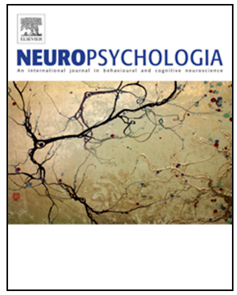
Journal Pre-proof



Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes

Trevor Brothers, Liv J. Hoversten, Shruti Dave, Matthew J. Traxler, Tamara Y. Swaab

PII: S0028-3932(19)30268-4

DOI: https://doi.org/10.1016/j.neuropsychologia.2019.107225

Reference: NSY 107225

To appear in: *Neuropsychologia*

Received Date: 15 May 2019

Revised Date: 5 October 2019

Accepted Date: 8 October 2019

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes

Trevor Brothers[α], Liv J. Hoversten[β], Shruti Dave[γ], Matthew J. Traxler[Ω], & Tamara Y. Swaab[Ω]

[α] Tufts University

[β] Basque Center on Cognition, Brain and Language

[γ] Feinberg School of Medicine, Northwestern University

[Ω] University of California, Davis

Short Title: Flexible Prediction

Address Correspondence to:
Trevor Brothers
Department of Psychology, Tufts University
490 Boston Avenue
Medford, MA 02155
Trevor.brothers@tufts.edu
919-539-4007

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

ABSTRACT

During listening comprehension, the identification of individual words can be strongly influenced by properties of the preceding context. While sentence context can facilitate both behavioral and neural responses, it is unclear whether these effects can be attributed to the pre-activation of lexico-semantic features or the facilitated integration of contextually congruent words. Moreover, little is known about how statistics of the broader language environment, or information about the current speaker, might shape these facilitation effects. In the present study, we measured neural responses to predictable and unpredictable words as participants listened to sentences for comprehension. Critically, we manipulated the *reliability* of each speaker's utterances, such that individual speakers either tended to complete sentences with words that were highly predictable (*reliable speaker*) or with words that were unpredictable but still plausible (*unreliable speaker*). As expected, the amplitude of the N400 was reduced for locally predictable words, but, critically, these context effects were also modulated by speaker identity. Sentences from a reliable speaker showed larger facilitation effects with an earlier onset, suggesting that listeners engaged in enhanced anticipatory processing when a speaker's behavior was more predictable. This finding suggests that listeners can implicitly track the *reliability* of predictive cues in their environment and use these statistics to adaptively regulate predictive processing.

1. Introduction

Human language is highly generative and capable of expressing potentially infinite meanings (Chomsky, 1957). At the same time, language is also highly structured, containing statistical regularities at the semantic and syntactic level that can help listeners efficiently interpret expressions. In everyday speech, we typically hear a new word every 400ms. In this short time-period, we must decode the raw acoustic input into phonemes, link these sounds to a known word, and integrate this word's syntactic form and meaning into an ongoing representation of the sentence. To help reduce this sensory processing burden, listeners can use their prior linguistic knowledge and the constraints of the current sentence context to anticipate upcoming words before they appear in the actual speech signal (Christiansen & Chater, 2016; Elman, 1991).

While probabilistic prediction can provide one potential solution to this processing bottleneck, it also brings a new set of computational challenges – specifically, the problem of variable linguistic environments. While all well-formed English sentences are based on a fundamental set of linguistic rules, speakers can also vary considerably, both acoustically (in how they realize individual speech sounds) and also in their semantic and syntactic preferences. An important question in language comprehension is how listeners are able to detect changes in their linguistic environment and adapt their internal models to continue to predict upcoming input accurately. In the current experiment, we investigated the flexibility of anticipatory mechanisms during natural speech comprehension. Specifically, we investigated if listeners track the successes and failures of predictive processing across different speakers and use this information to adjust the strength of contextually-based predictions.

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

1.1 Prediction – more or less

The effects of contextual predictability on stimulus processing are typically robust and ubiquitous, particularly in online language comprehension (e.g. Altmann & Kamide, 1999; Kutas & Federmeier, 2011; Kuperberg & Jaeger, 2016). For example, words are identified more quickly and accurately when they are presented in a predictable semantic context (*The mother gave her dirty child a* *bath*/*lecture*; Schwanenflugel & Shoben, 1985). This is true both in reading and listening comprehension (Fischler & Bloom, 1979; Ehrlich & Rayner, 1981; Staub, 2015; Traxler & Foss, 2000). Human event-related potential (ERP) studies suggest that comprehenders also show differential neural responses to predictable and unpredictable words. For example, the N400 (a negative-going ERP component peaking 300-500ms after word onset) has been shown to be highly sensitive to contextual predictability, with a roughly linear reduction in N400 amplitude as lexical predictability increases (Kutas & Hillyard, 1984; Wlotko & Federmier, 2012).

