Authors: Holger Mitterer, Susanne Brouwer, & Falk Huettig

Title: How important is prediction for understanding spontaneous speech?

Abstract: The idea that prediction plays a central role in language processing is currently very popular. Moreover, it has been argued that prediction is especially important for word recognition when the input is suboptimal. Conversational speech contains many phonological reductions and is hence a suboptimal signal. We evaluate this claim both on a conceptual level and on the basis of the available empirical data. We conclude that, given the current data, prediction does not seem to play an important role for word recognition in natural interactions. Prediction may nevertheless be important in streamlining natural interactions such as turn-taking. We close by discussing what kind of empirical data would be necessary to illuminate the role of prediction in word recognition.

Author Affiliation, Address and Email Address:
Holger Mitterer; University of Malta, Faculty of Media and Knowledge Sciences, Department of Cognitive Science; Msida, Malta, MSD 2080; holger.mitterer@um.edu.mt
Susanne Brouwer, Radboud University, Nijmegen, The Netherlands; Postbus 9103, 6500 HD Nijmegen, The Netherlands; S.Brouwer@let.ru.nl
Falk Huettig, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands; PO Box 310, 6500 AH Nijmegen, The Netherlands; Falk.Huettig@mpi.nl

1. Introduction

The idea that *prediction* is an important component of language comprehension has gained considerable ground over recent years (Altmann and Mirkovic, 1999; Christiansen and Chater, 2016; Dell and Chang, 2014; Huettig, 2015; Pickering and Garrod 2013). It is interesting therefore that prediction in language processing is considered a non-starter in the generative-linguistics view (Jackendoff 2002, p. 59). In this theoretical framework, creating a sentence is a highly creative process that selects words in a sentence from an infinite number of possibilities. According to generative linguistics, the creative speaker cannot be predicted. However, this noble view of the

creative speaker is somewhat weakened by findings that spoken language is not quite as complex. For instance, much psycholinguistic work has focussed on grammars that allow multiple center-embedded structures , such as relative clauses, with typical examples like “The dog that the cat that the man bought scratched ran away” (Crystal 2011). The reality of language use turns out to be more prosaic, and such multiple centre-embedding appears to be almost non-existent in normal, spoken dialogue (Karlsson 2007). Given that language use is typically simpler and hence more predictable, it is thus conceivable that prediction may be a key process in comprehension. Indeed, there is now a large body of experimental evidence which suggests that one reason why language processing tends to be so effortless, accurate, and efficient is that mature (e.g., Altmann and Kamide, 1999; Kamide et al., 2003; DeLong, Urbach, and Kutas, 2005; Federmeier and Kutas, 1999; Rommers et al., 2013; Van Berkum, Brown, Kooijman, Zwitserlood, and Hagoort, 2005; Wicha, Moreno, and Kutas, 2004) and developing (e.g., Borovsky, Elman, and Fernald, 2012; Nation, Marshall, and Altmann, 2003; Mani and Huettig, 2012, 2014) language users anticipate upcoming language input.

At this stage, it is important to clarify what we mean by prediction. We define prediction as any pre-activation of linguistic representations before any bottom-up input is processed. Note that this definition does not exclude the use of statistical properties of language (such as bigram frequencies) for language processing (though there is little evidence currently available, which unequivocally links statistical learning to prediction, see Huettig and Mani, 2016, for further discussion).

An important short-coming of most of the work on the use of prediction in language comprehension is that it has used almost exclusively “lab speech”, which is considerably different from the casual speech we typically encounter during our daily interactions. In this chapter, we evaluate the possible role of prediction for language comprehension during a normal, spontaneous dialogue.

2. Prediction during natural interactions: Challenges

Over the last decade, corpus studies have consistently shown that a large amount of spontaneous dialogue contains “phonological reductions”. With phonological reductions, we mean realizations which differ in their phonological make-up from the canonical form. Note that this differs from some usages of the term *reduction*, which may include somewhat shorter, unaccented productions of a word, in which all segments are present. Phonological reductions are ubiquitous in spontaneous speech. About every other word differs in one phoneme or more from its canonical form (Johnson 2004) and most words have 5 or more phonemically different forms (Keating 1997). The word *handbag,* for instance, may be pronounced alternatively as *hanbag* or *hambag* or even *ambag*. While this indicates that comprehension faces quite different challenges in the auditory and

visual domain, much of the research on processes in language comprehension, especially on the sentence level, treats this difference as negligible. It is, for instance not difficult to find journal articles on sentence processing in psycholinguistics for which it is impossible to tell from the abstract whether stimuli were presented in spoken or written form. Moreover, models of written- and spoken-word recognition are often very similar with regard to the proposed architecture, be it based on connectionist (McClelland and Elman 1986; McClelland and Rumelhart 1981) or Bayesian principles (Norris 2006; Norris and McQueen 2008). One reason for this treatment of modality differences may be the prevalence of what has been called "lab speech" in psychology, which, if not otherwise controlled, may be overly clear (for discussion, see Xu 2010). As a consequence, spoken words in typical psycholinguistic experiments are pronounced in line with their canonical form so that the difference between spoken and written language diminishes and does indeed become a minor methodological detail.

Interestingly, prediction has been considered to be especially useful if the signal is degraded. Spontaneous speech is, compared to lab speech, a degraded signal. However, while spontaneous speech is definitely more of a challenge for perception than careful lab speech, it is somewhat odd to include conversational speech as an example of a "degraded signal", when, in fact, it is the normal usage for which language actually has evolved (Dunbar 1998; Enfield 2003). Note that in the field of psycholinguistics most studies start with recording a speaker who reads from a script. Even Xu (2010), who defended the use of lab speech, noted that especially reading from an alphabetic script may give rise to an overall hyper-articulated speech pattern. In a way, it may hence be more appropriate to view read speech as hyper-articulated rather than to view the perception of spontaneous speech as an example of "Speech perception under adverse conditions" (Mattys et al. 2012). However, independent of what one views as normal and what as aberrant (then either "hyperarticulated" or "reduced"), it is clear that evidence for the use of prediction comes from studies using clear lab speech (or even written text).

Gagnepain, Henson, and Davis (2012) for instance argued that listeners predict phonemes based on the input and their lexical knowledge and that this forms the basis of word recognition rather than the more conventional assumption that word candidates compete for recognition. For instance, on hearing [fo:mj …], listeners predict that the next segment is [u] since only the word *formula* fits the input [fo:mj]. They tested this account by training participants on new words such as *formubo*. Critically, these words increase the lexical competition, but do not increase the uncertainty in predicting the [u] after hearing g [fo:mj …]. This gives rise to differential predictions for a lexical-competition account and a segment-prediction account. Adding the new word should increase lexical competition before the "deviation point" (i.e., the point at which the new newly learned word

differs from the existing word), which predicts “more processing” due to the presence of the new competitor. However, the prediction of an /u/ after [fo:mj …] is still possible, even with the new competitor *formubo* added, so that “more processing” is only necessary at the new deviation point. As an operationalization of “more processing”, the authors used the global field power of the MEG signal, using the assumption that larger conflict (either in terms of lexical competition or prediction error) leads to a larger MEG response. Under this assumption, the results supported a segment-prediction account. After learning the new words, there was no evidence of increased competition before the deviation point—despite the added lexical competition from the new words. However, there was evidence that after the deviation point there was more processing going on, possibly due to the increased prediction error after hearing [fo:mju], since now the listener cannot rely anymore on the prediction that the next segment must be a [l].

