabilisation en cours et variations lexicales. *Journal of French Language Studies*, *4*, 25–54. https://doi.org/10.1017/S0959269500001964

Hasson, U., & Honey, C. J. (2012). Future trends in neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*, 1272–1278. https://doi.org/10.1016/j.neuroimage.2012.02.004

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405. https://doi.org/10.1016/j.wocn.2003.09.006

Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE*, *9*, e98598. https://doi.org/10.1371/journal.pone.0098598

Hendrix, P., Bolger, P., & Baayen, R. H. (2017). Distinct ERP signatures of word frequency, phrase frequency, and prototypicality in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 128–149. https://doi.org/10.1037/a0040332

Herd, W., Jongman, A., & Sereno, J. (2010). An acoustic and perceptual analysis of /t/ and /d/ flaps in American English. *Journal of Phonetics*, *38*, 504–516. https://doi.org/10.1016/j.wocn.2010.06.003

Herrmann, F., P. Whiteside, S., & Cunningham, S. (2009). An acoustic investigation into coarticulation and speech motor control: High vs. low frequency syllables. In *Proceedings of Meetings of the Acoustical Society of America*, *4*, 1–11. https://doi.org/10.1121/1.3085742

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, *29*, 2–20. https://doi.org/10.1080/01690965.2013.834370

Hinskens, F. (2011). Lexicon, phonology and phonetics. Or: Rule-based and usage-based approaches to phonological variation. In P. Siemund (Ed.), Linguistic universals and language variation (pp. 416–456). Berlin, Germany: Mouton de Gruyter.

Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data. *Language and Speech*, *36*, 235–260.

Indefrey, P. (2014). Time course of word production does not support a parallel input architecture. *Language, Cognition and Neuroscience*, *29*, 33–34. https://doi.org/10.1080/01690965.2013.847191

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jacobs, C. L., Yiu, L. K., Watson, D. G., & Dell, G. S. (2015). Why are repeated words produced with reduced durations? Evidence from inner speech and homophone production. *Journal of Memory and Language*, *84*, 37–48. https://doi.org/10.1016/j.jml.2015.05.004

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Language Sciences*, 230. https://doi.org/10.3389/fpsyg.2013.00230

Jaeger, T. F., & Buz, E. (2016). Signal reduction and linguistic encoding. In E. M. Fernández & H. S. Cairns (Eds.), Handbook of psycholinguistics (pp. 38–81). Oxford, UK: Wiley-Blackwell.

James, C. J., & Hesse, C. W. (2005). Independent component analysis for biomedical signals. *Physiological Measurement*, *26*, R15–R39. https://doi.org/10.1088/0967-3334/26/1/R02

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*, 33202. https://doi.org/10.1371/journal.pone.0033202

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843. https://doi.org/10.1037/0278-7393.20.4.824

Jescheniak, J. D., Schriefers, H., & Hantsch, A. (2003). Utterance format affects phonological priming in the picture–word task: Implications for models of phonological encoding in speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 441–454. https://doi.org/10.1037/0096-1523.29.2.441

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), Proceedings of the 1st session of the 10th international symposium (pp. 29–54). Tokyo, Japan: National International Institute for Japanese Language.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), Frequency and the emergence of linguistic structure (pp. 229–254). Amsterdam, The Netherlands: Benjamins.

Kager, R. W. J. (1997). Rhythmic vowel deletion in Optimality Theory. In I. Roca (Ed.), Derivation and constraints in phonology (pp. 463–499). Oxford, UK: Oxford University Press.

Kahn, J. M., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language*, *67*, 311–325. https://doi.org/10.1016/j.jml.2012.07.002

Kahn, J. M., & Arnold, J. E. (2015). Articulatory and lexical repetition effects on durational reduction: Speaker experience vs. common ground. *Language, Cognition and Neuroscience*, *30*, 103–119. https://doi.org/10.1080/01690965.2013.848989

Kemp, J., Pupier, P., & Yaeger, M. (1980). A linguistic and social description of final consonant cluster simplification in Montreal French. In R. W. Shuy & A. Shnukal (Eds.), Languages and the uses of language (pp. 12–40). Washington, DC: Georgetown University Press.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition*, *14*, 185–209. https://doi.org/10.1016/0010-0277(83)90029-X

Kilbourn-Ceron, O. (2017). Speech production planning affects phonological variability: A case study in French liaison. In *Proceedings of the 2016 Annual Meetings on Phonology* (Vol. 4). https://doi.org/10.3765/amp.v4i0.4004

Kim, M., Horton, W., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Journal of Laboratory Phonology*, *2*, 125–156.