While there is clear evidence that comprehenders are able to use a wide range of contextual information to anticipate upcoming meaning, less is known about the flexibility of these anticipatory mechanisms in a changing language environment. If the prediction of upcoming words requires the transmission of information from higher to lower levels of representation, then this process must incur some basic metabolic costs (i.e. increased neural firing, Clark, 2013; Kuperberg and Jaeger, 2016). The costs incurred by active prediction may, in certain environments, outweigh the benefits, particularly when prediction errors are relatively frequent. Based on this logic, it has been hypothesized that – in the same way that comprehenders can adapt to acoustic or phonological variation across speakers (Johnson, 2005;

Liberman, Delattre, & Cooper, 1952) – comprehenders may also adapt to varying levels of prediction error by altering the strength of anticipatory processing across different contexts (Huettig, 2015; Lupyan & Clark, 2015).

Some of the strongest evidence for flexible anticipatory mechanisms in language comprehension comes from single-word semantic priming tasks (den Heyer, Briand, & Dannenbring, 1983; Holcomb, 1988; Lau, Holcomb & Kuperberg, 2013). In these studies, participants show faster response times and smaller N400 amplitudes to words that are preceded by a semantically related prime (doctor- NURSE) relative to an unrelated prime (taco - NURSE). Critically, the magnitude of this semantic priming effect is enhanced when participants encounter a high proportion of related prime-target pairs within an experiment, relative to when this proportion is low (for a thorough review of this *relatedness proportion effect*, see Neely, 1991). The effects of semantic priming found in low proportion blocks are thought to reflect automatic spreading activation among semantically associated words, while the enhanced semantic priming in high proportion blocks is thought to reflect an independent *expectancy mechanism*, whereby participants use primes to generate active predictions about upcoming target words.

Considering that these relatedness proportion effects are mainly observed with long intervals between primes and targets (Posner & Snyder, 1975; Stolz & Neely, 1995), and diminished for participants with low executive function (Hutchison, 2007), it has been claimed that these relatedness proportion effects are too strategic or resource-intensive to influence online sentence comprehension (Seidenberg, Waters, Sanders & Langer, 1984; see also Traxler & Foss, 2000). Some evidence to the contrary was obtained in a recent reading time study

(Brothers, Swaab & Traxler, 2017, Experiment 2), in which the likelihood of encountering a confirmed or disconfirmed lexical prediction was manipulated across sentence contexts. In this experiment, participants in the "high validity" condition read a large proportion of sentences with a highly predictable completion, while participants in the "low validity" condition encountered a high proportion of sentences that violated a strong lexical prediction. Similar to the relatedness proportion paradigm, the authors observed an interaction between word predictability and environmental context, such that the benefits of a highly predictable sentence context were greatly diminished when the global validity of predictive cues was low.

While these results are highly suggestive of flexible anticipatory processing, they must be interpreted with some caution. First, the effects of contextual predictability were assessed using a self-paced reading task in which readers were only able to view one word of a sentence at a time (Just, Carpenter & Woolley, 1982). It is currently unclear whether this pattern of results would also hold in a more natural comprehension task. More importantly, self-paced reading time data provide a relatively coarse estimate of the time-course of this prediction validity effect. To determine whether the effects of prediction validity and lexical predictability were influencing the same stages of lexical processing, it is critical to use a technique with better temporal resolution such as ERPs.

In the current experiment we used a direct measure of neural activity to further investigate the flexibility of language prediction mechanisms during auditory sentence comprehension. In previous studies, it has been shown that listeners are sensitive to speaker-specific variation at multiple levels of linguistic representation, including phonology (Kraljic & Samuel, 2006; Norris, McQueen, & Cutler, 2003), syntax (Chang, Dell, & Bock, 2006; Hanulikova, van Alphen, van Goch, & Weber, 2012; Kamide, 2012), and pragmatics (Grodner & Sedivy, 2011). Moreover, speech carries important identity information (Belin, Fecteau & Bedard, 2004), which listeners can use inform online language comprehension (Lattner & Friederici, 2003; Van Berkum, van den Brink, Tesink & Hagoort, 2008). In this study we manipulated the global validity of predictive cues across two different speakers in order to investigate how and when "speaker reliability" might influence neural responses to contextual constraint during word processing.