Regarding the prediction of phonemes, one problem should be immediately obvious, even by looking at the example chosen by Gangepain et al. (2012). They argue that listeners can be sure that after [fo:mj …], the next segment has to be [u]. This led them to define the concept of prediction error, that is, the amount to which a listener can predict the next segment. The prediction error is supposed to be high when the lexicon still allows many possible words with different following segments that are still in line with a given input. They calculate the prediction error using a standard electronic dictionary. Defining this prediction error is, however, a bit more complex in normal, spontaneous interaction. Given the example [fo:mjulɑ], in spontaneous speech, the vowel [u] would be likely to be a schwa-like vowel rather than a full vowel, given its post-stress position. Importantly, there is evidence that listeners sometimes store alternative representations of alternative phonological forms of a given word (Bürki and Gaskell 2012; Connine, Ranbom and Patterson 2008; Pitt 2009). The computation of the prediction error is hence far from straightforward. Do alternative pronunciations count as well? What happens if a British speakers hears the word *formula* from a North-American speaker who produces the post-vocalic /r/? All these questions show that it is difficult to define a prediction error once we go beyond dictionary citation forms. Simply using an electronic dictionary with canonical forms does not live up to the challenge of recognizing normal, conversational speech. Although this is a problem for all models of word recognition, it is especially relevant for a model that makes assumptions about possible continuations given the segments already heard. It is again important to remember that Keating (1997) found that most words have about 5 different phonemic transcriptions. Although some of these may be easy to recover from (for example, the prediction process may not care about the difference between the “two schwas”, see Keating, 1997) it seems evident that once we take reductions into account it becomes less straightforward to define what the prediction error should

be. Note that similar problems exists for model of predictive language processing that focus on other levels of representation. Levy (2008) proposed a model of syntactic parsing that makes use of predictability to allocate processing resources. According to this model, resources are mostly allocated when the next word class is unpredictable. Just as the presence of segments is more difficult to predict in spontaneous speech, this type of model has problems with false starts, hesitations, and repetitions, that often occur in spontaneous speech (e.g., "it is on on the left side").

An additional problem arises if we consider that the model proposed by Gangepain et al. (2012) seems to necessitate a strict order of phoneme slots on which to base a prediction for what the next slot is. However, deciding whether a phoneme is there or not is not straightforward in spontaneous speech. In the Kiel Corpus of Spontaneous Speech (IPDS 1994), the manual gives an additional transcription symbol for phonemes which are deleted, but for which some residual may be heard in the signal. Manuel (1992) provides the example of the minimal pair *sport* and *support*, which may sound close-to-identical in spontaneous speech, but with some residual evidence that the intended word was *support*, even if there is no vowel-like segment between [s] and [p]. While such "lost phonemes" may sometimes be recoverable (Spinelli and Gros-Balthazard 2007), this is unlikely to be the case for all examples. Moreover, signal-based cues such as those discussed by Manuel (1992) are likely to be probabilistic, that is, there remains some uncertainty whether there is an additional phoneme or not. This gives rise to the question how a listener would be able to predict the next phoneme if s/he is not even certain how many phonemes have been heard already.

This indicates that it may be a significant challenge to scale up a phoneme-prediction model to the challenge of normal, spontaneous speech. To reiterate, while other models of word recognition also have difficulties with accounting for word recognition given reductions, this problem is particularly problematic for a prediction model because of the following reasons. First, the variation in normal spontaneous speech makes it difficult to even define a prediction error, a very basic concept in such a prediction model. Second, the model requires certainty about the number of phonemes, which again is unlikely to be the case in spontaneous speech.

This is not just a problem that arises at phonetic/phonological levels of representation. A similar problem also arises for prediction on a word-by-word basis. Just as there sometimes is uncertainty about the number of phonemes that have been in the input, sometimes there is uncertainty about the number of words in the input (Dilley and Pitt 2010). Only in exceptional circumstance do we become aware of such processes. A well-known case are the words uttered by Neil Armstrong while leaving the Apollo spacecraft to set foot on the moon: "That's one small step for [a] man, one giant leap for mankind". Note that the presence of the indefinite article [a] makes quite a difference. Without the [a], the quote seems internally contradictory, because "for man" can

also be interpreted as “for mankind”, so that his step would be, for mankind, at the same time a small and giant one. Importantly, “for a” and “for” are difficult to distinguish in spontaneous speech, as there is no clear “two-syllable” structure with a dip in intensity between *for* and *a*, especially for speakers from Central Ohio (such as Neil Armstrong). These speakers tend to produce both *for* and *for a* as something like "fer", the distinction being based on the duration of one vowel-like segment. Dilley et al. (2013) measured the actual duration of the vowel-like segment in *for (a)* in the Armstrong quote and found that it is, after taking into account overall speech rate, in the ambiguous range between typical “for” and typical “for a”, but leaning towards a “for a” interpretation. This exercise indicates that predicting words may, just as predicting phonemes, be more difficult in normal, spontaneous speech. If one is not sure what has just been said (as often will be the case), how can one predict the next word?

The problem is greatly aggravated if one looks at the role of prediction in the perception of continuous spontaneous speech. Note that the crucial question here is not really whether context helps the recognition of reduced forms. It has been demonstrated multiple times that severely reduced words are in fact only recognizable with sufficient context (Ernestus, Baayen and Schreuder 2002; Janse and Ernestus 2011); the question is whether reduced word forms in fact benefit *more* from sentence context than canonically produced words, and especially whether such effects are attributable to an active prediction process rather than post-lexical integration. Given the prominence that has been given to the idea that prediction may be especially important if the bottom-up signal is sub-optimal, it is surprising that only very few studies actually allow to evaluate such a claim. In the next section, we focus on five studies that allow to directly investigate what the role of prediction in spontaneous speech may be;[1] two of these used “lab speech” (Mitterer and Russell 2012; Viebahn, Ernestus and McQueen 2015), that is sentences were constructed especially for the given study while three other studies used samples from a corpus of conversational speech (Brouwer, Mitterer and Huettig 2013; Magyari and de Ruiter 2012; van de Ven, Ernestus and Schreuder 2012).

