Discourse context and the recognition of reduced and canonical spoken words

SUSANNE BROUWER and HOLGER MITTERER
Max Planck Institute for Psycholinguistics

FALK HUETTIG
Max Planck Institute for Psycholinguistics and Donders Institute for Brain, Cognition, and Behaviour

ADDRESS FOR CORRESPONDENCE
Susanne Brouwer, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: s-brouwer@northwestern.edu

ABSTRACT
In two eye-tracking experiments we examined whether wider discourse information helps the recognition of reduced pronunciations (e.g., "puter") more than the recognition of canonical pronunciations of spoken words (e.g., "computer"). Dutch participants listened to sentences from a casual speech corpus containing canonical and reduced target words. Target word recognition was assessed by measuring eye fixation proportions to four printed words on a visual display: the target, a "reduced form" competitor, a "canonical form" competitor, and an unrelated distractor. Target sentences were presented in isolation or with a wider discourse context. Experiment 1 revealed that target recognition was facilitated by wider discourse information. It is important that the recognition of reduced forms improved significantly when preceded by strongly rather than by weakly supportive discourse contexts. This was not the case for canonical forms: listeners' target word recognition was not dependent on the degree of supportive context. Experiment 2 showed that the differential context effects in Experiment 1 were not due to an additional amount of speaker information. Thus, these data suggest that in natural settings a strongly supportive discourse context is more important for the recognition of reduced forms than the recognition of canonical forms.

Casual speech used in everyday conversations is highly variable and contains many phonological reductions (Ernestus, 2000; Johnson, 1997, 2004). For example, during a casual conversation a speaker of Dutch may pronounce the word *beneden* [bəneːdə] "downwards" as [məneːə]. Reduced forms can thus be substantially different from their canonical counterparts. Surprisingly, however, reductions do not seem to hinder the communication between speaker and listener. A likely reason for this is that phonological and sentential context may help listeners to recognize reduced forms. Reduced words in a sentential context, however, are still misidentified in almost 10% of the cases (Ernestus, Baayen, & Schreuder, 2002).

The question therefore arises how listeners recognize the meaning of reduced forms successfully. In the present research, we test whether a supportive wider discourse context is a key factor for successful recognition. More specifically, we examine the hypothesis that a strongly supportive discourse context is more important for the recognition of reduced forms (e.g., [mǝne:ǝ]) than it is for the recognition of canonical forms (e.g., [bǝne:dǝ]).

The question how reduced forms can be recognized at all has received considerable attention recently. Classically, models of the mental lexicon have assumed that the entry for a given word consists of one abstract representation of that word (McClelland & Elman, 1986; Norris, 1994), similar to a dictionary. This makes it difficult to account for the recognition of reduced forms, because the input that does not match the form in the mental lexicon, which is based on a canonical pronunciation. Within the abstractionist account, there are two ways to still account for the recognition of reduced forms. One is to assume that prelexical processes are able to partly undo the effects of reductions (Gaskell & Snoeren, 2008; Gow, 2002; Mitterer & Blomert, 2003), just as prelexical processes normalize for speaking rate and speaker differences (Ladefoged & Broadbent, 1957; Newman & Sawusch, 1996). Another possibility is to assume that features that are reduced do not form a part of the lexical representation. This assumption of phonological underspecification (Lahiri & Reetz, 2002) is, however, difficult to reconcile with many findings on spoken-word recognition (Mitterer, 2011). Another account with a radically different view also assumes that storage in the mental lexicon is crucial for the recognition of reduced forms (Goldinger, 1998; Johnson, 1997, 2004). According to so-called episodic models of the mental lexicon, the mental lexicon consists of stored exemplars of spoken words. A cloud of exemplars will then contain both reduced and canonical versions of a given word and the recognition of reduced forms can be achieved by a simple pattern matching. Evidence for the role of episodes stems from findings that word-form frequencies influence recognition. Connine and coworkers showed that recognizing a reduced form was easier if that particular word is often reduced (Connine, Ranbom, & Patterson, 2008; Ranbom & Connine, 2007). Pitt (2009) provided very clear evidence for lexical storage by using a learning paradigm. Participants learned new words (e.g., *senty)*, and only recognized a reduced variant of a newly learned word (e.g., *senny)*, if that word had been heard in that reduced form before. Even though the reduction was quite regular (cf., *twenty*, which is often produced as *twenny)*, and participants would have phonological knowledge that /nt/ sequences can be produced as [n], they needed exposure to the reduced form.

Our current question is, however, orthogonal to the question of the basic mechanisms that support the recognition of reduced forms. The accounts above do not differ in their assumption how discourse context can influence word recognition. Most of the past research using carefully pronounced laboratory speech has investigated the effect of discourse context for the *prediction* of upcoming words rather than the effect of the wider discourse context on the *recognition* of spoken words (e.g., Altmann & Kamide, 1999, 2007, 2009; Kamide, Altmann, & Haywood, 2003). Altmann and Kamide (1999), for example, showed that when listeners hear a sentence such as "The boy will eat the cake" in the context of a scene depicting a boy and a cake (and other things), they shift their eye gaze toward

the cake even before "cake" starts to acoustically unfold. Altmann and Kamide interpreted this finding as evidence that selectional information conveyed by a verb can be used to predict an upcoming theme. Kamide et al. (2003) further explored whether the combination of verb information with the preceding grammatical subject can be used for prediction. They found increased fixations to a motorbike when participants heard "The man will ride . . .," but increased fixations to a carousel when they heard "The girl will ride . . . " Therefore, information provided by the grammatical subject and by the verb can jointly constrain anticipation (at least when a visual context is present; see also Kamide, Scheepers, & Altmann, 2003, for evidence that case marking can be used for prediction).

Most studies investigating the effects of (semantically predictable) context have used isolated sentences only (e.g., Federmeier, 2007). Van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort's study (2005), however, investigated how wider discourse context (i.e., more than one sentence) can be used to predict an upcoming noun (e.g., as in "The burglar had no trouble locating the secret family safe. Of course, it was situated behind a . . . "). Event-related potentials (ERPs) to determiners and adjectives were measured for prediction-consistent (e.g., "big$_{NEU}$ painting$_{NEU}$") and prediction-inconsistent nouns (e.g., "big$_{COM}$ bookcase$_{COM}$"). The results showed an N400 effect for adjectives inconsistent with the discourse-predictable noun relative to adjectives consistent with the discourse-predictable noun. The N400 component is typically associated with difficulty during semantic integration of words in a sentence context (Kutas & Hillyard, 1984). When these stories were presented in a self-paced reading task, prediction-inconsistent adjectives also slowed readers down. These data suggest that people use wider discourse context to predict upcoming words.