Kirchner, R., Moore, R. K., & Chen, T.-Y. (2010). Computing phonological generalization over real speech exemplars. *Journal of Phonetics*, *38*, 540–547. https://doi.org/10.1016/j.wocn.2010.07.005

Kirov, C., & Wilson, C. (2012). *The specificity of online variation in speech production*. Paper presented at the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan

Kochetov, A., & Pouplier, M. (2008). Phonetic variability and grammatical knowledge: An articulatory study of Korean place assimilation. *Phonology*, *25*, 399–431. https://doi.org/10.1017/S0952675708001553

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 54–81. https://doi.org/10.1016/j.cognition.2007.07.013

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kuijpers, C., & van Donselaar, W. (1998). The influence of rhythmic context on schwa epenthesis and schwa deletion in Dutch. *Language and Speech*, *41*, 87–108. https://doi.org/10.1177/002383099804100105

Kuijpers, C., van Donselaar, W., & Cutler, A. (1996). Phonological variation: Epenthesis and deletion of schwa in Dutch. In H. T. Bunnell (Ed.), Proceedings of the Fourth International Conference on Spoken Language (pp. 94–97). New York, NY: IEEE Press.

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *Journal of the Acoustical Society of America*, *121*, 2261–2271. https://doi.org/10.1121/1.2537393

La Heij, W. L., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture–word tasks. *Psychological Research*, *61*, 209–219. https://doi.org/10.1007/s004260050026

Labov, W. (1966). Social stratification of English in New York City. Washington, DC: Center for Applied Linguistics.

Labov, W. (1968). A study of the non-standard English of Negro and Puerto Rican speakers in New York City: Vol. I. Phonological and grammatical analysis. New York, NY: Columbia University.

Laganaro, M., & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, *55*, 178–196. https://doi.org/10.1016/j.jml.2006.05.001

Laganaro, M., Python, G., & Toepel, U. (2013). Dynamics of phonological–phonetic encoding in word production: Evidence from diverging ERPs between stroke patients and controls. *Brain and Language*, *126*, 123–132. https://doi.org/10.1016/j.bandl.2013.03.004

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, N. Warner, & T. Rietveld (Eds.), Laboratory Phonology VII (pp. 637–676). Berlin, Germany: Mouton de Gruyter.

Laks, B. (1977). Contribution empirique à l'analyse socio-différentielle de la chute de /r/ dans les groupes consonantiques finals. *Langue Française*, *34*, 109–125.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*, 1137–1146. https://doi.org/10.3758/MC.38.8.1137

Lam, T. Q., & Watson, D. G. (2014). Repetition reduction: Lexical repetition in the absence of referent repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 829–843. https://doi.org/10.1037/a0035780

Lavoie, L. M., & Cohn, A. C. (1999). Sesquisyllables of English: The structure of vowel–liquid syllables. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A.C. Bailey (Eds.), Proceedings of the XIVth International Congress of Phonetic Sciences (pp. 109–112). Berkeley, CA: University of California, Linguistics Department.

Léon, P. (1971). Essai de phonostylistique. Ottawa, ON: Didier.

Levelt, W. J. M. (1989). Speaking: From intention to articulation. Cambridge, MA: MIT Press.

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

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269. https://doi.org/10.1016/0010-0277(94)90030-2

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–187. https://doi.org/10.1177/002383096300600306

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Amsterdam, The Netherlands: Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

Linke, M., Bröker, F., Ramscar, M., & Baayen, R. H. (2017). Are baboons learning "orthographic" representations? Probably not. *PLOS ONE*, *12*, e0183876. https://doi.org/10.1371/journal.pone.0183876

Luce, P. (1986). Neighbourhoods of words in the mental lexicon (Research on Speech Perception, Technical Report No. 6). Bloomington, IN: Indiana University.

Magnuson, J. S, Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Malécot, A. (1976). The effect of linguistic and paralinguistic variables on the elision of the French mute-e. *Phonetica*, *33*, 93–112.