Two questions are important for the evaluation of the claim that active prediction is especially important for word recognition in natural interaction when dealing with reduced forms. First, is the benefit for reduced forms really larger than for canonical forms? Second, are reduced words particularly predictable from their context?

[1] There is a considerable literature that information structure influences syllable duration (e.g., Aylett & Turk 2004). However, such findings do not necessarily show that there is segmental reduction, as deaccented versions of a word may still contain all segments. In fact, this literature may be better viewed as investigating strengthening of new information rather than reduction of given information.

3. Prediction during natural interactions: Data

The two studies using "lab speech" evaluated how Dutch listeners recognize past participles in which the prefix has undergone schwa reduction. That is, the participle *gelacht* (Engl., *laughed*) is produced as [xlɑxt] rather than [xəlɑxt]. As this concerns lab speech, these data are relevant to answer the question whether prediction has a particularly strong effect on reduced as compared to canonical forms. The main question of Mitterer and Russell (2013) was whether such words are recognized more easily if the past participle is a frequent word form (note that lemma and word-form frequency were highly correlated for this item set, so that these two factors cannot be separated). To that end, they created sentences in which the last word had to be a past participle that was either low- or high-frequent (e.g., *Afgelopen nacht heeft hij veel gedroomd/gedronken*, Engl. *Last night he has dreamt/drunk a lot*). These past participles were presented in either full or reduced form, and the efficiency of word recognition was measured with the printed word version of the visual-world eye-tracking paradigm (Huettig and McQueen 2007; McQueen and Viebahn 2007). That is, while listening to these sentences, participants saw four printed words on the screen, containing printed versions of both participles. Note that the pairs of participles were purposefully chosen to also overlap in the stem part (gedr*oomd* – gedr*onken*) in order to maximize the uncertainty and the lexical competition between the items. The main finding of this study was that the reduction costs (i.e., the difference in fixation proportions on the correct word given a full or reduced pronunciation) were stronger for the low frequency participles than the high frequency participle. This mirrors effects in production; schwa reduction is more likely in high-frequency than low-frequency Dutch past participles (Pluymaekers, Ernestus and Baayen 2005). More important for the current topic, Mitterer and Russell also performed a cloze test with their items to test how predictable each target word was. The results revealed only a small collinearity between lexical frequency and predictability, so that their respective effects could be distinguished. The results showed that predictable items were recognized better, that is, the fixations converged on the target quicker if the target was predictable. However, there was no indication of an interaction of predictability with reduction: Full forms benefitted as much from predictability as reduced forms. This finding contradicts the assumption that prediction is especially important if the bottom-up signal is sub-optimal. While the study of Mitterer and Russell (2013) was not explicitly designed to test the role of predictability in sub-optimal conditions, it provides a first clue that such effects may not be strong and hence not easy to find.

This conclusion is supported by the findings of Viebahn, Ernestus, and McQueen 2015), who designed their study to test the role of syntactic predictability. They also used the visual-world paradigm with printed words. In their case, predictability was manipulated by syntactic means as

follows. In Dutch subordinate clauses with an auxiliary verb and a past participle, there are two possible word orders. The auxiliary verb may precede or follow the main verb (*Ik weet zeker dat hij {heeft geleund|geleund heeft} op de houten tafel*, Engl., *I know for sure that he has leaned against the wooden table*). Crucially, if the auxiliary verb comes first it *must* be followed by the past participle, so that the past participle is predictable in that situation. In fact, this is a rather strong implementation of predictability, especially given how the visual-world experiment was set-up. For the sentence with the target *geleund*, a phonological competitor without schwa *gleuven* /xløvə/ (Engl., *grooves*) was used, which perfectly matches the onset of a reduced form of *geleund*, which is [xlønt]. Note that the use of *gleuven* is not syntactically licensed after an auxiliary, so that in fact it can be excluded as a potential competitor on syntactic grounds. Despite this strong manipulation of predictability, Viebahn et al. did not find consistent interactions of predictability and phonological reductions over a series of three experiments. Nevertheless, there were robust effects of both predictability and phonological reduction. If the words were reduced, participants took longer to look towards and click on the past-participle targets than if they were presented in full form. If the past participles were predictable, they were clicked on and looked at faster than if they were not predictable. However, no consistent interaction between the predictability and the reduction costs emerged in the three experiments by themselves. Only when the data from the three experiments were merged, did Viebahn et al. find an interaction of these two robust main effects, with predictability moderating the costs of phonological reduction. It is particularly interesting in which measures this interaction manifested itself. Viebahn et al. defined different time windows[2] in their eye-tracking analysis: one time window before the onset of the word to test whether there is active prediction, one while listening to the word, and two additional time windows for the interval after word offset to the click response. The interaction of predictability and phonological reduction was only significant in the reaction times and the latest of the four time windows in the analysis from the eye-tracking data. That is, the early processing of the reduced word form was not influenced by predictability. Note that this pattern is opposite to what one would expect if listeners were indeed actively predicting an upcoming target in situations in which bottom-up input is sub-optimal. Such an account would predict that effects arise early in the recognition of reduced forms. Instead the results seem to show that sentence context effects may come in only late. These data thus suggest

[2] These time windows took into account a 200ms time lag for initiating a saccade based on the acoustic input. This is a relatively large estimate for this lag, so that the failure to find early effects cannot be attributed to the way the windows were defined.

that it only plays a role at a post-lexical integration stage, clearly not in line with the assumption that active prediction plays an elevated role in the perception of phonologically reduced words.

Moreover, one potential problem with these studies is their use of "lab speech", that is, the sentences were created to be predictable—mostly in written form—and then read out loud by a speaker in front of a microphone. Even if, under these circumstances, prediction is possible, this might not extend to natural dialogues. Take, for instance, the example of stimuli used by Viebahn et al. (2015), in which the crucial part is the auxiliary verb *heeft* (Engl., *has*). In spontaneous speech, it may be produced as [ef] rather than the canonical [heft], since both /h/ and word-final /t/ are prone to deletion (Mitterer and Ernestus 2006; Pierrehumbert and Talkin 1992). Would prediction then still be possible, when the listener has a hard time to recognize the auxiliary to begin with? To be fair, Viebahn et al. were aware of the problem and noted in their method section that the speaker was explicitly asked to produce speech in a casual way. This instruction was effective so that the speaker produced examples of prefix reduction in *geleund* (/xəlʚnt/ → [xlʚnt]) spontaneously. However, when stimuli are generated by writing them down first, this may give rise to word choices which are unusual in spoken dialogue (cf. Hayes 1988). Thus it remains an open question whether prediction is important in the context of natural interaction.

On a more general level, studies with custom-made stimuli fail to address the second of the two questions we posed as critical for this section: How predictable are words, and especially reduced words, really, in a natural interaction? While there is evidence that statistical properties, such as mutual information and lexical frequency, influence the amount of reduction (Ernestus 2014), it is far from clear that reduced words are actively predictable from their context.