The focus of our present research is not on whether people can use such context for prediction (evidently they can) but on the effect of wider discourse context on the actual *recognition* of spoken words. How context affects the recognition of canonical words, in written or auditory form, has been the topic of a large number of investigations. In a classic paper, Swinney (1979) tested the influence of sentential context on the interpretation of homonyms such as *bank*. The experiments made use of cross-modal priming, the participants heard a sentence containing *bank* and then made lexical decisions to words related to one of the meanings of *bank* just after the word had been heard (e.g., "money" or "river"). The results suggested that the context had no influence on initial word recognition because significant priming was observed even if the sentential context supported the contextually inappropriate meaning of *bank*. Some later studies, however, came to a different conclusion. Tabossi (1988), for instance, found that context biasing toward the dominant meaning of an ambiguous word primed associates of the dominant (but not the subordinate) meaning. In contrast, a context biasing toward the subordinate meaning of the ambiguous word primed associates of both the dominant and the subordinate meaning. Tabossi's (1988) results suggested that context can suppress meaning activation of (at least) the subordinate meanings of ambiguous words, but that in general context has quite a limited role in meaning activation.

Later studies used more time-sensitive methods such as ERPs or eye tracking. It is well established that semantically anomalous words trigger an ERP N400 effect (Kutas & Hillyard, 1984). As mentioned above, Van Berkum et al. (2005),

for instance, found a reduced N400 for adjectives consistent with the discourse-predictable noun. It is, however, still a matter of debate if the reduction of the N400 indicates ease of word recognition. A number of studies investigated this issue by exploiting the temporal unfolding inherent to speech (Connolly & Phillips, 1994; Van den Brink, Brown, & Hagoort, 2001). In these studies, the sentence context was highly constraining (e.g., "The gambler had a streak of bad . . ." and the final unexpected word has in some cases an initial overlap with the expected word (e.g., *luggage*). The results showed an earlier negativity if the final unexpected word had no initial overlap with the expected word. Although this result can be taken to suggest that context influences the recognition of spoken words at an early level, the results can also be explained by assuming a matching between expected words and candidate words at a postlexical stage (cf. Van den Brink et al., 2001).

Like ERP studies, visual world eye-tracking studies do not provide a clear picture. Huettig and Altmann (2007) found that participants' attention was directed toward a visual referent (a needle) that was similar in shape to the dominant referent of a heard homonym (e.g., pen-writing implement) even though there was a picture of the subordinate referent (pen-cage) present, and the linguistic context biased the subordinate meaning of the homonym. Their data thus provided evidence for the activation of the inappropriate dominant meaning of the word "pen." This suggests that the visual-shape representations of the contextually inappropriate dominant referent were accessed even though the contextually more appropriate subordinate (but not the dominant) referent was depicted in the scene. Dahan and Tanenhaus (2004) tested whether semantic constraints provided by a verb can influence word recognition in Dutch sentences such as "Nog nooit klom een bok zo hoog" (English, "Never before climbed a goat so high"). Participants heard such sentences and saw four pictures on the screen with the instruction to click on pictures of words that appeared in the sentence. It was critical that the display contained not only a picture of a goat but also a picture of a bone (an onset competitor of goat, *bok*, in Dutch, *bot*). If the sentence did not contain verb-based constraints, participants looked more at the picture of the onset competitor than at the picture of unrelated distractors, but this effect disappeared if the participants had heard the semantically constraining verb before. The authors argued that their results "favor models in which mapping from the input onto meaning is continuous over models in which contextual effects follow access of an initial form-based competitor set" (p. 498; but see Huettig, Rommers, & Meyer, 2011, for an alternative explanation of their Experiment 1). Dahan and Tanenhaus' (2004) second experiment, however, suggested that the use of context information does not override bottom-up information. Dahan and Tanenhaus used a cross-splicing manipulation to make the input slightly and temporally more similar to the Dutch word for *bone*. This resulted in a reemergence of looks to the picture of a bone. These data suggest that contextual information may be used during early stages of word recognition, but that such context information is easily ignored if there is bottom-up evidence suggesting a different interpretation.

On balance, therefore, there seems to be some evidence that the recognition of canonical forms of spoken words can be influenced by sentential and discourse context information. These influences, however, seem to be quite restricted. Context information appears not to override perceptual processing, and words are

recognized on the basis of their match with input, even if they are not supported by context.

Note that all these studies used carefully recorded stimuli. In particular, very few studies have looked at the use of context for the recognition of *reduced forms* during casual speech. An exception is a study by Ernestus et al. (2002), who selected samples from a spontaneous speech corpus to examine how listeners recognize highly reduced forms (e.g., [mɔk] from [mɔxələk] *mogelijk* "possible") in Dutch. Participants listened to such forms in a sentential context (e.g., [zo snɛl mɔk na ə:] *zo snel mogelijk naar eh* "as fast as possible to uhm"), in a phonetic context (e.g., [ɛl mɔk na] *el mogelijk naa* "ast possible to"), and without any context (e.g., [mɔk] *mogelijk* "possible"), and were asked to write down the form they heard. The results showed that listeners hardly recognize reduced forms on the basis of the acoustic signal for that word alone. Identification performance increased when highly reduced forms were presented in a phonetic context. However, only when presented in a sentential context performance for highly reduced forms improved substantially. Nevertheless, listeners still misidentified reduced forms in almost 10% of the cases (see also Kemps, Ernestus, Schreuder, & Baayen, 2004).

We extend the results of Ernestus et al. (2002) in three respects. First, we present the wider discourse context of reduced forms and not only the surrounding sentence to participants. Second, we include conditions with canonically produced words in order to compare the recognition of reduced forms with the recognition of canonical forms. Because reduced forms are more difficult to recognize, it is conceivable that discourse information aids reduced and canonical forms to a different degree. A third extension with regard to Ernestus et al. (2002) is the use of an online method. Ernestus and colleagues (2002) used an offline task (self-paced listening); here, we investigate target recognition online by using eye tracking. Participants listen to target sentences while four printed words are displayed on the screen: the target word (e.g., *beneden* "downwards"), a phonologically unrelated distractor (e.g., *vakantie* "holiday"), a "canonical form" competitor (e.g., *benadelen* "to disadvantage"), and a "reduced form" competitor (*meneer* "mister"). The critical experimental manipulation was whether the target sentences were preceded by discourse context.

