# 8 The Effect of Usage Predictability on Phonetic and Phonological Variation

RORY TURNBULL

## Introduction

In English, an optional phonological rule permits the deletion of unstressed schwa in certain environments. For example, *memory* /ˈmɛməɹi/ can be pronounced with the schwa, as [ˈmɛməɹi], or without, as [ˈmɛmɹi]. Likewise, *camera* /ˈkæməɹə/ is equally acceptable as [ˈkæməɹə] or [ˈkæmɹə]. The word *mammary* /ˈmæməri/ is almost identical to *memory* in phonological form. The only difference is the stressed vowel, which is the same as in the word *camera*. Despite *memory* and *camera* both being acceptable targets of the schwa deletion rule, the word *mammary* only very rarely undergoes schwa deletion. That is, *mammary* is nearly always [ˈmæməri] and never [ˈmæmri]. Hooper (1976) explained this discrepancy in terms of usage frequency—*mammary* is a low-frequency word, whereas *memory* and *camera* are both much more common.

This example serves as an introduction to the central question of this chapter—how do usage-based factors, such as frequency and predictability more broadly, interact with phonetic and phonological variation? Why do these relationships exist? This chapter surveys work on these questions and provides some perspectives on current and future research. This chapter is necessarily selective, due to the size of this topic and length restrictions of this review, but I've tried to highlight suggestions for further reading where relevant.

## Defining predictability

Predictability can be defined as the extent to which the presence of a linguistic element—such as a word or a phoneme—can be predicted by an idealized observer. Usually, this prediction relies on properties such as patterns of language use or current discourse context. In this section, I introduce frequency, semantic predictability, and discourse mention as specific operationalizations of predictability.

*The Handbook of Usage-Based Linguistics*, First Edition. Edited by Manuel Díaz-Campos and Sonia Balasch.

## *Frequency*

Frequency has long been associated with usage-based approaches to linguistics, with early work leaning heavily on evidence and reasoning associated with lexical frequency (e.g., Hooper, 1976; Phillips, 1981, 1984). Lexical frequency values are usually calculated from a corpus by counting how often each word appears. These values are then often normalized by dividing against some value to give word frequency per million words, which is an interpretable measure that can be meaningfully compared between corpora.

## *Semantic predictability*

Also called *contextual predictability*, this variable denotes how predictable a target word is given the semantic content of the rest of the sentence. Classically, this variable has been manipulated using stimulus sentences with "(un)predictable" keywords. Consider (1) and (2) (Kalikow, Stevens, & Elliott, 1977). Both sentences end in the word *bay*, predictably in (1) but unpredictably in (2). The meaning of the sentences constrains the plausible completions in (1), while there are many more potential completions in (2).

(1) The boat sailed across the bay.
(2) Mr. Smith knew about the bay.
(3) The dealer shuffled the cards.

This manipulation is intuitive—English speakers generally agree which sentences are predictable or unpredictable—but typically binary and unquantified. Consider the final word *cards* in (3), another predictable sentence from Kalikow et al.'s (1977) set. By far, *cards* is the most plausible completion of this sentence, along with *deck*. In contrast, *bay* is just one of many probable completions for (1)—words such as *river*, *ocean*, *sea*, *lake*, *harbor*, and many others are just as likely. The binary classification ignores this intuition, however, and treats (1) and (3) as equally predictable.

It is possible to overcome this limitation by directly quantifying the predictability of the target word on some continuous scale. The two most common methods of quantification are subjective human judgments via a cloze task, and deriving probability estimates via a language model trained on a corpus. The cloze task is time- and labor-intensive but may provide a closer match to language users' linguistic representations than probabilities derived from a language model (Smith & Levy, 2011).

## *Discourse mention*

This variable looks beyond the sentence context and considers the entire discourse. The concept of "mention," and the closely related notions "givenness" and "salience," are intended to capture how well-established a particular word is in a conversation.

Mention is usually simply defined as the number of times a word has been used over some time domain—typically an interaction or monologue. For example, in (4), the first occurrence of *memorize* is the first mention and the second occurrence is the second mention. This definition is one based purely on form, rather than meaning. A consequence of this definition is that the instances of the word *book* in (5) are first

and second mention respectively, even though they have different real-world referents. Similarly, the word *door* in (6) is classified as first mention, even though it may already be salient to the participants of an interaction.

(4) There I'd sit and memorize arithmetic tables and memorize state capitols and major exports of many lands.
(5) I'll take this book back to the library and get a new book.
(6) You can leave the door open, thanks.

Mention is therefore a rather crude measure. More nuanced approaches to discourse mention directly examine the meaning of words in their proper contexts. This examination usually involves classifying the referents of words or phrases as "given" or "new." Givenness has a long history in both the functional linguistic and formal semantic traditions (e.g., Kruijff-Korbayová & Steedman, 2003; Rochemont, 2016). Whether the referent of a word or utterance is "given" varies between theories. For example, a referent or proposition may be "given" if it is "anaphorically recoverable" (Halliday, 1967, p. 208) or if "it is entailed by the prior discourse" (Schwarzschild, 1999, p. 147). Generally, the idea is that "given words" refer to items which are salient in the prior discourse, while "new words" do not.