A study by Van der Ven, Schreuder, and Ernestus (2012) addressed this issue. They investigated whether severely reduced words from a corpus of spontaneous speech were predictable from the context. The target words were modifying expressions that tend to be ubiquitous in natural interactions (such as the English *like*, which can often be added to a sentence). The context—as found in the corpus—was presented either in auditory or written form and either the complete sentence or just the preceding context was provided. They found that participants were generally better in guessing the correct word (from four options) in the auditory condition: With just the preceding context, predictability rose from 33% for the written condition to 40% in the auditory condition (chance performance is 25%), and with both preceding and following context, performance rose from 45 to 51%. What complicates the comparison among these measures is the fact that the options given to participants were different between the context conditions. With just the preceding context, participants were given four different words, so that all choices were qualitatively similar. However, when both preceding and following context were presented, one

option was “no word missing” (remember that the target words were optional modifying expressions), which is qualitatively different from the three other choices. This option (which was never correct) was also dispreferred by the participants so chance-performance level was above 25% (one out of four) but below 33% (one out of three).

Just as the comparison of the performance between the context conditions (i.e., preceding context only or both preceding and following context) is difficult, the comparison of presentation mode confounds two factors. The type of modifying expressions used by Van der Ven et al. tend to be more likely in spoken than in the written domain (Hayes 1988). The higher success rate in the auditory modality may hence simply reflect that listeners took into account that these words were more likely to occur in the spoken than in the written modality. The authors, however, focus on the fact that presenting the sentence auditorily means to also present the listeners with some coarticulatory cues about the identity of the omitted word (see, e.g., Salverda, Kleinschmidt and Tanenhaus 2014). While this is certainly the case, it remains difficult to attribute the better predictability of the reduced words fully to the presence of auditory cues rather than the difference in spoken versus written usage frequencies.

Nevertheless, the results of Van der Ven et al. (2012) show that strongly reduced words are difficult to predict from the preceding context. Even in a multiple-choice cloze task, predicting from the preceding context alone is only slightly above chance (33% correct with 25% chance performance). This would probably mean that in a free cloze test, the reduced target words would have hardly ever been mentioned. This brings us back to a point mentioned in the introduction: Many generative linguists considered prediction not to be useful because language is often difficult to predict. The study by Van der Ven et al. reinforces this point based on samples of natural interactions. It is noteworthy that the best evidence for prediction in language processing comes from studies with carefully controlled materials (Altmann and Kamide 1999; Borovsky, Elman and Fernald 2012; DeLong, Urbach and Kutas 2005; Kamide, Altmann and Haywood 2003; Mani and Huettig 2012; Mani and Huettig 2014; Nation, Marshall and Altmann 2003; Van Berkum et al. 2005). The study of Van der Ven et al. (2012) indicates that, in natural interaction, language may often be quite unpredictable. Nevertheless, reductions still occur and appear to be tolerated by listeners, even if the reduced word is relatively unpredictable from the context.

A similarly mixed picture for the role of prediction for the recognition of reduced forms comes from a study by Brouwer, Mitterer, and Huettig (2013). They also used the visual-world paradigm with printed words (just as Viebahn et al., 2015, and Mitterer and Russel 2013), but as auditory stimuli, they used excerpts from a corpus of spontaneous speech (just as Van der Ven et al., 2012). As such, this study addressed both issues, namely, whether reduced words are really

predictable and whether prediction is especially important for reduced words. The excerpts were chosen because they contained strongly reduced words (e.g., [pjutər] for *computer* or [mənejə] for *beneden* [bənedə(n)]). As control stimuli, they also found utterances in the corpus containing the same words in unreduced form. In these experiments, participants were instructed to click on a printed word if it occurred in the spoken sentence, and the main variable was how much and how quickly participants looked at the respective targets. Predictability of the target word varied both with experimental manipulation and by exploiting differences between the corpus utterances. As an experimental manipulation, Brouwer et al. (2013) presented the words just with the sentence they were uttered in or with an additional context of a duration of at least 5s. These contexts were either the actual contexts in which the target utterances occurred or a random sample of 5s of speech from the same speaker. The latter experimental condition may seem odd, since presenting a random sample of speech may be a superfluous condition, which should pattern with the control condition of not providing any context. This, however, fails to take into account the well-known effects of speaker adaptation. There are numerous papers that show that listeners tune their speech perception to the particular acoustic patterns of a given speaker. Some of these effects seem to occur on a very early, auditory level (Sjerps, McQueen and Mitterer 2013; Sjerps, Mitterer and McQueen 2011), while others may be modulated by experience with certain speech styles (Bradlow and Bent 2008; Sumner and Samuel 2009). The condition that provides speaker but no relevant context information is hence necessary, because the comparison of no context and actual context would overestimate the role of the content of the context, as it would include any effect of speaker adaptation.

An additional covariate used by Brouwer et al. was how well the word fitted the wider context. This was operationalized with the help of a pre-test: Another group of participants was asked how well the target utterance fitted in the wider discourse context. This addresses the issue whether reduced words are in fact occurring in especially predictable contexts, the second of our two questions. The results of the pre-tests ran counter to this expectation. The reduced forms did not occur in especially coherent dialogues, as measured in the pre-test. The contextual fit was overall comparable for reduced and canonical forms. The data hence fail to support the assumption that speakers reduce especially in those circumstances in which a word may be predictable, as assumed by various theoretical accounts of variation in speech (Aylett and Turk 2004; Lindblom 1990). Nevertheless, there was quite some variation in these judgements for both reduced and full forms, making it useable as a covariate. The randomly picked contexts were overall judged to be neither fitting nor unfitting, just as one would expect after random sampling.

To summarize, the highly reduced words were presented with no discourse context, an unfitting discourse context that supplied information about the speaker, or the actual context which showed some natural variation in how well they predicted the target words. It is worthwhile to consider what the predictions of a prediction account of words recognition are in this case. First of all, since a focus on prediction as a means of comprehension highlights the importance of top-down information, we should expect relatively little benefit from the randomly sampled contexts in comparison to the benefits provided by the real contexts. This difference between actual and random contexts should be especially large for reduced forms. Additionally, we should see that the difference caused by the differences in coherence of the natural dialogue (as established in the pre-test) should affect reduced forms more than canonical forms.

