Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon

Benjamin V. Tucker[1] and Mirjam Ernestus[2, 3]

[1]Department of Linguistics, University of Alberta

[2]Centre for Language Studies, Radboud University

[3]Max Plank Institute for Psycholinguistics

Address for correspondence:

Benjamin V. Tucker

Univ. of Alberta, Dept. of Linguistics, 2-40 Assiniboia Hall, Edmonton, Alberta. T6G 2E7, Canada

bvtucker@ualberta.ca

ABSTRACT

The majority of studies addressing psycholinguistic questions focus on speech produced and processed in a careful, laboratory speech style. This ‘careful’ speech is very different from the speech that listeners encounter in casual conversations. This article argues that research on casual speech is necessary to show the validity of conclusions based on careful speech. Moreover, research on casual speech produces new insights and questions on the processes underlying communication and on the mental lexicon that cannot be revealed by research using careful speech. This article first places research on casual speech in its historic perspective. It then provides many examples of how casual speech differs from careful speech and shows that these differences may have important implications for psycholinguistic theories. Subsequently, the article discusses the challenges that research on casual speech faces, which stem from the high variability of this speech style, its necessary casual context, and that casual speech is connected speech. We also present opportunities for research on casual speech, mostly in the form of new experimental methods that facilitate research on connected speech. However, real progress can only be made if these new methods are combined with advanced (still to be developed) statistical techniques.

Keywords:
Casual speech, conversational speech, experimental paradigms, statistical analyses, pronunciation variability

One of the central goals of psycholinguistic research is understanding the way in which language is recognized, produced and represented in the mental lexicon. As researchers have worked to achieve this goal, a substantial collection of hypotheses, theories, and computational models of lexical representation, comprehension and production have been generated and tested in (psycholinguistic) experiments and corpus analyses. Importantly, nearly all work focusing on speech has investigated read aloud speech or speech elicited under formal conditions (often in the laboratory, in front of a microphone, and directed towards strangers). Words are carefully articulated and occur in isolation, in simple sentences, or in sentences that reflect those sentences typically found in written language (books, newspapers, etc.). In this article, we use the umbrella term "careful speech" for all speech that is not spontaneously uttered or that has not been carefully recorded in order to reflect everyday speech. Warner (2011) surveyed the *Journal of Phonetics* (v.36, 2008) to verify the amount of research being performed on non-careful speech. In that survey of 36 articles, only 4 articles used non-careful speech; two of which investigated infant babbling and isolated word productions, and two studied North American indigenous languages, and used several speech genres. Wagner, Trouvain, & Zimmerer (2015) reported that in the 2007 and 2011 conference proceedings of the International Congress of Phonetic Sciences, only 9% and 19% of the papers respectively reported data which they classified as "unscripted" speech.

In this article, we argue that in order to truly understand the cognitive processes underlying communication, psycholinguistic research requires a concerted shift to more casual, spontaneous types of speech, which indicates, at least, whether our models, theories and hypotheses are also valid for the type of speech speakers and listeners most frequently use. Moreover, as we show in this paper, studies of spontaneous speech also raise new questions which have to be answered in

order to establish a clear understanding of speech production, speech comprehension, and the mental lexicon.

Research on the processing of conversational speech requires experimental paradigms in which target words can be embedded in sentences, in all sentence positions. The focus on careful speech, however, has resulted in experimental paradigms suited for the study of words presented without context or with very little context. For instance, many studies have used the auditory lexical decision paradigm which is not well suited for investigating the processing of words in linguistic context. Other popular experimental paradigms such as phoneme monitoring, shadowing or word spotting use sentences rather than isolated words. However, many questions have been raised about these paradigms because they can induce strong task effects and participants rely heavily on the orthography of the words they hear to perform the task (e.g., Bates & Hiu, 1996; Connine & Titone, 1996; McQueen, 1996). Today, only a few experimental paradigms allow for the investigation of production and recognition of words in sentences: the map task, cross-modal priming, eye tracking, pupilometry, and recording brain activity. However, all these methods have serious shortcomings. Below, we discuss these shortcomings as well as introducing new opportunities for the research on casual speech.

So far, we have used the terms “spontaneous”, “conversational”, “casual speech”, “non-careful speech” and “unscripted speech, as opposed to “careful speech”. We believe that there is a continuum spanning the very careful, hyper-articulated styles of speech to the fast flowing, often hypo-articulated conversations with a close friend (e.g., Ernestus, Hanique, & Verboom, 2015; Tucker, 2007; Warner 2011; Wagner et al., 2015) and that all these different speech styles should be addressed in psycholinguistic research (see also Warner, 2012). For convenience, we focus on the extremes of the continuum: careful and casual speech. That is, we compare

unscripted conversational speech uttered in informal conversations with careful speech, which we have defined as an umbrella term referring to carefully articulated speech often produced in formal situations.

Careful speech comprises read aloud speech, which has, among other things, the same semantic and syntactic properties of written text. Because our focus is not on the difference between written language versus spoken language, but on (unscripted) speech uttered in formal settings versus more spontaneously produced speech, we only consider read aloud speech if it provides information about careful speech in general (for instance with respect to pronunciation).

We first explain why casual speech has received so little attention in the past, and why the time is now ripe to start studying this form of speech. We then demonstrate the importance of investigating casual speech for psycholinguistics on the basis of different types of speech data and their implications. This section is not a thorough review of the literature, but a selection of the literature illustrating our points. We finally discuss some of the challenges facing research on casual speech, ways in which these challenges can be met, as well as some of the opportunities arising from these challenges. We conclude this paper with a short section on future directions for psycholinguistics and casual speech.

Why casual speech has attracted so little attention within psycholinguistics

Research on first language acquisition has, for many years, shown that studying spontaneously uttered speech can provide valuable insight into language representation, comprehension and production. This research field was one of the first to build a large corpus of spontaneously uttered speech (MacWhinney, 2000). Similarly, the field of sociolinguistics has

long studied spontaneously uttered dialogues (e.g., Labov, 1972) with interesting results. Nevertheless, research on adult speech processing has neglected casual speech. We identify several (related) explanations for the general research focus on careful speech.

First, research on how language users produce and process carefully articulated words in isolation and in simple sentences provides insights into the basic mechanisms underlying speech production and comprehension that are likely to play a part in everyday language communication. That is, in order to learn about speech production and processing in conversational settings, research on casual speech is not always strictly necessary. The advantage of investigating careful speech over casual speech is that with highly controlled speech stimuli, we can conduct better controlled experiments which allow for firmer conclusions (Xu, 2007). It is thus tempting to only investigate the processing of careful speech as it already provides so much information about the linguistic system.