## Predictability and phonetic and phonological variation

### *Continuous variation*

Continuous variation occurs along a spectrum, with an infinite range of possible values. For example, Chafe (1974, p. 112) suggested that given material may be attenuated "by pronouncing the items that convey such material with lower pitch and weaker stress." Chafe's example was impressionistic, and it was not based on experimental data, but there is now a growing evidence base of experimental work on various variables. In this section, we cover speech intelligibility, duration, and spectral variation in vowels.

### *Speech intelligibility*

There is a robust body of evidence suggesting that unpredictable speech tends to be less intelligible than predictable speech. In terms of semantic predictability, words in predictable contexts, such as the word *bay* in (1), have long been known to be more intelligible than words in unpredictable contexts, such as the word *bay* in (2) (Kalikow et al., 1977, *inter alia*). This effect is partly driven by phonetic differences between words in high- and low-predictability contexts, but the effect remains after the phonetic factors are controlled for. For example, Clopper (2012) compared the intelligibility of the final word of high-predictability sentences like (7) with the final word of semantically anomalous sentences like (8):

(7) The judge is sitting on the bench.
(8) The arm is riding on the bench.

In these stimuli, the instances of the word *bench* were acoustically identical—taken from the same original recordings. Clopper (2012) showed that the intelligibility for words in high-predictability sentences like (7) was around 30 percentage points higher than in semantically anomalous sentences like (8). Because the words under comparison were acoustically identical, the only factor influencing this difference in intelligibility is the predictability of the word given the context.

Other dimensions of predictability can influence perception too. Ranbom and Connine (2007) examined the phenomenon of flapping in American English, which can apply optionally in intervocalic /nt/ sequences, e.g., *center* [sɛɾ̃ɚ], *gentle* [ʤɛɾ̃l̩]. In a series of lexical decision tasks, they first confirmed that the flapped (reduced) stimuli take longer for the listener to process than the canonical stimuli. Next, they compared words which have a high rate of flapping (e.g., *center)* to words which have a lower rate of flapping (e.g., *gentle)*. The frequently-reduced words had a smaller flapping disadvantage than the infrequently-reduced words. That is, the processing slowdown from flapping was attenuated by virtue of the words commonly being flapped. (See also Bürki, Ernestus, & Frauenfelder, 2010, for converging evidence from French schwa deletion.) Although this experiment examines a very different domain from that of Clopper (2012), the common thread remains: more predictable content, whether that be lexical identity or the presence or absence of flapping, is easier for perceivers to process.

## *Duration*

Zipf (1949) famously observed that high-frequency English words tend to have fewer letters than low-frequency words. This relationship has long been observed in the phonetic domain too (Aylett & Turk, 2004; Whalen, 1991; Wright, 1979), where higher-frequency words tend to have shorter durations than lower-frequency words. This effect has been observed in several languages in addition to English (Pluymaekers, Ernestus, & Baayen, 2005; Strunk et al., 2020, Tyrone & Mauk, 2010, 2012).

A popular explanation for this pattern appeals to motor routines. The high-frequency words are, essentially, more well-practiced, and the articulators can produce the word more fluently and more quickly (Bybee, 2006). However, Gahl (2008) and Lohmann (2018) demonstrated frequency effects even among homophones (e.g., *time* and *thyme*) such that the higher-frequency lemma was produced with shorter duration than the lower-frequency lemma. This finding cannot easily be explained in terms of motor routinization, suggesting that there are additional factors at play.

In addition to frequency, word duration has been observed to negatively correlate with semantic predictability (Engelhardt & Ferreira, 2014; Moore-Cantwell, 2013; Turnbull, 2019) and positively correlate with lexical informativity (Seyfarth, 2014). Again, the overall pattern is that the more predictable (i.e., less informative) a word is, the shorter its duration is.

Several studies have examined segment duration, especially vowel duration. Like the word duration findings, vowel duration tends to be negatively correlated with word frequency, such that vowels in high-frequency words tend to be shorter than vowels in low-frequency words (e.g., Burdin & Clopper, 2015; Goldrick & Blumstein, 2006). Vowel duration has also been observed to be negatively correlated with semantic

predictability (Shaw & Kawahara, 2019; Turnbull, 2019) and segment informativity (Cohen Priva, 2008; 2015). For consonants, Van Son and Van Santen (2005) found that consonant durations in American English are negatively correlated with the frequency of the consonant class—e.g., the class of labials—but not lexeme frequency (see also Van Son & Pols, 2003, for converging evidence from Dutch).

Finally, discourse mention has a strong effect on word and vowel durations, with second and subsequent mentions of words being shorter than first mentions (Baker & Bradlow, 2009; Burdin, Turnbull, & Clopper, 2014; Fowler & Housum, 1987; Turnbull, 2019). In addition to English, second mention reduction has been reported in Dutch (Pluymaekers et al., 2005), Thai (Vajrabhaya & Kapatsinski, 2011), Mandarin (Wiener, Speer, & Shank, 2012), Dutch Sign Language (Hoetjes, Krahmer, & Swerts, 2014), Korean (Lam & Marian, 2015), and Papuan Malay (Kaland & Himmelmann, 2020).