Using the eye-tracking method allowed Brouwer et al. (2013) to test whether listeners could predict the target word, as one can evaluate looks to the target words before they have been uttered. Figure 1 shows the amount to which participants were able to predict the target, depending on the predictability and the phonological form of the upcoming target. The results are somewhat surprising as they show an interaction of predictability and phonological form. Prediction only was possible when the context was fitting well *and* the target word was produced in its canonical form. It is somewhat surprising that the form of the *upcoming*—hence not yet presented—target word influences whether a word can be predicted or not. Therefore, the authors evaluated how clearly the other words in the same sentence were produced and found that the sentences containing these reduced target words also carried more reduced words overall elsewhere in the sentence. This appears not to be restricted to the materials used by Brouwer et al. (2013). Viebahn, Ernestus, and McQueen (2012) report in a corpus study that reduced forms tend to co-occur in natural conversations, even if the influence of speech rate is partialled out. This suggest the following: Listeners' ability to predict may depend strongly on how sure they are about what they just heard. However, when speech carries a large number of phonological reductions, listeners are often not able to predict what comes next, possibly because they are unsure what they just heard. This hence may indicate a catch-22 situation for the use of prediction in normal, spontaneous interaction. While it makes sense a-priori to assume that prediction may help especially in these sub-optimal circumstances, these data seem to indicate that exactly in these circumstances prediction may not be possible because the data on which to base a prediction are not good.

- -

Insert Figure 1 about here

- -

Predictability hence mattered only for canonical forms when we consider active prediction before the actual target word is heard. Interestingly, the role of a predictive context changed substantially after the target word has been heard. The data of this time window is shown in Figure 2. As already described above, Brouwer et al. (2013) also presented listener with a context from the same speaker, but from a different part of the same conversation. The results show, unsurprisingly, that reduced forms are more difficult to recognize than unreduced forms. Surprisingly, however, the randomly selected contexts led to the same overall benefits as the actual contexts in which the target words occurred (left panel of Figure 2). Even more surprisingly, this pattern is quite similar (and statistically indistinguishable) for both full and reduced forms. Exposure to a given speaker turns out to be extremely beneficial for listeners in order to decode which words are uttered. Even if the context was not relevant at all, content-wise, it was still a great help for listeners. In fact, for full forms, it did not matter at all how well the context fitted the conversation, the benefit did not differ between the three conditions. These data do not fit well with the focus on top-down information that follows from the prediction-account of word recognition, because speaker adaptation is typically considered to be a bottom-up process.

- -

Insert Figure 2 about here

- -

It is worthwhile at this juncture to digress and note that this result is also interesting in the debate about the importance of speaker adaptation. In order to rule out that the effects of context are simply due to speaker adaptation, we used speech material from the same speaker but from a different part of the conversation. We found that these random contexts still had a quite strong beneficial effect on word recognition of the targets. This is interesting in light of the long debate to what extent we need to adapt our perception to perceive the correct phonological categories, especially for vowels. Classically, studies here focus on identification accuracy. Nearey (1989) provides a meta-analysis and concludes that speaker normalization may not be pivotal in vowel perception due to vowel inherent cues. Our eye-tracking data provide additional data that speaker normalization may nevertheless hugely speed up the efficiency of speech recognition, an effect that may not be visible in the untimed identification responses that Nearey based his conclusion on.

Returning to the issue of prediction and phonological reduction, the only effect of predictability arises when we take into account the natural variation in predictability afforded by the different actual context from the corpus sampled by Brouwer et al. (2013). The right panel shows the effect of adding the actual context for reduced and full forms additionally split up by how well the target utterances fitted into their actual discourse context. While the overall benefit from providing

the actual context did not differ, the efficiency of word recognition—measured as fixation proportion—depends on the amount of coherence for the reduced forms but not the full forms. This is in line with the assumption that prediction is especially important for the recognition of reduced word forms. The reduced forms were strongly influenced by their contextual fit, while the canonical forms were recognized equally well. This suggests that the overall logic arguing for the use of prediction—context information may be more influential when the bottom-up signal is sub-optimal—is sound. However, it remains problematic that, overall, adding more of the actual context in which these forms occurred did not affect canonical and reduced forms differently, the overall benefit was similar for both types of forms. It also remains problematic that the reduced forms do not only occur in high-predictability contexts. Recall that the difficult-to-recognize, relatively unpredictable reduced forms were presented in their actual discourse context. This result indicates that quite often, the reduced word may not be very predictable, leaving little leverage for active prediction to help comprehension of reduced forms.

It is important to emphasize here again that the contextual fit of the target word had mirror-reversed effects in the two time windows. For the canonical forms, it mattered only in the pre-target window but for reduced forms, it only mattered in the post-target window. This suggests that prediction is only beneficial when speech is pronounced carefully, arguably a listening situation when prediction may not be that important for word recognition anyhow. In contrast, when prediction would be useful—that is when words are reduced—the overall amount of reduction makes it difficult for the listener to predict anything. In the time window of target-word recognition, our data suggest that the actual prediction ceases to have any positive impact on word recognition if there is a good bottom-up signal. Stated pointedly, in aiding spoken-word recognition, prediction only works when it is in fact not particularly useful.

If one accepts that prediction in language processing is quite limited, there are two possible consequences. First, prediction may be overrated as a mechanism in language processing (cf. Huettig and Mani, 2016). A second possibility is that prediction serves a different purpose. Evidence for the latter possibility comes from a study by Magyari and De Ruiter (2012). They tested how well the final words of utterances from a spontaneous speech corpus could be predicted, so their data is relevant for the question how predictable spontaneous conversations really are. Their focus, however, was not word recognition but the role of predictability in turn-taking during a dialogue. Turn-taking here refers to the fact that, in a dialogue, one usually functions as both a speaker and a listener. An amazing aspect here is the smoothness with which these transitions occur. Stivers and colleagues

(Stivers et al. 2009) analysed corpora from ten different languages, many of which were linguistically unrelated. Stivers et al. measured the floor-transfer offset (FTO), which is the time difference between one interlocutor's end of a turn and the start of the next interlocutor's turn. Note that FTOs may be negative, indicating an overlap between the two speakers. Stivers et al. found that all languages have a mode in the range of 0-100ms, usually with two thirds of the FTOs occurring in the -150 – 250ms windows. This suggests that a no gap-no overlap FTO seems to be a nearly universally accepted target in spontaneous dialogue (for discussion, see Heldner 2011). The question pursued by Magyari et al. was how this amazing precision is achieved. They used samples from a corpus which had previously been used in a perception experiment (De Ruiter, Mitterer and Enfield 2006). Based on the results of this experiment, Magyari and De Ruiter selected turns with endings that could be predicted well or not. They then used a cloze test with these natural speech samples and found that turns whose endings could be well projected (as established in the experiment of De Ruiter et al., 2006) were also those turn endings in which either the words itself or at least the number of words still to be uttered in the current could be predicted by listeners.

These results show that prediction may indeed have an important role to play for functioning in a natural interaction, as assumed for instance by Pickering and Garrod (2013). However, in contrast to Pickering and Garrod we argue that the main role of prediction is to support turn transitions in natural interaction rather than facilitating word recognition. One may argue that this marginalises the role of prediction, but this is not necessarily the case. This is because producing a topical, well-timed contribution to a fast-paced natural interaction is constant "threat to face" (Levinson and Brown 1978) for the language user. We conjecture that contributing to the solution of this problem is in no way a minor feat.