Comparing the recognition of a given word in a sentence context with recognition in a wider discourse context leads to a possible confound. The preceding contexts contain not only additional discourse information but also additional speaker information. Many studies have shown that speaker information can be an important aid for the listener to recognize spoken words (e.g., Bradlow, Nygaard, & Pisoni, 1999; Mullenix, Pisoni, & Martin, 1989; Nygaard, Sommers & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993). Furthermore, a large body of research has shown that listeners adapt to speaker-specific characteristics on the time scale of minutes (e.g., Eisner & McQueen, 2005; Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2005; Norris, McQueen, & Cutler, 2003) and even seconds (Ladefoged & Broadbent, 1957). For instance, Mitterer (2006) showed that adaptation to a speaker is stronger when more information about this speaker is available. Therefore, it is essential to show that the advantage in the processing of reduced (and possibly canonical) forms in a wider discourse context over the processing of the same form in the sentence context is not solely due to more efficient adaptation

to the speaker. After all, by presenting wider discourse context, we also expose the listener longer to a given speaker. In Experiment 2, we thus presented the same target sentences with different contexts. Instead of using the actual context in which the word occurred, we selected another arbitrarily chosen sample from the same speaker. These control contexts provided the same amount of speaker information but no matching discourse information. Experiment 2 therefore allows us to measure how much benefit speaker information provides for the recognition of reduced and canonical forms.

In sum, the present research examines the effects of wider discourse context on the recognition of reduced forms. To judge whether the effects found in our paradigm are specific to reduced forms, we also collect data for canonical forms. Our critical prediction is that the recognition of reduced forms relies more on strongly supportive contexts than the recognition of canonical forms. To assess how contextually supportive the different contexts were (i.e., both the actual discourse context and the control contexts), we first performed a pretest with these materials.

PRETEST

In the present research we use ecologically valid examples of reductions in casual speech. To be able to do this, we have to work with stretches of speech extracted from a spontaneous speech corpus. A downside of using spontaneous speech materials is the lack of control one has over such stimuli. We extracted target sentences and the discourse context directly preceding these target sentences. The discourse contexts provided minimally 5 s of speech of the target speaker. We conducted a pretest to examine whether the selected samples provide supportive discourse information to listeners, which they can use to recognize targets successfully. A second purpose of the pretest was to empirically confirm that the "speaker-only" contexts, which were to be used in Experiment 2, actually do not contain any supportive discourse information.

In this pretest, listeners were asked to rate how well the contexts preceding the target sentences (e.g., *Ja, dat is echt uh . . . Nou we hebben daar ook nog gestaan. Ik heb daar ook nog gefilmd. En dan komt dat water komt echt zo naar je toe en dan* "Yes, that is really, uhm . . . Well, we were also standing there. I also made a movie there. And then the water really approaches you and then") matched with the target sentences (e.g., *buigt het zo af en dan valt het naar beneden, dat is echt* "it bends like this and then it falls down, that is really").

METHOD

Eighteen members of the Max Planck Institute subject pool participated in the pretest for which they were paid. None of them reported any hearing disorders and all had normal or corrected to normal vision. Listeners were tested individually in a sound-attenuated booth. The presentation of the stimuli was controlled by Presentation software. The auditory stimuli were presented to the participants over headphones.

Table 1. *Mean (standard deviation) ratings in the pretest*

	Word Form	
Information	Reduced	Canonical
Discourse (Exp. 1)	1.90 (1.99)	2.19 (2.11)
Weakly supportive	0.63 (1.14)	1.14 (1.62)
Strongly supportive	3.16 (0.50)	3.24 (0.40)
Control (Exp. 2)	−0.52 (2.44)	−0.91 (2.28)

Note: The scale is from mismatch (−5) to match (+5).

We presented 112 preceding discourse contexts followed by their accompanying target sentence. Half of the items were experimental items (context A), whereas the other half of the items were control items (context B, to be used in Experiment 2). For the control items we selected random contexts from the same speaker that did not directly precede the target sentences. This presentation mode created an AXBX task, in which A and B were the preceding contexts and X the target sentences. The presentation of A and B was counterbalanced. Each participant received a different random ordering of the stimuli, but started with the same three practice trials to familiarize themselves with the task. A fixation cross appeared for 300 ms between the presentation of the preceding context and the target sentence. This fixation cross was an indication for the participant that the target sentence would start. After the presentation of the target sentence, participants saw a vertical line crossing out a horizontal bar on the screen. The horizontal bar represented a continuum from mismatch (−5) to match (+5). Participants were asked to indicate with the scroll wheel on the computer's mouse whether the preceding contexts matched with the target sentences or not. The scroll wheel enabled participants to move the vertical line on the continuum to the left (−5) or to the right (+5). Once participants made a decision, they had to confirm the position of the vertical line on the continuum with the left mouse button. After they clicked on the left mouse button, the next trial initiated. Participants were put under no time pressure to perform this action. There was a short pause half way through the experimental list. The pretest lasted about 35 min.

RESULTS

Table 1 shows the rating scores for word form (canonical vs. reduced) and for information (discourse vs. control). A mixed effect logistic regression model was used to test whether the target sentences were rated to match better with the discourse contexts than with the control contexts. This was the case ($\beta_{information} = -2.76$, $p_{MCMC} = .0001$). We found no main effect of word form or an interaction between information and word form (all $p_{MCMC}s > 0.1$). The results indicate that our stimuli selection was appropriate: contexts with discourse information (to be used in Experiment 1) provide more useful information for listeners than our control contexts (to be used in Experiment 2).

Note that the range of ratings for our selected experimental items was wide for both target types (for reduced targets, range = −2.11 to 3.83; for canonical targets, range = −2.67 to 4.11). This shows that some contexts were strongly supportive, whereas other contexts were only weakly supportive. Thus, as in real-world situations, not all discourse contexts provide supportive information to a similar extent. We therefore took into account how supportive a given context is for a given item in our data analysis. For visualization purposes, we used a median split to label the canonical and reduced items below the median as weakly supportive contexts and those above the median as strongly supportive contexts. For the statistical data analysis, we used the degree of support as a covariate to examine whether this influences target recognition as measured by fixation proportions.