## *Spectral variation in vowels*

Vowels vary in multiple dimensions in addition to duration. Overall, it appears that vowels which are in more predictable words are pronounced less precisely—i.e., they are more schwa-like—than vowels in less predictable words. The precision of vowel articulation is usually measured as the vowel's distance from the center of the vowel space. This pattern has been observed for English vowels in words with high lexical frequency (Burdin et al., 2014, Munson & Solomon, 2004) and for vowels in words which are contextually predictable (Aylett & Turk, 2004; Clopper & Pierrehumbert, 2008; but see Clopper, Turnbull, & Burdin, 2018, for a more complex set of results).

Another spectral property of vowels is their fundamental frequency. There is evidence of a relationship between predictability and pitch peak height in American English (Burdin & Clopper 2015; Turnbull, 2017; Watson, Arnold, & Tanenhaus, 2008) and Japanese (Hashimoto, 2020): generally, higher-predictability words tend to have lower pitches than less predictable words. Taken together, these results demonstrate an association between high predictability and phonetic reduction.

## *Discrete variation*

Discrete variation is categorical rather than continuous, such as the presence versus absence of a particular phoneme, or the application or non-application of a phonological process. Most prior work on discrete variables has examined phoneme deletion. Perhaps one of the most well-studied phenomena is /t/ and /d/ deletion in English (e.g., Coetzee, 2004; see, in this volume, Kapatsinski, Chapter 5; and Brown, Chapter 10). Bybee (2000) reported high-frequency words in Chicano English, such as *just*, as having higher /t/-deletion rates than low-frequency words, such as *jest*. Similar findings were reported by Coetzee and Kawahara (2013) for Columbus English. Further work by both Raymond, Dautricourt, and Hume (2006), and Raymond, Brown, and Healy (2016) has challenged the direct link between deletion and word frequency, arguing instead for the role of contextual predictability. Regardless of the precise origin of the effect, however, the trend is clear: deletion is more common in more predictable contexts than in less predictable ones.

Similarly, Cohen Priva (2008, 2015) found segment informativity to be a reliable predictor of deletion (see also Cohen Priva, 2017). Turnbull (2018) likewise analyzed phoneme deletion in American English and Japanese, replicating prior findings that sounds in high-frequency words undergo deletion more often than in low-frequency ones.

Other discrete variables include the outcome of variable phonological rules. In many varieties of English, the phonemes /t/ and /d/ can be produced as alveolar plosives [t] and [d], or they can neutralize to a voiced alveolar flap [ɾ]. Biro, Olmstead, and Viswanathan (2022) found that speakers can modulate their application of this rule in response to communicative demands. Specifically, in contexts where a word was misheard, such as *petal* being misperceived as *pedal*, experimental participants were more likely to change their production from [ɾ] to [t], relative to other kinds of mishearing errors. Nevertheless, there was still overall a high proportion of flap productions, and there was considerable variation in how likely each participant was to alter their production.

American English /t/ can also be glottalized—realized as [ʔ]—in the same environment as [ɾ] (Eddington & Channer, 2010). Kilbourn-Ceron, Clayards, and Wagner (2020) investigated the distribution of these two variants in word final-position, examining phrases such as *out of* and *quite easy*. Their analysis showed more glottalization on lower-predictability words than on higher-predictability ones. This finding appears to be inconsistent with other trends observed in the literature: generally, higher-predictability items are the ones that undergo reduction. Kilbourn-Ceron et al. framed their interpretation in terms of the Production Planning Hypothesis, which claims that a word's predictability is not the proximate cause of reduction.

Finally, voiced geminates in Japanese only exist in loanwords and can be optionally devoiced in words with another voiced obstruent (e.g., *doggu~dokku* "dog") (Kawahara, 2006). In addition to many other factors, Kawahara (2011) has demonstrated that word frequency is correlated with devoicing rates. The application of the devoicing rule, which can be construed as a form of reduction, is more common in high-frequency forms.

## Current approaches

Clopper and Turnbull (2018) classified approaches to phonetic variation and predictability into three broad groups: (1) "listener-oriented"; (2) "talker-oriented"; and (3) "passive evolutionary" models. The terminology of "listener" versus "talker" orientation was first used in this context by Bradlow (2002), although Lindblom (1990, p. 433) wrote of the "tug-of-war and balancing of production-oriented and listener-oriented forces." Indeed, the opposition between the perspective of the producer and the perceiver in the analysis of language dates to at least Von Savigny's (1976) account of conventional meaning.[1]

Unfortunately, the words "listeners" and "talkers" inadvertently exclude signed languages and focus only on the oral modality. As these models are fruitfully applied to both signed and spoken languages, this exclusion is likely unintentional rather than principled. Indeed, while signed languages are understudied, there is preliminary evidence that they are subject to similar pressures with regards to frequency and predictability effects (e.g., Hoetjes et al., 2014; Occhino, 2016; Tyrone & Mauk, 2010;

Tkachman et al., 2019). It is not always clear that the authors of these models even considered signed languages in their theorizing; nevertheless, I propose that these theories make predictions about signed languages to the same extent that they make predictions about spoken languages. Accordingly, I adopt modality neutral language where appropriate, changing Clopper and Turnbull's "listener" to "perceiver" and their "talker" to "producer."