This focus on turn transition brings us to another issue not yet raised. Proposals that argue for the use of prediction in comprehension (Chang, Dell and Bock 2006; Dell and Chang 2014; Federmeier 2007; 2013) also often assume that the production system is used for prediction. Pickering and Garrod (2013), for instance, cite several neuro-imaging papers that find motor activation in a speech-perception task. It has to be noted, first, that it is not universally agreed that these papers provide evidence for the importance of motor activation for spoken-word recognition (for discussion, see Hickok, Holt and Lotto 2009). In fact, a recent paper indicates that motor involvement may only be important if a task has a "phoneme-awareness component" but not if the listener is simply trying to understand the meaning of an utterance (Krieger-Redwood et al. 2013). Secondly, the issue of turn-taking raises a conceptual issue about the use of the production system for spoken-word recognition. Let us assume for a moment that the production system is indeed used for perception. If we now also take into account that, in natural dialogue, one often produces no

gap-no overlap FTOs, and the planning of an utterance takes 600ms (Indefrey and Levelt 2004), this raises the following question: Which production system is in fact producing these turns, since the (other?) production system seems to be busy predicting which words the speaker will utter next? Starting your own turn with the same word in which the last turn ended might be a solution, but it is not one that speakers actually employ. In fact, analysis of corpora show that the likelihood of a content word from a given turn to be repeated in the following turn is in fact only 10%. While we are not the first to note this conundrum (Scott, McGettigan and Eisner 2009), it is still the case that this issue is not well developed in theories that assume a role of the production system for comprehension.

To summarize, the data on the role of predictability in the processing of spontaneous speech indicate that active prediction is difficult in casual conversation given the uncertainty of the information on which the prediction is to be based: speech containing phonological reductions. Instead we see huge benefits for providing bottom-up information, that is, having heard the speaker for some time makes a huge difference for the efficiency of word recognition (see Figure 2). Prediction is important for functioning in a dialogue sFigure 1ituation nevertheless. Prediction is a crucial part of the turn-taking system, which in turn, is one of the foundations of natural interaction (Sacks, Schegloff and Jefferson 1974). Highlighting this function is important, because it has been assumed that the evidence for prediction in lab speech must mean that it is helpful for word recognition; for what other reason should listeners otherwise predict?

The current data are hence not in line with the assumption that active prediction is an important part of the puzzle when it comes to recognizing reduced words in a natural interaction. However, as the number of studies on this issue is still low, and most of these studies have not been designed to investigate this issue directly, it would be premature to dismiss the idea outright. What should future studies look like?

4. Further directions

An important lacuna in our understanding is how predictable words are in a natural interaction. This brings us back to the introduction, where we stated that prediction seems to be a non-starter from a generative-linguistics point of view. With some psychologists being fundamentally opposed to the generative paradigm (e.g., Christiansen and Chater 2008), it seems that the opposite position which is currently en vogue—language is predictable—may have been adopted too easily. The data by Van der Ven et al. (2012) and Brouwer et al. (2012) seem to indicate that natural interactions may not always be easily predictable (a fact, arguably, that we should be happy about as language users). However, a research paradigm on the importance of prediction in language processing would require more data on how predictable language in a natural interaction

really is. Such an effort cannot rely on corpus studies alone, because the human perceiver may not be able to store all N-gram probabilities as well as a computational model. It is important to point out here that exploiting statistical regularities of language is something quite different from actually predicting what comes next, even though these things are sometimes conflated (cf. Huettig and Mani, 2016).

Another issue is the distinction between the use of lab speech and conversational speech. We do not oppose the use of lab speech in principle. Xu (2010), for instance, makes several valid points that show that the use of lab speech can be useful. Lab speech is not necessarily unnatural and affords much better control of experimental variables than the use of material from a corpus. For instance, the studies of Mitterer and Russell (2013) and Viebahn et al. (2015) were able to actively and precisely manipulate lexical frequency and syntactic predictability in conjunction with presenting both a reduced and a full form in the same sentence. Achieving a similar *ceteris paribus* would be impossible to achieve with excerpts from a corpus. Such well-designed experiments with a clear cut prediction remain the best tool for establishing causal relationships, even in the age of big data (cf. Shadish, Cook and Campbell 2002). The use of lab speech is necessary for such experiments.

This, however, does not mean that all is well and one can simply record any speaker without additional considerations. For instance, we disagree with Xu's statement that "It is not true, in my own experience, that everyone would uncontrollably adopt a careful manner of speaking as soon as they are in front of a microphone." (2010, p. 330) Everything we know from social psychology is that our behaviour is quite often influenced by external primes (Kahneman 2011), and that these primes often do their work without us being aware of it. There is hence good reason that putting someone in front of a microphone—a clear example of an external prime—does indeed lead them to adopt a careful speaking style in an uncontrollable, that is, automatic and unconscious, fashion. Despite this disagreement about the default mode that speakers adopt in such a situation, we agree with Xu (2010) that the speaking style can be controlled, even in a lab-speech style recording situation. In our experience, it helps, for instance, to provide truncated forms (e.g., *wanna* instead of *want to*) in the written form of these sentences (see also Warner 2012, for more discussion). Another more thorny issue is the generation of the stimuli, which should try to approximate the typical usage of spoken rather than written language. A sentence such as "the secret was whispered" (Friederici, Gunter and Mauth 2004) may be grammatical, but such sentences are not typical of spoken language. Even though the examples of Van Berkum et al. (2005, see above) are somewhat more "down to earth", they still feel more like typical written narrative rather than spontaneously produced utterance. While there is no hard and fast rule, it should become an important step to check how likely a given sentence would be in natural interaction. That is, authors should describe

the steps they took to generate materials that are representative for language use outside the lab. This is getting easier by the day. Vast amounts of data are coming online as we speak (such as spoken parts of the British National Corpus, see Renwick et al. 2013), which allows researchers to see whether their constructed sentences structures are likely to occur in natural interactions as well. Raising methodological awareness to these issues as a basic step in constructing materials and pragmatism seems more important than an ideological debate whether one should only use carefully controlled materials or only excerpts from natural interactions.

