3 Signal Reduction and Linguistic Encoding

T. FLORIAN JAEGER AND ESTEBAN BUZ
University of Rochester

Introduction

Human languages provide speakers with a remarkable degree of flexibility in how to linguistically encode near-meaning equivalent messages. This chapter focuses on what is arguably the most pervasive type of flexibility: flexibility in the amount or quality of the signal that encodes the speaker's message. Figure 3.1 illustrates this for English (inspired by Friedman, 2013). For example, an element may be mentioned or omitted (e.g., the optional complementizer *that*, or the argument *the World Cup*), or the articulatory realization of an element may be more or less detailed (e.g., producing a more centralized vowel or shortening the duration of a word). Such flexibility has been of central interest in psycholinguistic research: speakers' preferences to encode a message with a more or less reduced signal serve as a window into the architecture underlying the language production system.

Although speakers typically do not become of aware of this flexibility while talking, the choice between more or less reduced linguistic forms or signals is ubiquitous within and across languages. Alternations like those illustrated in Figure 3.1 exist across many, if not all, languages. Languages differ, however, in the specific alternations that they afford. For example, many languages allow omission of grammatical subjects in certain contexts (e.g., Italian, Japanese, Russian, and Yucatec Maya), whereas this omission is considered ungrammatical—or restricted to colloquial registers—in other languages (e.g., English). Other examples of reduction include optional mention of case-marking (e.g., in Japanese, Korean, and Turkish) or optional head-marking morphology (e.g., in many languages of the Balkan sprachbund), neither of which are available in English.

The Handbook of Psycholinguistics, First Edition. Edited by Eva M. Fernández and Helen Smith Cairns.

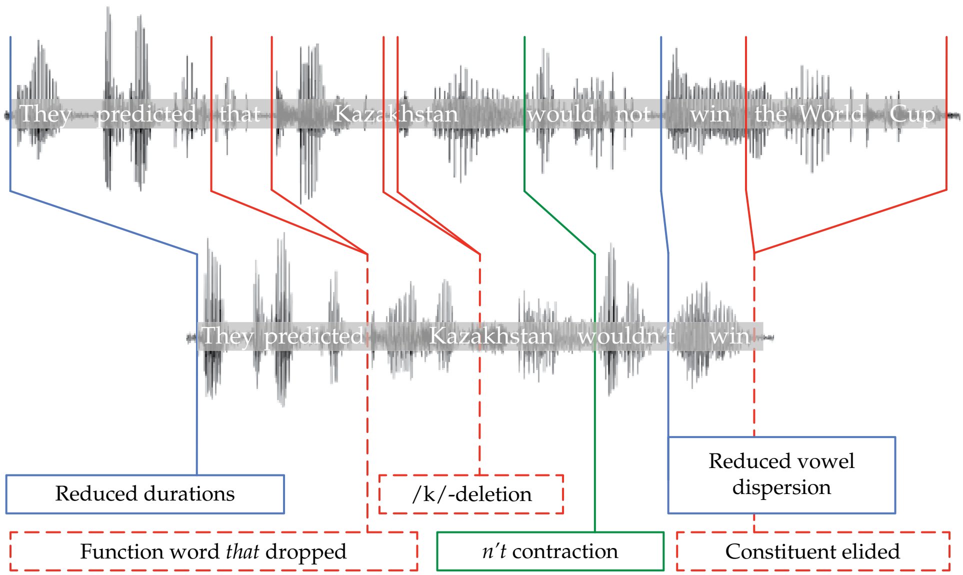

Figure 3.1 Illustration of a few types of implicit decisions speakers make during linguistic encoding that affect the degree of signal reduction. Reduction can be caused by decisions at multiple levels of linguistic encoding, including sentence, lexical, and phonological planning, as well as articulation. In the appropriate context, the upper and lower utterances encode the same message; yet the lower utterance contains shorter linguistic forms and is realized with a much reduced speech signal, compared to the upper utterance. Inspired by Friedman (2013).

Reduction constitutes the empirical focus of this chapter. Specifically, we focus on *probabilistic reduction*: a large body of work has found that speakers tend to produce shorter linguistic forms and more reduced signals for contextually predictable parts of the message. To the best of our knowledge, a systematic review of research on reduction across different levels of linguistic representations has so far been lacking. We thus begin with a summary of this literature and the questions it raises for future research.

The second part of this chapter reviews competing theories and accounts of the empirical findings discussed in the first part. Although we focus on probabilistic reduction, the discussion bears on more general architectural questions. In particular, we discuss competing views of the link between production and comprehension, as well as the link between online processing and biases implicitly encoded in linguistic representations. We distinguish between three broad classes of accounts. One hypothesis holds that flexibility in encoding a message allows speakers to navigate the attentional and memory demands of language production. This type of explanation is sometimes referred to as "production-internal" (Arnold, 2008), "production-based" (Gahl, Yao, & Johnson, 2012), "production-oriented" (Lindblom, 1990a), or "production-centered" (Watson, Arnold, & Tanenhaus, 2010).

This has been contrasted with the idea that production is affected by communicative considerations. According to the latter view, the mechanisms underlying linguistic encoding are—directly or indirectly—affected by comprehension (e.g., Brennan & Clark, 1996, Clark & Fox Tree, 2002; Jaeger, 2006; Lindblom, 1990a). This alternative idea is variously referred to as, for example, "listener-oriented" (Arnold, 2008), "comprehension-facilitation" (Arnold, Kahn, & Pancani, 2012), "intelligibility-based" (Gahl *et al.*, 2012), or "audience design" (Clark & Murphy, 1982; Galati & Brennan, 2010). Here we use the labels *production ease* and *communicative* accounts to refer to these two views.

Independent of what (mixture of) pressures ultimately drive speakers' preferences, there are questions about whether these pressures operate on-line, directly affecting speakers' preferences during incremental linguistic encoding, or off-line, changing linguistic *representations* and thus only indirectly affecting incremental encoding. Explanations that focus on the latter possibility constitute a third type of account, which we will refer to as *representational* accounts (e.g., Pierrehumbert, 2001; Wedel, 2006). For each of these accounts, we review specific proposals and isolate some challenges we consider particularly pressing for future research. The picture that emerges from this discussion is one in which probabilistic reduction is not driven by any single factor, but rather the result of multiple mechanisms.

A few terminological clarifications Throughout this chapter, we refer to such differences as *reduction* (and to *reduced* variants), without meaning to imply a directionality of this process: for many phenomena we discuss, it is an open question whether they are better understood as reduction or enhancement. For example, although it might seem more intuitive to think of the complementizer *that* in Figure 3.1 as being optionally omitted, there are also arguments as to why it is better thought of as being optionally mentioned.[1] In conversational American English, for example, the complementizer *that* is absent in about 83% of all complement clauses (Jaeger, 2010, p. 29). Even when the most frequent complement clause embedding verbs are excluded, omission is more frequent (53%) than mention of complementizer *that* (Jaeger, 2010, Table 1). This makes it difficult to determine whether this alternation is better understood as optional mention or optional omission. Similarly, word durations may undergo reduction (i.e., shortening) or enhancement (i.e., lengthening).

We also distinguish between message components, linguistic forms, and their realization in the linguistic signal. Message components are parts of the message speakers wish to convey (e.g., a specific lexical meaning). Linguistic forms are instances of linguistic categories, such as phonological segments, words, and syntactic structures. These forms are not directly observable. Rather, they underlie the observable linguistic signal. The linguistic signal can be acoustic (in the case of speech) or visual (in the case of gestures, sign language, or writing). We sometimes refer to more or less reduced *forms* to highlight that reduction goes beyond gradient manipulation of the *signal* and includes cases where language provides speakers with several more or less reduced linguistic forms (e.g., mentioning or omitting *the world cup* in Figure 3.1).

Probabilistic reduction: Contextual predictability and signal reduction

As shown in Figure 3.1, reduction can take place at different levels of linguistic encoding. Reduction at many of these levels has been found to be correlated with contextual predictability, so that more probable (and less informative) message components tend to be realized with reduced signal.[2] We begin with a summary of work on phonetic and phonological reduction. Then we summarize work at successively higher levels of linguistic encoding, including morphological contraction, the omission of optional function words, and the realization of referring expressions. We close this section with an overview of open empirical questions.

Phonetic and phonological reduction and omission

A large number of studies have investigated the articulatory or acoustic reduction of phonemes, syllables, and words. This research has found that contextually predictable instances of words tend to be produced with shorter duration (e.g., Aylett & Turk, 2004; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Tily *et al.*, 2009) and less articulatory detail (e.g., Aylett & Turk, 2006; Gahl *et al.*, 2012; Son & Santen, 2005). Such probabilistic phonetic reduction has been observed in conversational speech corpora (e.g., Arnon & Cohen Priva, 2014; Aylett & Turk, 2004; Bell *et al.*, 2003; Gahl *et al.*, 2012; Pluymaekers, Ernestus, & Baayen, 2005a) and in the lab, including read speech (e.g., Arnon & Cohen Priva, 2013; Gahl & Garnsey, 2004; Kurumada, 2011) and unscripted speech (e.g., Watson *et al.*, 2010).

For example, contextually predictable instances of words tend to have more reduced vowels (e.g., Aylett & Turk, 2006; but see null results in Bürki, Ernestus, Gendrot, Fougeron, & Frauenfelder, 2011; Clopper & Pierrehumbert, 2008; Gahl *et al.*, 2012; Scarborough, 2010) and consonants (Rose, 2017: Ch. 3; Torreira & Ernestus, 2009, 2012). Aylett and Turk investigated predictability based reduction in a corpus of citation speech. They measured syllable durations and first and second formant values of vowels within those syllables and binned syllables into high and low predictability based on unigram, bigram, and trigram probabilities. They found that syllables with high predictability were shorter in duration and vowels within those syllables showed more centralization. Contextually predictable words are also more likely to undergo phonological weakening or deletion (Bell *et al.*, 2009, 2003). As an example, many varieties of English favor the reduction of complex codas in some phonological environments. A specific case of this is *t/d*-deletion, where a *t* or *d* that is present in citation form is not produced. Such *t/d*-deletion is more common in predictable words (Gahl, Jurafsky, & Roland, 2004; Jurafsky, Bell, Gregory, & Raymond, 2001; see also Bybee & Hopper, 2001). Other research has further found that a segment's informativity[3] about the word affects the segment's realization even after the word's predictability is taken into account (e.g., van Son & Pols, 2002; van Son & van Santen, 2005).

Similar reduction effects have also been observed as a function of previous mention of a word (Bard *et al*., 2000; Bell *et al*., 2009; Pluymaekers, Ernestus, & Baayen, 2005b; Watson *et al*., 2010). Since the statistics of human language are such that previous mention generally increases the probability of being mentioned again (e.g., Rosenfeld, 1996, Section 2.3), these effects could at least in part be mediated through effects of previous mention on a word's contextual predictability (for evidence, see J. R. Heller & Pierrehumbert, 2011; for discussion, see Kahn & Arnold, 2012).

While phonological weakening or deletion has been studied extensively, less is known about phonological insertion. One example comes from optional epenthesis. In epenthesis, speakers insert a reduced vowel into a consonant cluster (e.g., fil*e*m). Epenthesis enhances the signal and reduces syllable complexity compared to what would be expected under a faithful realization of the citation form. In a corpus study on conversational Dutch, Tily and Kuperman (2012) found that speakers were less likely to insert the schwa into words that were contextually predictable.

How big are the effects of contextual predictability? Bell *et al*. (2003) find that the top and bottom 5% most predictable instances of English function words (such as *the*, *I*, etc.) differ in their duration by about 20-30 ms, out of a mean duration of about 100 ms. For content words, the most predictable instances of words are sometimes more than 100 ms shorter than their least predictable instances (Demberg, Sayeed, Gorinski, & Engonopoulos, 2012, p. 364). These effect sizes mean that predictability effects tend to be somewhat smaller than, though sometimes comparable to, durational lengthening associated with differences in linguistic structure or meaning (such as contrastive prosodic accents, Berkovits, 1994; phrase final lengthening, Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991). At the same time, these effect sizes imply that at least some probabilistic reduction is clearly perceptible (cf. Beaver, Clark, Flemming, Jaeger, & Wolters, 2007, who report detection of 6 ms durational differences).[4] Indeed, although this is an area that has received surprisingly little attention, there is evidence that the phonetic reduction associated with contextual predictability does affect intelligibility (Bard & Anderson, 1983, 1994; see also Buz, 2016, Ch. 4 for related discussion).

In summary, there is ample evidence that a word's contextual predictability tends to be correlated with its reduction (for further references, see Ernestus & Warner, 2011; Ernestus, 2014). Here we have focused on evidence from English. This reflects the status of the field, with the majority of existing research on phonetic reduction coming from English and typologically related languages (e.g., Dutch: Kuperman, Pluymaekers, Ernestus, & Baayen, 2007; Pluymaekers *et al*., 2005a; van Son & van Santen, 2005; French: Bürki *et al*., 2011; Pellegrino, Coupé, & Marsico, 2011; Torreira & Ernestus, 2009; Italian: Pellegrino *et al*., 2011; Spanish: Torreira & Ernestus, 2012), with only a handful of comparable studies on other languages (e.g., Cantonese: Zhao & Jurafsky, 2009; Japanese: Kurumada, 2011; Vietnamese, among others: Pellegrino *et al*., 2011). By taking advantage of language-specific properties, future studies on languages other than English hold great promise for the study of probabilistic reduction (for the critical importance of cross-linguistic evaluations of psycholinguistic theory, see also Jaeger & Norcliffe, 2009; Norcliffe, Harris, & Jaeger, 2015). For example, Zhao and Jurafsky (2009) investigated the effects of frequency on the realization of lexical tone in Cantonese. They found that frequency influences pitch production: low frequency words are produced with tone contours that are more distinct from each other. Paralleling phonetic and phonological reduction in English, Cantonese speakers thus tend to produce more reduced—or less distinguishable—signals for contextually more expected—and thus less informative—message components.