## *Perceiver-oriented models*

One of the earliest explicit formulations of a perceiver-oriented model is Lindblom's (1990) Hyper- and Hypospeech (H&H) Theory, which states that predictability-based speech modulations are an adaptive feature that serve to enhance the communicative demands of a speech situation. H&H Theory posits that producers are trying to achieve two goals, which are sometimes at odds: maximizing intelligibility while minimizing effort. It follows that highly predictable items can "afford" to undergo phonetic reduction because their high predictability ensures that they are recoverable from context. Low predictability items, on the other hand, cannot be guessed from context and, therefore, must be produced in a clear (non-reduced) way. Speech is considered to be a delicate balance between the needs of the producer and those of the perceiver. Clopper and Turnbull (2018) classified this account (and others) as perceiver-oriented due to the focus on the needs of the perceiver. This classification is in contrast with other groups of models in which the perceiver's communicative needs are not relevant.

The idea that speech communication is essentially an optimization problem, minimizing difficulty for both producer and perceiver, can be found in Lindblom's earlier work too (e.g., Liljencrants & Lindblom, 1972; Lindblom, 1983). Indeed, the intuition that non-reduced speech exists to boost comprehension (thereby benefiting the perceiver) has been expressed in the literature several times. For example, Fowler and Housum (1987, p. 489) inferred that "talkers may attenuate their productions of words when they can do so without sacrificing communicative efficacy." Hunnicutt (1985, p. 53) framed hyperspeech as the producer "compensat[ing] for the lower-redundancy context, so that the resulting word in that context is more intelligible in isolation." Lieberman (1963, pp. 182–183) similarly framed his results in terms of explicit knowledge and goals on the part of the talkers, who "pronounced these words with less care and with less stress because *they know that listeners would be able to compensate* for the lack of acoustic information by making use of the contextual semantic and grammatical information of the entire sentence" (emphasis mine). An early nod to this trade-off between producer and perceiver effort can be seen in Malmberg's (1954, p. 65) observation that the producer "a tendance à obtenir le maximum d'effet avec un minimum d'effort" [tends to obtain the maximum effect with the minimum effort], and Zipf's (1949) "principle of least effort." Lindblom was simply one of the first to formalize the notion within a self-contained theory.

More recently, other models have been proposed which differ from H&H Theory in important ways yet retain the locus of explanatory power in the producer's orientation towards the needs of a(n) (idealized) perceiver. One such model is Message Oriented Phonology (Hall et al., 2016; Hall et al., 2018; Hume, Hall, & Wedel, 2016). This model emphasizes the role of phonology as an interface between the acoustic signal and *messages*, which Hall and colleagues define as meaning-bearing units (i.e., morphemes). This focus constitutes a reorientation in the study of phonology

away from phonemic units themselves and instead towards the role of these units in information transmission. The authors rely on Information Theory to model the transmission of messages, and Bayesian inference to model the reception of those messages by the perceiver. This model enjoys a degree of empirical support (e.g., Kawahara & Lee, 2018; Turnbull et al., 2018), although alternative interpretations are possible (e.g., Kilbourn-Ceron et al., 2020).

Before moving on, it is important to consider some of the assumptions that underlie this group of models. Perhaps most important is the idea that phonetically reduced speech is less intelligible than non-reduced speech and that high-frequency or high-predictability speech is more intelligible than low-frequency or low-predictability speech. These assumptions may seem to be intuitive, but some remarks on their empirical support are in order.[2]

Several studies have found that phonetically reduced speech is generally less intelligible in isolation than phonetically non-reduced speech (e.g., Ernestus, Baayen, & Schreuder, 2002; Hawkins & Warren, 1994; Tucker & Warner, 2007). Contextual support can help enormously, however. Ernestus et al. (2002) examined the perception of "massive reductions," such as the production of Dutch *natuurlijk* "of course" /naˈtyrlək/ as [tyk]. In isolation, these words were perceived correctly just over 50% of the time. When the words were presented in the context of their original recording, however, intelligibility increased to more than 90%. This effect also holds for words which are only slightly reduced in a speech-in-noise task (e.g., Duffy & Giolas, 1974; Kalikow et al., 1977). In sum, context can improve perception of otherwise completely unintelligible words (Janse & Ernestus, 2011).

The other assumption is that high-frequency or highly predictable speech is, all else being equal, more intelligible than low-frequency or unpredictable speech. There is a wide body of evidence supporting this assumption; however, note that it does not in itself lend credence to the perceiver-oriented perspective, but it does support its plausibility. These assumptions are also relevant for the passive evolutionary models. Next, we will examine the producer-oriented models, which provide a natural counterpoint to the perceiver-oriented perspective.

## *Producer-oriented models*

In contrast to perceiver-oriented models, producer-oriented models involve mechanisms which lack any consideration of a communicative partner at all. In these models, phonetic reduction is a consequence of interactions between different mechanisms within the speech production system (e.g., Bell et al., 2009). The specific details of these mechanisms vary from theory to theory but are unified by the idea that phonetic variation is not modulated by the needs of the (idealized) perceiver.