A second possible explanation for the focus on careful speech may simply be the common belief that language outside the laboratory is very similar to careful speech: Everyday speech may have some special characteristics, but the differences between careful speech and casual speech are not substantial. More than half a century ago, studies were published showing that the opposite may be true, demonstrating differences in speech rate (Pollack & Pickett, 1963) and the extent of vowel reduction (e.g. Lindblom, 1963), and documenting extreme assimilation (Hocket, 1955; Stampe, 1973) and syllable deletion (Richter, 1930; from Warner, 2011) for casual speech that is not present in careful speech. In addition, in 1988, Mehta and Culter indicated that the processing of casual speech may partly rely on mechanisms different from those underlying the processing of careful speech. They further argued for the need to study more casual speech styles in psycholinguistic research.

A third possible explanation for why modern psycholinguistic research has focused on careful speech is that historically this field has received inspiration from formal linguistic theories, many of which have traditionally focused on linguistic competence. How speakers produce words and sentences outside the laboratory and how listeners process words and sentences that do not represent careful speech were deemed uninteresting. This is because language users' performances in these situations may be heavily influenced by factors that are not part of their linguistic competence (Chomsky, 1965), for instance their working memory or their hearing acuity. A divide and debate around competence versus performance and their role in linguistics has been ongoing since the 1960s (e.g., Hymes, 1992). It is only very recently, especially with the rise in interest in individual differences and in language user groups other than highly educated young adults, that psycholinguists have formulated research programs that aim at developing complete theories of speech production and comprehension. These combine the processes and representations representing linguistic competence with factors that play a role in all types of behavior, such as working memory (e.g. Koch & Janse, 2016).

A final explanation for the neglect of casual speech may be that this speech style is hard to investigate. Traditional experimental paradigms are designed to test the production or processing of words in isolation, while the production or processing of casual speech requires researchers to test connected speech. Moreover, the laboratory environment in which most experiments are conducted does not favor the use of casual speech.

We believe that the time is now ripe to add casual speech as a research focus in speech production, speech comprehension, and the mental lexicon. There is increasing evidence indicating that casual speech is inherently different to careful speech, as we demonstrate in the next section. Moreover, we can make more progress if we extend our research focus to other

registers than to careful speech alone. Finally, we also note that to facilitate research on casual speech, new experimental and statistical techniques are required.

Relevance of differences between casual and careful speech for psycholinguistics

As briefly noted above, several studies have shown clear differences between careful and casual speech (e.g. Hocket, 1955; Pollack & Pickett, 1963; Lindblom, 1963; Richter, 1930; Warner, 2011), and that careful and casual speech may be processed differently (Mehta & Culter, 1988). The goal of this section is to demonstrate the value of studying casual speech to improve our understanding of language representation, production, and comprehension. We provide examples showing that casual and careful speech styles differ at many linguistic levels. Speech that is uttered in everyday situations may be very different in many respects from what the dictionary or grammar dictates should be said, from that what is often taught in an introductory linguistics course (which is typically based on careful speech), and from speech produced in an experimental or careful situation.

Differences in word choice

One way in which casual speech is different from careful speech is the word types that are used. There are many content words that are generally assumed not to be sufficiently 'proper' to be used in formal situations. For example, it is unlikely that the American English *dude*, Dutch *vent* 'guy', Romanian *măi* 'dude' or French *pote* 'friend' occur in formal settings. Similarly, Torreira, Adda-Decker and Ernestus (2010) showed that swear words occur much more frequently in conversational French (e.g. *putain* 'whore' has a frequency of 0.79 occurrences per thousand

words in the Nijmegen Corpus of Casual French) than in French broadcasted on the radio (as incorporated in the ESTER corpus, Galliano et al., 2005; where *putain* has a frequency of zero). This difference between careful and casual speech does not only hold for content words, but also for words that function more as discourse markers: in one conversation of Western Canadian English, the multifunctional *like* was used 181 times in a 25-minute conversation by one of two interlocutors (Podlubny, Geeraert, & Tucker, 2015).

More generally, nouns, articles, adjectives and prepositions are more frequent in careful than in casual speech situations (e.g., Heylighen and Dewaele, 2002; Biber, 1988; Biber et al. 1998). Careful speech, in contrast, is said to be characterized by relatively higher frequencies of interjections, verbs, adverbs and pronouns. Furthermore, previous research documented that in careful speech, word type/token ratios are higher, and longer words are used more often (e.g. Biber 1988; Biber et al. 1998). It therefore comes as no surprise that Bentum, Ernestus, ten Bosch and van den Bosch (submitted) show that the different subcorpora of the Spoken Dutch Corpus (Oostdijk 2001), with each subcorpus representing a different speech register, can be distinguished based on the likelihood of the words given the immediately preceding words.

These differences at the lexical level raise several (related) important questions about the mental lexicon, speech production and speech comprehension. A first general question is: what do these differences mean for lexical representation? Is the information about the frequencies of occurrence of the words lexically stored per speech register and, if so, is this frequency information stored in the form of a continuum, in line with our idea that speech registers form a continuum? Second, does the frequency of occurrence of a word in a given speech style just result from the lexical storage of frequency information specific to this speech register, or is there (in addition) active inhibition during speech production for specific speech styles (e.g. for a word like *dude* in careful

speech)? A third general question relates to speech comprehension. We know that listeners make use of the predictabilities of words. For instance, lexical frequency is a good predictor for how easily listeners recognize words in isolation, while the conditional probability of a word given the preceding word is a better predictor if the word occurs in sentence context. Are listeners able to use the register-specific predictabilities of words, as they seem able to use modality specific probabilities of words (e.g., Gaygen & Luce, 1998)?

Differences at the syntactic level

Another way in which casual speech differs from careful speech concerns syntactic structures. Unsurprisingly, casual speech is replete with false starts, filled pauses, hesitations, repetitions, and unfinished sentences. These are absent in careful speech and are typically avoided as much as possible in casual speech produced in formal situations (e.g., Torreira et al., 2010). Further, sentence structure tends to be more complex in careful speech than in casual speech (e.g., Biber 1988; Biber et al. 1998). Sentence length, for instance, helps to identify the different components of the Spoken Dutch Corpus (Wiggers & Rothkrantz, 2007). Other types of sentence structures may be more frequent in casual speech than in careful speech. To give an example, the pervasive use of dislocations, where one phrase, usually the subject, is set apart from the main clause, and within that clause, a pronoun is used (as in *Le chocolat, c'est bon* 'The chocolate, that is good'), is a key characteristic of spoken French (e.g., De Chat, 2007).