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*, 177–192. https://doi.org/10.1016/j.cognition.2010.04.010

Martin, R. C., Miller, M., & Vu, H. (2004). Lexical–semantic retention and speech production: Further evidence from normal and brain-damaged participants for a phrasal scope of planning. *Cognitive Neuropsychology*, *21*, 625–644. https://doi.org/10.1080/02643290342000302

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 539–553. https://doi.org/10.1037/0278-7393.29.4.539

McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*, 243–260. https://doi.org/10.1016/j.cognition.2010.08.019

McQueen, J. M., Dahan, D., & Cutler, A. (2003). Continuity and gradedness in speech processing. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production. Differences and similarities* (pp. 39–78). Berlin, Germany: Mouton de Gruyter.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*, *29*, 524–545. https://doi.org/10.1016/0749-596X(90)90050-A

Meyer, A. S. (1996). Lexical access in phrase and sentence production: Results from picture–word interference experiments. *Journal of Memory and Language*, *35*, 477–496. https://doi.org/10.1006/jmla.1996.0026

Meyer, A. S., Huettig, F., & Levelt, W. J. M. (2016). Same, different, or closely related: What is the relationship between language production and comprehension? *Journal of Memory and Language*, *89*, 1–7. https://doi.org/10.1016/j.jml.2016.03.002

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition*, *66*, B25–B33. https://doi.org/10.1016/S0010-0277(98)00009-2

Meyer, A. S., & van der Meulen, F. F. (2000). Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychonomic Bulletin & Review*, *7*, 314–319. https://doi.org/10.3758/BF03212987

Michel Lange, V., & Laganaro, M. (2014). Inter-subject variability modulates phonological advance planning in the production of adjective–noun phrases. *Frontiers in Language Sciences*, *5*, 43. https://doi.org/10.3389/fpsyg.2014.00043

Miozzo, M., & Caramazza, A. (1999). The selection of lexical–syntactic features in noun phrase production: Evidence from the picture–word interference paradigm. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *25*, 907–922. https://doi.org/10.1037/0278-7393.25.4.907

Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of stimulus onset asynchrony? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 388–391. https://doi.org/10.1037/0278-7393.28.2.388

Mitchell, D. B., & Brown, A. S. (1988). Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 213–222. https://doi.org/10.1037/0278-7393.14.2.213

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103. https://doi.org/10.1016/j.wocn.2005.03.003

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*, 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 244–263. https://doi.org/10.1037/a0012730

Mousikou, P., & Rastle, K. (2015). Lexical frequency effects on articulation: A comparison of picture naming and reading aloud. *Frontiers in Psychology*, *6*, 1571. https://doi.org/10.3389/fpsyg.2015.01571

Munding, D., Dubarry, A.-S., & Alario, F.-X. (2015). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*, 441–462. https://doi.org/10.1080/23273798.2015.1071857

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighbourhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058. https://doi.org/10.1044/1092-4388(2004/078)

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*, 790–804.

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. https://doi.org/10.1016/j.wocn.2017.01.008

Nevins, A. (2011). Phonologically-conditioned allomorph selection. In C. Ewen, E. Hume, M. van Oostendorp, & K. Rice (Eds.), *Blackwell companion to phonology* (pp. 2357–2382). Oxford, UK: Wiley Blackwell.

Nguyen, N. (2012). Representations of speech sound patterns in the speaker's brain: Insights from perception studies. In A. C. Cohn, C. Fougeron, & M. K. Huffman (Eds.), *Oxford handbook of laboratory phonology* (pp. 359–368). Oxford, UK: Oxford University Press.

Nguyen, N., Wauquier, S., Lancia, L., & Tuller, B. (2007). Detection of liaison consonants in speech processing in French. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (Vol. 282, pp. 3–23). Amsterdam, The Netherlands: Benjamins. https://doi.org/10.1075/cilt.282.03ngu

Niebuhr, O., Clayards, M., Meunier, C., & Lancia, L. (2011). On place assimilation in sibilant sequences—Comparing French and English. *Journal of Phonetics*, *39*, 429–451. https://doi.org/10.1016/j.wocn.2011.04.003

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*, 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Nolan, F. (1992). The descriptive role of segments: Evidence from assimilation. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 261–280). Cambridge, UK: Cambridge University Press.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*, 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition, and Neuroscience*, *30*, 1187–1208. https://doi.org/10.1080/23273798.2015.1006238

Ohala, J. J. (1993). Coarticulation and phonology. *Language and Speech*, *36*, 155–170.