Morphological contraction and omission

Effects resembling probabilistic phonetic reduction have been observed in speakers' preferences between near-meaning equivalent morphological forms. For example, Frank and Jaeger (2008) investigate morphological contraction in American English conversational speech. Specifically, they focus on NOT (e.g., *isn't* versus *is not*), auxiliary BE (e.g., *he's* versus *he is*), and auxiliary HAVE (e.g., *I've done that* versus *I have done that*). They find that the rate of morphological contraction increases with the predictability of the meaning of the contractible element (e.g., NEGATION in *isn't* versus *is not*). This effect holds while controlling for potentially confounding factors such as speech rate, the type of host word preceding the contractible element, and the complexity of the material following the contractible element (see also Frank & Jaeger, 2008). These effects are confirmed by other studies on morphological contraction in conversational English (Bresnan & Spencer, 2016; Bybee & Scheibman, 1999).

More recent research has investigated alternations in which a bound morpheme can be either mentioned or omitted under near-meaning equivalence (Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2014). For example, Kurumada and Jaeger (2015) investigate optional case-marking in Japanese. Like in other case-marking languages, Japanese has case-marking morphology on the arguments of the verb that encode the grammatical function assignment. For example, the direct object of a transitive verb is marked with the suffix *-o*. Case-marking is important in understanding Japanese sentences, since Japanese has flexible word order, allowing both subject-before-object and object-before-subject ordering in transitive sentences. Unlike languages in which case-marking is obligatory (e.g., German), informal spoken Japanese allows speakers to omit the case marker without loss of near-meaning equivalence (see also Fry, 2001). In fact, case omission is frequent in informal Japanese (e.g., up to 51% of object case markers are omitted, Fry, 2001). In a spoken recall study, Kurumada and Jaeger find that speakers are more likely to omit the direct object case marker *-o* when the sentence makes the intended grammatical function assignment contextually predictable (e.g., *grandma* is more likely to be case marked in *The doctor sued the grandma* than in *The doctor treated the grandma*). Related corpus-based research has found that the rate of case-marking depends on how typical an argument is for the grammatical function it carries in the sentence (e.g., for Japanese: Fry, 2001; Korean: H. Lee, 2006; for further evidence from artificial miniature language learning, see Fedzechkina, Jaeger, & Newport, 2012, Fedzechkina, Newport, & Jaeger, 2016). For example, in conversational Korean, which also has optional case-marking, definite subjects are less likely to be case marked than indefinite subjects, whereas definite objects are more likely to be case marked than indefinite objects (H. Lee, 2006, Table 4). Since subjects are more likely to be definite than objects are, these findings suggest that case is more likely to be omitted when the meaning it encodes is predictable from context (for discussion, see Kurumada & Jaeger, 2015).

Another example comes from a recent study on optional head-marking in Yucatec Maya (Norcliffe, 2009; Norcliffe & Jaeger, 2014). In head-marking languages, grammatical function assignment and other information is encoded through bound morphology or clitics attached to the verb (rather than the arguments, as in the case of case-marking). In Yucatec some of this morphology is optional in certain environments. Norcliffe and Jaeger (2014) provide evidence that this optional morphology follows similar patterns as described for case-marking in Japanese above.

In sum, existing cross-linguistic evidence suggests that speakers' preferences in morphological reduction environments (i.e., contraction and omission) are affected by contextual predictability in ways that are at least qualitatively similar to phonetic reduction. However, compared to phonetic and phonological reduction, relatively little is known about the pressures driving optional morphological contraction and omission. Research on morphological production in morphologically rich languages seems a particularly promising venue for future work.

Omission of optional function words

Probabilistic reduction has also been documented for morphologically free function words. For example, English allows the omission of complementizer *that*, as in sentences like *She certainly knew (that) this was a required test* (Elsness, 1984; Huddleston & Pullum, 2002). This phenomenon is sometimes referred to as optional complementizer *that*-mention or *that*-omission. Speakers are more likely to produce the optional complementizer *that*, when the complement clause is less predictable given the matrix verb (e.g., *knew* in the example above). This effect has been observed in conversational speech (Jaeger, 2010) as well as production experiments (e.g., in written sentence completion, Garnsey, Pearlmutter, Myers, & Lotocky, 1997, Table 5; spoken or written recall, Ferreira, 2008; Jaeger & Grimshaw, 2013).

Optional function word omission is also observed in certain types of relative clauses. For example, in Standard American English, both finite non-subject-extracted non-pied-piped relative clauses (e.g., *That's the way (that) it is done*) and passive subject-extracted relative clauses (e.g., *These are the type of people (who are) not taken seriously*) allow similar omissions. For these environments, too, speakers have been found to be more likely to omit the optional function words the more predictable the constituent they introduce is in context (Jaeger, 2010, 2011; Wasow, Jaeger, & Orr, 2011; see also Melnick, 2011; Wiechmann, 2015).

In sum, speakers' preference to mention or omit optional function words seems to exhibit sensitivity to contextual predictability in ways that resemble phonetic reduction. However, beyond *that*-omission, the sensitivity of optional function word omission to contextual predictability has remained under-explored. Alternations similar to optional complementizer *that* exist in other languages (e.g., in Danish), though omission is sometimes accompanied by constituent order changes (e.g., in German). English, too, contains a number of additional environments that support optional omission of function words, such as the omission of *to* after verbs like *help* (Rohdenburg, 2004) or in the DO-BE construction (e.g., *all I want to do is (to) go to work*, Flickinger & Wasow, 2013). Additional examples are observed in non-standard varieties of American English, such as optional copula omission in African American Vernacular English (e.g., *You done yet?*; Bender, 2000, p. 85) or relativizer omission in subject-extracted relative clauses in, for example, the English of the British Isles (e.g., *And there were a wee alarm clock sat on the window*; Tagliamonte & Smith, 2005, p. 87).

It is thus an open question whether the effects of contextual predictability observed in research on *that*-omission in Standard American English will generalize to these similar phenomena and across languages. Preliminary evidence comes from ongoing research on the DO-BE construction (Wasow, Levy, Melnick, Juzek, & Zhu, 2015). Wasow and colleagues find that speakers are more likely to omit *to* in the DO-BE construction in lexical contexts that frequently co-occur with the DO-BE construction.

Reduction and omission of referring expressions

Another domain in which languages typically provide multiple near meaning-equivalent forms with more or less reduced signals is referring expressions. For example, in many contexts speakers can choose between a pronoun (e.g., *he*), name (e.g., *John*), or a full lexical noun phrase (e.g., *a colleague of mine*) to refer to the same referent.[5]

It has long been hypothesized that the choice between these different ways of encoding a reference depends on the referents "accessibility" in context (e.g., Ariel, 1999; Givón, 1983). This includes several factors that make referents more predictable (Arnold, 1998, 2010). For example, previous mention of a referent makes it more likely that it will be referred to in subsequent utterances. Previous mention also makes it more likely that a more reduced form is chosen (Bard *et al.*, 2000 and references therein). Moreover, the probability that a previously mentioned referent is referred to again decreases with increasing distance from its last mention. Similarly, the preference for a pronoun over a longer referring expression decreases with increasing distance from the last mention of a referent (Arnold, 1998; Arnold, Bennetto, & Diehl, 2009; as summarized in Arnold, 2010, p. 190).

Recent work has more directly assessed the effect of contextual predictability on the realization of referring expressions, paralleling research on probabilistic phonetic reduction. Tily and Piantadosi (2009) employed a type of Shannon guessing game (Shannon, 1951) to obtain estimates of the contextual predictability of over 2,000 references in a newspaper corpus. In their version of the Shannon guessing game, raters saw story fragments up to the next referring expression. Their task was to guess which of the previously introduced referents (or possibly a new referent) the next expression would refer to. Almost 500 raters provided a total of over 70,000 guesses. This made it possible to calculate estimates of the contextual predictability of the actual references made in the corpus. Tily and Piantadosi found that writers had indeed been more likely to use longer linguistic forms (e.g., names rather than pronouns) when the intended reference was less expected given the preceding context. This effect held beyond the effects of previous mention and other previously documented effects (for related results, see also Rohde & Kehler, 2014).

Mahowald, Fedorenko, Piantadosi, and Gibson (2013) investigated speakers' preference between full and reduced lexical forms with the same meaning, such as *mathematics* and *math*. Mahowald and colleagues found that speakers' preference for the shorter form increases with the contextual predictability of the concept encoded by either form. In a corpus study, the average informativity (measured as Shannon information) of long forms was significantly higher than for short forms suggesting that short forms tend to be used in contexts where they conveyed less information. In a sentence completion study, Mahowald and colleagues further found that participants chose the short form for sentences with supportive contexts (e.g., *Susan loves the apes at the zoo, and she even has a favorite …*) as compared to non-supportive contexts (e.g., *During a game of charades, Susan was too embarrassed to act like a …*). This preference closely mirrors the preference observed for contractible auxiliaries and negation (Bresnan & Spencer, 2016; Bybee & Scheibman, 1999; Frank & Jaeger, 2008).

A similar preference to produce reduced linguistic signals for contextually more predictable referents is also observed for optional argument omission (Kravtchenko, 2014; Resnik, 1996). In certain lexical environments, speakers of English can decide to omit an entire argument (e.g., *the semi-finals* in *Germany lost (the semi-finals)*), while maintaining near meaning-equivalence. In his seminal corpus study, Resnik found that verbs that contained more information about the types of arguments they take, thereby making the arguments following them (on average) more predictable, also are associated with a higher rate of argument omission (Experiment 4, Resnik, 1996).

Recent work on optional subject omission in Russian builds on these results (Kravtchenko, 2014). While considered non-standard or ungrammatical in English, many languages allow omission of contextually inferable subjects, sometimes referred to as pro-drop (Dryer, 2013). Using the version of the Shannon guessing game developed by Tily and Piantadosi (2009), Kravtchenko (2014) obtained estimates of the contextual predictability of over 700 subject noun phrases from a Russian corpus. Paralleling the results for the realization of referential expressions in English, Kravtchenko found that Russian subjects are more likely to be omitted when they are contextually predictable.

Reduction beyond the level of the clause

The majority of psycholinguistic research has focused on linguistic encoding at the level of the clause or below. A few more recent studies have begun to investigate reduction beyond the clause. For example, Asr and Demberg (2015) investigated the realization of coherence relations in English (see also Asr & Demberg, 2012). Simplifying somewhat, coherence relations are discourse relations between propositions. Asr and Demberg (2015) focused on the so-called *Chosen Alternative* relation and the coherence marker *instead* in environments in which it is optional (e.g., *They didn't panic during the first round of selling. (Instead,) they sold into the strength, which kept the market orderly*). Asr and Demberg found that *instead* was more likely to be omitted in the presence of a contextual cue to the Chosen Alternative relation (but see Anibel, 2010, for a failure to find such effects for other types of coherence relations).

Another environment in which speakers have the choice between providing more or less linguistic material to encode a near meaning-equivalent message was investigated by Gallo and colleagues (Gallo, 2011; Gallo, Jaeger, & Furth, 2010; Gallo, Jaeger, & Smyth, 2008). For example, Gallo *et al*. (2008) had speakers participate in a version of the Map Task (A. H. Anderson *et al*., 1991; see Pardo, this volume, for a description). Speakers instructed another (confederate) participant to replicate on their screen a specific arrangements of objects seen only by the speaker. Gallo and colleagues coded whether speakers used one or two sentences to convey the same message. For example, participants could say *Move the triangle to Central Park* or use a more verbose message like *Take the triangle. Now move it to Central Park*. Gallo and colleagues found that speakers were more likely to split the message across two clauses when the object (e.g., *the triangle*) consisted of less predictable words (for similar evidence from Spanish, see Gallo *et al*., 2010). These effects held beyond effects of previous mention, which is known to be correlated with the choice between pronoun versus lexical NPs (cf. Tily & Piantadosi, 2009).

Of the areas summarized here, production planning (including preferences regarding reduction) beyond the clause-level is probably the least understood. Further work is required to see whether the tentative evidence summarized here will confirm that principles similar to those observed in phonological, lexical, and syntactic reduction also operate during planning of larger linguistic chunks.

Summary and open questions

Language provides speakers with an astonishing degree of flexibility in the linguistic encoding of messages. Many of the options available to speakers differ in the amount of signal produced by the speaker. Across all stages of production summarized here, speakers' preferences between different ways of realizing the same message seem to be affected by a similar bias, reflected in a correlation between contextual predictability and reduction. More specifically, it seems that it is the predictability of a linguistic form or *message component* (roughly, part of the meaning a speaker wishes to convey) that correlates with a preference for shorter linguistic forms at the next lower level and more reduced linguistic signals. For example, the predictability of NEGATION following a lexical context (e.g., *President Clinton did …*) correlates with an increased preference for morphological contraction (i.e., saying *President Clinton didn't …* rather than *President Clinton did not …*, Frank & Jaeger, 2008). Similarly, it seems to be the predictability of a complement clause that correlates with an increased preference to omit the relativizer *that* (Jaeger, 2010) and the predictability of a lemma that correlates with the reduction of its word form (Aylett & Turk, 2004; Jurafsky *et al.*, 2001). In this context, a particularly intriguing piece of evidence comes from research on homophones, such as *time* and *thyme*. While *time* and *thyme* have the same phonological citation form, the actual realization of the two words tends to differ subtly (Gahl, 2008). Speakers tend to produce the more frequent lemma (*time*) with a more reduced speech signal, compared to the less frequent lemma (*thyme*). To the best of our knowledge, comparable work on the effects of contextual predictability on homophone pronunciation has yet to be conducted. Still, this type of effect suggests that it is at least partly the predictability of a message component (in this case the lemma or its meaning) that drives the extent to which its realization in the linguistic signal is reduced (see also Jaeger, 2006, Study 6).