These differences at the syntactic level raise similar questions to the differences noted above at the lexical level. For instance, what are the implications for the mental lexicon, and what exactly are the mechanisms underlying these differences? Furthermore, listeners have been shown to

expect certain word types (e.g. a past participle) given the preceding sentence structure (Viebahn, Ernestus & McQueen, 2015). Do listeners adapt these expectations as a function of speech style?

In addition, these differences have implications for research on casual speech. They show that if we would like to contrast the processing of careful and casual speech, we cannot simply present participants with exactly the same sentences pronounced either in a careful or in a casual way. Participants may show different processing when listening to syntactically careful sentences pronounced in a casual way, as opposed to listening to syntactically casual speech.

Differences in tones and intonation

Casual speech may also differ from careful speech in how lexical tones and intonation are realized. In recent work, Brenner (2013) showed that the acoustic cues for individual tones in Mandarin casual speech are greatly reduced, and many of the standard cues for tone (like pitch) are often unreliable. Brenner (2015) investigated the perceptual implications of these results using two perception studies. The first study compared careful and casual speech with the pitch replaced by a synthetic whisper or hum and found that casual speech was more difficult to recognize. In a second experiment, participants were asked to transcribe either the full resynthesized utterance or a synthesized whispered form from casual speech. Brenner found that the utterances with pitch cues slightly reduced the transcription errors (e.g., from 16% to 12.5% for character recognition). The results from Brenner's studies indicate that recognizing tones in casual speech is more difficult than in careful speech. He also shows that tone perception in casual speech is a complex interaction of acoustic cues and not just based on pitch, which is in line with previous careful speech research (e.g., Taft & Chen, 1999; Fu and Zeng, 2000; Liu and Samuel, 2004; Chen & Tucker, 2013). Brenner's results suggest that in everyday conversations,

listeners may rely on different cues for language processing than would be expected based on laboratory studies alone.

Differences between casual and careful speech have also been documented for intonation. De Ruiter (2015) showed that in German the information status of discourse referents is marked differently in unscripted narrations of picture stories than in read aloud speech. For instance, whereas in the scripted speech given referents were consistently de-accented, they were not in the unscripted speech. In addition, where scripted speech showed low boundary tones, unscripted speech showed low pitch accents in combination with high boundary tones to indicate continuity. The unscripted speech investigated by de Ruiter could also be classified as careful, because they are monologues elicited in the laboratory. This may indicate that the differences between read aloud careful speech and casual speech are even more substantial.

In three corpus studies of casual German and English, Schweitzer et al. (2015) investigated whether the exact intonation contour is affected by the identities of the words. They found that the frequency with which a given word occurs with a given accent type affects the amplitude of the accent. Moreover, this absolute frequency increases the variability in pitch accent shape, while the relative frequency with which a word occurs with a given accent has the opposite effect. These findings are unexpected in a production model which predicts that intonation contours are determined independently of the words. Schweitzer and colleagues therefore argued that intonational contours are stored as part of the lexical representations of words.

These studies have direct implications for our view of the mental lexicon. Most importantly, they raise questions about what is exactly stored in the mental lexicon. For instance, are lexical tones stored as lexical tones or are they specified in other ways that better correspond to their role in casual speech? Are words specified for their most likely intonation contour? The answers

to these questions necessarily have direct implications for our models of speech production and comprehension.

Phonological assimilation

According to the phonological descriptions of many languages, mostly based on introspection and careful speech, categorical assimilation is a frequent phenomenon. That is, many languages contain productive processes that categorically change the feature of one segment to match a neighboring segment. As a consequence, a sentence like *a quick ru*[m] *picks you up* is completely ambiguous as to whether it refers to alcohol or to sports. This observation has initiated several theories and experimental work on how listeners cope with complete assimilation (e.g., Gaskell & Marslen-Wilson, 1998; Lahiri & Reetz, 2002). Recent work, however, casts doubt on the wide-spread assumption that categorical, fully productive assimilation is prevalent in languages (for an overview, see Ernestus, 2012). Many categorical assimilation processes are less frequent than previously assumed. Moreover, many processes are not always categorical: often the segment that is assumed to completely assimilate to its neighboring segment still contains some of its original acoustic properties (Ernestus, 2012). This shows that the strategies that listeners have been assumed to apply in order to cope with categorical assimilation may not play an important role in casual speech comprehension. Experiments testing how listeners process categorical assimilation show what the perception system does if put under pressure, but they may not reflect casual speech processing.

Regressive voice assimilation in Dutch is one of the processes that is more complex than previously thought, and that consequently raises questions about speech production. This process is generally assumed to be categorical: every obstruent followed by a voiced plosive is

completely voiced. However, the study by Ernestus, Lahey, Verhees, & Baayen (2006), which is based on lively read aloud stories for the blind (and thus not even on casual speech), shows that this assumption is not correct; in no less than 25% of the productions where regressive voice assimilation should have occurred according to received wisdom (e.g. in *wetboek*, consisting of the parts *wet* /ʋɛt/ 'law' and *boek* /buk/ 'book'), the segment sequence showed progressive voice assimilation (/ʋɛtpuk/ rather than /ʋɛdbuk/). Furthermore, the study showed a correlation between the frequency of the word (often a compound) and the degree and direction of voicing. A finding like this has direct implications for models of speech production that are fully decompositional, such as Weaver++ (Levelt, Roelofs, & Meyer, 1999). In these models, effects of frequency are assigned at the morpheme level and they therefore have difficulty accounting for frequency effects across two lexeme boundaries of compounds.

Single sound substitutions and the absence of sounds

Not only assimilation, but also other processes that affect or delete single sounds may be more frequent in casual speech than in careful speech. Dilts' (2013) dissertation used the phonetic transcriptions (hand corrected force-aligned segmentation) provided in the Buckeye Corpus of Conversational Speech (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume, & Fosler-Lussier, 2007) to examine reduction and found that 29,888 of the 137,319 content word tokens (i.e., 22%) were transcribed as lacking one or more sounds from the citation form. More generally, 38% of the content word tokens were transcribed in a form that did not match the citation forms provided by the corpus.