In the current study, participants listened to individual sentences spoken by a male and a female speaker while EEG was recorded from the scalp. Some of these sentences contained a predictable or unpredictable critical word (Predictable: *Eric sued the taxi driver and took him to court…* Unpredictable: *Eric picked up his friend and took him to court…*), which we expected to produce differences in the amplitude of the N400 (Kutas & Hillyard, 1984; Federmeier & Kutas, 2011). While holding these critical sentences constant, we used a separate set of highly constraining filler sentences to manipulate the reliability of the two speakers. In these filler sentences, the reliable speaker always used a predictable completion (*The volleyball shot barely made it over the net*), while the unreliable speaker always violated this strong lexical prediction (*The volleyball shot barely made it over the car*).

We tested whether listeners would 1) track the statistical regularities of these two speakers' utterances, and 2) use this information to alter the strength of their lexical predictions across the two speaker contexts. Specifically, we hypothesized that the same set of critical sentences would show larger N400 facilitation effects when they were spoken by a reliable speaker compared to an unreliable speaker.

We should note that the hypothesized interaction between contextual predictability and speaker reliability depends on two theoretical assumptions: 1) that listeners can use semantically constraining sentence contexts to generate implicit lexical predictions for upcoming content, and 2) that this implicit pre-activation can influence the difficulty of lexico-semantic retrieval, as reflected in the amplitude of the N400 (Lau, Phillips & Poeppel, 2008; Kutas & Federmeier, 2011). There are also alternative accounts of N400 context effects, which suggest that this ERP component reflects the *integration* of a word's meaning into the preceding context, which can be influenced by factors such as coherence and plausibility (Osterhout & Holcomb, 1992; Brown & Hagoort, 1993; Hagoort, Baggio & Willems, 2009). Because a speaker's prior history of expected or unexpected utterances should have little influence on the semantic coherence of the current message, a pure integration account would predict no differences in the N400 effect across speakers.

2. Materials and methods

2.1 Participants

40 undergraduate students from UC Davis participated in return for credit towards an experimental participation course requirement. All participants were right-handed, native English speakers with no history of language issues or neurological impairments. The average age of the participants was 19.4 (range 18-28).

2.2 Stimuli

For this study we prepared two sets of sentence materials: 1) a set of critical sentences that manipulated predictability of a specific word, and 2) a set of highly constraining filler sentences that were used to establish the reliability of our two speakers.

For the set of critical sentences, we first selected 128 words with an average length of 5 letters, and an average SUBTLEX-US log per million frequency of 1.35. For each word we generated two sentence contexts: a constraining context that was highly predictive of the critical word (*Eric sued the taxi driver and took him to court*...) and a non-constraining context in which the same critical word was unpredictable but still plausible (*Eric picked up his friend and took him to court*...). The position of the critical word was always matched within each pair (mean = 11 words), and on average, two words prior to the critical word were held constant across the two contexts (range: 1-5). A short continuation was also added to each pair of critical sentences, to ensure that the critical word was always sentence-medial (...*late last week*).

To assess the cloze probability of these critical words, two groups of 60 UC Davis undergraduates performed an offline sentence completion task. In this task, participants read individual sentence frames (*Eric sued the taxi driver and took him to…*) and were asked to fill in the first continuation that came to mind. Each participant only saw one context from each pair. On average, 95% of participants provided the expected critical word in the high-constraint context (range: 79% - 100%) while less than 1% of participants provided this critical word as a completion following the low-constraint context (range: 0% - 2%).

For the filler sentences, we generated a separate set of 202 highly constraining sentence contexts (*At the nude beach, I was reluctant to be seen completely*…). Two completions were generated for each sentence: an expected ending that was highly predictable in context (*naked*) and an alternative ending that was unexpected (*alone*). Critically these unexpected endings were still categorized as plausible, based on the intuitions of the authors. These two endings were matched on word frequency (SUBTLEX-US log per-million: 1.38 vs. 1.28, t = 1.4) and length

(5.0 vs. 5.2 characters, t < 1). In an offline cloze norming task (N = 60), the predictable completions had an average probability of 95% (range: 80% - 100%), while the unpredictable completions were generated less than 1% of the time on average, (range: 0% - 6%). The final set of filler and experimental items contained 660 sentences in total.