In one well-known study, Baese-Berk and Goldrick (2009) performed a series of experiments in English designed to probe the nature of phonetic variation in response to communicative contexts. They observed voice onset time (VOT) enhancement in words such as *cod*, which forms a voicing pair with *god*, relative to words such as *cog*, which is not part of a minimal pair with **gog*. This VOT enhancement serves to make the words in the minimal pair more distinct from each other, which at first appears to support a perceiver-oriented account of speech production. However, this enhancement was observed even when the communicative context made confusion between *cod* and *god* impossible. Why bother enhancing *cod* to make it distinct from

*god* when the context makes it clear that *god* is not possible? Baese-Berk and Goldrick (2009) explained their results in terms of lexical competition during speech production: it is the presence of *god* in the lexicon, not communicative pressures, that causes the enhancement of *cod*. Specific details of this model were further explicated by Goldrick, Baker, Murphy, and Baese-Berk (2011).

Another prominent producer-oriented account is the Production Planning Hypothesis (Kilbourn-Ceron, 2017; Tanner, Sonderegger, & Wagner, 2017; Wagner, 2012). Originally developed to account for locality effects in sandhi phenomena (Wagner, 2011), the hypothesis carries implications for a wide variety of word boundary effects. Central to the hypothesis is the idea that speech is planned in small "chunks," often only a few words at a time, and two adjacent words may or may not be members of the same planning chunk (Levelt, Roelofs, & Meyer, 1999; Wheeldon, 2012). When words are members of the same chunk, sandhi effects such as flapping or assimilation are much more likely. When words are members of different chunks, boundary enhancement may occur. The connection here with predictability is indirect: highly predictable sequences of words are much more likely to be planned in the same chunk than in separate chunks, which then leads to predictable sequences undergoing reduction (e.g., assimilation) more than unpredictable sequences. The relationship with predictability is therefore epiphenomenal, and the true source of the effect is the architecture of the speech production mechanism.

## *Passive evolutionary models*

The final category of models classified by Clopper and Turnbull (2018) is the passive evolutionary one. These models are *passive* in that they do not posit an active cognitive force or mechanism that effects the change, and they are *evolutionary* in that they rely on a mechanism similar to that of natural selection in biological evolution. These models typically rely on an exemplar framework, whereby perceivers' mental representations of words consist of detailed auditory traces of every token of speech perceived (see Pierrehumbert, 2016). This approach contrasts with classical lexical representations in the generative tradition, where it is assumed that the lexicon is structured to minimize long term memory requirements.

Here, I focus on Pierrehumbert's (2002) exemplar-based production model. Further refinements and extensions to these ideas have been made by Blevins and Wedel (2009), Todd, Pierrehumbert, and Hay (2019), Tupper (2014), Wedel (2006), and others. As is standard in exemplar approaches, the perceiver's perception of each word is stored in memory as an *exemplar* with detailed sensory information. Production is intrinsically linked to perception—the process of production involves sampling from the (generalized) exemplar cloud. In an exemplar framework, you are what you eat.

Speech perception is subject to "top-down" influences such as context, and also "bottom-up" influences such as phonetic reduction. All else being equal, a high-frequency word should be more predictable—and therefore more guessable—than a low-frequency word, and a clearly pronounced word should be more perceptible than an unclearly pronounced word. Together, these relationships yield a hierarchy of perceptibility with clearly pronounced high-frequency words at the top and unclearly pronounced low-frequency words at the bottom. Because word tokens that are less perceptible are less likely to influence the exemplar category (how can a token be added to the relevant category when the perceiver was not able to classify it?),

the consequence of this hierarchy is that unclearly pronounced low-frequency word tokens are less likely to be perceived, and thereby less likely to be produced. Using the terminology of natural selection, they have a lower "fitness" than other tokens. High-frequency word tokens, on the other hand, can rely on contextual support and do not need to be pronounced clearly.

This model sketches a scenario where we observe the standard frequency effects—high-frequency words are commonly reduced while low-frequency words are not—without recourse to any *active* phonological or cognitive processes that exert an influence on speech production. Instead, the observed relationship is simply a consequence of the mechanisms of speech production and perception within an exemplar framework. Summarizing these ideas, Silverman (2012, p. 147) described apparent linguistic pressures towards anti-homophony and similar processes as "a *passive* result of the pressures that inherently act upon the interlocutionary process."

## Pending issues

This penultimate section considers residual issues with the three sets of current approaches outlined above. First, we consider some outstanding issues that remain unresolved, both theoretical and empirical. Then we discuss the potential role of "hybrid" models that seek to combine aspects of these accounts, and, finally, we consider the question of the explanatory scope of these accounts.

### *Hybrid models*

Some scholars have called for "hybrid" models that incorporate different aspects of each class of models into a single, unified one. Watson's (2010) "Multiple Sources" view attributes different kinds of phonetic reduction to different cognitive sources and enjoys some degree of empirical support (Lam & Watson, 2010, 2014). Turnbull (2015, pp. 183–184) criticized technical aspects of Watson's model but repeated the call for hybrid models to be investigated more seriously. The difficulty with these models in principle is in determining which processes and mechanisms apply in which contexts: "there should be a principled way of predicting when the listener-oriented constraint is operating, and when the speaker-oriented constraint is operating" (Tabain, 2001, p. 58, attributed to Anders Löfqvist).