While the inverse correlation between predictability and linguistic signal is now firmly established, many questions remain about the nature of this relation. The perhaps most pressing questions regard the processes underlying probabilistic reduction and, in particular, the relation between production planning and the realization of the linguistic signal. Before we address these questions in the second part of this chapter we briefly summarize outstanding empirical questions about probabilistic reduction. One question that deserves further attention is the relation between reduction at different levels of linguistic encoding (e.g., phonetic vs. syntactic reduction). Simply put, what determines the level of linguistic encoding at which speakers reduce or enhance the signal? This question has received some attention in research on phonetic reduction and phonological deletion (e.g., Bürki, Ernestus, & Frauenfelder, 2010; Bürki *et al.*, 2011; Hanique, Ernestus, & Schuppler, 2013; Torreira & Ernestus, 2011). For example, some cases of omission might be better understood as extreme cases of gradient phonetic reduction, while others are better understood as originating in categorical phonological representations.

Another open question is what types of cues affect probabilistic reduction. The majority of previous research on probabilistic reduction has focused on the immediately surrounding lexical context. For example, for phonetic reduction most research has estimated the word's predictability based on its surrounding trigram context (e.g., Aylett & Turk, 2004; Bell *et al.*, 2009, 2003; Gahl *et al.*, 2012; van Son & Pols, 2003; van Son & van Santen, 2005). Arnon and Cohen Priva (2013) find that the predictability of the final word in a 4gram (e.g., *tea* in *a cup of tea*) is correlated with phonetic reduction, even after bi-, tri-, and unigram frequencies are accounted for (see also Arnon & Cohen Priva, 2014; Demberg *et al.*, 2012). Similarly, most research on reduction at higher levels of linguistics encoding has employed local lexical cues (e.g., Frank & Jaeger, 2008; Jaeger, 2010; Mahowald *et al.*, 2013; Resnik, 1996).

There are, however, also some studies that have found less local or more abstract cues to affect reduction. For example, phonetic reduction has been found to be correlated to the word's predictability given its semantic (Sayeed, Fischer, & Demberg,

2015) or syntactic context (or larger *n*grams, Demberg *et al*., 2012; Kuperman & Bresnan, 2012; Kurumada, 2011; Tily *et al*., 2009; but see Rose, 2017: Ch. 4).[6] Less local cues have also been found to affect the omission of optional function words (syntactic context, Jaeger, 2006, Study 5; Levy & Jaeger, 2007; Wasow *et al*., 2011) as well as the reduction of referring expressions (cloze completions, Kravtchenko, 2014; Tily & Piantadosi, 2009), although some of these studies have not tested whether the same effects could be attributed to more local cues.

A closely related question is whether different types of cues are weighted differently depending on the level of linguistic encoding (e.g., phonological versus morphological contraction). This would arguably be expected under most accounts discussed below. Even accounts of linguistic encoding that assume that information from lower levels can affect earlier stages of production generally assume that these influences are weaker than influences from the current or earlier stages of production (e.g., Dell, 1986; Dell, Chang, & Griffin, 1999; Janssen & Caramazza, 2009). For example, segmental phonological properties generally only weakly affect syntactic preferences (Jaeger, Furth, & Hilliard, 2012a; McDonald, Bock, & Kelly, 1993). *Supra*segmental phonological preferences, on the other hand, have been found to affect syntactic production. For example, speakers prefer to insert optional function words or reorder constituents so as to avoid adjacent stressed syllables (Anttila, Adams, & Speriosu, 2010; Jaeger *et al*., 2012a; M.-W. Lee & Gibbons, 2007). Similar asymmetries in the factors that drive variation have been observed between the phonetic reduction of segments and their omission (for results and discussion, see Bürki *et al*., 2011; Hanique *et al*., 2013). Whether similar asymmetries are reflected in what cues affect probabilistic reduction is a question for future research (for preliminary results, see an unpublished study by Jaeger, Snider, Staum, & Jurafsky, 2006, who compared the phonetic reduction and optional omission of complementizer and relativizer *that*).

Another questions is whether and how speakers integrate multiple cues to the same target (e.g., the same word). For example, does such integration follow similar principles that have been observed in comprehension, where comprehenders seems to be able to integrate multiple sources of information (e.g., Hare, McRae, & Elman, 2004; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995)? To the best of our knowledge, there is so far no published work that addresses this question. A few studies have compared the effect of predictability (or surprisal) estimates based on different types of cues (e.g., Demberg *et al*., 2012, p. 364; Sayeed *et al*., 2015). But these studies have not directly compared the *objective* information contained in these cues to their relative importance in the *subjective* language models that speakers implicitly draw on during linguistic encoding. Preliminary evidence comes from an unpublished study on phonetic reduction in speech (Post & Jaeger, 2010). Post and Jaeger integrated multiple lexical and syntactic cues into a single estimate of a word's predictability. They found that both types of cues contributed to a word's phonetic reduction and that they did so proportionally to their contribution to the word's predictability. If confirmed by future work, results like these would suggest that probabilistic reduction draws on multiple contextually available cues, weighted by their relative informativity (see also Jaeger, 2006, Studies 3 and 4, for related evidence for optional complementizer and relativizer *that*).

Theoretical positions

Psycholinguistic accounts of probabilistic reduction tend to come in three broad flavors: production ease, communicative, and representational accounts. Production ease accounts attribute variation in speakers' preferences to the demands of incremental linguistic encoding. Below we discuss three related classes of proposals about *how* production ease affects linguistic encoding. Following that, we discuss accounts of linguistic reduction that refer to communicative goals. This includes a discussion of research on *audience design*. We also discuss more recent communicative accounts that either draw on information theoretic considerations (cf. Shannon, 1948) or the concept of rational (J. R. Anderson, 1990) or boundedly rational cognition (e.g., Simon, 1990).

Production ease and communicative accounts share a focus on online processes that affect production as it is unfolding. This contrasts with *representational* accounts, which have focused on changes in the phonetic representations of words over longer periods of time (e.g., the lifetime of a speaker or even generations of speakers). The majority of psycholinguistic work on reduction and omission has interpreted speakers' preferences in alternations as providing a window into the mechanisms underlying language production, thereby more or less explicitly assuming the former (e.g., Arnold *et al.*, 2012; Baese-Berk & Goldrick, 2009; Bard *et al.*, 2000; Ferreira & Dell, 2000; Gahl *et al.*, 2012). In research on speech production, however, phonological and phonetic reduction is often described as the result of changes to phonological representations (e.g., Bybee & Hopper, 2001; Kohler, 1990; Pierrehumbert, 2001, 2002; Wedel, 2006; Zipf, 1929; for additional references, see Ernestus, 2014). Following our discussion of production ease and communicative accounts, we turn to this third type of account of reduction mentioned above, representational accounts. We discuss the relation between such offline accounts and online accounts of reduction.

Before we turn to these different accounts, we begin with an important caveat.

Production ease versus communicative goals: Mutually exclusive?

Although it is helpful for the purpose of exposition to group accounts of reduction into broad classes of competing positions, production ease and communicative accounts are arguably better seen as defining a continuum of perspectives. For example, some communicative accounts do not argue against the idea that the resource demands inherent to linguistic encoding affect speakers' production preferences. Rather, speakers' preferences are assumed to *also* be affected by communicative considerations. Specifically, a long-standing idea holds that language production is subject to competing pressures—on the one hand, speakers want to achieve their communicative goals, on the other, they have limited resources (e.g., planning time, memory capacity) to achieve these goals (Zipf, 1949; see also e.g., Jaeger, 2006, 2013; Lindblom, 1990a, 1990b). These types of accounts thus do not predict that speakers will always be able to (or even intend to) rank their communicative goals above production ease.

Similarly, production ease accounts are typically not intended to completely rule out communicative effects on linguistic encoding. Such a position would be untenable in light of existing evidence. Rather, only generic knowledge about common situations is assumed to affect language production (e.g., such as talking louder in a noisy place, Lombard, 1911; child-, hearing impaired-, or foreigner-directed speech, Kuhl *et al.*, 1997; Stern, Spieker, Barnett, & MacKain, 1983; Uther, Knoll, & Burnham, 2007; Picheny, Durlach, & Braida, 1986). In this view, *specific* knowledge about the current situation and interlocutors, including perspective taking, is assumed to have no effect on language production (cf. Dell & Brown, 1991). Later work has further softened this claim to suggest that listener-specific[7] audience design is possible, but very resource-demanding and thus easily abandoned (e.g., Horton & Keysar, 1996; Wardlow Lane & Ferreira, 2008). Some accounts are further only intended for specific aspects of language production. For example, some proposals hold that lexical encoding can be subject to audience design, whereas decisions during phonological encoding and articulation are unaffected by listener-specific audience design (Arnold, 2008; Bard *et al.*, 2000; but see Buz *et al.*, 2016). Other proposals hold that speakers can avoid ambiguity when it is apparent *prior* to linguistic encoding (e.g, when referring to a small bat, when there is also a large bat on the screen), but not when the ambiguity only becomes apparent after lexical retrieval (e.g., when referring to a baseball bat, when there is also an animal bat on the screen, Ferreira, Slevc, & Rogers, 2005). Similarly, some proposals hold that grammatical encoding is unaffected by listener-specific audience design (Ferreira, 2008; Ferreira & Dell, 2000; MacDonald, 2013; but see Jaeger, 2013).

With these caveats in mind, the remainder of this section sets out to isolate core differences in theoretical perspectives. We do so because it is those differences in focus or perspective that often end up driving researchers' decision to conduct a particular study. During our discussion of communicative accounts, we return to questions about the specificity of audience design, and review the available evidence. We refer to Gambi & Pickering (this volume) for additional discussion of these issues.

Production ease

With these clarifications in mind, we now introduce three types of production ease accounts. These accounts differ in whether they attribute speakers' preferences—including reduction—to the planning of previous, the current, or following material.

Production ease: Planning of upcoming material affects the realization of current material One influential proposal is *availability-based production* (Bock, 1987; Ferreira & Dell, 2000). Availability-based production holds that speakers prefer linguistic forms that let them articulate whatever material is fully planned, while maintaining grammaticality (Ferreira, 1996, 2008; Ferreira & Dell, 2000). Language production is, in this sense, greedy, presumably because the attentional and memory resources available for sentence planning are limited. By sending completed linguistic plans to articulation, resources are freed up and articulation continues. There is clear evidence that some such principle is at work during sentence production. Availability-based production is reflected in a variety of strategies that speakers seem to employ in order to avoid suspension of articulation. This includes the insertion of filled pauses (Clark & Fox Tree, 2002; Shriberg, 1996), restarts (Clark & Wasow, 1998), and constituent reordering (for review, see Jaeger & Norcliffe, 2009). For example, speakers tend to produce more easily retrievable word forms and word forms associated with more easily retrievable concepts or constituents earlier in a sentence (Arnold, Losongco, Wasow, & Ginstrom, 2000; Bock, 1982, 1987; Branigan, Pickering, & Tanaka, 2008; Bresnan, Cueni, Nikitina, & Baayen, 2007; Ferreira, 1996; Ferreira & Yoshita, 2003; Kempen & Harbusch, 2004; Rosenbach, 2008; Tanaka, Branigan, McLean, & Pickering, 2011).

Similarly, there is evidence that availability affects reduction. Most of this evidence comes from research on phonetic reduction and research on the omission of optional function words. For example, speakers slow down their speech rate before points of production difficulty (e.g., complex, infrequent, or novel words, Fox Tree & Clark, 1997; Watson, Buxó-Lugo, & Simmons, 2015). Speakers are also more likely to produce optional complementizer or relativizer *that* before difficult-to-retrieve clause onsets (Ferreira & Dell, 2000; Jaeger, 2010; Jaeger & Wasow, 2006; Race & MacDonald, 2003; Roland, Elman, & Ferreira, 2006; Temperley, 2003). These types of findings show that the availability of *upcoming* material can cause reduction or omission preceding that material.

One open question is whether such planning effects can also account for the link between the contextual predictability of a linguistic form and its own realization (rather than the realization of preceding forms). For example, consider the case of phonetic reduction. While the majority of studies has not controlled for the availability of upcoming material (e.g., Aylett & Turk, 2004, 2006; Bell *et al.*, 2009, 2003; Gahl *et al.*, 2012), there is, in fact, some evidence that the effect of contextual predictability on phonetic reduction is independent of the availability of upcoming material (Kidd & Jaeger, 2008; Post & Jaeger, 2010; Watson *et al.*, 2010). Similar evidence exists for morphological reduction (Frank & Jaeger, 2008) and morphological or optional function word omission (Jaeger, 2006, 2010; Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2014).

Could availability nevertheless account for probabilistic reduction? To some extent, this is a question of *granularity* (i.e., at what level of linguistic representation availability is taken to apply). One recent proposal holds that probabilistic phonetic reduction is driven by the incremental availability of the word's *segments* (Watson *et al.*, 2015). Indeed, when a word is contextually predictable, so is—on average—the sequence of its segments. Thus, if the contextual predictability of a segment contributes to its availability, it seems plausible that a segment-level availability account could explain word-level probabilistic phonetic reduction. This proposal does, however, face at least two challenges. The first challenge is empirical: we know of no studies that directly test whether word-level effects of contextual predictability can be reduced to segment-by-segment predictability effects. The second challenge is integrating this view with existing theories of lexical production: standard models of lexical production assume that the units at the interface between phonological encoding and articulation are syllables (Dell, 1986; Levelt, 1989). That is, articulation is assumed to proceed syllable-to-syllable, not segment-to-segment. It is thus unclear how standard models could explain probabilistic phonetic reduction of mono-syllabic words (which is observed, e.g., Bell *et al.*, 2003; Gahl *et al.*, 2012).