EXPERIMENT 1

Method

Participants.   Forty-eight participants from the Max Planck Institute's subject pool, undergraduates at the Radboud University in Nijmegen were paid to participate in this experiment. All participants were native speakers of Dutch and had normal or corrected to normal vision. No participant reported any (history of) hearing problems. None of the participants took part in the pretest.

Materials.   Twenty-eight polysyllabic, mid- to high-frequency content words were selected from the Spoken Dutch Corpus (Oostdijk, 2000) as target words, of which 28 were realized canonically (e.g., [bəneːdə] for *beneden* "downwards") and 28 were pronounced in a reduced way (e.g., [məneːə]). We selected Dutch recordings (and not Flemish), because this variant of Dutch is most familiar to the participants in our subject pool. Recordings with background noise or overlapping speech were excluded. Two independent raters transcribed the target words using the software package PRAAT (Boersma, 2001) such that they could observe the signal in auditory and visual, spectrographic form. In case of disagreement between the independent ratings, the transcribers conferred to achieve an unanimous transcription. Canonical targets were selected if all segments were (almost) fully realized, whereas their reduced counterparts were selected if one or more segments were missing.

Each target was embedded in a sentence. For each of the target sentences, we searched in the spontaneous speech corpus for the discourse context directly preceding the target sentence. A preceding context was included in the study so that the speech of the target speaker in the preceding context consisted of a minimum duration of 5 s. Participants listened to the target sentence alone (sentence-only condition) or to the additional context and the target sentence (wider discourse condition).

We used the printed-word variant (Huettig & McQueen, 2007; McQueen & Viebahn, 2007) of the visual world paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). On each trial, participants were presented with a visual display containing four printed words. Each display contained the

printed target word (e.g., *beneden* "downwards"), a "canonical form" competitor (e.g., *benadelen* "to disadvantage"), a "reduced form" competitor (e.g., *meneer* "mister"), and a phonologically unrelated distractor (e.g., *vakantie* "holiday"). We included competitors of the target word to make the task ("Click on the target word that appears in the sentence") more challenging to participants. A "canonical form" competitor shared more onset overlap with the canonical target (e.g, *benadelen* "to disadvantage" [bənaːdeːlə] for [bəneːdə]), whereas the initial segments of a "reduced form" competitor overlapped better with the reduced target (e.g., *meneer* "mister" [məneːr] for [məneːə]). In such a display, there are always two to three phonologically related words, of which one was the target. We therefore masked this pattern by adding filler items, which we also selected from the spontaneous speech corpus. Each filler trial also consisted of two to three phonologically similar words and one to two unrelated words, but half of the time one of the dissimilar words was the target and half of the time one of the similar words was the target. In this way, listeners were discouraged to limit their attention to the phonologically related words only.

Two lists were created. One list contained 28 reduced targets and 28 fillers, whereas the other list contained 28 canonical targets and 28 fillers. Within each list, half of the words were presented in only their sentence context, and the other half in the complete discourse context. The number of times each item appeared with or without a discourse context was counterbalanced over participants. The order of each list was then randomized for each participant separately. This randomization did not only affect the order of the auditory stimuli but also the position of the four types of printed words on the screen. We have often observed that participants will first look to the item in the upper left quadrant of the screen. It is hence important to make sure that the randomization also varies the position of the target words, so that each item appears on all positions over participants and that for each participant, the target (and its competitors) occur in the same proportions on the four quadrants of the computer screen. To familiarize participants with the task, the experimental run started with a warm-up session containing six practice items. These items were also selected from the corpus.

Procedure. Participants were tested individually in a sound-attenuated booth. They were seated at a comfortable viewing distance from the computer screen. Eye movements were monitored at a sampling rate of 1 kHz with an SR-Research EyeLink1000 eye tracker (used in the tower-mount version). The presentation of the auditory and visual stimuli was controlled with SR-Research program Experiment builder. The auditory stimuli were presented to the participants over headphones.

Participants received written instructions on the screen. They were instructed that they would first see a cross in the middle of the screen. During the presentation of this cross they either listened to an auditory fragment (i.e., the preceding context) or to a 300-ms silence. After the auditory fragment or the silence, the target sentences were presented. During this presentation, the four printed words appeared in a 24-point Courier font on the screen. The centers of the printed words corresponded, independently of the length of the words, to the centers of the quadrants on the screen. The participants had to use the computer's mouse to

click on the printed word that appeared in the target sentence. After they clicked with the mouse on one of the words, the next trial initiated. Participants were put under no time pressure to perform this action. A central fixation cross appeared centered on the screen after every 10 trials, permitting for drift correction in the calibration.

Each participant first completed six practice trials. After that, one of the two lists was presented in random order. The experimental session lasted about 15 min.

Design and analysis. Reduced targets were presented to half of the participants and canonical targets to the other half of the participants. Click responses and eye movements were the dependent variables. For the click responses we calculated the percentage of correct clicks to the target and the percentage of incorrect clicks to the competitors and the distractor. Participants made no errors in any of the experiments. Statistical analyses on the errors were therefore not carried out.

For the eye movement data we discarded blinks and saccades. In order to assess the effect of the wider discourse context on the actual recognition of reduced and canonical forms, we analyzed our data from 200 ms onward, because of estimates that it takes 200 ms to program and launch a saccadic eye movement (e.g., Matin, Shao, & Boff, 1993). Thus, before 200 ms after word onset fixations are unlikely to be driven by acoustic information from the critical target word. As Figure 1 illustrates, fixations to the competitors in the wider discourse condition converged with the distractor at around 1000 ms after word onset. We therefore choose to statistically analyze fixation proportions during the 200–1000 ms time window after the acoustic onset of the target word.