This kind of approach is appealing due to empirical and theoretical problems with each of the three sets of accounts. For instance, some of the perceiver-based models explicitly require that each language user have a model of their interlocutor's mental state, updated in real time; this requirement is likely unrealistic (Turnbull 2019). For some producer-based models, the nature of the relationship between lexical activation and articulation is left unspecified, making specific theoretical claims difficult to verify. See Clopper and Turnbull (2018) for more discussion of these issues. Hybrid models, which selectively adapt portions of each account which are consistent and appealing, have a strong potential for guiding our future understanding.

### *Mismatch of levels*

Language use is a complex entity, and there are multiple levels at which we can analyze a given phenomenon. Much of the work reviewed above has assumed that

the word-level is the basic level for predictability effects: we speak of predictable and unpredictable words. Yet the concept of predictable phonemes or phrases is entirely coherent, and early work on multi-word units has shown promising results (e.g., Arnon & Cohen Priva, 2013; Kilbourn-Ceron, 2017).

A related question is that of intentionality and awareness. For example, Biro et al. (2022) showed that at least some speakers will directly modify their speech in response to explicit feedback that suggests a miscommunication. Similarly, studies on "speaking style" have often induced "clear speech" (as opposed to "plain" or "lab" speech) by asking participants to speak to an (imagined or real) interlocutor with a hearing impairment (e.g., Smiljanić & Bradlow, 2009). This is clearly a perceiver-oriented effect, but it is consciously and deliberately controlled, and so is qualitatively quite different from the subtle effects of frequency on phonetic reduction. Is the perceiver-oriented account intended to cover both automatic and deliberate phenomena? What about cases where the intentionality is less clear, such as infant- or pet-directed speech where the talkers may be less aware of their altered speech style (cf. Burnham, Kitamura, & Vollmer-Conna, 2002; Uther, Knoll, & Burnham, 2007)?

We can ask similar questions of the other accounts. How far does the producer-oriented account extend? Are there truly no perceiver-oriented aspects to speech production? Can the passive evolutionary model be scaled up to cover the (non-)application of phonological processes?

## Final remarks

Predictability, broadly construed, is in a complex relationship with many aspects of phonetic and phonological variation. The overall trend is that linguistically predictable items tend to be more phonetically reduced, and more likely to undergo processes of markedness reduction, than less predictable items; however, there are exceptions. This chapter has reviewed three categories of theoretical models dealing with these relationships: perceiver-oriented, producer-oriented, and passive evolutionary models. While none of these approaches are, alone, able to account for all of the empirical findings in the field, they lay the groundwork for future research on hybrid models and alternative explanations.

## NOTES

1 Von Savigny describes his article as an explication of Wittgenstein's *Philosophical Investigations*.

2 These assumptions can even be found in antiquity. Horace wrote in *Ars Poetica* (25–26) "Brevis esse laboro // obscurus fio" [I labor to be brief, and I become obscure]. While he may not have been talking about phonetic reduction per se, the point remains: brevity leads to unintelligibility. I am grateful to an anonymous reviewer for noting that Horace's adage also suggests that the act of reduction is effortful, contrary to contemporary thinking.

## REFERENCES


Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371.

Aylett, M. ,& Turk, A. E. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

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

Baker, R., & Bradlow, A. R. (2009). Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech*, *52*(4), 391–413.

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*(1), 92–111.

Biro, T., Olmstead, A. J., & Viswanathan, N. (2022). Talker adjustment to perceived communication errors. *Speech Communication*, *138*, 12–25.

Blevins, J. & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, *26*, 143–183.

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (vol. 7, pp. 241–273). Berlin: Mouton de Gruyter.

Burdin, R. S., & Clopper, C. G. (2015). Phonetic reduction, vowel duration, and prosodic structure. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (p. 378). Glasgow: The University of Glasgow.

Burdin, R. S., Turnbull, R., & Clopper, C. G. (2014). Interactions among lexical and discourse characteristics in vowel production. *Proceedings of Meetings on Acoustics*, *22*(1), 060005.

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.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, *296*(5572), 1435–1435.

Bybee, J. (2000) The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford, CA: CSLI.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711–733.

Chafe, W. L. (1974). Language and consciousness. *Language*, *50*(1), 111–133.

Clopper, C. G. (2012). Effects of dialect variation on the semantic predictability benefit. *Language and Cognitive Processes*, *27*(7–8), 1002–1020.

Clopper, C. G., & Pierrehumbert, J. B. (2008). Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America*, *124*(3), 1682–1688.

Clopper, C. G., & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 25–72). Berlin: de Gruyter.

Clopper, C. G., Turnbull, R., & Burdin, R. S. (2018). Assessing predictability effects in connected read speech. *Linguistics Vanguard*, *4*(S2), 2017044.

Coetzee, A. W. (2004) *What it means to be a loser: Non-optimal candidates in Optimality Theory*. [PhD thesis, University of Massachusetts Amherst].

Coetzee, A. W., & Kawahara, S. (2013). Frequency biases in phonological variation.