Production ease: Planning of material affects its realization An alternative type of production ease account holds that the planning of a linguistic unit (e.g., a word) is directly reflected in its *own* realization (e.g., Arnold *et al.*, 2012; Baese-Berk & Goldrick, 2009; Bard *et al.*, 2000; Goldrick, Vaughn, & Murphy, 2013; J. R. Heller & Goldrick, 2014). Unlike availability-based production, which links reduction to the planning of upcoming material, these proposals thus link reduction of a linguistic form to the planning of that form. Whereas availability-based production has been investigated at the phonetic, lexical, and syntactic level (see references above), the alternative account has so far only been applied to the link between lexical planning and articulation (i.e., phonetic reduction). Within work on phonetic reduction, the idea that activation levels during phonological or lexical planning predict articulatory reduction is receiving increasing attention (e.g., see also Arnold *et al.*, 2012; Ernestus, 2014; Gahl *et al.*, 2012; Kahn & Arnold, 2012; Seyfarth, 2014; Watson *et al.*, 2015). We thus focus on phonetic reduction.

A specific instance of this proposal is the *competition* account introduced by Baese-Berk & Goldrick (2009). Baese-Berk and Goldrick investigated the articulation of words with minimal pair neighbors (e.g., the articulation of the/p/in *pin*, which has the minimal pair neighbor *bin*, compared to the/p/in *pipe*, which lacks the minimal pair neighbor *bipe*). The existence a minimal pair neighbor is assumed to lead to competition during lexical planning. This, in turn, is taken to mean that the target word will reach higher activation before being selected for articulation, and this increased activation is assumed to be correlated with hyper-articulation of the target word. In the words of Baese-Berk and colleagues: “[t]he higher activation level for words […] will lead to more active phonetic representations and consequently more extreme articulatory realizations” (Baese-Berk & Goldrick, 2009, p. 531).

This competition account makes two predictions that have been investigated. First, it predicts the hyper-articulation of words with minimal pair neighbors. This prediction has received support from experiments on isolated word production (Baese-Berk & Goldrick, 2009; Fox, Reilly, & Blumstein, 2015; Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011) and the acoustic realization of words in conversational speech (Wedel & Sharp, 2014). For example, Baese-Berk & Goldrick (2009) found that the/p/in in *pin* is hyper-articulated with longer voice onset timing (making it more clearly unvoiced and thus distinguishable from a/b/) compared to the/p/in *pipe*.[8]

Second, if a minimal pair neighbor receives contextual support, competition accounts predict further competition with the target word, thus leading to increased hyper-articulation. A number of studies have addressed this question for onset minimal pair neighbors (e.g., *pin-bin*). This work has consistently found hyper-articulation of the contrasting phonetic features (except where phonetic constraints impede hyper-articulation; see e.g., the lack of hyper-articulation for voiced onsets Goldrick *et al*., 2013). For example, speakers tend to hyper-articulate the/p/in the word *pin* with longer voice onset timing when *pin* is displayed on the same screen as its neighbor *bin*, compared to when that neighbor is not displayed (e.g., Baese-Berk & Goldrick, 2009; Buz, Jaeger, & Tanenhaus, 2014; Kirov & Wilson, 2012; Seyfarth, Buz, & Jaeger, 2016).

The empirical picture is less clear when the critical contrast is on the vowel (Kirov & Wilson, 2012; Schertz, 2013) or a coda consonant (e.g., *coat-code;* De Jong, 2004; Goldrick *et al*., 2013; Seyfarth *et al*., 2016). For example, Goldrick *et al*. (2013) investigate voicing of plosive codas in word with minimal pair coda neighbors (e.g., *coat*, which has the neighbor *code*) and words without such neighbors (e.g., *rap*). After controlling for phonological confounds between their conditions, Goldrick and colleagues find no evidence for hyper-articulation of the voicing contrasts in the coda in words with minimal pair neighbors as compared to words without such neighbors (but see Seyfarth *et al*., 2016).

The possibility of an asymmetry between onset and rhyme minimal pairs is intriguing in light of studies on the effect on onset versus coda overlap on lexical planning. These works have generally found that onset overlap between adjacent words is associated with planning difficulty, whereas coda overlap is associated with facilitation (or at least, less difficulty, O'Seaghdha & Marin, 2000; Jaeger, Furth, & Hilliard, 2012b; Smith & Wheeldon, 2004; Rapp & Samuel, 2002; Sevald & Dell, 1994; Wheeldon, 2003). If future studies confirm that onset neighbors lead to production difficulty during lexical planning, whereas coda neighbors lead to facilitation, and that these differences are reflected in articulation, this would provide support that competition affects articulation. We consider such studies critical: without a clearer understanding of the conditions under which phonological neighbors inhibit or facilitate lexical planning, evoking these processes as an explanation of articulation (including reduction) risks ad-hoc meaning (for further discussion, see also Buz & Jaeger, 2016; Chen & Mirman, 2015; Gahl, 2015).

Future work on competition accounts will also need to elaborate on the linking function between activation levels during planning and reduction during articulation. Baese-Berk and colleagues link higher activation levels during planning to *hyper*-articulation. At the same time, accounts of probabilistic reduction would seem to assume the opposite: contextual predictability is generally assumed to lead to increased activation; yet, contextually predictable words are reduced compared to less predictable words. That is, accounts attributing the reduction of predictable words directly to activation levels during production planning implicitly or explicitly assume that higher activation leads to *hypo*-articulation (this assumption is made explicit in, e.g., Arnold, 2008, p. 506, referring to Arnold (1998); Arnold *et al*., 2012, p. 506, Bard *et al*., 2000, p. 17). If the competition account is to explain probabilistic reduction effects, future research will need to address these *prima facie* conflicting assumptions. One possibility is to distinguish effects on activation and effects on the activation threshold required for a word to be selected for articulation. Another possibility is that effects of *pre*-activation due to anticipatory effects (e.g., contextual predictability given preceding material) differ from effects of competition that arise later during lexical planning and phonological encoding.

Production ease: Planning of previous material affects the realization of current material A third type of production ease account attributes reduction to *previously* encountered production difficulty (Bell *et al.*, 2009). We will refer to this type of account as *compensatory reduction*. The compensatory reduction account shares with availability-based production models that it attributes reduction to processes that coordinate linguistic planning at higher levels (e.g., lexical planning) with the execution of the articulation in order to maintain fluency. For example, Bell and colleagues submit that phonetic reduction could in part be attributed to the coordination of articulation plans with the planning of the smallest prosodic units sent off to articulation (e.g., phonological words, Wheeldon & Lahiri, 1997): if a slow down is experienced during the planning of a prosodic unit, this is assumed to trigger a slow down in the execution of articulation for the following phrase in order to maintain information flow between these two stages of production (Bell *et al.*, 2009, p. 106).

By itself the compensatory reduction account is likely insufficient to explain the full range of phonetic reduction: probabilistic reduction is observed, even when local speech rate on preceding syllables is controlled for, thus suggesting that reduction is not exhaustively caused by compensatory changes in speech rate (e.g., for phonetic reduction, Gahl *et al.*, 2012; Kidd & Jaeger, 2008; for morphological reduction, Frank & Jaeger, 2008; for function word omission, Jaeger, 2006, 2010).

Another challenge that the compensatory reduction account faces is that there is as of yet no independent evidence for compensatory reduction. We know of no studies that have investigated effects of previous complexity on subsequent articulation, while holding constant the complexity of following material.

Summary We have summarized three production ease proposals that have been put forward to explain probabilistic reduction. While summarized as alternatives, it is possible that multiple production ease mechanisms jointly explain the observed range of probabilistic reduction phenomena. One challenge that any unified account of reduction in terms of production ease will have to address is how to account for reduction beyond the phonetic level. For example, it is unclear the *omission* of optional elements (e.g., optional arguments, adjuncts, or function words) could be a consequence of these elements being easy to produce (for similar arguments, see also Jaeger, 2006, 2010; Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2014).

Communicative accounts

Communicative accounts hold that speakers' preferences in reducing or enhancing the linguistic signal are affected by their communicative goals. This includes a bias for robust—or even efficient—message transmission. This view is closely related, but not identical, to questions about *audience design*—the idea that "the speaker designs each utterance for specific listeners" (Clark & Murphy, 1982, p. 287).

Unlike this (strong) audience design hypothesis, communicative accounts are not necessarily committed to the claim that speakers design their utterances for a *specific* type of audience (cf. generic vs. listener-specific audience design, Dell & Brown, 1991). We return to this issue below.

Here we refer to message transmission in the most general sense. For example, we include the transmission of non-literal and social meaning (for related discussion, see also Pardo, this volume). The goal of communication in the broad sense intended here is to cause a change in the interlocutors' state of mind. Even under this broad definition, communication is arguably not the only function of language (cf. Chomsky, 2000; Jenkins, 2000), but it is a common and important function.

A variety of communicative accounts have been proposed. For example, some accounts focus on production effort, whereas others focus on fluency; some accounts focus on accurate message transmission, whereas others focus on the speed of accurate transmission. The scope of this chapter prevents us from discussing these differences. Instead, we focus on the idea—shared more or less explicitly between most communicative accounts—that the understanding of production preferences requires reference to the *goals* of language use (see also Tanenhaus & Brown-Schmidt, 2008), and in particular, the goal to transmit specific intended messages.

Central to communicative accounts as defined here is the idea that production ease and communicative are often in competition. We begin by briefly reviewing this notion, and a major challenge it faces. Then we turn to the notion of audience design. We conclude this section with a few key considerations for future work on audience design.

Trading off production ease and communicative goals According to one influential view, speakers' production preferences arise from the competition between, on the one hand, a bias for robust message transmission and, on the other hand, a bias for production ease or effort (an idea that goes back to at least, Zipf, 1949). The former bias favors better signal quality. The latter bias favors shorter and less clearly articulated signals. In such communicative accounts, probabilistic reduction arises because contextual predictability increases the *a priori* accuracy and speed of message transmission, thereby allowing speakers to produce less costly signal. For contextually less predictable message components, on the other hand, speakers are expected to provide better signals in order to facilitate comprehension (e.g., Jaeger, 2013; Kohler, 1990; Kurumada & Jaeger, 2015; Lindblom, 1990a; Piantadosi, Tily, & Gibson, 2011; see also Aylett & Turk, 2004; Gibson, Bergen, & Piantadosi, 2013; Jaeger, 2010; Levy & Jaeger, 2007; Pate & Goldwater, 2015, though not all of these accounts do necessarily commit to the exact trade-off described here).

Three points deserve clarification as they have led to frequent confusion in the literature. First, the goal of facilitating comprehension is not to be confused with altruism: as described here, facilitating comprehension serves the speaker's communicative goals. Second, successful communication as defined here is not identical to the facilitation of comprehension of a particular linguistic unit (such as a phone, word, or syntactic structure). Rather, the successful recognition of a linguistic unit is taken to be relevant only to the extent that it serves the speaker's communicative goals (for a discussion of this for phonology, see Hall, Hume, Jaeger, & Wedel, submitted). Third, as formulated here, communicative accounts share with production ease accounts that production is assumed to be inherently costly. Therefore signal enhancement is not always expected, but rather expected when its benefits outweigh its costs (for further discussion, see Kurumada & Jaeger, 2015; Pate & Goldwater, 2015). As a consequence, the communicative accounts described here do not argue against the idea that production planning affects production preferences (see, e.g., Zipf, 1949; Jaeger, 2006, 2013; Lindblom, 1990a, 1990b). In fact, we are not aware of any account that claims that production ease does not affect linguistic encoding.[9]

What makes a signal 'better'? A central challenge that communicative accounts still have to address is what constitute a better signal. Presumably, a better signal is one that makes it more likely that the speaker achieves their communicative goals (e.g., the goal to be understood). A common assumption is that less reduced signals generally facilitate comprehension. In general this assumption seems warranted. For example, reduced durations or reduced or missing segments can make a word harder to comprehend out of context (e.g., Ernestus, Baayen, & Schreuder, 2002; see also van de Ven, Tucker, & Ernestus, 2011). There is also evidence that probabilistic reduction affects the out of context intelligibility of words: words that are more predictable in context tend to be pronounced in a way that makes them less intelligible out of context (e.g., Bard & Anderson, 1983, 1994; for similar evidence from *that* omission, see Race & MacDonald, 2003).

It is important to note, however, that signal enhancement is not *always* expected to facilitate comprehension. According to ideal observer models of comprehension (Kleinschmidt & Jaeger, 2015; Levy, 2008; Norris & McQueen, 2008), the property of the signal that determines how much it facilitates comprehension is its likelihood under the intended message, i.e., how likely the signal is to have arisen under the intended message (as compared to alternative messages). This means that extreme hyper-articulation might sometimes impede comprehension, because it produces percepts that are not expected under *any* message (including the intended one). More generally, signal enhancement should facilitate comprehension only to the extent that it increases the relative probability that this signal is observed under the intended message. Since the distribution of signals for any given message is itself conditioned on the context (the same message is realized differently in different contexts), this implies that, in the right context, reduced signals should be processed at least as easily and accurately as less reduced signals. Most pertinent to the current discussion, reduction is to be *expected* in contexts where the message is predictable. In such context, producing an enhanced signal should not necessarily facilitate comprehension and might even do the opposite or trigger additional inferences (see, e.g., Arnold, Fagnano, & Tanenhaus, 2003; Arnold, Kam, & Tanenhaus, 2007; Kravtchenko & Demberg, 2015; see also the discussion of *distinctiveness* versus *formality* in Rischel, 1992, pp. 387–388). In line with this reasoning, there is evidence that more signal does not facilitate comprehension when it occurs in context where reduction would be expected (e.g., Caballero & Kapatsinski, 2014; Jaeger, 2007; Race & MacDonald, 2003).[10]

A related challenge for future work is empirical. Most studies on this question have tested communicative accounts against *production* data, rather than intelligibility measures. This work is generally based on the assumption that more distinct acoustic signals will likely facilitate comprehension. For the reasons outlined above, this is, however, not always to be expected. Future work will thus have to assess whether variations in pronunciation that are attributed to communicative goals indeed facilitate comprehension. In doing so, it will be important to keep in mind that speakers might be willing to increase their production effort even if this results in only a small increase in the *average* probability of successful communication (for further discussion see Buz, Tanenhaus, & Jaeger, 2016; Buz, 2016).