The dependent variable was the fixation proportions to the target word. For the analysis we transformed the fixation proportions with the empirical logit function (cf. Barr, 2008). We tested whether target fixations were influenced by the presence versus absence of wider discourse information using linear mixed effects models (Baayen, Davidson, & Bates, 2008), with participants and items as random effects and in which discourse information was coded as a numeric contrast (−0.5 and 0.5; cf. Barr, 2008). The sentential context only condition was coded as −0.5 and the wider discourse condition as 0.5. The amount of support provided by the wider discourse context as obtained in the pretest was used as a covariate. We estimated *p* values by using Markov chain Monte Carlo simulations (Baayen et al., 2008).

Results

Figure 1 shows the proportions of fixations over time from −200 until 1200 ms after target word onset for both conditions in strongly and weakly supportive discourse contexts. We plotted the two competitors together by taking the average instead of each of them separately because the competitors did not differ significantly from each other. We first analyzed strongly and weakly supportive discourse contexts together to examine whether listeners benefit in general from the presence of the discourse context. In the 200–1000 time window, we found a main effect of discourse information for the reduced targets ($\beta_{\text{discourse information}} = 1.08$,
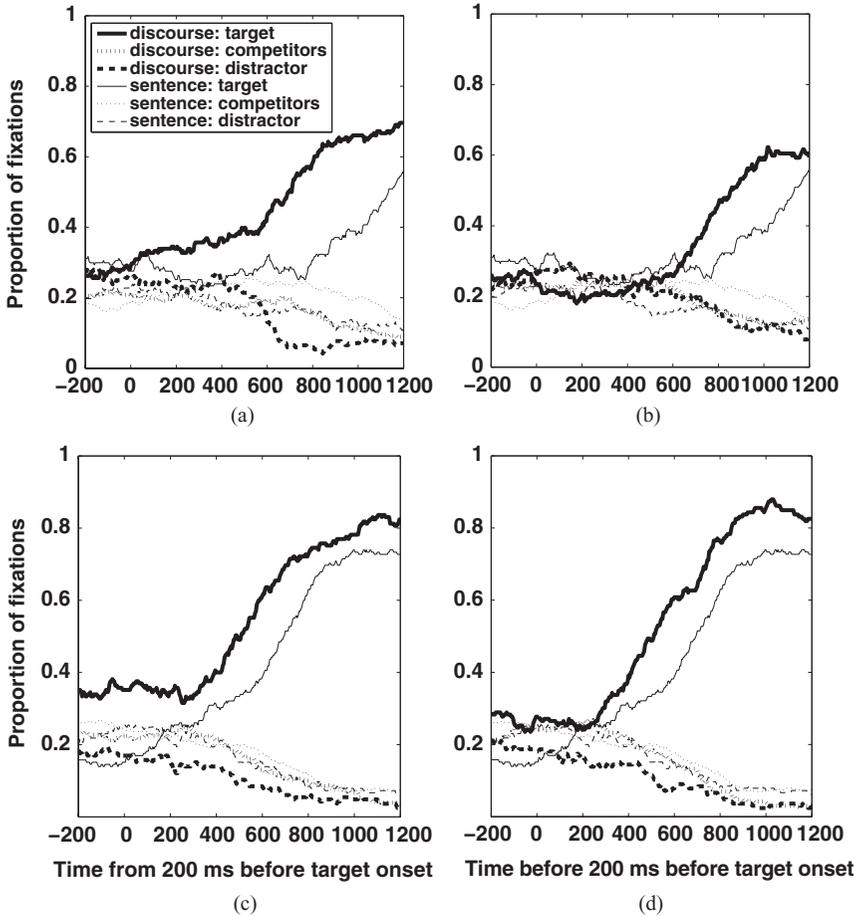
Figure 1. Fixation proportions over time from 200 ms before until 1200 ms after target word onset to targets, averaged competitors, and distractors for the discourse condition (bold lines) and the sentence-only condition (thin lines) in strongly and weakly supportive discourse contexts. (a) Reduced forms in strongly supportive discourse contexts, (b) reduced forms in weakly supportive discourse contexts, (c) canonical forms in strongly supportive discourse contexts, and (d) canonical forms in weakly supportive discourse contexts.

$p_{MCMC} < .0001$) and for the canonical targets ($\beta_{discourse\ information} = 0.99$, $p_{MCMC} = .0001$). The positive betas indicate more looks to both types of targets (reduced and canonical) when the discourse context was present than when it was absent.

Next, we added degree of support as a covariate to the target analysis on the reduced forms. This analysis again showed a main effect of discourse information ($\beta_{discourse\ information} = 1.08$, $p_{MCMC} < .001$), a main effect of the degree of support ($\beta_{degree\ of\ support} = 0.45$, $p_{MCMC} < .01$), and an interaction effect of

Discourse Information $\times$ Degree of Support ($\beta_{\text{Discourse Information}\times\text{Degree of Support}}$ = 0.59, $p_{\text{MCMC}} < .001$). This interaction shows that the ratings, indicating how supportive the discourse contexts were, influenced target fixations only when the discourse contexts were actually presented. This shows there was nothing inherently different between the sentences that happened to occur after strongly versus weakly supportive discourse contexts. The positive beta weight of the interaction shows that the presence of a strongly supportive discourse context (2A) aided word recognition for reduced forms more than the presence of a weakly supportive discourse context (2B).

Degree of support was also added as a covariate to the target analysis on the canonical forms. This analysis showed only a main effect of discourse information ($\beta_{\text{discourse information}}$ = 0.99, $p_{\text{MCMC}} = .0001$). Neither a main effect of degree of support nor an interaction between discourse information and degree of support was found ($p_{\text{MCMC}} > .1$). For the canonical forms, the benefit provided by the presence of a discourse context was therefore independent of how supportive the context actually was (see Figure 1c, d).

An inspection of Figure 1 (i.e., are there more looks to the target already in the baseline period?) suggests that the supportive discourse context allowed listeners to predict the target word in the canonical form condition. Therefore, we tested whether there was increased attention to the target during a time window ranging from 200 ms before to 100 ms after target word onset (i.e., before eye gaze could be influenced by acoustic information from the target word). For the canonical forms, this analysis revealed an interaction effect of Discourse Information $\times$ Degree of Support ($\beta_{\text{Discourse Information}\times\text{Degree of Support}}$ = 0.48, $p_{\text{MCMC}} < .01$). This interaction shows that our participants predicted the target if they heard a supportive discourse context, but that there were no inherent differences in predictability between the items if only the preceding sentence was heard. Given this result, we also tested the same time window for reduced forms, but there was no effect of any predictor on target looks in the early time window ($\beta_{\text{Discourse Information}\times\text{Degree of Support}}$ = -0.08, $p_{\text{MCMC}} < .2$).