Another example concerns the realization of voiced and voiceless alveolar stops as flaps in North American English. In contrast to what may be concluded on the basis of introspection

(which is likely to be influenced by orthography), the flap is the most common production in words like *atom* [ærəm]: Warner and Tucker (2011) show that 99% of tokens of word-medial alveolar stops are flapped when the stop is followed by an unstressed vowel, even in a carefully articulated word list. In fact, a production containing an aspirated /t/ sounds awkward and foreign in North American English and the flap represents the careful realization. In line with this result, in a series of six experiments, McLennan, Luce, & Charles-Luce (2003) observed that listeners recognize words more easily if they are pronounced with a flap than with a careful stop.

One question raised by this variability is whether the different pronunciation variants (like /t^h/ versus flap in American English) result from some "phonological" rule operating during speech production and the inverse during comprehension, or whether the different pronunciation variants may be stored in lexical memory. Ranbom & Connine, (2007) addressed this question by investigating how quickly listeners recognize the /t^h/ and the flap variants of words, contrasting words that are mostly produced with /t^h/ with those that are mostly produced with a flap. They found that listeners recognize a word's flap variant more easily if the word occurs more often in the flap variant, and they take this as evidence that the different word pronunciation variants are lexically stored, together with their lexical frequencies. Similar frequency effects were observed for the processing of the variants of French words lacking the word-medial schwa (e.g., /fnɛtr/ for /fənɛtr/ *fenêtre* 'window' and /rnar/ for /rənar/ *renard* 'fox' (Brand & Ernestus, submitted)). Bürki, Ernestus & Frauenfelder (2010) showed that speakers need less time to start producing the reduced variant of a word (i.e. the variant without word-medial schwa), the more frequent the occurrence of this variant relative to the corresponding full variant. Brand & Ernestus (submitted) showed that these same frequency effects hold for perception: listeners recognize the reduced variant of a French schwa word more quickly if that

word occurs more often without schwa. Following Ranbom & Connine (2007), both Bürki and colleagues and Brand and Ernestus interpret these frequency effects as supporting the hypothesis that a word may be lexically stored with several pronunciation variants.

This hypothesis has important implications for theories of the mental lexicon. It raises questions, for instance, about the number of pronunciation variants that are typically stored for a single word, about which types of pronunciation variants are likely to be stored (only those that differ from the full variants at the segmental level or also those that differ from the full variants in the exact realization of these segments?), and about how the pronunciation variants of a single word are lexically organized. These questions are directly relevant for theories on speech production and speech comprehension. For instance, if many pronunciation variants are stored, what role is left for computation (phonological rules)? Moreover, do the different pronunciation variants stored in the lexicon compete with each other during word recognition? This latter question is a good example of a question that would not have been raised if all research was restricted to careful speech.

Continua of pronunciation variation

In casual speech, there is often a continuum between one realization of a segment and another, which again raises questions about lexical representation, speech production and speech comprehension. For instance, Warner and Tucker (2011) showed that, in addition to /nt/ and flap realizations, there are many pronunciation variants of the North American English flap that differ in strength. Importantly, the weakened flaps spanning this continuum between a real flap and an approximant-like consonant are very common in casual speech.

Whole words can also show continua of pronunciation variants. Greenberg (1999) reported an investigation of the Switchboard corpus of casual conversations (Godfrey, Holliman, & McDaniel, 1992) and found that the word *that* is produced in 117 different ways, with the form [ðæ] being the most frequent at 11%. Greenberg also found that the word *people* was produced 45 times in the corpus with 21 different realizations. The word *and* was produced 87 different ways; the top seven realizations for this word are [æn], [ɛn], [ɪn], [ən], [n̩], [n], with [ænd] in seventh position. Similar findings have been reported for other languages. For instance, Ernestus (2000: 141) reported 14 pronunciations for Dutch *natuurlijk* 'of course'. In a laboratory setting this range of variation found in casual speech is unlikely to occur.

This variation raises the question of how listeners arrive at lexical access. There is no known phonological model available that would allow for the listener to decode all these variants into their phonological forms as a form of allophonic variation. There are many possible alternative explanations. For instance, there could be a pattern matching system that simply finds the "best fit" to the acoustic input. It is unclear, however, whether such a simple mechanism can explain all findings (e.g. why is Dutch [ɛik] so easily recognized as *eigenlijk* 'actually'/ɛixələk/ even though there is also the word *eik* /ɛik/ 'oak tree'). Moreover, if we just assume a simple pattern matching mechanism for comprehension, we have to find another mechanism to explain the (consistent) variation in production. If we assume storage of all pronunciation variants, the same questions arise as those concerning the storage of pronunciation variants of words differing in single segments, including which pronunciation variants are stored and how pronunciation variants of the same word are treated during word recognition.

Challenges and opportunities for research on the production of casual speech

In the previous section, we discussed some of the many differences between careful speech and casual speech. We also illustrated some of the questions arising from this research in relation to our understanding of psycholinguistics. Taken together, we show that it is imperative that more research be performed on casual speech, however we recognize that is not an easy process. In this section, we describe some of the challenges and opportunities for research on casual speech production.

Experimental paradigms

Speech production in psycholinguistics is commonly investigated with experimental paradigms like picture naming or sentence reading. These paradigms seldom elicit casual speech, as they are conducted in formal settings (typically a laboratory) and speakers have to perform a (formal) task. Because speakers tend to only produce casual speech if they feel free to say what they want, in the way they want, in informal settings, the elicitation of casual speech is extremely difficult in highly controlled experiments. This especially excludes all experimental paradigms that aim at measuring how much time speakers need to proceed from concept to the articulation of a word.

Many studies have attempted to draw information about the cognitive processes underlying speech production from the detailed acoustic characteristics of words (e.g. Bürki et al., 2011; Pluymaekers, Ernestus & Baayen, 2006). Variables that affect the acoustic properties of words must have a role in the speech production process. For instance, Gahl, Yao and Johnson (2012) argued on the basis of corpus research that words with many phonological neighbors tend to be phonetically reduced (shortened in duration and produced with more centralized vowels) in connected speech, in contrast to what has been found for more careful speech (Munson &

Solomon, 2004). They argue that since words with many neighbors are harder to understand, reduction does not benefit the listener and is therefore likely to be speaker-driven. Further research is necessary to confirm this hypothesis.

A very different approach to research on speech production can be found in studies using ultrasound tongue imaging and electromagnetic articulography (e.g., Schönle, Gräbe, Wenig, Höhne, Schrader, & Conrad, 1987; Stone 1990). These techniques provide direct information on the position of the articulators during speaking. Although the speakers and their interlocutors can easily see and feel the equipment involved (in the case of articulography, the sensors, being small, only interfere minimally with articulation), they tend to accommodate their productions relatively quickly (e.g., Fowler and Turvey, 1980), especially when they are free to move their heads (Gick, 2002; Wrench & Scobbie 2011). As a consequence, the speech recorded with this type of equipment can also be said to be casual.