*Natural Language & Linguistic Theory*, *31*(1), 47–89.

Cohen Priva, U. (2008). Using information content to predict phone deletion. In *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 90–98).

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, *6*(2), 243–278.

Cohen Priva, U. (2017). Informativity and the actuation of lenition. *Language*, *93*(3), 569–597.

Duffy, J. R., & Giolas, T. G. (1974). Sentence intelligibility as a function of key word selection. *Journal of Speech and Hearing Research*, *17*(4), 631–637.

Eddington, D., & Channer, C. (2010). American English has go? a lo? of glottal stops: Social diffusion and linguistic motivation. *American Speech*, *85*(3), 338–351, doi:http://doi.org/10.1215/00031283-2010-019

Engelhardt, P. E., & Ferreira, F. (2014). Do speakers articulate over-described modifiers differently from modifiers that are required by context? Implications for models of reference production. *Language, Cognition and Neuroscience*, *29*(8), 975–985.

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.

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*(5), 489–504.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Goldrick, M., Baker, H. R., Murphy, A., & Baese-Berk, M. (2011). Interaction and representational integration: Evidence from speech errors. *Cognition*, *121*, 58–72.

Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649–683.

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2016). The message shapes phonology. Preprint. https://doi.org/10.31234/osf.io/sbyqk

Hall, K. C., Hume, E., Jaeger, T. F., & Wedel, A. (2018). The role of predictability in shaping phonological patterns. *Linguistics Vanguard*, *4*(S2), 20170027.

Halliday, M. A. K. (1967). Notes on transitivity and theme in English (Part 2), *Journal of Linguistics*, *3*, 199–244.

Hashimoto, D. (2020). Pitch peak and word predictability: Results from CSJ corpus. In *Proceedings of the 10th International Conference on Speech Prosody 2020* (pp. 444–448).

Hawkins, S., & Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, *22*, 493–511.

Hoetjes, M. W., Krahmer, E. J., & Swerts, M. G. J. (2014). Do repeated references result in sign reduction? *Sign Language & Linguistics*, *17*(1), 56–81, doi:10.1075/sll.17.1.03hoe

Hooper, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. M. Christie Jr. (Ed.), *Current progress in historical linguistics* (pp. 95–105). Amsterdam: North-Holland.

Hume, E., Hall, K. C., & Wedel, A. (2016). Predicting perceptually weak and strong unmarked patterns: A message-based approach. In *Proceedings of the Annual Meetings on Phonology* (vol. 3).

Hunnicutt, S. (1985). Intelligibility versus redundancy—conditions of dependency. *Language and Speech*, *28*(1), 47–56.

Janse, E., & Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, *39*(3), 330–343.

Kaland, C., & Himmelmann, N. P. (2020). Repetition reduction revisited: The prosody of repeated words in Papuan Malay. *Language and Speech*, *63*(1), 31–55.

Kalikow, D. N., Stevens K. N., & Elliott L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*, 1337–1351.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language*, *82*, 536–574.

Kawahara, S. (2011). Aspects of Japanese loanword devoicing. *Journal of East Asian Linguistics*, *20*, 169–194.

Kawahara, S., & Lee, S. J. (2018). Truncation in message-oriented phonology: A case study using Korean vocative truncation. *Linguistics Vanguard*, *4*(S2), 20170016.

Kilbourn-Ceron, O. (2017). *Speech production planning affects variation in external sandhi*. [PhD thesis, McGill University].

Kilbourn-Ceron, O., Clayards, M., & Wagner, M. (2020). Predictability modulates pronunciation variants through speech planning effects: A case study on coronal stop realizations. *Laboratory Phonology*, *11*(1), 5. https://doi.org/10.5334/labphon.168

Kruijff-Korbayová, I., & Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language, and Information*, *12*(3), 249–259.

Lam, T. Q., & Marian, V. (2015). Repetition reduction during word and concept overlap in bilinguals. *Journal of Memory and Language*, *84*, 88–107.

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition*, *38*(8), 1137–1146.

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*(3), 829.

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

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–187.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839–862.

Lindblom, B. (1967). Vowel duration and a model of lip mandible coordination. *Speech Transmission Laboratory: Quarterly Progress and Status Report*, *8*(4), 1–29.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed), *The production of speech* (pp. 217–245). New York: Springer-Verlag.

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). Dordrecht: Kluwer.

Lohmann, A. (2018). Time and thyme are not homophones: A closer look at Gahl's work on the lemma-frequency effect, including a reanalysis. *Language*, *94*(2), e180–e190. https://doi.org/10.1353/lan.2018.0032.

Malmberg, B. (1954). *La Phonétique*. Paris: Presses Universitaires de France.

Moore-Cantwell, C. (2013). Syntactic predictability influences duration. *Proceedings of Meetings on Acoustics*, *19*, 060206.

Munson, B., & Solomon, N. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058.

Occhino, C. (2016). *A cognitive approach to phonology: evidence from signed languages*. [PhD thesis, University of New Mexico].

Phillips, B. S. (1981). Lexical diffusion and Southern *tune*, *duke*, *news*. *American Speech*, *56*, 72–78.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–342.