Two audio recordings of each sentence were created, one by a male speaker and one by a female speaker. These recordings were conducted in a sound-attenuated booth using a *Blue* condenser microphone (44,100Hz, 16 bit). All sentences were spoken at a normal rate, and the onset and duration of each critical word were determined by visualizing individual speech waveforms. For the set of critical sentences, the average onset of the critical words was 3060ms (SD = 607ms) in the predictable condition and 2872ms (SD = 591) in the unpredictable condition. Critical word durations were matched across the two versions of each sentence (Expected: 489ms, SD = 84; Unexpected: 489ms, SD = 81). For the 202 pairs of filler sentences, the average onset of the sentence-final word was 3164ms, with average word durations of 562ms in the predictable condition (SD = 100ms) and 581ms in the unpredictable condition (SD = 102ms). The full set of sentence materials are available upon request.

2.3 Design

In our main ERP experiment, we employed a 2 x 2 within-subjects design, manipulating word predictability (Predictable vs. Unpredictable) and speaker reliability (Reliable vs. Unreliable). The final stimulus set was counterbalanced across 8 experimental lists manipulating: 1) whether each critical word appeared in a constraining or non-constraining sentence context, 2) whether each filler sentence was spoken by the reliable or unreliable speaker, and 3) whether the male or female speaker was reliable/unreliable. This counterbalancing scheme ensured that no critical words or contexts were repeated within an experimental session. In addition, by counterbalancing the same audio files across participants, we were able to compare ERP responses to identical auditory stimuli across the two speaker contexts. Experimental and filler items were presented in a pseudo-random order, with the requirement that at least one filler sentence appeared between critical sentences.

In the main ERP experiment, each participant heard 330 sentences divided equally across the two speakers (165 each). For the reliable speaker, 133 of these sentences (both filler and critical sentences) contained a highly expected continuation. For the unreliable speaker only 32 of these sentences had a highly expected continuation. This resulted in probabilities of approximately 80% and 20% of encountering a highly predictable word across the two speakers.

2.4 Procedure

During the ERP session, participants were seated in an electrically-shielded, sound-attenuated booth. At the beginning of each trial, a small fixation cross appeared in the center of the computer screen. After a 500ms delay, participants heard a sentence presented binaurally through *Beyer* dynamic headphones. On 25% of trials, participants answered a True/False comprehension question about the preceding sentence by pressing one of two buttons on a keyboard. Participants were told that they would hear a variety of sentences from two different speakers but that their only task was to listen attentively and to answer the comprehension questions to the best of their ability. Participants were not informed of the speaker reliability manipulation.

The experimental session was split into four blocks of equal length, and short rest breaks were provided between each block. Participants heard sentences from one speaker in the first and fourth block, and from the other speaker in the second and third block. The order of presentation of the reliable and unreliable speakers was counterbalanced across participants.

2.4.2 EEG Recording. The electroencephalogram (EEG) was recorded from 29 tin electrodes, embedded in an elastic cap (Electro-Cap International; Eaton, OH). Horizontal and vertical electro-oculograms were additionally recorded to monitor eye movements and blinks. All electrode impedances were kept below 5 kOhm. The EEG signal was amplified using a Synamps Model 8050 Amplifier (band-pass 0.05-100 Hz) and digitally recorded at a sampling rate of 250 Hz. All channels were initially referenced to the right mastoid and later re-referenced to the average of the right and left mastoids.

After EEG recording, independent component analysis (ICA) was used to isolate and remove EEG artifacts due to blinks. The raw EEG data was separated into epochs time-locked to the onset of each critical word of interest, which included a 600ms baseline and 1400ms of activity post-stimulus onset. Individual epochs with remaining EEG artifact due to electrode drift, muscle artifacts, or eye movements were rejected prior to analysis (5% of trials). Finally, event-related potentials (ERPs) were averaged separately according to the critical word's predictability in context (Expected vs. Unexpected) and the reliability of the speaker (Reliable vs. Unreliable).