Compiling and working with corpora of casual speech

Corpora of casual speech that best approximate the speech obtained in controlled experiments are those in which the participants have to perform a task. One task that may make participants produce requested words in a (semi-) casual speech style is the map task, in which one participant has to describe a route indicated on a map to another participant, who has a slightly mismatching map without the route (e.g., Anderson et al., 1991; Engen, Baese-Berk, Baker, Choi, Kim, Bradlow, 2010). This task may elicit casual speech, for instance if the participants know each other very well and the setting is as informal as possible. If the map contains specific landmarks (e.g. a cathedral, a tree), these landmarks are likely to be mentioned and the researcher can thus elicit and compare specific words. The words mentioned in the route

descriptions, however, are likely to differ in many respects due to their immediate contexts (e.g. in prosody, speech rate, the predictability of following word), which makes direct comparisons of the word tokens difficult.

Recently, many corpora have been compiled of truly casual conversations in many different languages (see, e.g. Ernestus & Baayen, 2011; Warner, 2012 for descriptions of some of these speech corpora). These corpora allow researchers to quickly identify bits of speech and investigate their syntactic constructions, acoustic characteristics, their words' frequencies of occurrence, and many other characteristics. Analyzing a recording of free casual conversation is even more difficult than analyzing recordings made with a map task because the speakers are free to say what they want and have no restrictions regarding their production. Nevertheless, the many studies based on these corpora show that those that are currently available are a valuable asset to psycholinguistics, but that more are necessary to truly understand the full range of variation possible across speech styles and languages (including signed languages) and the cognitive processes underlying them. For example, most of the data available are limited in terms of speech register; only a few data sets allow direct comparisons of more than two or three speech registers.

Many of these corpora have also begun to include video along with the audio and transcripts to enable the analysis and investigation of gestures during casual speech. Related to these gestural aspects of communication, it is also necessary to investigate the nature of casual signed languages, where differences between casual and careful language should also be explored, as in Tyrone and Mauk (2010). One possible advantage of investigating casual signed languages is that the articulatory aspects of signed languages are simpler to access.

Unfortunately, corpora come at a high cost in terms of their creation: it is relatively simple to record the data; the labor intensive aspect is manually tagging the speech (e.g., orthographic transcription, segmenting words and phonemes) for future analysis. Tagging speech data can, however, at least partly be done automatically. Increasingly, researchers use automatically generated transcriptions of speech corpora to speed up their research. This is a promising development, although the output of automatic transcribers has to be considered with care. Automatic speech recognizers available for research purposes still produce many errors in their orthographic transcriptions of casual conversations, among others, due to the overlap between speakers (e.g., Çetin & Shriberg, 2006) and the extensive variation between and within speakers (Benzeghiba et al., 2007). However recent advances indicate that significant improvement is on the horizon (Xiong et al., 2016).

The automatic generation of phonemic transcription consists of the machine choosing a given word in the orthographic transcription from the pronunciation variants listed in the machine's lexicon that best matches the acoustic signal (forced alignment). The resulting phonemic transcriptions are often as (un)reliable as those produced by human transcribers (for a discussion, see, e.g. Ernestus & Baayen, 2011). Note that this procedure implies that the automatically generated transcription can only contain word pronunciation variants that are listed in the lexicon and therefore cannot reveal new pronunciation variants.

The use of a corpus that is automatically phonemically transcribed makes it possible to quickly extract a large amount of data, which is extremely helpful when analyzing casual speech. Researchers can do this without having ever listened to or visually inspected the data. This can be a major disadvantage, however, because the researcher can miss important aspects of the pronunciation variation under investigation. We believe that research on casual speech requires

that the researcher listens to the actual conversations. In addition, researchers should visually investigate the signal's acoustic characteristics because often the researcher, who is also a listener, reconstructs the full form of the speech when listening to casual, connected speech (Kemps, Ernestus, Schreuder & Baayen, 2004). This can make it very difficult to identify reductions such as [dʲætə] for 'do you have to…".

Statistical modeling of casual speech data

The lack of control in production studies of casual speech determines the need for advanced statistical analysis. For instance, if we would like to know whether a given variable has an effect on the duration of vowels, the effects of all other possible variables on vowel duration (e.g. speech rate, the presence of accent, the word's frequency of occurrence) have to be partialed out. This may not be easy because the resulting statistical analyses, with many control variables, may result in data overfitting. Moreover, a variable that was not considered relevant at the time of the study may be found to be so later on, which could make the result of the study difficult to interpret. Note, however, that the latter also holds for highly controlled experiments where words are produced in isolation: if an important variable is not taken into account, the different conditions in the experiment may not be well matched and the results may consequently be difficult to interpret.

The last decade has seen a tremendous increase in the use of linear mixed effects models, also called multi-level models, in the analysis of language data (e.g., Baayen, 2008). These models make it possible to extract more patterns from noisy data than was possible in the past. The resultant models are complex, and new journal articles are regularly published indicating how these models should be applied in order to avoid overfitting and Type 1 errors (e.g. Bates, Kliegl,

Vasishth & Baayen, submitted; Wurm & Fisicaro, 2014). These articles sometimes contradict each other and it is therefore still unclear what the best procedure for model fitting is. We recommend that researchers explicitly describe their fitting procedure and report only those statistical results that also emerge if the fitting procedure is (slightly) changed (e.g. with or without random slopes). This is also true for generalized additive mixed models, which some researchers have also started to use (e.g. Kryuchkova et al., 2012; Baayen, van Rij, de Cat, & Wood, to appear; Hastie & Tibshirani, 2002; Wood, 2006). In addition to the methods exemplified here, there are several other methods that can facilitate the analysis of casual speech, including Bayesian modeling (e.g., Kruschke, 2010; 2014) and classification techniques (e.g., Dilts, 2012; Tagliamonte & Baayen, 2011). While statistical approaches offer many opportunities, we do not believe that statistical solutions will overcome all the challenges presented in the analysis of casual speech.

Challenges and opportunities for research on the comprehension of casual speech

Like research on speech production, those researching the comprehension of casual speech face many challenges, although they can benefit from many new opportunities. This section describes both the challenges and opportunities for different aspects of casual speech comprehension research.

The stimuli in comprehension experiments

As is true for any study of speech comprehension, studies of the comprehension of casual speech must be based on sound knowledge of the characteristics of the type of speech presented.