Pierrehumbert, J. B. (2002). Word-specific phonetics. *Laboratory Phonology*, *7*(1), 101–140.

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

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America*, *118*(4), 2561–2569, doi:10.1121/1.2011150

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.

Raymond, W. D., Brown, E. L., & Healy, A. F. (2016). Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Language Variation and Change*, *28*(2), 175–202.

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*(1), 55–97.

Rochemont, M. (2016). Givenness. In C. Féry & S. Ishihara (Eds.), *The Oxford handbook of information structure*. Oxford: Oxford University Press.

Schwarzschild, R. (1999). Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, *7*, 141–177.

Seyfarth, S. (2014). Word informativity influences acoustic duration: effects of contextual predictability on lexical representation. *Cognition*, *133*(1), 140–155.

Shaw, J. A., & Kawahara, S. (2019). Effects of surprisal and entropy on vowel duration in Japanese. *Language and Speech*, *62*(1), 80–114.

Silverman, D. (2012). *Neutralization*. Cambridge: Cambridge University Press.

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass*, *3*(1), 236–264.

Smith, N. J., & Levy, R, (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1637–1642). Austin, TX: Cognitive Science Society.

Strunk, J., Seifart, F., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., & Bickel, B. (2020). Determinants of phonetic word duration in ten language documentation corpora: Word frequency, complexity, position, and part of speech. *Language Documentation & Conservation*, *14*, 423–461.

Tabain, M. (2001). Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production. *Language and Speech*, *44*(1), 57–93.

Tanner, J., Sonderegger, M., & Wagner, M. (2017). Production planning and coronal stop deletion in spontaneous speech. *Laboratory Phonology*, *8*(1), 15, doi:http://doi.org/10.5334/labphon.96

Tkachman, O., Hall, K. C., Fuhrman, R., & Aonuki, Y. (2019). Visible amplitude: Towards quantifying prominence in sign language. *Journal of Phonetics*, *77*, 100935.

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

Tucker, B. V., &Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken* (pp. 1949–1952).

Tupper, P. F. (2014). Exemplar dynamics models of the stability of phonological categories. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1628–1633).

Turnbull, R. (2015). Assessing the listener-oriented account of predictability-based phonetic reduction. PhD thesis, Ohio State University.

Turnbull, R. (2017). The role of predictability in intonational variability. *Language and Speech*, *60*(1), 123–153.

Turnbull, R. (2018). Effects of lexical predictability on patterns of phoneme deletion/reduction in conversational speech in English and Japanese. *Linguistics Vanguard*, *4*(S2), 20170033.

Turnbull, R. (2019). Listener-oriented phonetic reduction and theory of mind. *Language, Cognition, and Neuroscience*, *34*(6), 747–768.

Turnbull, R., Seyfarth, S., Hume, E., & Jaeger, T. F. (2018). Nasal place assimilation trades off inferrability of both target and trigger words. *Laboratory Phonology*, *9*(1), 15.

Tyrone, M. E., & Mauk, C. E. (2010). Sign lowering and phonetic reduction in American Sign Language. *Journal of Phonetics*, *38*(2), 317–328.

Tyrone, M. E., & Mauk, C. E. (2012). Phonetic reduction and variation in American Sign

Language: A quantitative study of sign lowering. *Laboratory Phonology*, *3*(2), 425–453.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*(1), 2–7.

Vajrabhaya, P., & Kapatsinski, V. (2011). There is more to the story: First-mention lengthening in Thai interactive discourse. In *Proceedings of the Seventeenth International Congress of the Phonetic Sciences*.

Van Son, R., & Pols, L. C. W. (2003). Information structure and efficiency in speech production. In *Proceedings of Interspeech 2003*.

Van Son, R., & Van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*, 100–123.

Von Savigny, E. (1976). Listener-oriented versus speaker-oriented analysis of conventional meaning. *American Philosophical Quarterly*, *13*(1), 69–74.

Wagner, M. (2011). Production planning constraints on allomorphy. *Canadian Acoustics*, *39*(3), 160–161.

Wagner, M. (2012). Locality in phonology and production planning. *McGill Working Papers in Linguistics*, *22*(1), 1–18.

Watson, D. G. (2010). The many roads to prominence: Understanding emphasis in conversation. *Psychology of Learning and Motivation*, *52*, 163–183.

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*(3), 1548–1557.

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

Whalen, D. H. (1991). Infrequent words are longer in duration than frequent words. *The Journal of the Acoustical Society of America*, *90*(4), 2311–2311 doi:10.1121/1.401072

Wheeldon, L. (2012). Producing spoken sentences: The scope of incremental planning. In S. Fuchs, M. Weirich, D. Pape, & P. Perrier (Eds.), *Speech planning and dynamics* (pp. 97–118). Bern: Peter Lang.

Wiener, S., Speer, S. R., & Shank, C. (2012). Effects of frequency, repetition and prosodic location on ambiguous Mandarin word production. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of the 6th International Conference on Speech Prosody* (pp. 528–531), Shanghai.

Wright, C. E. (1979). Duration differences between rare and common words and their implications for the interpretation of word frequency effects. *Memory & Cognition*, *7*(6), 411–419.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.