Discussion

The results of Experiment 1 revealed opposite patterns for canonical and reduced forms. In an early time window, before information about the target could influence eye gaze, participants used supportive discourse context to predict the upcoming target word in the canonical-target condition but not in the reduced-target condition. This effect of discourse context completely reversed in a later time window, when the target was actually heard. In the later time window strongly supportive discourse contexts helped listeners more to recognize reduced targets than weakly supportive discourse contexts. In contrast, strongly and weakly supportive discourse contexts both improved the recognition of canonical targets to a similar degree.

The latter result suggests that strongly supportive discourse contexts are especially important for the *recognition* of reduced forms. Such a result may not come as a surprise, but it is noteworthy that a similar effect was not observed for canonical targets. A somewhat more surprising result is that strongly supportive

discourse context led to the *prediction* of the target word only in the sentences with the canonical word form. One reason may be that the sentences with the canonical and the reduced targets did not only differ in the amount of reduction on the target but also in the amount of reduction overall (much less reduction overall in the casual-speech sentences with canonical forms; 93% of the phonemes realized) than in the sentences with the reduced forms: 78% of the phonemes realized; $t(27) = 10.86$, $p < .001$. This may have resulted in increased difficulty during the recognition of the words preceding the target word, thereby making prediction of upcoming target words more demanding. Nevertheless, this reduced context still allowed better recognition of the target word. It is important to note that we are not the first to report such a dissociation between recognition and prediction effects. Schiller, Horemans, Ganushchak, and Koester (2009), for instance, found that the task relevance of ambient speech influences the amount of prediction, but not the use of the context for the recognition of a given word. Similarly, in the current study we observed that reduction in the context interferes with prediction but nevertheless *supports* actual word recognition.

Our findings could be taken to suggest that prediction plays not much of a role for actual word recognition. What is prediction during language comprehension for then? Certainly, more research could usefully be directed at this issue. One purpose of prediction for instance may be for turn-taking (see De Ruiter, Mitterer, & Enfield, 2008, for experimental evidence supporting this claim). Furthermore, notwithstanding the striking demonstrations of, for example, Altmann and Kamide (1999) and Van Berkum et al. (2005), there is also recent evidence that prediction may play less of a critical role during *sentence processing* than typically assumed. Huettig, Singh, and Mishra (2010), for instance, found that Hindi speakers with high literacy (15 mean years of formal education) initiated eye movements to target objects well before the onset of the target words (using semantic and syntactic information in the preceding sentence for prediction). Hindi speakers with low literacy (on average 2 years of formal education) initiated eye movements much later, usually after the onset of the target word. These findings suggest that formal literacy supports language-mediated anticipatory mechanisms and converges with evidence from other areas of cognitive psychology that prediction requires relevant expertise (e.g., Aglioti, Cesari, Romani, & Urgesi, 2008; Mori, Ohtani, & Imanaka, 2002). In sum, clearly more theoretical and experimental work needs to be done on the role of prediction in language comprehension.

Before concluding we need to consider another factor that may have influenced the results. The data of Experiment 1 show that the presence of a discourse context was beneficial in all conditions, from the "easiest" condition (canonical forms in supportive contexts) to the most difficult condition (reduced forms in less supportive contexts). As discussed in the Introduction, there is a potential caveat in attributing these benefits to the given discourse information. Previous research has shown that exposure to a speaker's voice is a helpful source in the recognition and adaptation to carefully pronounced canonical forms. Thus, Experiment 2 was conducted to measure to what extent the context effects in Experiment 1 were not in fact effects of speaker adaptation.

In Experiment 2 we presented the same target sentences as in Experiment 1, but now they were preceded by the control contexts from our pretest. These control

contexts provided the voice of the same speaker as the one in the target sentence but no matching discourse information.

EXPERIMENT 2

Method

Participants.   Forty-eight participants from the Max Planck Institute's subject pool, undergraduates at the Radboud University in Nijmegen, were paid to participate in this experiment. They did not participate in the pretest or in Experiment 1. All participants were native speakers of Dutch and had normal or corrected to normal vision. None of the participants reported any (history of) hearing problems.

Materials and procedure.   Experiment 2 used the same target sentences as in Experiment 1, but the target sentences were preceded by the control discourse contexts of the pretest. The control discourse contexts were randomly selected from the corpus, consisted of a minimum duration of 5 s, and contained the same speaker as the one who spoke in the target sentence. Hence, the control discourse contexts provided speaker information but no matching discourse. Participants listened to the target sentence (sentential context only condition) or to the target sentence and the additional "discourse" context (wider "discourse" condition). The procedure was identical to Experiment 1.

Design and analysis.   Reduced targets were presented to half of the participants, whereas canonical targets were presented to the other half of the participants. The analyses were similar to Experiment 1.

Results and discussion

Figure 2 presents the proportions of fixations over time from -200 until 1200 ms after target word onset for both conditions. As in Experiment 1, we plotted the two competitors together by taking the average instead of each of them separately, because the competitors did not differ significantly from each other. Given the absence of an effect of degree of support of the contexts on target fixations (see below), we did not plot the fixation proportions separately for strongly and weakly supportive discourse contexts.

In the 200- to 1000-ms time window we first analyzed whether participants pay more attention to the targets in one of the conditions. We found that listeners looked more often to the reduced targets ($\beta_{\text{'discourse' information}} = 0.91$, $p_{\text{MCMC}} < .001$) and to the canonical targets ($\beta_{\text{'discourse' information}} = 0.91$, $p_{\text{MCMC}} < .001$) when additional speaker information was present than when it was absent. This result demonstrates that hearing more of the same speaker facilitates the recognition of reduced and canonical targets.