Whereas, for careful speech styles, this knowledge can be based on introspection or pronunciation dictionaries, this is not the case for casual speech, as argued above. Comprehension studies with laboratory-made stimuli must therefore follow production studies. An example is the work by Tucker (2011) on reduced word-medial stops. Tucker created stimuli in the laboratory that reflected the production findings of Warner & Tucker (2011). He then used a lexical decision task with words in isolation to compare the comprehension of reduced and unreduced stops.

We often do not know which properties of a casually produced sentence or word are important for speech comprehension. For instance, if we are interested in how listeners process left dislocation in casual speech, what do the stimulus sentences have to sound like for participants to accept them as representing real casual speech? One way to overcome this problem is to present listeners with stretches from spontaneously produced speech. Examples of these types of studies include: Brenner (2015), described above; Ernestus, Baayen & Schreuder (2002), who investigated how the amount of context influences the recognition of highly reduced word forms; Van de Ven, Ernestus & Schreuder (2012), who investigated what type of context listeners use most when processing reduced pronunciation variants of words; Podlubny, Nearey, & Tucker (2011) and Bernhard & Tucker (2015), who both investigated the contribution of specific acoustic cues such as duration, amplitude and pitch to the recognition of reduced words.

This solution, however, implies that researchers surrender some level of control when designing their experiments. It is very unlikely that they will find two stretches of speech in a corpus that only differ in the characteristic under investigation. Moreover, they may not find examples of, for instance, the words they are interested in because, while the variation in casual speech is much greater in some respects, it is smaller in others (as discussed above). As a result,

a certain amount of noise is inherently introduced into these experiments. This noise can often be dealt with by incorporating additional control variables in the statistical modeling, as described above, including properties of the word, or information about the intonation contour. This, however, leads to the same statistical challenges discussed above. Moreover, this statistical solution requires large amounts of perception data to accurately test against the research questions.

Experimental paradigms

Another challenge for research on the comprehension of casual speech concerns the experimental paradigm. An experimental paradigm has to allow the researcher to investigate the comprehension of linguistic units (e.g., syntactic structures, words) in their natural context and it has to allow participants to listen to the stimuli as in a normal conversation. Traditional paradigms, however, present words in isolation or ask participants to perform unnatural metalinguistic tasks (e.g. lexical decision or phoneme monitoring).

One possibility that fulfills one of the two requirements is to play back a casual monologue and have participants perform a type of cross-modal identity priming during parts of the monologue. This would allow the researcher to test the listener's processing at predetermined points in the monologue where a phenomenon of interest occurs. So far, we have little experience with this experimental method and, as a consequence, it is unclear what the best interval is between the word in the acoustic signal and the word on the screen and how this interval should vary as a function of speech style. It is also unclear how sensitive this method may be. Moreover, we suspect that the metalinguistic task that participants have to perform may cause them to listen to the speech in a different way than they do in everyday informal settings.

Another experimental paradigm that can be used with spontaneous speech is the visual world paradigm, where an individual's eye-gazes are recorded while they are listening to speech and looking at words or objects presented on a computer screen. Brouwer, Mitterer and Huettig (2012), for instance, presented stretches of casual speech to participants, who had to indicate which of the four words presented on the screen appeared in the sentence. The researchers showed that if words are reduced, listeners penalize acoustic mismatches less strongly when listening to casual speech than when listening to fully articulated speech. One of the advantages of this experimental paradigm is that participants do not have to perform a metalinguistic task.

There are, however, also disadvantages. The visual world paradigm primes the listeners for the words presented on the computer screen (in the form of pictures or the printed orthographic transcriptions), which affects speech processing. Furthermore, the statistical modeling of eye movements may be an even greater challenge than the modeling of acoustic data (discussed above) because eye movement data do not consist of independent single points (e.g., vowel durations, single reaction times) but of time series of dependent data points. To overcome this problem, eye movements are traditionally reconverted to simple data points by averaging the percentage of looks to a given object over a given time window. This analysis method does not allow the researcher to spot differences in the exact shapes of the eye movement curves. Fortunately, new methods are being developed that can take the exact shapes of the curves into account (e.g., Baayen, et al., to appear; Mirman, Dixon, & Magnuson, 2008; Oleson, Cavanaugh, McMurray, & Brown, 2015).

Eye tracking equipment can also be used to measure the size of the pupil. Pupilometry (measuring pupil dilation or size) may provide information about the processing of casual speech as changes in pupil size have been claimed to reflect changes in cognitive load or the amount of

effort put forth during a task. This technique has provided interesting information about many domains of cognition (e.g., Goldinger & Papesh, 2012), including language processing. A typical auditory task involves listening to a stimulus and responding to that stimulus, like repeating the word that was heard (e.g., Klingner, Tversky & Hanrahan, 2011; Zekveld, Kramer & Festen, 2010). One of the major challenges for experimental paradigms using this technique is that the pupils require time to return to a baseline or "resting" dilation diameter; this is often up to 3 seconds. Postponing presenting the next stimulus until after the reset time results in an unnatural experiment and makes it difficult to investigate the natural processing of casual speech. Several recent studies have, nevertheless, used pupilometry with connected sentences (Koch & Janse, 2016; van Rij et al., submitted). Van Rij et al. (submitted) present several statistical solutions for resolving many of the challenges of analyzing pupil dilation over the course of a sentence, which could be applied to stimuli of casual speech.

With some significant investment, it may also be possible to analyze the processing of casual speech with electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), functional magnetic resonance imaging (fMRI), or magnetoencephalography (MEG). Neurolinguistics has become a very popular area in psycholinguistics, but the number of studies using these techniques to investigate the processing of casual speech is extremely low. The probable reason is that these data are very difficult to analyze and, according to accepted wisdom, these studies should therefore contrast extremely well-controlled stimuli such they that are (nearly) identical in the different experimental conditions except for the single feature under investigation. Moreover, if target words are presented in sentences, they should occur in sentence-final position so that the processing of the target word does not coincide with the processing of the following word. These restrictions obviously exclude the study of casual

speech. We are aware of only one project that, in contrast to this tradition, has investigated natural connected speech instead of highly controlled sentences. The study presented read aloud stories in a careful style of speech to participants in an fMRI experiment and investigated the effect of a word's predictability on processing (Willems, Frank, Nijhof, Hagoort and van den Bosch, 2015). We are convinced that neuro-imaging will provide useful data once reliable statistical methods are available to analyze this extremely noisy data, and we are excited by the attempts to solve these problems (e.g., Baayen et al., to appear; Mulder, ten Bosch & Boves, submitted).