To conclude, we have argued here that prediction is much less useful for the recognition of reduced words in spontaneous interaction than typically assumed. This raises the question of the use of prediction for language processing more generally. We partly tried to resolve this tension by proposing turn-taking as function that requires prediction, but other functions are also possible. Elman (2009), for instance, suggest that prediction is an important learning mechanism (see also Chang, Dell and Bock 2006; Dell and Chang 2014; but see Huettig and Mani, 2016). Such an assumption dovetails well with findings that populations that have problems in learning to decode written language (e.g., dyslexics, Huettig and Brouwer, 2015 or less opportunity to learn to read and write (illiterates, see Huettig and Mishra 2014; Mishra et al. 2012) show decreased anticipatory language processing. Given the scarcity of evidence on prediction in spontaneous speech, it is too early to close the door on the assumption that prediction may be important for word recognition in natural interactions. However, we conjecture that data with speech materials that are ecologically valid (see Warner 2012, for suggestions) are necessary to "resurrect" the idea that prediction is important for word recognition in natural interactions.

5. References

Altmann, Gerry & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3). 247–264.

Aylett, Matthew & Alice Turk. 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. (29 December, 2014).

Berkum, Jos J. A. Van, Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Peter Hagoort. 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(3). 443.

Borovsky, Arielle, Jeffrey L. Elman & Anne Fernald. 2012. Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of experimental child psychology* 112(4). 417–436.

Bradlow, Ann R. & Tessa Bent. 2008. Perceptual adaptation to non-native speech. *Cognition* 106. 707–729.

Brouwer, Susanne, Holger Mitterer & Falk Huettig. 2013. Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics* 34. 519–539. doi:10.1017/s0142716411000853.

Bürki, Audrey & M. Gareth Gaskell. 2012. Lexical representation of schwa words: two mackerels, but only one salami. *Journal of Experimental Psychology. Learning, Memory, and Cognition* 38(3). 617–631. doi:10.1037/a0026167.

Chang, Franklin, Gary S. Dell & Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113(2). 234–272. doi:10.1037/0033-295X.113.2.234.

Christiansen, Morten H. & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31. 489–509. doi:10.1017/s0140525x08004998.

Christiansen, M. H., & Chater, N. (2016). The now-or-neverbottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences,* 39, e62. doi: 10.1017/S0140525X1500031X

Connine, Cynthia M., Larissa J. Ranbom & David J. Patterson. 2008. Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics* 70. 403–411. doi:10.3758/PP.70.3.403.

Crystal, David. 2011. *Dictionary of linguistics and phonetics*. . Vol. 30. John Wiley & Sons.

Dell, Gary S. & Franklin Chang. 2014. The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 369(1634). 20120394. doi:10.1098/rstb.2012.0394.

DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience* 8(8). 1117–1121.

De Ruiter, Jan Peter, Holger Mitterer & Nick J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*. 515–535. (30 September, 2013).

Dilley, Laura, Melissa M. Baese-Berk, Stephanie Schmidt, Jesse Nagel, Tuuli Morrill & Mark Pitt. 2013. One small step for (a) man: Function word reduction and acoustic ambiguity. *Proceedings of Meetings on Acoustics* 19(1). 060297. doi:10.1121/1.4800664.

Dilley, Laura & Mark A. Pitt. 2010. Altering context speech rate can cause words to appear or disappear. *Psychological Science* 21(11). 1664–1670. (21 February, 2014).

Dunbar, Robin. 1998. *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Elman, Jeffrey L. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive science* 33(4). 547–582.

Enfield, Nicholas J. 2003. The definition of< i> what-d'you-call-it</i>: semantics and pragmatics of recognitional deixis. *Journal of Pragmatics* 35(1). 101–117.

Ernestus, Mirjam. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142. 27–41.

Ernestus, Mirjam, Harald R. Baayen & Rob Schreuder. 2002. The recognition of reduced word forms. *Brain and Language* 81. 162–173. doi:10.1006/brln.2001.2514.

Federmeier, Kara D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4). 491–505.

Friederici, Angela D., T.C. Gunter & K. Mauth. 2004. The relative timing of syntactic and semantic processes in sentence comprehension. *Neuroreport* 15. 165–169.

Gagnepain, Pierre, Richard N. Henson & Matthew H. Davis. 2012. Temporal Predictive Codes for Spoken Words in Auditory Cortex. *Current Biology* 22(7). 615–621. doi:10.1016/j.cub.2012.02.015.

Hayes, Donald P. 1988. Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language* 27(5). 572–585. (24 December, 2014).

Heldner, Mattias. 2011. Detection thresholds for gaps, overlaps, and no-gap-no-overlaps. *The Journal of the Acoustical Society of America* 130(1). 508–513. (23 December, 2014).

Hickok, G., L. L. Holt & A. J. Lotto. 2009. Response to Wilson: What does motor cortex contribute to speech perception? *Trends in Cognitive Sciences* 13(8). 330–331. doi:10.1016/j.tics.2009.05.002.

Huettig, Falk.2015. Four central questions about prediction in language processing. Brain Research 1626. 118-135. doi:10.1016/j.brainres.2015.02.014.
Huettig, Falk & Brouwer, Susanne. 2015. Delayed anticipatory spoken language processing in adults with dyslexia - Evidence from eye-tracking. *Dyslexia* 21(2). 97-122.
Huettig, Falk, & Mani, Nivi. 2016. Is prediction necessary to understand language? Probably not. Language, Cognition and Neuroscience, 31(1). 80-93. doi: 10.1080/23273798.2015.1072223
Huettig, Falk & James. M. McQueen. 2007. The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language* 57. 460–482.
Huettig, Falk & Ramesh K. Mishra. 2014. How literacy acquisition affects the illiterate mind–a critical examination of theories and evidence. *Language and Linguistics Compass* 8(10). 401–427.
Indefrey, P. & W. J. M. Levelt. 2004. The spatial and temporal signatures of word production components. *Cognition* 92(1-2). 101–144. doi:10.1016/j.cognition.2002.06.001.
IPDS. 1994. *The Kiel Corpus of Spontaneous Speech*. Kiel Germany: Universität Kiel.
Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press. http://books.google.com/books?hl=de&lr=&id=d9O9-w1c1j4C&oi=fnd&pg=PP5&dq=ray+jackendoff+foundations&ots=6mom6N_wyb&sig=yN2OR6I-31n6pqiVgVEuktOTvxg (14 April, 2014).
Janse, Esther & Mirjam Ernestus. 2011. The roles of bottom-up and top-down information in the recognition of reduced speech: evidence from listeners with normal and impaired hearing. *Journal of Phonetics* 39(3). 330–343.
Johnson, Keith. 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*, 29–54. Tokyo, Japan: The National International Institute for Japanese Language.
Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.
Kamide, Yuki, Gerry Altmann & Sarah L. Haywood. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49(1). 133–156.
Karlsson, Fred. 2007. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43. 365–392. doi:10.1017/S0022226707004616.
Keating, Patricia. 1997. Word-level phonetic variation in large speech corpora. *an issue of ZAS Working Papers in Linguistics, ed. Berndt Pompino-Marschal. Available as http://www. humnet. ucla. edu/humnet/linguistics/people/keating/berlin1. pdf*. http://www.linguistics.ucla.edu/people/keating/berlin1.pdf (12 December, 2014).
Krieger-Redwood, Katya, M. Gareth Gaskell, Shane Lindsay & Elizabeth Jefferies. 2013. The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. *Journal of cognitive neuroscience* 25(12). 2179–2188.
Levinson, Stephen C. & Penelope Brown. 1978. *Politeness: Some universals in language usage*. Cambridge, UK: Cambridge University Press.
Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177. doi:10.1016/j.cognition.2007.05.006.
Lindblom, Björn. 1990. Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (eds.), *Speech production and speech modelling*, 403–439. Dordrecht, The Netherlands: Kluwer.
Magyari, Lilla & J. P. de Ruiter. 2012. Prediction of Turn-Ends Based on Anticipation of Upcoming Words. *Frontiers in Psychology* 3. doi:10.3389/fpsyg.2012.00376. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3483054/ (30 September, 2013).
Mani, Nivedita & Falk Huettig. 2012. Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance* 38(4). 843.