The strong audience design hypothesis Broadly speaking, audience design refers to any aspect of production that serves to adapt productions so as to increase the probability of successful communication. The debates in the literature, however, have often focused on a much stronger variant of this hypothesis. The strong audience design hypothesis holds 1) that speakers integrate *listener-specific* information during linguistic encoding and 2) that they do so *immediately* when the information becomes available.

Evidence regarding the first criterion—listener-specificity—is mixed and seems to depend on the level of linguistic encoding. Listener-specific information can be seen as knowledge that speakers have about their interlocutors' perspective (such as information about which referents in a display are visually accessible to an interlocutor). For phonetic reduction, several studies have failed to find evidence for listener-specific audience design (e.g., Arnold *et al*., 2012; Bard *et al*., 2000; Kahn & Arnold, 2015). Other studies, however, have found that phonetic reduction is at least in part sensitive to listener-specific audience design (Galati & Brennan, 2010). The evidence is similarly mixed for the reduction or enhancement of prosodic boundaries. Some studies have found that speakers can strengthen the cues to prosodic phrasing (thereby signaling syntactic structure), thereby facilitating comprehension of utterances with temporary syntactic ambiguities (Price *et al*., 1991; Schafer, Speer, Warren, & White, 2000; but see Allbritton, McKoon, & Ratcliff, 1996; Snedeker & Trueswell, 2003). However, later studies have found that such strengthening takes place regardless of whether the utterance is actually ambiguous in the current context, thus arguing for listener-generic rather than listener-specific audience design (Kraljic & Samuel, 2005).

The picture seems to differ somewhat for lexical encoding and, specifically, the selection of referential expressions. For example, speakers' preference between pronouns (e.g., *he*), names (e.g., *John*), and full lexical noun phrases (e.g., *my colleague*) have been found to be at least partly affected by audience design considerations, though certainly not exclusively (Arnold, 2008, 2010; Arnold & Griffin, 2007; Bard *et al*., 2000). Speakers also seem to be capable of taking into consideration interlocutors' perspective and knowledge state (e.g., whether an interlocutor knows the label for a referent), when choosing between different ways of referring to same entity, though again not without fail (Brennan & Hanna, 2009; Ferreira *et al*., 2005; D. Heller, Gorman, & Tanenhaus, 2012).

For syntactic encoding, research on audience design has mostly focused on the question of whether speakers avoid temporary syntactic ambiguities that are known to lead to processing difficulty ("garden paths"). This work has returned little evidence that grammatical encoding is affected by ambiguity avoidance (see also Arnold, Wasow, Asudeh, & Alrenga, 2004; Ferreira, 2008; Wasow & Arnold, 2003; but see Haywood, Pickering, & Branigan, 2005; Roche, Dale, & Kreuz, 2010). For example, Temperley (2003) reports that writers are more likely to produce an optional relativer *that*, if it helps to avoid temporary ambiguity, but other production experiments (Ferreira & Dell, 2000) and corpus studies on written language (Jaeger, 2011; Roland, Dick, & Elman, 2007) and conversational speech (Jaeger, 2006) have failed to replicate this effect.

The second criterion for strong audience design has received comparatively little attention, although it is arguably critical in understanding failures to exhibit audience design. For example, in order to engage in effective audience design, speakers need to notice (though not necessarily become consciously aware of) that there is a potential comprehension problem. Speakers would also have to find a solution to this problem, such as determining which of several ways of encoding the intended message is most likely to successfully convey the intended meaning. This is likely to be computational costly, as has been recognized in arguments against strong audience design (Bard *et al.*, 2000; Ferreira, 2008; Shintel & Keysar, 2009). What has received less appreciation in the literature, however, is that this makes audience design a problem of *learning* how to best communicate in a given situation (but see Buz *et al.*, 2016; Galati & Brennan, 2010; Jaeger & Ferreira, 2013): when confronted with a novel situation—as is the case in, for example, most production experiments—speakers need to *infer* which variant is most likely to achieve their communicative goals (Buz *et al.*, submitted). This requires inference under uncertainty about what is in common ground, what will cause difficulty for the comprehender, and so on. Thus, even a mathematically ideal speaker (in the sense of ideal observers; J. R. Anderson, 1991) that engages in audience design would not be expected to immediately arrive at an optimal solution to the communicative problem. Rather, speakers would be expected to rely on prior expectations and adapt or learn from the perceived communicative success or failure of previous utterances (see also Brennan & Hanna, 2009; Buz *et al.*, 2016). While this prediction has, to the best of our knowledge, not received much attention in previous work, there is some evidence in support of it. For example, several recent studies have found that speakers adapt subsequent productions toward less reduced variants, if previous use of more reduced variants resulted in communicative failure (Buz *et al.*, 2016; Roche *et al.*, 2010; Schertz, 2013; Stent, Huffman, & Brennan, 2008). Such findings challenge accounts that attribute reduction solely to production ease.

Future research on audience design A few considerations deserve particular attention in future research on audience design. First, asking whether speakers are in principle capable of audience design is different from determining the conditions under which speakers actually engage in audience design. For example, some experiment that have failed to find audience design employed informed confederates as interlocutors (Dell & Brown, 1991). When the same experiments were repeated with interlocutors that believably benefitted from audience design, participants indeed engaged in audience design (Lockridge & Brennan, 2002). Interlocutors' believability, crucially, seems to be in large parts based on the absence of unnatural communicative *behavior*. For example, speakers and listeners seem to be exquisitely sensitive to the timing with which linguistic information is delivered, including speech rates, back-channels, etc. When interlocutors act in a way that violates these expectations, audience design effects are often suspended (for an excellent review, see Kuhlen & Brennan, 2013).

Second, in interpreting behavioral evidence for or against audience design, it can be helpful to ask whether audience design "makes sense" in a given situation—that is, whether it is a *rational* or *boundedly rational* behavior (see Buz *et al*., 2016; Jaeger, 2013; Kurumada & Jaeger, 2015). Specifically, it can be helpful to think about the *utility* of audience design in the current context: if, for example, listener-specific audience design increases attentional or memory demands, speakers might implicitly weigh these costs against the expected benefits of audience design. These expected benefits in turn depend on how likely audience design is to increase the probability of achieving one's communicative goals. This has several immediate consequences. One of them is believability, as discussed in the previous paragraph: if interlocutors act in a way suggesting that they will not benefit from audience design, these decrease the utility of audience design. More generally, when there is little incentive for successful communication, this decreases the utility of audience design (see also Tanenhaus, 2013). Even if there is an incentive to be understood, the utility of audience design is low when there is little *need* for it in the current context (e.g., when context provides sufficient information, Ferreira, 2008; Jaeger, 2010; Wasow & Arnold, 2003). Finally, thinking about utility requires consideration of the *available alternatives* to *a priori* audience design. For example, in some situations, it might be more effective for speakers to *repair* miscommunication after it has occurred (a form of a *posteriori* audience design), rather than to try to avoid miscommunication *a priori*.

Representational accounts

Our discussion so far has focused on the processes and architecture involved in online language production. A largely independent line of research in phonetics and speech production has focused on how the phonological representations of words become reduced over time (e.g., Bybee & Scheibman, 1999; Bybee & Hopper, 2001; Bybee, 2006; Johnson, 1997; Pierrehumbert, 2001, 2002; Wedel, 2006), including changes to the phonological system (e.g., phonological mergers, Wedel, Jackson, & Kaplan, 2013; Wedel, Kaplan, & Jackson, 2013; lenition, Cohen Priva, 2012, 2015). The primary goal of this line of research is to understand how phonological representations change over multiple generations, thereby explaining language change.

This research has investigated how the alternating forms available to speakers synchronically arise through historical processes (though most empirical evaluations have rested on synchronic data; for exceptions, see Hay, Pierrehumbert, Walker, & LaShell, 2015; Kanwal, Smith, Culbertson, & Kirby, 2017; Sóskuthy & Hay, 2017). Consider for example, speakers' selection between alternatives like *math* and *mathematics*, discussed above. While psycholinguistic research has identified factors that affect speakers' preference between these alternatives in a given context (Mahowald *et al*., 2013), historical research has focuses on changes in the availability of these forms over time (for a review of this and related work, see Hall, Hume, Jaeger, & Wedel, submitted).

It might, therefore, be tempting to consider the two lines of research as entirely orthogonal. However, representational accounts also offer explanations of reduction during online language production (for discussion, see Baese-Berk & Goldrick, 2009; Ernestus, 2014; Seyfarth, 2014). To illustrate this point, we first provide some background. We start with a brief overview of exemplar-based and related models (Johnson, 1997; Pierrehumbert, 2001). We then explain how these representational assumptions can explain the reduction of frequent words and discuss whether a similar account could be applied to probabilistic reduction beyond frequency effects.

The nature of phonological representations Research suggests that phonological representations go beyond abstract knowledge about a language's phonology. Rather, listeners seem to have rich knowledge of the specific acoustic realizations of words previously experienced. Some influential accounts of phonology hold that, each time a word is heard, its perceptual input—or at least some subphonemic representation of that input—is stored as an "episode" (e.g., Goldinger, 1996) or "exemplar" (e.g., Johnson, 1997), along with the context it is experienced in. The phonological representation of a word is then taken to be the cloud of all the exemplars that were recognized as that word in previous experience. Later exemplar-based accounts revise this assumption slightly by introducing a process called *entrenchment*. Entrenchment 'compresses' the cloud of exemplars toward the center of the exemplar cloud (Pierrehumbert, 2001). In these accounts, listeners do not necessarily maintain *all* previously experienced exemplars.

Under any of these accounts, the phonological representation of a word is inherently distributional (the cloud of exemplars), word-specific, and context-sensitive. These accounts have received strong support from research on speech perception (for reviews, see Foulkes & Hay, 2015; Pardo & Remez, 2006; Weatherholtz & Jaeger, 2016). For example, context-sensitivity is evidenced by studies finding that the recognition of speech sounds is improved if it occurs in the same context that it was previously experienced in (e.g., Drager, 2011; Goldinger, 1996; A. Walker & Hay, 2011). There is also evidence for word-specificity (for review, see Pierrehumbert, 2002). For example, the typical realization of words can change over time in ways that go beyond what is expected by their abstract phonology (i.e., their phonological citation form). A striking example of this are differences in the realization of homophones (i.e., words that have the same phonological citation form, such as *time* and *thyme*). Homophones with higher frequency tend to have shorter duration, compared to their less frequent homophone partner (Gahl, 2008). More generally, words with higher frequency and higher *average* contextual predictability tend to have not only fewer phonological segments (Manin, 2006; Piantadosi *et al*., 2011; Zipf, 1949), but also shorter average duration even after segment counts and types are taken into account (Seyfarth, 2014). Similarly, the rate with which a word exhibits final consonant deletion (e.g., *t*/*d*-deletion) increases with the *average* predictability of the deleted segment even after the predictability in the current context is taken into account (Cohen Priva, 2008, 2015). This suggests that at least some phonetic and phonological reduction effects are encoded in the word-specific phonetic representations. As we explain next, exemplar-based and related accounts not only offer an explanation for such word-specific effects, but also can explain why higher usage frequency leads to reduced forms.

Explaining the reduction of frequent words Pierrehumbert (2002) extends exemplar-based accounts of speech perception to production and shows that such a model predicts reduced realizations of frequent words under very general assumptions (though some of these assumptions are called into question by a recent diachronic study of sound change, see Hay *et al*., 2015). Specifically, production is assumed to consist of sampling from the cloud of previously stored exemplars, while being biased toward more reduced forms. This bias toward reduced forms is taken to follow from a general bias to minimize effort (see also Zipf, 1949). These two assumptions alone predict that words would get further and further reduced, the more often they are used. This correctly predicts more reduced forms for more frequent words, but also incorrectly predicts that words should quickly become reduced to no form at all.

There thus needs to be a competing bias to prevent arbitrary degrees of reduction—very much like in the communicative accounts discussed above. In the exemplar-based account, however, this bias is assumed to operate during *comprehension*: listeners are taken to store the inputs they receive only if understood with sufficient certainty (see also Lindblom, 1990a; Ohala, 1988). Since more frequent words are more likely to be correctly recognized by chance, even with a deteriorated acoustic signal as compared to less frequent words, the stored exemplars corresponding to a frequent word will—over time—contain increasingly more reduced forms. This shifts the *average* realization of the word toward a more reduced form, but not arbitrarily so (for similar ideas, see Guy, 1996; Labov, 1994, pp. 580–588; Ohala, 1988).

This exemplar-based account thus provides a cognitive plausible explanation of the inverse correlation between usage frequency and phonological form. It shares with the communicative accounts discussed above, that competing communicative biases affect what speakers produce. However, whereas the communicative accounts discussed above place both competing biases within production, the exemplar-based accounts distributes the biases across the production-perception loop.

Probabilistic reduction beyond the reduction of frequent words The argument made above for frequency readily extends to the *average* effect of contextual predictability on word-specific phonetics (for discussion, see Seyfarth, 2014). But can an exemplar-based account explain the effects of contextual predictability beyond those average effects (i.e., the effects of the *current* context on that word's realization)? The answer to this question depends on what assumptions are being made about the *granularity* of the exemplars that are stored by listeners. For example, as discussed above, there is evidence from production and comprehension that language users' implicit linguistic knowledge includes knowledge of *n*gram statistics (Arnon & Snider, 2010; Arnon & Cohen Priva, 2013; Bannard & Matthews, 2008; for discussion, see also Arnon & Cohen Priva, 2014; Baayen, Hendrix, & Ramscar, 2013). If producers sample from their previous experience in a way that takes into account the lexical *n*gram context, an exemplar-based model could thus explain the correlation between a word's *n*gram predictability and its reduction. More generally, if whatever cues correlate with a word's phonetic reduction are assumed to be stored along with the perceptual exemplar, it would seem that an exemplar-based model can, in principle, account for probabilistic phonetic reduction.