Where Do We Think Psycholinguistic Research Should Go?

This article has put forward the claim that real progress in theories of speech production, speech comprehension, and the mental lexicon can only be achieved if more studies investigate the processing of speech in its most natural context, that is, in casual conversations. These studies will provide information that cannot be revealed by studies on careful speech and will raise new and important questions.

Moreover, studies have to focus on casual speech in order to investigate whether hypotheses that are supported by studies on careful speech or on written language also hold for everyday language situations. We thus see the extension to casual speech as the last step in a research path that establishes the relevance of a given cognitive process, first in highly controlled conditions and then in conditions that are increasingly natural. This path can start, for instance, with research on the processing of printed words in isolation, subsequently broaden to printed words in sentences, printed words in stories, single word utterances, single word utterances reflecting

the reduction found in casual speech, sentence-final words, sentence medial words, and end in research on words in casual conversations. We find that in nearly all of the psycholinguistic literature the last step is left out and it is this step that we explored and discussed in the present article, focusing specifically on what can be learned and gained from the investigation of casual speech in research on language representation, comprehension and production.

We provided several examples of experimental paradigms that have already been used to investigate casual speech. We also discussed several new possible experimental methods such as cross-modal identity priming, the visual-world paradigm, pupilometry, and neurolinguistic methods that may also provide useful avenues for researching casual speech. We believe that there are many other ways that production and comprehension could be explored in casual speech and encourage researchers to explore other alternative avenues.

This article focused on casual speech and discussed experimental methods in which speech is the only means of communication. In everyday conversations, however, speakers also convey messages with their facial expressions and their gestures. Moreover, speech may be produced differently and comprehended differently when the object the speech refers to is visibly present. The next step in the ecological validity path should therefore be casual speech in its multi-modal context (e.g. the work described by Drijvers & Özyürek, in press). Moreover, we expect that this type of speech will also produce new questions and insights not revealed by studies that only focus on casual speech.

Our exploration of how we can investigate everyday language processing showed that this is only possible with advanced statistical techniques. We therefore encourage researchers to learn and develop new statistical procedures, for instance, based on techniques developed in other

disciplines, and apply them to language data. We believe that the field will profit greatly from strong collaborations between psycholinguists and mathematicians.

In conclusion, the study of speech production and comprehension and of the mental lexicon is awaiting rewarding challenges. Substantial extension to casual speech, that is, to language in an ecologically valid setting, will enrich our knowledge about daily language behavior. Doing so is likely to show that what is currently known based on careful speech is at best only half of the story.

Acknowledgements

The writing of this article was partly funded by a Canadian Social Sciences and Humanities Research Council Insight grant to the first author and a vici grant from the Netherlands Organization for Scientific Research to the second author.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., & Sotillo, C. (1991). The HCRC map task corpus. *Language and Speech*, 34(4), 351-366.

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

Baayen, R. H., van Rij, J., de Cat, C. and Wood, S. N. (to appear). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In Speelman, D., Heylen, K. and Geeraerts, D. (eds), *Mixed Effects Regression Models in Linguistics*. Berlin, Springer. Retrieved from http://arxiv.org/abs/1601.02043

Bates, D., Kliegl, R., Vasishth, S. and Baayen, R. H. (submitted). Parsimonious mixed models.

Bates, E., & Liu, H. (1996). Cued shadowing. *Language and Cognitive Processes*, 11(6), 577–582.

Bentum, M., Ernestus, M., ten Bosch, L. & van den Bosch, A. (submitted). How do speech registers differ in the predictability of words?

Benzeghiba, M., De Mori, R. Deroo, O., Dupont, S., Erbes, T., Jouvet, D. Fissore, L., Laface, P., Mertins, A., Ris, S., Rose, R., Tyagi, V., & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech Communication*, 49(10), 763-786.

Bernhard, D., & Tucker, B. (2015). The effects of duration on human processing of reduced speech. *Canadian Acoustics*, *43*(3).

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.

Brand, Sophie & Ernestus, Mirjam (submitted). How do native listeners and learners of French comprehend French word pronunciation variants?

Brenner, D. (2013). The acoustics of Mandarin tones in careful and conversational speech. *The Journal of the Acoustical Society of America*, *134*(5), 4246.

Brenner, D. S. (2015). The phonetics of Mandarin tones in conversation. Retrieved from http://arizona.openrepository.com/arizona/handle/10150/578721

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, 27(4), 539-571.

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), 3980-3991.

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

Çetin, Ö, & Shriberg, E. (2006). Speaker overlaps and ASR errors in meetings: Effects before, during, and after the overlap. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (Vol. 1).

Chen, T.-Y., & Tucker, B. V. (2013). Sonorant Onset Pitch as a Perceptual Cue of Lexical Tones in Mandarin. *Phonetica*, 70(3), 207–239.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Connine, C. M., & Titone, D. (1996). Phoneme monitoring. *Language and Cognitive Processes*, *11*(6), 635–646.

De Chat, C. (2007) *French Dislocation. Interpretation, syntax, acquisition.* Oxford Studies in Theoretical Linguistics, 17. Oxford University Press, Oxford, (288pp).

Dilts, P. C. (2013). *Modelling phonetic reduction in a corpus of spoken English using Random Forests and Mixed-Effects Regression* (Thesis). Retrieved from https://era.library.ualberta.ca/downloads/5425k999s

Drijvers, L., & Özyürek, A. (in press). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research*.

Engen, K. J. V., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of Native-and Foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology-phonetic interface*. Holland Institute of Generative Linguistics, Utrecht.

Ernestus, M. (2012). Message related variation: Segmental within speaker variation. In: A.C. Cohn, C. Fougeron, & M. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*, 92–102. OUP, Oxford.

Ernestus, M. & R.H. Baayen (2011). Corpora and exemplars in phonology. In: J. Goldsmith, J. Riggle, & A. Yu (eds.), *The Handbook of Phonological Theory* (2nd ed.), pages 374-400. Wiley-Blackwell, Chichester, West Sussex.

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

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics, 48,* 60–75.

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

Fowler, C. A., & Turvey, M. T. (1981). Immediate compensation in bite-block speech. *Phonetica, 37*(5–6), 306–326.

Fu, Q.; Zeng, F. (2000). Identification of temporal envelop cues in Chinese tone recognition. *Asia Pacific Journal of Speech Language and Hearing*, 5: 45–57.

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.