Oppermann, F., Jescheniak, J. D., & Schriefers, H. (2010). Phonological advance planning in sentence production. *Journal of Memory and Language*, *63*, 526–540. https://doi.org/10.1016/j.jml.2010.07.00

Ouyang, G., Sommer, W., Zhou, C., Aristei, S., Pinkpank, T., & Abdel Rahman, R. (2016). Articulation artifacts during overt language production in event-related brain potentials: Description and correction. *Brain Topography*, *29*, 791–813. https://doi.org/10.1007/s10548-016-0515-1

Oviatt, S., Levow, G. A., Moreton, E., & MacEachern, M. (1998). Modeling global and focal hyperarticulation during human–computer error resolution. *Journal of the Acoustical Society of America*, *104*, 3080–3098.

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

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*, 2254–2264. https://doi.org/10.3758/APP.72.8.2254

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language*, *69*, 183–195. https://doi.org/10.1016/j.jml.2013.06.002

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

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. A corpus analysis of variant frequency in American English flap production. *Phonetica*, *58*, 254–275. https://doi.org/10.1159/000046178

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, *60*, 45–69. https://doi.org/10.1159/000070454

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2010). Phonological neighbourhood effects in spoken word production: An fMRI Study. *Journal of Cognitive Neuroscience*, *23*, 593–603. https://doi.org/10.1162/jocn.2010.21489

Pham, A. V., Fine, J. G., & Semrud-Clikeman, M. (2011). The influence of inattention and rapid automatized naming on reading performance. *Archives of Clinical Neuropsychology*, *26*, 214–224. doi:https://doi.org/10.1093/arclin/acr014

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam, The Netherlands: Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7: Phonology and phonetics* (pp. 101–139). Berlin, Germany: Mouton de Gruyter.

Pierrehumbert, J. B., Beckman, M., & Ladd, D. R. (2000). Conceptual foundations of phonology as a laboratory science. In P. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge* (pp. 273–303). Oxford, UK: Oxford University Press.

Pierrehumbert, J. B., & Clopper, C. (2010). What is LabPhon? and where is it going. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée

(Eds.), Laboratory phonology 10 (pp. 113–132). Berlin, Germany: Mouton de Gruyter.

Pinnow, E., & Connine, C. M. (2014). Phonological variant recognition: Representations and rules. *Language and Speech, 57*, 42–67. https://doi.org/10.1177/0023830913479105

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 118*, 2561–2569. https://doi.org/10.1121/1.2011150.

Porcaro, C., Medaglia, M. T., & Krott, A. (2015). Removing speech artifacts from electroencephalographic recordings during overt picture naming. *NeuroImage, 105*, 171–180. https://doi.org/10.1016/j.neuroimage.2014.10.049

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25*, 143–170. https://doi.org/10.1016/j.newideapsych.2007.02.001

Racine, I., & Grosjean, F. (2002). La production du E caduc facultatif est-elle prévisible? Un début de réponse. *Journal of French Language Studies, 12*, 307–326. https://doi.org/10.1017/S0959269502000340

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*, 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Rapp, B., Buchwald, A., & Goldrick, M. (2014). Integrating accounts of speech production: The devil is in the representational details. *Language, Cognition and Neuroscience, 29*, 24–27. https://doi.org/10.1080/01690965.2013.848991

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t, d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change, 18*, 55–97. https://doi.org/10.1017/S0954394506060042

Raymond, W. D., Fisher, J. A., & Healy, A. F. (2002). Linguistic knowledge and language performance in English article variant preference. *Language and Cognitive Processes, 17*, 613–662. https://doi.org/10.1080/01690960143000380

Rialland, A. (1986). Schwa et syllabes en français. In L. Wetzels, & E. Sezer (Eds.), Studies in compensatory lengthening (pp. 187–226). Dordrecht, The Netherlands: Foris.

Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language, 107*, 102–113. https://doi.org/10.1016/j.bandl.2008.01.008

Roelofs, A. (1997). The WEAVER Model of word-form encoding in speech production. *Cognition, 64*, 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2014). Integrating psycholinguistic and motor control approaches to speech production. *Language, Cognition, and Neuroscience, 29*, 35–37. https://doi.org/10.1080/01690965.2013.852687

Roon, K. D., & Gafos, A. I. (2015). Perceptuo-motor effects of response–distractor compatibility in speech: beyond phonemic identity. *Psychonomic Bulletin & Review, 22*, 242–250. https://doi.org/10.3758/s13423-014-0666-6

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237. https://doi.org/10.3758/PBR.16.2.225

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighbourhood effects in speech production through single trial analysis. *Cognitive Psychology, 68*, 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Scarborough, R. (2010). Lexical and contextual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), Laboratory phonology 10 (pp. 557–586). Berlin, Germany: Mouton de Gruyter.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighbourhood density effects in speech production and perception. *Journal of the Acoustical Society of America, 134*, 3793–3807. https://doi.org/10.1121/1.4824120

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics, 41*, 249–263. https://doi.org/10.1016/j.wocn.2013.03.007

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language, 48*, 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9

Schnur, T. T. (2011). Phonological planning during sentence production: Beyond the verb. *Frontiers in Language Sciences, 2*, 319. https://doi.org/10.3389/fpsyg.2011.00319

Schriefers, H. (1992). Lexical access in the production of noun phrases. *Cognition, 45*, 33–54. https://doi.org/10.1016/0010-0277(92)90022-A

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 841–850.

Schweitzer, A., & Möbius, B. (2004). Exemplar-based production of prosody: Evidence from segment and syllable durations. In B. Bel & I. Marlien (Eds.), Proceedings of Speech Prosody 2004 (pp. 459–462). SProSIG.

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition, 133*, 140–155. https://doi.org/10.1016/j.cognition.2014.06.013

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustical Society of America, 139*, EL31–EL37. https://doi.org/10.1121/1.4942544

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper, & E. C. T. Walker (Eds.), Sentence processing: Psycholinguistic studies presented to Merrill Garrett (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*, 422–429. https://doi.org/10.3758/BF03194890

Skousen, R. (1989). Analogical modeling of language. Dordrecht, The Netherlands: Kluwer.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73*, 205–246. https://doi.org/10.1016/S0010-0277(99)00053-0

Snoeren, N. D., Hallé, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics, 34*, 241–268. https://doi.org/10.1016/j.wocn.2005.06.001

Spalek, K., Bock, K., & Schriefers, H. (2010). A purple giraffe is faster than a purple elephant: Inconsistent phonology affects determiner selection in English. *Cognition, 114*, 123–128. https://doi.org/10.1016/j.cognition.2009.09.011

Spieler, D. H., & Griffin, Z. M. (2006). The influence of age on the time course of word preparation in multiword utterances. *Language and Cognitive Processes, 21*, 291–321. https://doi.org/10.1080/01690960400002133

Spinelli, E., & Ferrand, L. (2005). Psychologie du langage: L'écrit et le parlé, du signal à la signification. Paris, France: Armand Colin.

Spinelli, E., & Gros-Balthazard, F. (2007). Phonotactic constraints help to overcome effects of schwa deletion in French. *Cognition, 104*, 397–406. https://doi.org/10.1016/j.cognition.2006.07.002

Stemberger, J. P. (1983). Speech errors and theoretical phonology: A review. Bloomington, IN: Indiana University Linguistics Club.

Stemberger, J. P., & Lewis, M. (1986). Reduplication in Ewe: Morphological accommodation to phonological errors. *Phonology Yearbook, 3*, 151–160.

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, *50*, 163–178. https://doi.org/10.1016/j.specom.2007.07.005

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 484–503. https://doi.org/10.1080/23273798.2015.1120878

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage*, 163. https://doi.org/10.1016/j.neuroimage.2017.09.041

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338. https://doi.org/10.1016/j.jml.2004.11.004

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, *37*, 276–296. https://doi.org/10.1016/j.wocn.2009.03.004

Torreira, F., & Ernestus, M. (2011). Vowel elision in casual French: The case of vowel /e/ in the word c'était. *Journal of Phonetics*, *39*, 50–58. https://doi.org/10.1016/j.wocn.2010.11.003

Trask, R. L. (Ed.). (2000). Dictionary of historical and comparative linguistics (1st ed.). Edinburgh, Scotland: Routledge.

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), Perspectives on formulaic language: Acquisition and communication (pp. 151–173). London, UK: Continuum.