There is, however, evidence that at least the type of production-perception loop described above, which operates *between* interlocutors (Pierrehumbert, 2001, 2003), is insufficient to explain probabilistic reduction. This evidence comes from studies finding that speakers adjust their productions on the perceived communicative success of their previous utterances. One example of such research are perturbation studies on articulation (e.g., Houde & Jordan, 1998; Tourville, Reilly, & Guenther, 2008; Villacorta, Perkell, & Guenther, 2007). In perturbation studies, speakers' productions are manipulated online and played back to the speaker with a non-detectable delay. For example, a speaker might be producing the word *pen*, but hear herself produce something more like *pin* (Frank, 2011; Tourville *et al.*, 2008). These studies provide evidence that speakers can rapidly adjust their articulation if they perceive their own productions to deviate from their intended production. There is also evidence for similar adaptation based on feedback from interlocutors that, critically, does not involve the target word (Buz *et al.*, 2016; Schertz, 2013; Stent *et al.*, 2008). This includes non-verbal indication from interlocutors that they did not successfully understand the speaker. In recent work we find that speakers increased the hyper-articulation of minimal pair onset neighbors when their interlocutor failed to understand them (Buz *et al.*, 2016, submitted) for evidence of similar adaptation to syntactic production, see Roche *et al.*, 2010). The production-perception loop between interlocutors cannot explain this effect. Instead, it seems that speakers can also learn from their own productions, adapting subsequent productions to be better suited for the current context.

Summary and open questions While insufficient as a sole account of phonetic reduction, the production-perception loop and exemplar-based accounts likely form part of an explanation of probabilistic phonetic reduction. It remains to be seen how these accounts can be extended to reduction at other levels of linguistics representation.

Beyond the specifics of exemplar-based accounts, the relation between offline and online accounts of probabilistic reduction and related phenomena has only relatively recently become a target of research (Baese-Berk & Goldrick, 2009; Buz, 2016; Buz *et al.*, 2016; Cohen Priva, 2008; Ernestus, 2014; Seyfarth, 2014). This is an area we consider of particular importance in advancing the understanding of linguistic reduction: representational and online accounts offer qualitatively different, though related and mutually compatible explanations for probabilistic reduction. Studying these different explanations in isolation of each other risks missing the bigger picture.

Finally, speakers' ability to learn from their own productions raises an interesting question: are such adaptations stored, so that they remain available for similar future occasions? Research on speech *perception* has found that listeners can learn expectations about talker-specific pronunciations (Kraljic & Samuel, 2007; Norris, McQueen, & Cutler, 2003) and even generalization across groups of talkers (e.g., Bradlow & Bent, 2008; Baese-Berk, Bradlow, & Wright, 2013; Weatherholtz, 2015; for review, see Weatherholtz & Jaeger, 2016). Once learned, talker- and group-specific expectations do not seem to be lost, but can be maintained over longer periods of time (e.g., Eisner & McQueen, 2006; Goldinger, 1996; Hay, Warren, & Drager, 2006; Kraljic & Samuel, 2006; Niedzielski, 1999; Strand, 1999; for review, see Foulkes & Hay, 2015; Kleinschmidt & Jaeger, 2015). It is thus possible that linguistic encoding is sensitive to this or similar talker-, group-, or situation-specific knowledge. Some support for this hypothesis comes from studies suggesting that speakers can maintain multiple phonetic representations for different dialects or registers (Clopper & Pierrehumbert, 2008). It remains to be seen whether speakers develop and store even more specific (talker-, situation-, or task-specific) representations. If this view receives further support, it offers a way to reconcile the presence of context-specific communicative effects on language production with the apparent limitations of listener-specific audience design: when confronted with a novel situation, speakers first need to learn effective communicative behaviors for that situation. Once learned and reinforced, these behaviors—such as targeted hyper-articulation of a specific phonological segment (Buz *et al.*, 2016; Kirov & Wilson, 2012; Schertz, 2013; Seyfarth *et al.*, 2016)—might be stored and thus more easily available in similar future situations.

Conclusion

Human languages provide speakers with alternative means of expressing near meaning-equivalent messages. The competing variants of such alternations differ in the amount and quality of linguistic signal that is provided to listeners. Research on different types of reduction has often proceeded in separate lines of research. This applies in particular to quantitative study of reduction, with most studies focusing on one type of reduction (but see Finegan & Biber, 2001). Here we have focused on one generalization that seems to apply to reduction at various levels of linguistic representation, *probabilistic reduction*. We have summarized evidence that the reduction of form and signal is sensitive to contextual predictability and that this generalization seems to hold across many levels of linguistic representation. Specifically, speakers prefer comparatively reduced realizations of contextually more predictable message components. This relation between the predictability or informativity of message components and the amount of form and signal provided in encoding them has long intrigued language researchers.

The mechanisms that underlie such probabilistic reduction are still under debate. There is, however, converging evidence that both production ease and communicative goals mediate reduction (see also Arnold *et al.*, 2012, p. 506), and that reduction can become entrenched over time, leading to reduced canonical forms (see also Ernestus, 2014). A better understanding of how these factors interact will inform research on the architecture of the language production system, the effect of language use and linguistic representations on production, and language change.

Acknowledgments

The authors wish to thank Jennifer Arnold, Ellen Bard, Alan Bell, Vic Ferreira, Susanne Gahl, Matt Goldrick, Chigusa Kurumada, Janet Pierrehumbert, Uriel Cohen Priva, Andy Wedel, an anonymous reviewer, and the editors of the volume, Eva Fernández and Helen Cairns, for helpful feedback on earlier versions of this article. We also thank Meghan Sumner and members of the Human Language Processing Lab for helpful related discussions. This work was partially supported by NSF CAREER grant IIS-1150028 to TFJ and a graduate fellowship (NIH training grant T32 DC000035) as well as an NRSA pre-doc (F31HD083020) to EB. T. Florian Jaeger may be contacted at fjaeger@ur.rochester.edu.

NOTES

1 In conversational American English, for example, the complementizer is absent in 82.5% of all complement clauses (Jaeger, 2010, p. 29). Even when the most frequent complement clause embedding verbs are excluded, omission is more frequent (53%) than mention of complementizer *that* (Jaeger, 2010, Table 1).

2 The informativity of, for example, a word can be measured as its Shannon information (also called surprisal). The Shannon information of a word is defined as the logarithm of the inverse of the word's probability in context (Shannon, 1948).

3 For example, the information that hearing a [t] after an [s] adds to the recognition of the word *street*.

4 Investigations of phonetic reduction in conversational speech have reported that about 1–2% or less of the variance in word duration is attributable to contextual predictability (Aylett & Turk, 2004; Bell *et al.*, 2009). This would suggest that contextual predictability

has only a small effect on word durations (for example, word frequency is reported to account for about 10% of the variance, Aylett & Turk, 2004). This is, however, misleading since word durations are strongly affected by the phonological canonical form of the word. It is therefore not surprising that factors that vary *between* word forms (such as word frequency) account for more variance than those that vary *within* word forms (such as contextual predictability).

5 This is not to say that the distributions of these forms is unconstrained. For example, both pronouns and lexical noun phrases are subject to constraints of Binding Theory—including, potentially, categorical ones. However, similarly strong constraints apply to optional function word omission or morphological reduction (for examples and references, cf. Jaeger, 2006). Our point here is that there *are* contexts in which speakers can choose between the different forms while maintaining near meaning-equivalence.

6 Some of these effects of syntactic context might be due to predictability effects on prosodic phrasing, with less expected prosodic boundaries being realized with longer phrase final lengthening (Gahl & Garnsey, 2004, 2006; Kurumada, 2011).

7 Or, in the terminology of Dell and Brown (1991), *listener-particular* audience design. We use the term listener-specific to highlight parallels to talker-specific expectations during *comprehension* (e.g., in speech perception, Bradlow & Bent, 2008; lexical processing, Creel, Aslin, & Tanenhaus, 2008; and sentence processing, Kamide, 2012; Kraljic & Samuel, 2007).

8 More recent experiments suggest that it is the number of *any* type of neighbors, rather than minimal pair neighbors, that is correlated with hyper-articulation of the/p/(Fox *et al*., 2015; for discussion, see also Peramunage *et al*., 2011). This might suggest that the hyper-articulated voice onset timing observed in Baese-Berk & Goldrick (2009) was part of more general across-the-board hyper-articulation for words with many neighbors (which has independently been observed, see e.g., Buz & Jaeger, 2015; Gahl, 2015; but see Gahl *et al*., 2012; Munson, 2007). Regardless of the specifics though, hyper-articulation of words with phonological neighbors is compatible with competition accounts.

9 The opposite position—that at least certain aspects of production are too 'automatic' to be directly affected by communicative goals—*has been* proposed (Bard & Aylett, 2005; e.g., Bard *et al*., 2000; Ferreira, 2008).

10 To further complicate the picture, it is possible that what constitutes a good signal can—at least to some extent—change dynamically as a result of accommodation and alignment processes (see, e.g., Pardo, 2017).

REFERENCES

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(3), 714–735. doi: 10.1037/0278-7393.22.3.714

Anderson, A. H., Badger, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., … Weinert, R. (1991). The HCRC Map Task corpus. *Language and Speech*, 34(4), 351–366. doi: 10.1177/002383099103400404

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, 14(03), 471–485. doi: 10.1017/S0140525X00070801
Anibel, B. (2010). *Optional Marking of Coherence Relations: Production of Coherence Markers in Sentence Completion Tasks*. Master's thesis, University of Rochester.
Anttila, A., Adams, M., & Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes*, 25(7-9), 946–981. doi: 10.1080/01690960903525481
Ariel, M. (1999). Cognitive universals and linguistic conventions: The case of resumptive pronouns. *Studies in Language*, 23(2), 217–269.
Arnold, J. E. (1998). *Reference form and discourse patterns* (doctoral dissertation). Stanford University.
Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527. doi: 10.1080/01690960801920099
Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Linguistics and Language Compass*, 4, 187–203. doi: 10.1111/j.1749-818X.2010.00193.x
Arnold, J. E., & Griffin, Z. M. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56(4), 521–536. doi: 10.1016/j.jml.2006.09.007
Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism. *Cognition*, 110, 131–146. doi: 10.1016/j.cognition.2008.10.016
Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of psycholinguistic research*, 32(1), 25–36. doi: 10.1023/A:1021980931292
Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, 19(3), 505–512. doi: 10.3758/s13423-012-0233-y
Arnold, J. E., Kam, C. L. H., & Tanenhaus, M. K. (2007). If You Say Thee uh You Are Describing Something Hard: The On-Line Attribution of Disfluency During Reference Comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 914–930. doi: 10.1037/0278-7393.33.5.914
Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28–55. doi: 10.1353/lan.2000.0045
Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, 51(1), 55–70. doi: 10.1016/j.jml.2004.03.006
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. doi: 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. *The Mental Lexicon*, 9(3), 377–400. doi: 10.1075/ml.9.3.01arn
Arnon, I., & Snider, N. E. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67–82. doi: 10.1016/j.jml.2009.09.005
Asr, F. T., & Demberg, V. (2012). Implicitness of discourse relations. In M. Kay & C. Boitet (Eds.), *COLING* (pp. 2669–2684). Mumbai: Indian Institute of Technology Bombay.
Asr, F. T., & Demberg, V. (2015). Uniform information density at the level of discourse relations: Negation markers and discourse connective omission.

In *Proceedings of the 11th international conference on computational semantics* (pp. 118–128). London, UK: Association for Computational Linguistics.

Aylett, M. P., & 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. doi: 10.1177/00238309040470010201

Aylett, M. P., & Turk, A. E. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, 119(5), 3048–3058. doi: 10.1121/1.2188331

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(3), 329–347. doi: 10.1177/0023830913484896

Baese-Berk, M. M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and Cognitive Processes*, 24(4), 527–554. doi: 10.1080/01690960802299378

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133(3), EL174–EL180. doi: 10.1121/1.4789864

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The Effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19(3), 241–248. doi: 10.1111/j.1467-9280.2008.02075.x

Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language*, 10(02), 265–292. doi: 10.1017/S0305000900007777

Bard, E. G., & Anderson, A. H. (1994). The unintelligibility of speech to children: Effects of referent availability. *Journal of Child Language*, 21(03), 623–648. doi: 10.1017/S030500090000948X

Bard, E. G., & Aylett, M. P. (2005). Referential form, word duration, and modeling the listener in spoken dialogue modeling listeners while speaking. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 173–191). Cambridge: MIT Press.

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

Beaver, D., Clark, B. Z., Flemming, E. S., Jaeger, T. F., & Wolters, M. K. (2007). When semantics meets phonetics: Acoustical studies of second occurrence focus. *Language*, 83(2), 245–276. doi: 10.1353/lan.2007.0053

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. S. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92–111. doi: 10.1016/j.jml.2008.06.003

Bell, A., Jurafsky, D. S., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America*, 113(2), 1001. doi: 10.1121/1.1534836

Bender, E. M. (2000). *Syntactic variation and linguistic competence: The case of AAVE copula absence*. Doctoral dissertation, Stanford University.

Berkovits, R. (1994). Durational Effects in Final Lengthening, Gapping, and Contrastive Stress. *Language and Speech*, 37(3), 237–250. doi: 10.1177/002383099403700302

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review*, 89(1), 1–47. doi: 10.1037/0033-295X.89.1.1

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language*, 26, 119–137. doi: 10.1016/0749-596X(87)90120-3

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. doi: 10.1016/j.cognition.2007.04.005

Branigan, H. P., Pickering, M. J., & Tanaka, M. N. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua*, 118(2), 172–189. doi: 10.1016/j.lingua.2007.02.003

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. doi: 10.1037/0278-7393.22.6.1482

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, 1(2), 274–291. doi: 10.1111/j.1756-8765.2009.01019.x

Bresnan, J., & Spencer, J. (2016). Frequency effects in spoken syntax: "have" and "be" contraction. In S. Grondelaers & R. van Hout (Eds.), *In new ways of analyzing syntactic variation*. Berlin: De Gruyter Mouton.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Edita-The Publishing House of the Royal Netherlands Academy.

Buz, E. (2016). *Speaking with a (slightly) new voice: Investigating speech changes, communicative goals, and learning from past communicative success*. Doctoral dissertation, University of Rochester.