Next, degree of support was added as a covariate to the target analysis on the reduced and the canonical forms. We found no significant interaction between speaker information and degree of support for reduced forms
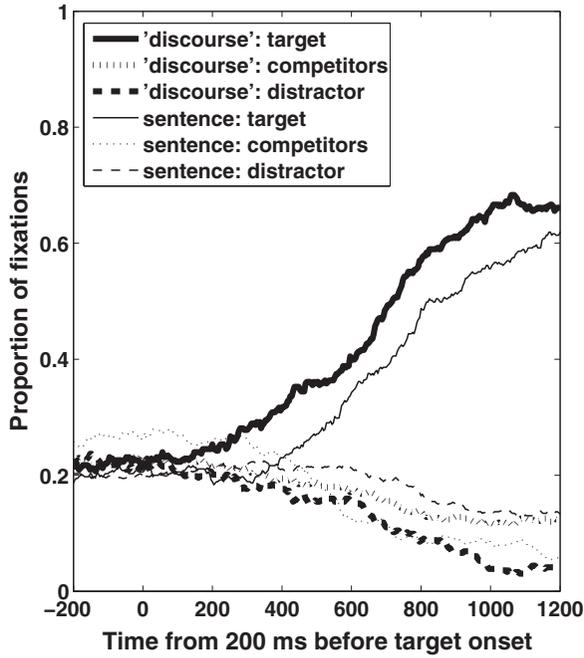
(a)
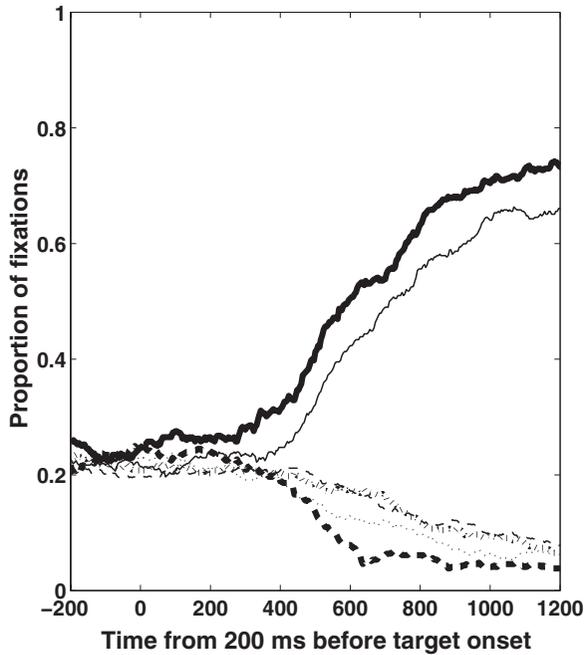
(b)

Figure 2. Fixation proportions over time from 200 ms before until 1200 ms after target word onset to targets, averaged competitors, and distractors for the discourse condition (bold lines) and the sentence-only condition (thin lines). (a) Reduced forms and (b) canonical forms.

($\beta_{\text{'Discourse' Information}\times\text{Degree of Support}} = 0.20$, $p_{\text{MCMC}} > .1$) nor for canonical forms
($\beta_{\text{'Discourse' Information}\times\text{Degree of Support}} = -0.02$, $p_{\text{MCMC}} > .5$).

In sum, the results of Experiment 2 showed that the effects of speaker adaptation were similar for canonical and reduced forms, and, were not modulated by the degree of support provided by the context. This indicates that the benefits for the wider discourse condition consist of two effects: speaker adaptation and discourse information. Experiment 2 shows that speaker adaptation benefits the recognition of reduced and canonical forms to the same degree.

GENERAL DISCUSSION

We investigated the extent to which wider discourse context contributes to the recognition of reduced and canonical forms. Perhaps unsurprisingly, Experiment 1 showed that target recognition of both canonical and reduced forms improved when listeners were exposed to discourse information. This result nevertheless extends the findings of Ernestus et al.'s (2002) study. They found that sentential context helps the recognition of reduced forms; here we have shown that wider discourse information helps even more. Of more importance, however, we observed that strongly supportive contexts help the recognition of reduced forms more than weakly supportive contexts, a pattern that was not observed for canonical forms. For canonical forms, the degree of support by wider discourse context allowed participants to *predict* the target word. However, once there was bottom-up information, the degree of support from the discourse context seized to play any detectable role.

Experiment 2 revealed that the benefits in Experiment 1 are composed of two separate effects: a basic effect of speaker adaptation (which is similar for reduced and canonical forms) and an effect of discourse information (which differentially affects canonical and reduced forms). When comparing the results for canonical forms between Experiment 2 and Experiment 1, it is noteworthy that the magnitude of the effects is similar, about 1 logit unit.

This is a surprising finding: for canonical forms, we observed no benefit at all of wider discourse context on word recognition but a strong beneficial effect of having heard a random excerpt of the same speaker. The fact that the benefit from an arbitrary sample from the same speaker is as large as the benefit from the overall discourse context is in a way a logical continuation of the results of Experiment 1. There, the degree of contextual support did not modify the large recognition benefit provided by hearing this context. Experiment 2 adds to this finding by revealing that really any context can provide the same benefit. This suggests that the benefits for canonical forms seem to be largely due to speaker adaptation. Several types of speaker adaptation effects have been reported in the literature. First, there seem to be rather simple and direct effects of adaptation to a speaker's speaking rate and vowel space (e.g., Ladefoged & Broadbent, 1957; Newman & Sawusch, 1996). Although most of these adaptations should be sufficiently supported by hearing only a few words, Mitterer (2006) showed that vowel normalization is stronger if the sentence contains more information about the vowel space of a given speaker. This indicates that vowel normalization may actually be more efficient with a longer excerpt from a speaker than just one sentence. In a similar vein, Floccia,

Goslin, Girard, and Konopczynski (2006) showed that adaptation to a regional accent requires several sentences. Talker familiarity effects are not restricted to voices with a regional accent though; Nygaard and Pisoni (1998) showed that listeners can benefit from exposure to any talker. Specifically relevant in relation to our current result is their third experiment that found that perceptual learning of novel voices from sentence-length utterances, similar to our discourse contexts, improved speech intelligibility for words in sentences, similar to our measure of target word recognition in a sentence. Part of this benefit may stem from an adaptation to idiosyncratic pronunciations (Mitterer, Chen, & Zhou, 2011; Norris et al., 2003; for a review see Samuel & Kraljic, 2009). This line of experiments showed that listeners can adapt to unfamiliar pronunciations by a speaker on the basis of 10 (and maybe even less) examples of an unusual pronunciation.