Tremblay, A., & Tucker, B. (2011). The effects of *N*-gram probabilistic measures on the recognition and production of four-word sequences. *Mental Lexicon*, *6*, 302–324. https://doi.org/10.1075/ml.6.2.04tre

Turk, A. (1992). The American English flapping rule and the effect of stress on stop consonant durations (pp. 103–134). Ithaca, NY: Cornell University Press.

Urigüen, J. A., & Garcia-Zapirain, B. (2015). EEG artifact removal-state-of-the-art and guidelines. *Journal of Neural Engineering*, *12*, 031001. https://doi.org/10.1088/1741-2560/12/3/031001

Van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, *12*, 1–21. https://doi.org/10.1016/0167-6393(93)90015-D

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

van Son, R. J. J. H., Bolotova, O., Lennes, M., & Pols, L. C. W. (2004). Frequency effects on vowel reduction in three typologically different languages (Dutch, Finnish, Russian). In Proceedings of Interspeech (pp. 1277–1280). Seoul, South Korea: ISCA Archive. www.isca-speech.org/archive/interspeech_2004

van Son, R. J. J. H., & Pols, L. C. W. (1999). Acoustic description of consonant reduction. *Speech Communication*, *28*, 125–140. https://doi.org/10.1016/S0167-6393(99)00009-6

van Son, R. J. J. H., & Pols, L. C. W. (2003). An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness. In M. J. Solé, D. Recasens, & J. Romero (Eds.), Proceedings of the International Congress of Phonetic Sciences (pp. 2141–2144). ICPhS Archive. www.internationalphoneticassociation.org/icphs/icphs2003

Verluyten, S. P. (1988). La phonologie du français. Amsterdam, The Netherlands: Benjamins.

Vigario, R., & Oja, E. (2008). BSS and ICA in neuroinformatics: From current practices to open challenges. *IEEE Reviews in Biomedical Engineering*, *1*, 50–61. https://doi.org/10.1109/RBME.2008.2008244

Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*, 491–504. https://doi.org/10.3758/BF03196091

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*, 760–770. https://doi.org/10.1080/01690960500287196

Walsh, M., Schütze, H., Möbius, B., & Schweitzer, A. (2007). An exemplar-theoretic account of syllable frequency effects. In J. Trouvain & W. J. Barry (Eds.), Proceedings of the International Congress of Phonetic Sciences (pp. 481–484). Dudweiler, Germany: Pirrot.

Warner, N., Jongman, A., Cutler, A., & Mücke, D. (2001). The phonological status of Dutch epenthetic schwa. *Phonology*, *18*, 387–420. https://doi.org/10.1017/S0952675701004213

Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology*, *24*, 147–185.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic Tac Toe: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, *106*, 1548–1557. https://doi.org/10.1016/j.cognition.2007.06.009

Wolfram, W. A. (1969). A sociolinguistic description of Detroit Negro speech. Washington, DC: Center for Applied Linguistics.

Wood, S. A. J. (1996). Assimilation or coarticulation? Evidence from the temporal co-ordination of tongue gestures for the palatalization of Bulgarian alveolar stops. *Journal of Phonetics*, *24*, 139–164. https://doi.org/10.1006/jpho.1996.0009

Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381. https://doi.org/10.1006/jmla.1997.2517

Wright, R. A. (2004). Factors of lexical competition in vowel articulation. In J. Local, R. Ogden, & R. Temple (Eds.), Phonetic interpretation (pp. 75–87). Cambridge, UK: Cambridge University Press.

Wright, S., & Kerswill, P. (1989). Electropalatography in the analysis of connected speech processes. *Clinical Linguistics & Phonetics*, *3*, 49–57. https://doi.org/10.3109/02699208908985270

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/10.1371/journal.pone.0074746

Yuen, I., Davis, M. H., Brysbaert, M., & Rastle, K. (2010). Activation of articulatory information in speech perception. *Proceedings of the National Academy of Sciences*, *107*, 592–597. https://doi.org/10.1073/pnas.0904774107

Zue, V. W., & Laferriere, M. (1979). Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America*, *66*, 1039–1050.

Zwicky, A. (1986). The general case: Basic form versus default form. In Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society (pp. 305–314). Berkeley, CA: University of California, Berkeley Linguistics Society.