Buz, E., & Jaeger, T. F. (2016). The (in) dependence of articulation and lexical planning during isolated word production. *Language, Cognition and Neuroscience*, 31(3), 404–424. doi: 10.1080/23273798.2015.1105984

Buz, E., Jaeger, T. F., & Tanenhaus, M. K. (2014). Contextual confusability leads to targeted hyperarticulation. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Quebec City, Canada: Cognitive Science Society.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (submitted). Do speakers make inferences about the causes of miscommunication? Manuscript, University of Rochester.

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. doi: 10.1016/j.jml.2015.12.009

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(4), 421–437. doi: 10.1016/j.jml.2010.01.002

Bürki, A., Ernestus, M., Gendrot, C., Fougeron, C., & Frauenfelder, U. H. (2011). What affects the presence versus absence of schwa and its duration: A corpus analysis of French connected speech. *The Journal of the Acoustical Society of America*, 130(6), 3980. doi: 10.1121/1.3658386

Bybee, J., & Hopper, P. (2001). Introduction to frequency and the emergence of linguistic structure. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 9–32). John Benjamins Publishing.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, 37(4), 575–596. doi: 10.1515/ling.37.4.575

Caballero, G., & Kapatsinski, V. (2014). Perceptual functionality of morphological redundancy in Choguita Rarámuri (Tarahumara). *Language, Cognition and Neuroscience*, (July 2015), 1–10. doi: 10.1080/23273798.2014.940983

Chen, Q., & Mirman, D. (2015). Interaction between phonological and semantic representations: Time matters. *Cognitive science*, 39(3), 538–558. doi: 10.1111/cogs.12156

Chomsky, N. (2000). *New horizons in the study of language and the mind* (p. 230). Cambridge: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111. doi: 10.1016/S0010-0277(02)00017-3

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. L. Ny & W. Kintsch (Eds.), *Language and comprehension* (1st ed., pp. 287–300). Amsterdam: North-Holland Publishing Company.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, 37(3), 201–242. doi: 10.1006/cogp.1998.0693

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. doi: 10.1121/1.2953322

Cohen Priva, U. (2008). Using information content to predict phone deletion. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th West Coast Conference on Formal Linguistics* (pp. 90–98). Somerville, MA: Cascadilla.

Cohen Priva, U. (2012). *Sign and signal: deriving linguistic generalizations from information utility*. Doctoral dissertation, Stanford University.

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology*, 6(2), 243–278. doi: 10.1515/lp-2015-0008

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition*, 106(2), 633–664. doi: 10.1016/j.cognition.2007.03.013

De Jong, K. (2004). Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics*, 32(4), 493–516. doi: 10.1016/j.wocn.2004.05.002

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321. doi: 10.1037/0033-295X.93.3.283

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A swarthmore festschrift for lila gleitman* (pp. 105–130). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, 23(4), 517–542. doi: 10.1016/S0364-0213(99)00014-2

Demberg, V., Sayeed, A. B., Gorinski, P. J., & Engonopoulos, N. (2012). Syntactic surprisal affects spoken word duration in conversational contexts. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 356–367). Association for Computational Linguistics.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, 54(Pt 1), 99–121. doi: 10.1177/0023830910388017

Dryer, M. S. (2013). Expression of pronominal subjects. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (pp. 410–413). Leipzig: Max Planck Institute for Evolutionary Anthropology.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, 119(4), 1950–1953. doi: 10.1121/1.2178721

Elsness, J. (1984). That or Zero? A look at the choice of object clause connective in a corpus of American English. *English Studies*, 65(6), 519–533.

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

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3), 253–260. doi: 10.1016/S0095-4470(11) 00055-6

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and language*, 81(1-3), 162–173. doi: 10.1006/brln.2001.2514

Fedzechkina, M., Newport, E., and Jaeger, T. F. 2016. Balancing effort and information transmission during language acquisition: *Evidence from word order and case-marking. Cognitive Science*, 41(2), 416–446. doi: 10.1111/cogs.12346

Fedzechkina, M., Jaeger, T. F., and Newport, E. 2012. Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109(44), 17897–17902. doi: 10.1073/pnas.1215776109

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35(5), 724–755. doi: 10.1006/jmla.1996.0038

Ferreira, V. S. (2008). Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation*, 49, 209–246. doi: 10.1016/S0079-7421(08)00006-6

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296–340. doi: 10.1006/cogp.1999.0730

Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32(6), 669–692. doi: 10.1023/A:1026146332132

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, 96(3), 263–284. doi: 10.1016/ j.cognition.2004.09.002

Finegan, E., & Biber, D. (2001). Register variation and social dialect variation: The register axiom. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 235–267). Cambridge: Cambridge University Press.

Flickinger, D., & Wasow, T. (2013). A corpus-driven analysis of the Do-Be construction. In P. Hofmeister & E. J. Norcliffe (Eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag* (pp. 35–63). CSLI Publications.

Foulkes, P., & Hay, J. B. (2015). The emergence of sociophonetic structure. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 292–313). Oxford: Blackwell.

Fox, N. P., Reilly, M., & Blumstein, S. E. (2015). Phonological neighborhood competition affects spoken word production irrespective of sentential context. *Journal of Memory and Language*, 83, 97–117. doi: 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(2), 151–67. doi: 10.1016/ S0010-0277(96)00781-0

Frank, A. F. (2011). *Integrating linguistic, motor, and perceptual information in language production*. Doctoral dissertation, University of Rochester.

Frank, A. F., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive*

Science Society (pp. 939–944). Cognitive Science Society.
Friedman, L. F. (2013). Chart: The Art of "'Sup". Retrieved from https://www.psychologytoday.com/articles/201304/chart-the-art-sup
Fry, J. S. (2001). *Ellipsis and wa-marking in Japanese conversation*. Doctoral dissertation, Stanford University.
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. doi: 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. doi: 10.1016/j.wocn.2014.12.002
Gahl, S., & Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, 80(4), 748–775. doi: 10.1353/lan.2004.0185
Gahl, S., & Garnsey, S. M. (2006). Knowledge of Grammar Includes Knowledge of Syntactic Probabilities. *Language*, 82(2), 405–410. doi: 10.1353/lan.2006.0089
Gahl, S., Jurafsky, D. S., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, & Computers*, 36(3), 432–443. doi: 10.3758/BF03195591
Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, 66(4), 789–806. doi: 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(1), 35–51. doi: 10.1016/j.jml.2009.09.002
Gallo, C. G. (2011). *Incremental interpretation and production across clauses*. Doctoral dissertation, University of Rochester.
Gallo, C. G., Jaeger, T. F., & Furth, K. (2010). A Database for the Exploration of Spanish Planning. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, … D. Tapias (Eds.), *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 1–4).
Gallo, C. G., Jaeger, T. F., & Smyth, R. (2008). Incremental Syntactic Planning across Clauses. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1294–1299).
Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58–93. doi: 10.1006/jmla.1997.2512
Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(20), 8051–8016. doi: 10.1073/pnas.1216438110
Givón, T. (1983). *Topic Continuity in Discourse: A quantitative cross-language study* (p. 492). John Benjamins Publishing.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1166–1183. doi: 10.1037/0278-7393.22.5.1166
Goldrick, M., Vaughn, C., & Murphy, A. (2013). The effects of lexical neighbors on stop consonant articulation. *The Journal of the Acoustical Society of America*, 134(2), EL172–EL177. doi: 10.1121/1.4812821
Guy, G. R. (1996). Form and function in linguistic variation. In G. R. Guy, C. Feagin, D. Schiffrin, & J. Baugh (Eds.), *Towards a social science of language: Papers in honor of william labov. volume 1: Variation and change in language and society*

(pp. 221–252). Amsterdam: Benjamins Publishing Company.
Hanique, I., Ernestus, M., & Schuppler, B. (2013). Informal speech processes can be categorical in nature, even if they affect many different words. *The Journal of the Acoustical Society of America*, 133(3), 1644–1655. doi: 10.1121/1.4790352
Hare, M., McRae, K., & Elman, J. L. (2004). Admitting that admitting verb sense into corpus analyses makes sense. *Language and Cognitive Processes*, 19(2), 181–224. doi: 10.1080/01690960344000152
Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition*, 139, 83–91. doi: 10.1016/j.cognition.2015.02.012
Hay, J. B., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34(4), 458–484. doi: 10.1016/j.wocn.2005.10.001
Haywood, S. L., Pickering, M. J., & Branigan, H. P. (2005). Do speakers avoid ambiguities during dialogue? *Psychological Science*, 16(5), 362–366. doi: 10.1111/j.0956-7976.2005.01541.x
Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: shared knowledge affects referential form. *Topics in Cognitive Science*, 4(2), 290–305. doi: 10.1111/j.1756-8765.2012.01182.x
Heller, J. R., & Goldrick, M. (2014). Grammatical constraints on phonological encoding in speech production. *Psychonomic Bulletin & Review*, 21(6), 1576–1582. doi: 10.3758/s13423-014-0616-3
Heller, J. R., & Pierrehumbert, J. B. (2011). Word burstiness improves models of word reduction in spontaneous speech. In *Architectures and Mechanisms for Language Processing (AMLaP) Conference*. Paris, France.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117. doi: 10.1016/0010-0277(96)81418-1
Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. doi: 10.1126/science.279.5354.1213
Huddleston, R., & Pullum, G. K. (2002). *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.
Hall, K., Hume, E., Jaeger, T. F., & Wedel, A. (submitted). *The message shapes phonology*.
Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. Doctoral dissertation, Stanford University.
Jaeger, T. F. (2007). Rational speakers: Speakers help processing when it is most necessary. In *Architectures and Mechanisms for Language Processing (AMLaP) Conference*, Ghent, Belgium.
Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. doi: 10.1016/j.cogpsych.2010.02.002
Jaeger, T. F. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. M. Bender & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing. studies in honor of Tom Wasow* (pp. 161–197). Stanford, CA: CSLI Publications.
Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, 4, 230. doi: 10.3389/fpsyg.2013.00230
Jaeger, T. F., & Ferreira, V. S. (2013). Seeking predictions from a predictive framework. *Behavioral and Brain Sciences*, 36(4), 359–360. doi: 10.1017/S0140525X12002762
Jaeger, T. F., & Grimshaw, J. (2013). Information density affects both production and grammatical constraints. In *Architectures and Mechanisms for Language Processing (AMLaP) Conference*. Marseille, France.
Jaeger, T. F., & Norcliffe, E. J. (2009). The Cross-linguistic Study of Sentence Production. *Language and Linguistics Compass*, 3(4), 866–887. doi: 10.1111/j.1749-818X.2009.00147.x
Jaeger, T. F., & Wasow, T. (2006). Processing as a Source of Accessibility Effects on

Variation. In R. T. Cover & Y. Kim (Eds.), *Proceedings of the 31st annual meeting of the Berkeley Linguistic Society* (pp. 169–180). Ann Arbor, MI: Sheridan Books.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012a). Incremental phonological encoding during unscripted sentence production. *Frontiers in Psychology*, 3(November), 481. doi: 10.3389/fpsyg.2012.00481

Jaeger, T. F., Furth, K., & Hilliard, C. (2012b). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1439–1449. doi: 10.1037/a0027862

Jaeger, T. F., Snider, N. E., Staum, L., & Jurafsky, D. S. (2006). (In)dependence of lexical and syntactic production: that-reduction and omission in spontaneous speech. In *Proceedings of the 19th Annual CUNY Conference on Human Sentence Processing* (p. 42). New York, NY.

Janssen, N., & Caramazza, A. (2009). Grammatical and phonological influences on word order. *Psychological Science*, 20(10), 1262–1268. doi: 10.1111/j.1467-9280.2009.02429.x

Jenkins, L. (2000). *Biolinguistics: Exploring the biology of language*. Cambridge: Cambridge University Press.

Johnson, K. (1997). The auditory/perceptual basis for speech segmentation. *OSU Working Papers in Linguistics*, 50, 101–113.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*. Tokyo, Japan: The National International Institute for Japanese Language.

Jurafsky, D. S., Bell, A., Gregory, M., & Raymond, W. (2001). 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: John Benjamins Publishing Company.

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(3), 311–325. doi: 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(1–2), 103–119. doi: 10.1080/01690965.2013.848989

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's Law of Abbreviation and the Principle of Least Effort: Language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52. http://doi.org/10.1016/j.cognition.2017.05.001.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124(1), 66–71. doi: 10.1016/j.cognition.2012.03.001

Kempen, G., & Harbusch, K. (2004). A corpus study into word order variation in German subordinate clauses: Animacy affects linearization independently of grammatical function assignment. In T. Pechmann & C. Habel (Eds.), *Multidisciplinary approaches to language production* (pp. 173–182). Berlin: Walter de Gruyter.

Kidd, C., & Jaeger, T. F. (2008). Prosodic phrasing and function word pronunciation. In *Experimental and Theoretical Advances in Prosody*. Cornell University, Ithaca, NY.

Kirov, C., & Wilson, C. (2012). The specificity of online variation in speech production. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 587–592). Austin, TX: Cognitive Science Society.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, 122(2), 148–203. doi: 10.1037/a0038695

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 69–92). Boston/London: Kluwer.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178. doi: 10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2), 262–268. doi: 10.3758/BF03193841

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, 56(1), 1–15. doi: 10.1016/j.jml.2006.07.010

Kravtchenko, E. (2014). Predictability and syntactic production: Evidence from subject omission in Russian. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 785–790). Austin, TX: Cognitive Science Society.

Kravtchenko, E., & Demberg, V. (2015). Semantically underinformative utterances trigger pragmatic inferences. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1207–1212). Austin, TX: Cognitive Science Society.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science (New York, N.Y.)*, 277(5326), 684–686. doi: 10.1126/science.277.5326.684

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, 20(1), 54–72. doi: 10.3758/s13423-012-0341-8

Kuperman, V., Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2007). Morphological predictability and acoustic duration of interfixes in Dutch compounds. *The Journal of the Acoustical Society of America*, 121(4), 2261. doi: 10.1121/1.2537393

Kurumada, C. (2011). *Syntactic context affects probability estimation: Evidence from Japanese relative clauses*. Manuscript. Stanford University.