Overall, our data pattern reveals an interesting picture of the interplay between prior knowledge, such as the wider discourse context, and bottom-up information in speech perception. With a very clear speech signal, listener can and do predict upcoming words (cf. Altmann & Kamide, 1999; Van Berkum et al., 2005). However, this prediction appears not to influence word recognition once clear bottom-up information from the target word acoustically unfolds. This follows from the finding that the degree of contextual support does not seem to influence looks to the target words when the target word itself is heard. It is important to note that eye gaze during the acoustic lifetime of the target word is likely to reflect the combined contributions of recognition and prediction processes. However, given that the amount of looks is just as high in cases where no prediction is possible (Experiment 2, arbitrary contexts; see also Experiment 1, weakly supporting contexts), it follows that the effect of prediction on target recognition is zero (i.e., from prediction + recognition = recognition, it follows that prediction = 0). Our data thus suggest that the recognition of clearly articulated words is solely based on bottom-up processing. In other words, the contributions of discourse context to word recognition can all be ascribed to speaker adaptation processes which influence this bottom-up processing (Norris et al., 2003).

The situation is reversed for reduced forms. Here, our data indicate that prediction based on the discourse context is not possible (or at least less likely), maybe due to the poor phonetic quality of the input.[1] Discourse context, however, has a strong role to play in facilitating the recognition of the reduced forms. That is, when the bottom-up information is unclear, prior knowledge strongly influences word recognition.

Our data suggest that the presence of reduced forms in weakly supportive contexts increases the likelihood that word recognition will fail. This then offers an explanation of why speakers are more likely to use reduced forms in high predictability contexts than in contexts that are less predictable (e.g., Bell et al. 2003; Jurafsky, Bell, Gregory, & Raymond, 2001; Lieberman, 1963; Lindblom, 1990). Lieberman (1963), for example, showed that words are more carefully pronounced in unpredictable contexts than in predictable contexts (proverbs and adages). Words were generally shorter when they occurred in a highly predictable context than in an unpredictable context. This result is also in line with the so-called probabilistic reduction hypothesis: words are more often reduced when the context is highly predictable (Jurafsky et al., 2001).

Similarly, Lindblom (1990) argues in his hypo- and hyperspeech theory that speakers accommodate to a certain degree to listeners' communicative needs. In the hypo- and hyperspeech theory, speech production is characterized as an acoustic continuum to balance the speaker's aim to be understood and to minimize the speaker's effort by controlling the degree of reduction (hyper- and hypoarticulation) depending on the communicative context. If the listener is able to understand the message, the speaker may produce reduced speech (hypospeech), but if the listener appears to be unable to understand the message, the speaker is forced to use clear speech (hyperspeech). It should be noted, however, that our data do not allow us to conclude that speakers indeed reduce more if the discourse context strongly supports a given word. After all, our pretest did not show that the reduced forms happened to occur in more supportive contexts than the canonical forms. Note that our study was not designed to test whether reduction is more likely if the discourse context is strongly supportive for a given word. Our results thus suggest a need for future research to explore the conditions in which words are likely to be reduced.

In conclusion, the present study has used natural samples from a spontaneous speech corpus to investigate the extent to which wider discourse context helps the online recognition of spoken words. The present data demonstrate the importance of wider discourse context for the successful recognition of reduced forms during casual speech. A strong contextual match with the wider discourse is more important for the recognition of reduced than canonical pronunciations of spoken words in natural, communicative settings. Moreover, our data clarify the respective contributions of prior knowledge from wider discourse context and bottom-up information during spoken word recognition. When listening to canonically produced word forms, the bottom-up signal takes priority and discourse context does not seem to exert a strong influence on recognition. If there is no clear bottom-up signal, however, prior knowledge plays a major supportive role during word recognition.

NOTE

1. It is not difficult to conceive how phonetic reduction can make prediction difficult. Consider the German phrase "in der . . ." (English, "in the . . ."). In clear speech, one can predict that a noun must follow. In reduced speech, however, the phrase is often pronounced as "inner," allowing many other continuations (e.g., "innerhalb," "innerlich," "Innereien").

REFERENCES

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*, 1109–1116.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and visual world. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*, 502–518.

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye-movements and mental representation. *Cognition, 111*, 55–71.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Barr, D. J. (2008). Analyzing "visual world" eye tracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113*, 1001–1024.

Boersma, P. (2001). PRAAT, a system for doing phonetics by computer. *Glot International, 5*, 341–345.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61*, 206–219.

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

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84–107.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Evidence from immediate effects of verb-based constraints. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*, 498–513.

De Ruiter, J. P., Mitterer, H., & Enfield, N. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*, 515–535.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224–238.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus based study of the phonology-phonetics interface*. Utrecht: LOT.

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

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

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1276–1293.

Gaskell, M. G., & Snoeren, N. D. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1632–1647.

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

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163–179.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*, 985–1018.

Huettig, F., & McQueen, J. M. (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, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*, 151–171.

Huettig, F., Singh, N., & Mishra, R. K. (2010). *Language-mediated prediction is modulated by formal literacy.* Paper presented at the Brain, Speech, and Orthography Workshop, October 15–16, Brussels.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th International Symposium* (pp. 29–54). Tokyo: National International Institute for Japanese Language.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). Prediction and thematic information in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*, 37–55.

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 117–127.

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*, 51–81.

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 27*, 98–104.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, N. Warner, & T. Rietveld (Eds.), *Labphon 7* (pp. 637–676). Berlin: Mouton.

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

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

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception and Psychophysics, 53*, 372–380.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661–671.

Mitterer, H. (2006). Is vowel normalization independent of lexical processing? *Phonetica, 63*, 209–229.

Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 496–513.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics, 65*, 956–969.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical tone variation: Evidence from a learning paradigm. *Cognitive Science, 35*, 184–197.

Mori, S., Ohtani, Y., & Imanaka, K. (2002). Reaction times and anticipatory skills of karate athletes. *Human Movement Science, 21*, 213–220.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365–378.

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58*, 540–560.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Norris, D. G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60*, 335–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5*, 42–46.

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

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 309–328.

Pitt, M. 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.

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

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning in speech perception. *Attention, Perception & Psychophysics, 71*, 1207–1218.

Schiller, N. O., Horemans, I., Ganushchak, L., & Koester, D. (2009). Event-related brain potentials during the monitoring of speech errors. *NeuroImage, 44*, 520–530.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language, 27*, 324–340.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension, *Science, 268*, 1632–1634.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967–985.