Galliano, S., Georois, E., Mostefa, D., Choukri, K., Bonastre, J.-F., and Gravier, J. (2005). ESTER phase II evaluation campaign for the rich transcription of French broadcast news. *Proc. Interspeech 2005*, 2453–2456.

Gaskell, Gareth and William Marslen-Wilson (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 24: 380–396.

Gaygen, D. E., & Luce, P. A. (1998). Effects of modality on subjective frequency estimates and processing of spoken and printed words. *Perception & psychophysics*, *60*(3), 465–483.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association,* 32(02), 113-121.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *1992 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992. ICASSP-92* (Vol. 1, pp. 517–520).

Greenberg, S. (1999). Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176.

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, *21*(2), 90–95.

Hastie, T. J. & Tibshirani R.J. (2002). *Generalized additive models.* Vol. 43. CRC Press, 1990.

Heylighen, F., & Dewaele, J.-M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, *7*(3), 293–340.

Hockett, Charles F. 1955. *A manual of phonology.* Baltimore: Waverly Press.

Hymes, D. (1992). The concept of communicative competence revisited. Thirty years of linguistic evolution. *Studies in honour of René Dirven on the occasion of his sixtieth birthday*. 31-57.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, R. H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *19*, 117–127.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, 48(3), 323-332.

Koch, X., & Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, *139*(4), 1618–1636.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in cognitive sciences*, 14(7), 293-300.

Kruschke, J. K. (2014). Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press.

Lahiri, A., & Reetz, H. (2002). 'Underspecified recognition', in Carlos Gussenhoven, Natasha Warner, and Toni Rietveld (eds.), *Phonology & Phonetics: Laboratory Phonology VII*. Berlin, Mouton, pp. 637-676.

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

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, *35*(11), 1773–1781.

Liu, S.; Samuel, A.G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language & Speech*, 47: 109–138.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539–553.

McQueen, J. (1996). Word spotting. *Language and Cognitive Processes*, *11*(6), 695–699.

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. Language and Speech, 31 (Pt 2), 135–156.

Mirman, D. Dixon, J.A., & Magnuson, J.S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475-494.

Mulder, K., ten Bosch, L., & Boves, L. (submitted). Comparing different methods for analyzing ERP signals.

Munson, B., & Solomon, N. P. (2004). The Effect of Phonological Neighborhood Density on Vowel Articulation. *J Speech Lang Hear Res*, *47*(5), 1048–1058.

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2015). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical Methods in Medical Research*, 0962280215607411.

Oostdijk, N. (2000). The Spoken Dutch Corpus Project. *The ELRA Newsletter*, *5*, 4–8.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech (2nd release)[www. buckeyecorpus. osu. edu] Columbus, OH: Department of Psychology. *Ohio State University (Distributor)*.

Pluymaekers, M., Ernestus, M., & Baayen, R. (2006). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, 62(2-4), 146-159.

Podlubny, R., Geeraert, K., Tucker, B.V. (2015). It's all about, *like*, acoustics. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. Paper number 0477.

Podlubny, R., Tucker, B.V., & Nearey, T. (2011). 'Sorry, what was that?': The roles of pitch, duration, and amplitude in the perception of reduced speech. Poster presented at the Nijmegen Spontaneous Speech Workshop (Nijmegen, NL).

Pollack, I., & Pickett, J. M. (1963). Intelligibility of excerpts from conversational speech. *Language and Speech*, *6*, 165–171.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Richter, E. 1930. Beobachtungen über Anglitt und Abglitt an Sprachkurven und umgekehrt laufenden Phonogrammplatten. In Paul Menzerath (ed.) *Berichte über die I. Tagung der Internationalen Gesellschaft für experimentelle Phonetik*, 87–90. Bonn: Scheur.

Ruiter, de, L.E. (2015). Information status marking in spontaneous vs. read speech in story-telling tasks – Evidence from intonation analysis using GToBI. *Journal of Phonetics 48*, 29-44.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., & Conrad, B. (1987). Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract. *Brain and Language*, 31(1), 26-35.

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

Stampe, D. (1973) *A Dissertation on Natural Phonology*. PhD Diss. University of Chicago.

Stone, M. (1990). A three‐dimensional model of tongue movement based on ultrasound and X‐ray microbeam data. *The Journal of the Acoustical Society of America*, 87(5), 2207-2217.

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. *Advances in Psychology*, 90, 151-172.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178.

Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The Nijmegen Corpus of Casual French. *Speech Communication*, 52:201-221.

Tucker, B. V. (2007). *Spoken word recognition of the reduced American English Flap*. The University of Arizona. Retrieved from http://hdl.handle.net/10150/194987

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics, 39*(3), 312–318.

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

van Rij, J., Natalya, P., van Rijn, H., Wood, S. N., & Baayen, R. H. (submitted). Pupil dilation to study cognitive processing: challenges and solutions for time course analyses.

Ven, M. van de, Ernestus, M. & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology* 3, 455-481.

Viebahn, M., M. Ernestus, & J. McQueen (2015). Syntactic predictability in the recognition of carefully and casually produced speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41 (6), 1684-1702.

Wagner, P., Trouvain, J., & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics, 48*, 1–12.

Warner, N. (2011). Reduction. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology: General issues and segmental phonology* (Vol. 1, 1866–1891). John Wiley & Sons.

Warner, N. (2012). Methods for studying spontaneous speech. In A. Cohn, C. Fougeron, & M. Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press. 621-633.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America, 130*(3), 1606–1617.

Wiggers, P., & Rothkrantz, L. J. M. (2007). Exploratory analysis of word use and sentence length in the Spoken Dutch Corpus. In V. Matoušek & P. Mautner (Eds.), *Text, Speech and Dialogue* (366–373). Springer Berlin Heidelberg.

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Bosch, A. van den. (2016). Prediction during natural language comprehension. *Cerebral Cortex*, *26*(6), 2506–2516.

Wood, S. N. (2006). *Generalized Additive Models*. New York: Chapman & Hall/CRC.

Wrench, A. A., & Scobbie, J. M. (2011). Very high frame rate ultrasound tongue imaging. In Proceedings of the 9th international seminar on speech production (ISSP) (pp. 155-162).

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). Journal of Memory and Language, 72, 37-48.

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., … Zweig, G. (2016). The Microsoft 2016 Conversational Speech Recognition System. *arXiv:1609.03528 [Cs]*. Retrieved from http://arxiv.org/abs/1609.03528

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336.

Zekveld, A. A., Kramer, S. E., & Festen, J.M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, 31, 480-490.