Kurumada, C., & Jaeger, T. F. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, 83, 152–178. doi: 10.1016/j.jml.2015.03.003

Labov, W. (1994). *Principles of linguistic change, volume 1: Internal factors* (p. 664). Malden, MA: Wiley-Blackwell.

Lee, H. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean. *Journal of East Asian Linguistics*, 15(1), 69–96. doi: 10.1007/s10831-005-3004-1

Lee, M.-W., & Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition*, 105(2), 446–456.

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. doi: 10.1016/j.cognition.2007.05.006

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in*

Neural Information Processing Systems (Vol. 19, pp. 849–856). Cambridge, MA: MIT Press.

Lindblom, B. (1990a). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Kluwer Academic Publishers.

Lindblom, B. (1990b). On the communication process: Speaker-listener interaction and the development of speech. *Augmentative and Alternative Communication*, 6(4), 220–230. doi: 10.1080/07434619012331275504

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, 9(3), 550–557. doi: 10.3758/BF03196312

Lombard, E. (1911). Le signe de l'elevation de la voix. *Ann Maladies Oreille, Larynx, Nez, Pharynx*, 37(25), 101–119.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226. doi: 10.3389/fpsyg.2013.00226

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703. doi: 10.1037/0033-295X.101.4.676

Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2), 313–318. doi: 10.1016/j.cognition.2012.09.010

Manin, D. Y. (2006). Experiments on predictability of word in context and information rate in natural language. *Journal of Information Processes*, 229–236.

McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2), 188–230. doi: 10.1006/cogp.1993.1005

Melnick, R. (2011). *Judgment studies of information-driven syntactic reduction.* Master's thesis, Stanford University.

Munson, B. (2007). Lexical Access, Lexical Representation, and Vowel Production. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 201–228). Berlin: Mouton de Gruyter.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1), 62–85. doi: 10.1177/0261927X99018001005

Norcliffe, E. J. (2009). *Head-marking in usage and grammar: A study of variation and change in Yucatec Maya*. Doctoral dissertation, Stanford University.

Norcliffe, E. J., & Jaeger, T. F. (2014). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*, 1–39. doi: 10.1017/langcog.2014.39

Norcliffe, E. J., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience*, 30(9), 1009–1032. doi: 10.1080/23273798.2015.1080373

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. doi: 10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. doi: 10.1016/S0010-0285(03)00006-9

Ohala, J. J. (1988). Discussion of Bjoern Lindblom's "Phonetic Invariance and the adaptive nature of speech". In *Working models of human perception*. London, UK: Academic Press.

O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: sequential and nonsequential processes in word production. *Journal of*

Experimental Psychology: Human Perception and Performance, 26(1), 57–73. doi: 10.1037/0096-1523.26.1.57

Pardo, J. S. (2017). Parity and disparity in conversational interaction. In E. M. Fernández & H. S. Cairns (Eds.), *Handbook of psycholinguistics*. Wiley-Blackwell.

Pardo, J. S., & Remez, R. E. (2006). The Perception of Speech. In *The handbook of psycholinguistics, 2nd edition* (pp. 201–248).

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, 78, 1–17. doi: 10.1016/j.jml.2014.10.003

Pellegrino, F., Coupé, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language*, 87(3), 539–558. doi: 10.1353/lan.2011.0057

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. M. (2011). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience*, 23(3), 593–603. doi: 10.1162/jocn.2010.21489

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, 108(9), 3526–3529. doi: 10.1073/pnas.1012551108

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29, 434–446. doi: 10.1044/jshr.2904.434

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

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

Pierrehumbert, J. B. (2003). Probabilistic Phonology : Discrimination and Robustness. In R. Bod, J. B. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics*. Cambridge, MA: MIT Press.

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

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005b). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4), 146–159. doi: 10.1159/000090095

Post, M., & Jaeger, T. F. (2010). Word Production in Spontaneous Speech: Availability and Communicative Efficiency. In *The 23rd CUNY Sentence Processing Conference*. NYC, NY.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, 90(6), 2956. doi: 10.1121/1.401770

Race, D. S., & MacDonald, M. C. (2003). The use of "that" in the production and comprehension of object relative clauses. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 946–951). Mahwah, NJ: Cognitive Science Society.

Rapp, D. N., & Samuel, A. G. (2002). A reason to rhyme: Phonological and semantic influences on lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 564–571. doi: 10.1037/0278-7393.28.3.564

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61(1-2), 127–159. doi: 10.1016/S0010-0277(96)00722-6

Rischel, J. (1992). Formal linguistics and real speech. *Speech Communication*, 11, 379–392. doi: 10.1016/0167-6393(92) 90043-7

Roche, J. M., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during conversation: More than mere mimicry? In R. Camtrabone & S. Ohlsson (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (pp. 206–211). Austin, TX: Cognitive Science Society.

Rohde, H., & Kehler, A. (2014). Grammatical and information-structural influences on pronoun production. *Language, Cognition and Neuroscience*, 29(8), 912–927. doi: 10.1080/01690965.2013.854918

Rohdenburg, G. (2004). Grammatische Parallelen zwischen niederdeutschen Mundarten und Nichtstandardvarietäten im Englischen aus typologischer Sicht. *Niederdeutsches Jahrbuch*, 127, 85–122.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348–379. doi: 10.1016/j.jml.2007.03.002

Roland, D., Elman, J. L., & Ferreira, V. S. (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, 98(3), 245–272. doi: 10.1016/j.cognition.2004.11.008

Rose, D. E. (2017). Predicting plurality: An examination of the effects of morphological predictability on the learning and realization of bound morphemes (Unpublished PhD dissertation). University of Canterbury, Christchurch, New Zealand. Chapter 4.

Rosenbach, A. (2008). Animacy and grammatical variation: Findings from English genitive variation. *Lingua*, 118(2), 151–171. doi: 10.1016/j.lingua.2007.02.002

Rosenfeld, R. (1996). A maximum entropy approach to adaptive statistical language modeling. *Computer, Speech and Language*, 10, 187–228. doi: 10.1006/csla.1996.0011

Sayeed, A. B., Fischer, S., & Demberg, V. (2015). Vector-space calculation of semantic surprisal for predicting word pronunciation duration. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (pp. 763–773). Beijing, China: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/P15-1074

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

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of psycholinguistic research*, 29(2), 169–182. doi: 10.1023/A:1005192911512

Schertz, J. (2013). Exaggeration of featural contrasts in clarifications of misheard speech in English. *Journal of Phonetics*, 41(3-4), 249–263. doi: 10.1016/j.wocn.2013.03.007

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition*, 53, 91–127. doi: 10.1016/0749-596X(90)90011-N

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, 133(1), 140–155. doi: 10.1016/j.cognition.2014.06.013

Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). Dynamic hyperarticulation of coda voicing contrasts. *Journal of the Acoustic Society of America* 139(2), EL31–37. doi: 10.1121/1.4942544.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Mobile Computating and Communications Review*, 5(1), 3–55.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30(1), 50–64. doi: 10.1002/j.1538-7305.1951.tb01366.x

Shintel, H., & Keysar, B. (2009). Less Is More: A Minimalist Account of Joint Action in Communication. *Topics in Cognitive Science*, 1(2), 260–273. doi: 10.1111/j.1756-8765.2009.01018.x

Shriberg, E. (1996). Disfluencies in switchboard. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of International Conference on Spoken Language Processing* (pp. 11–14). Philadelphia, PA: IEEE.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41(1), 1–20. doi: 10.1146/annurev.ps.41.020190.000245

Smith, M., & Wheeldon, L. R. (2004). Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 675–686. doi: 10.1037/0278-7393.30.3.675

Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and language*, 48, 103–130. doi: 10.1016/S0749-596X(02)00519-3

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

Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, 50(3), 163–178. doi: 10.1016/j.specom.2007.07.005

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, 10(01), 1–15. doi: 10.1017/S0305000900005092

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology*, 18(1), 86–100. doi: 10.1177/0261927X99018001006

Tagliamonte, S., & Smith, J. (2005). No momentary fancy! The zero in English dialects. *English Language and Linguistics*, 9(02), 289. doi: 10.1017/S1360674305001644

Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65(3), 318–330. doi: 10.1016/j.jml.2011.04.009

Tanenhaus, M. K. (2013). All P's or mixed vegetables? *Frontiers in Psychology*, 4(May), 234. doi: 10.3389/fpsyg.2013.00234

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1493), 1105–1122. doi: 10.1098/rstb.2007.2162

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217–262). New York: Academic Press.

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language*, 79(3), 464–484. doi: 10.1353/lan.2003.0189

Tily, H., & Kuperman, V. (2012). Rational phonological lengthening in spoken Dutch. *The Journal of the Acoustical Society of America*, 132(6), 3935–40. doi: 10.1121/1.4765071

Tily, H., & Piantadosi, S. T. (2009). Refer efficiently: Use less informative expressions for more predictable meanings. In K. V. Deemter, A. Gatt, M. V. Gompel, & E. J. Krahmer (Eds.), *Proceedings of the Workshop on the Production of Referring Expressions: Bridging the Gap Between Computational and Empirical Approaches to Reference* (pp. 1–8).

Tily, H., Gahl, S., Arnon, I., Snider, N. E., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(02), 147–165. doi: 10.1515/LANGCOG.2009.008

Torreira, F., & Ernestus, M. (2009). Probabilistic effects on French [t] duration. *10th Annual Conference of the International Speech Communication Association 2009 (INTERSPEECH 2009)*, 448–451.

Torreira, F., & Ernestus, M. (2011). Vowel elision in casual French: The case of vowel/e/in the word c'etait. *Journal of Phonetics*, 39(1), 50–58. doi: 10.1016/j.wocn.2010.11.003

Torreira, F., & Ernestus, M. (2012). Weakening of intervocalic/s/in the Nijmegen Corpus of Casual Spanish. *Phonetica*, 69(3), 124–148. doi: 10.1159/000343635

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, 39(3), 1429–43. doi: 10.1016/j.neuroimage.2007.09.054

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49(1), 2–7. doi: 10.1016/j.specom.2006.10.003

van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, 39(7), 1301–1316. doi: 10.3758/s13421-011-0103-2

van Son, R. J. J. H., & Pols, L. C. W. (2002). Evidence for efficiency in vowel production. In *7th International Conference on Spoken Language Processing*. Institute of Phonetic Sciences, University of Amsterdam.

van Son, R. J. J. H., & Pols, L. C. W. (2003). How efficient is speech? *Proceedings Institute of Phonetic Sciences, University of Amsterdam*, 25, 171–184.

van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, 47(1-2), 100–123. doi: 10.1016/j.specom.2005.06.005

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122(4), 2306–2319. doi: 10.1121/1.2773966

Walker, A., & Hay, J. B. (2011). Congruence between "word age" and "voice age" facilitates lexical access. *Laboratory Phonology*, 2(1), 219–237. doi: 10.1515/labphon.2011.007

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1466–1481. doi: 10.1037/a0013353

Wasow, T. (1997). End-weight from the speaker's perspective. *Journal of Psycholinguistic Research*, 26(3). doi: 10.1023/A:1025080709112

Wasow, T., & Arnold, J. E. (2003). Post-verbal constituent ordering in English. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of grammatical variation in english* (pp. 119–154). Berlin: Mouton de Gruyter.

Wasow, T., Jaeger, T. F., & Orr, D. M. (2011). Lexical Variation in Relativizer Frequency. In H. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 175–195). Berlin: De Gruyter.

Wasow, T., Levy, R., Melnick, R., Juzek, T., & Zhu, H. (2015). Processing, prosody and optional *to*. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing. studies in honor of Janet Dean Fodor*. (pp. 133–158). Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-12961-7_8

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2010). Corrigendum to Tic Tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition*, 114(3), 462–463. doi: 10.1016/j.cognition.2010.01.007

Watson, D. G., Buxó-Lugo, A., & Simmons, D. C. (2015). The effect of phonological encoding on word duration: Selection takes time. In E. Gibson & L. Frazier (Eds.), *Explicit and implicit prosody in sentence processing* (pp. 85–98). Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-12961-7_5

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*. Doctoral dissertation, The Ohio State University.

Weatherholtz, K., & Jaeger, T. F. (2016). Speech perception and generalization across speakers and accents. In *Oxford Research Encyclopedia of Linguistics, Dec. 2016*. doi: 10.1093/acrefore/9780199384655.013.95.

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, 23, 247–274. doi: 10.1515/TLR.2006.010

Wedel, A., & Sharp, R. (2014). Phonetic cues distinguishing lexical competitors are hyperarticulated in natural speech. In *Linguistics Society of America (LSA) Annual Meeting* (Vol. 1). Minneapolis, MN.

Wedel, A., Jackson, S., & Kaplan, A. (2013). Functional Load and the Lexicon: Evidence that Syntactic Category and Frequency Relationships in Minimal Lemma Pairs Predict the Loss of Phoneme contrasts in Language Change. *Language and Speech*, 56(3), 395–417. doi: 10.1177/0023830913489096

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: a corpus study. *Cognition*, 128(2), 179–186. doi: 10.1016/j.cognition.2013.03.002

Wheeldon, L. R. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, 18(1), 81–109. doi: 10.1080/01690960143000470

Wheeldon, L. R., & Lahiri, A. (1997). Prosodic Units in Speech Production. *Journal of Memory and Language*, 37(3), 356–381. doi: 10.1006/jmla.1997.2517

Wiechmann, D. (2015). *Understanding relative clauses: A usage-based view on the processing of complex constructions* (Vol. 268, p. 270). Berlin: Walter de Gruyter.

Zhao, Y., & Jurafsky, D. S. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231–247. doi: 10.1016/j.wocn.2009.03.002

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 40, 1–95.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort* (p. 573). Oxford, UK: Addison-Wesley Press.