Mani, Nivedita & Falk Huettig. 2014. Word reading skill predicts anticipation of upcoming spoken language input: A study of children developing proficiency in reading. *Journal of experimental child psychology* 126. 264–279.
Manuel, S. Y. 1992. Recovery of "'deleted'" schwa. *Perilus: Papers from the Symposium on current phonetic research paradigms for speech motor control*, 115–118. Stockholm: University of Stockholm.
Mattys, Sven L., Matthew H. Davis, Ann R. Bradlow & Sophie K. Scott. 2012. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes* 27(7-8). 953–978.
McClelland, Jay L. & Jeffrey L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18. 1–86. doi:10.1016/0010-0285(86)90015-0.
McClelland, Jay L. & D. E. Rumelhart. 1981. An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review* 88. 375–407.
McQueen, James M. & Malte Viebahn. 2007. Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology* 60. 661–671. doi:10.1121/1.419865.
Mishra, Ramesh K., Niharika Singh, Aparna Pandey & Falk Huettig. 2012. Spoken language-mediated anticipatory eye movements are modulated by reading ability: Evidence from Indian low and high literates. *Journal of Eye Movement Research* 5(1). 1–10.
Mitterer, Holger & Mirjam Ernestus. 2006. Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics* 34. 73–103. doi:10.1016/j.wocn.2005.03.003.
Mitterer, Holger & Kevin Russell. 2012. How phonological reductions sometimes help the listener. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0029196.
Nation, Kate, Catherine M Marshall & Gerry T. M Altmann. 2003. Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology* 86(4). 314–329. doi:10.1016/j.jecp.2003.09.001.
Nearey, Terrance M. 1989. Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85. 2088–2113.
Norris, Dennis. 2006. The Bayesian reader: explaining word recognition as an optimal Bayesian decision process. *Psychological review* 113(2). 327.
Norris, Dennis & James M. McQueen. 2008. Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review* 115. 357–395. doi:10.1037/0033-295X.115.2.357.
Pickering, Martin J. & Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(04). 329–347.
Pierrehumbert, Janet & David Talkin. 1992. Lenition of/h/and glottal stop. *Papers in laboratory phonology II: Gesture, segment, prosody*. 90–117. (4 July, 2014).
Pitt, Mark A. 2009. How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language* 61. 19–36.
Pluymaekers, Mark, Mirjam Ernestus & Harald R. Baayen. 2005. Leixcal frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118. 2561–2569. doi:10.1121/1.2011150.
Renwick, Margaret E., Ladan Baghai-Ravary, Ros Temple & John S. Coleman. 2013. Assimilation of word-final nasals to following word-initial place of articulation in United Kingdom English. *ICA 2013 Montreal*, vol. 19. Montreal, Canada: ASA. doi:10.1121/1.4800279. http://link.aip.org/link/?PMA/19/060257/1 (12 June, 2013).
Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50(4). 696–735.
Salverda, Anne Pier, Dave Kleinschmidt & Michael K. Tanenhaus. 2014. Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of memory and language* 71(1). 145–163. (24 December, 2014).

Scott, Sophie K., Carolyn McGettigan & Frank Eisner. 2009. A little more conversation, a little less action: candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience* 10. 295–302. doi:10.1038/nrn2603.

Shadish, William R., Thomas D. Cook & Donald Thomas Campbell. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Sjerps, Matthias J., James M. McQueen & Holger Mitterer. 2013. Evidence for precategorical extrinsic vowel normalization. *Attention, Perception, & Psychophysics* 75(3). 576–587. doi:10.3758/s13414-012-0408-7 (19 June, 2013).

Sjerps, Matthias J, Holger Mitterer & James M McQueen. 2011. Listening to different speakers: on the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia* 49(14). 3831–3846. doi:10.1016/j.neuropsychologia.2011.09.044.

Spinelli, Elsa & Florent Gros-Balthazard. 2007. Phonotactic constraints help to overcome effects of schwa deletion in French. *Cognition* 104. 397–406.

Stivers, Tanya, Nick J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, et al. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106(26). 10587–10592.

Sumner, Meghan & Arthur G. Samuel. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language* 60. 487–501.

van de Ven, M. A. M., M. T. C. Ernestus & R. Schreuder. 2012. Predicting acoustically reduced words in spontaneous speech. http://dare.ubn.kun.nl/handle/2066/101944 (23 December, 2014).

Viebahn, Malte, Mirjam Ernestus & James M. McQueen. 2015. Syntactic predictability in the recognition of carefully and casually produced speech. *Journal of Experimental Psychology: Learning Memory and Cognition*. doi: 1684-1702. doi:10.1037/a0039326

Viebahn, Malte, Mirjam Ernestus & James M. McQueen. 2012. Co-occurrence of reduced word forms in natural speech. *INTERSPEECH*. http://20.210-193-52.unknown.qala.com.sg/archive/archive_papers/interspeech_2012/i12_2021.pdf (18 December, 2014).

Warner, Natasha. 2012. Methods for studying spontaneous speech. In Abigail C. Cohn, Cecile Fougeron & Marie K Huffman (eds.), *The Oxford Handbook of Laboratory Phonology*. Oxford, UK: Oxford University Press.

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

Figure Captions

Figure 1: Difference in fixation proportions to targets compared to the competitors and distractors just before target onset in Brouwer et al. (2013). The error bars indicated one standard error around the mean (estimated by subjects). The results show that prediction occurs mostly when the context is especially coherent and the speaking style is quite clear.

Figure 2: Fixation proportions for the target words in the different conditions of Brouwer et al. (2013) depending on whether (bar type, dark or light) and which type of context was presented. The error bars indicated one standard error around the mean (estimated by subjects). Panel A shows the results from the actual contexts compared with a random context from the same speaker. Panel B shows differences between the actual contexts that were highly or lowly coherent with the target sentence.