2.4.3 Analysis Methods. Except where indicated, statistical analyses were performed on mean ERP amplitudes averaged within each condition of interest. We performed two separate ANOVAs to assess the distribution of ERP differences across the scalp: one ANOVA at Midline electrode sites (AFz, Fz, Cz, Pz, POz) with a five-level factor of Anteriority, and one over lateral sites with a three level factor of Anteriority: *Frontal* (FP1/2, F7/8, F3/4), *Central* (FC5/6, FC1/2, C3/4, CP1/2, CP5/6) and *Posterior* (T5/6, P3/4, O1/2), and a two-level factor of Hemisphere (*Left*, *Right*). Significant interactions between topographic factors and experimental effects of interest were followed by pairwise comparisons.

To assess the timing of sentence context effects, we also compared *peak latency* (the time when an ERP effect reaches its maximum) and 25% peak latency (the time when an effect reaches 25% of its peak, Luck, 2014) on by-participant difference waveforms using the jackknife approach (Kiesel, Miller, Jolicoeur & Brisson, 2008). For all analyses with more than 1 degree of freedom, any reported p-values were first adjusted using the Greenhouse-Geisser correction for nonsphericity.

3. Results

3.1 Behavioral Results

On average participants answered 91% of comprehension questions correctly, with no difference in accuracy for sentences spoken by the reliable and unreliable speaker (Reliable: 91.6%, Unreliable: 90.3%, $t(39) = 1.4$, $p = 0.16$). This suggests that participants were listening carefully to the sentence materials and were equally attentive across the two speaker contexts.

3.2 ERP Results: N400 differences of Predictable vs. Unpredictable Fillers

To confirm that participants were generally sensitive to the violation of lexical predictions, we examined ERPs time-locked to the sentence-final word in the filler materials.

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

Recall that these filler sentences were always highly constraining and resolved to an *expected* ending when spoken by the reliable speaker (*The volleyball shot barely made it over the net)* and an *unexpected* ending when spoken by the unreliable speaker (*The volleyball shot barely made it over the car)*. Figure 1 shows grand-average event-related potentials for the two filler conditions.

These grand average ERPs showed a prominent N400 component, followed by a slow positive shift over posterior electrode sites. The amplitude of the N400 was clearly reduced for expected continuations generated by the reliable speaker. Beyond the N400 time-window, we also observed differences in the amplitude of a late, frontally-distributed positivity. Specifically, ERPs to unexpected words were more positive from approximately 800-1200ms post stimulus-onset. This finding is consistent with previous ERP studies of reading comprehension, which have demonstrated a frontally distributed "post-N400" positivity to unexpected but plausible continuations (Federmeier, Wlotko, Ochoa-Dewald & Kutas, 2007).

To confirm these differences statistically, we examined mean ERP amplitudes across the scalp in two time-windows of interest. In the N400 time-window (300-600ms), we observed robust effects of predictability (Midline: $F(1,39) = 40.3$, $p < .001$, Lateral: $F(1,39) = 27.1$, $p < .001$), particularly over posterior electrode sites (Cloze x Anteriority, Midline: $F(4, 156) = 28.2$, $p < .001$, Lateral: $F(2, 78) = 29.1$, $p < .001$). In a cluster of central-parietal electrode sites (*Pz*, *POz*, *CP1/2*, *P3/4*), N400 amplitudes were more negative for unexpected sentence completions (-1.67 μV) than expected completions (0.16 μV), $t(39) = 7.87$, $p < .001$.

In a later time-window (800-1200ms), we observed a significant interaction between predictability and anteriority (Midline: $F(4,156) = 6.12$, $p = .008$). This interaction was driven by more positive ERP amplitudes over frontal electrode sites (*AFz*, *Fz*, *F3/4*) for unexpected final words (-0.24 µV) compared to expected words (-0.84 µV), $t(39) = 3.16$, $p = .003$. While this late frontal component elicited by unexpected words has been well-established in studies of visual reading comprehension (Van Petten & Luka, 2012), to our knowledge, this is the first demonstration of this component in the auditory processing domain.

3.3 Expected vs Unexpected Critical nouns (300-600ms)

Next, we examined ERPs to the core set of critical sentences, which differed both in lexical predictability (*Predictable* vs. *Unpredictable*) and in the reliability of the speaker (*Reliable* vs. *Unreliable*). As can be seen Figure 2, we also observed a clear negative-going N400 component for unpredictable critical words and reduced N400 amplitudes for predictable words, particularly over the back of the scalp. Critically, the magnitude of this N400 context effect was approximately twice as large in sentences spoken by a reliable speaker, compared to the unreliable speaker. In addition to this difference in overall magnitude, the effects of contextual predictability also appeared to have an earlier onset in the reliable speaker context.

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

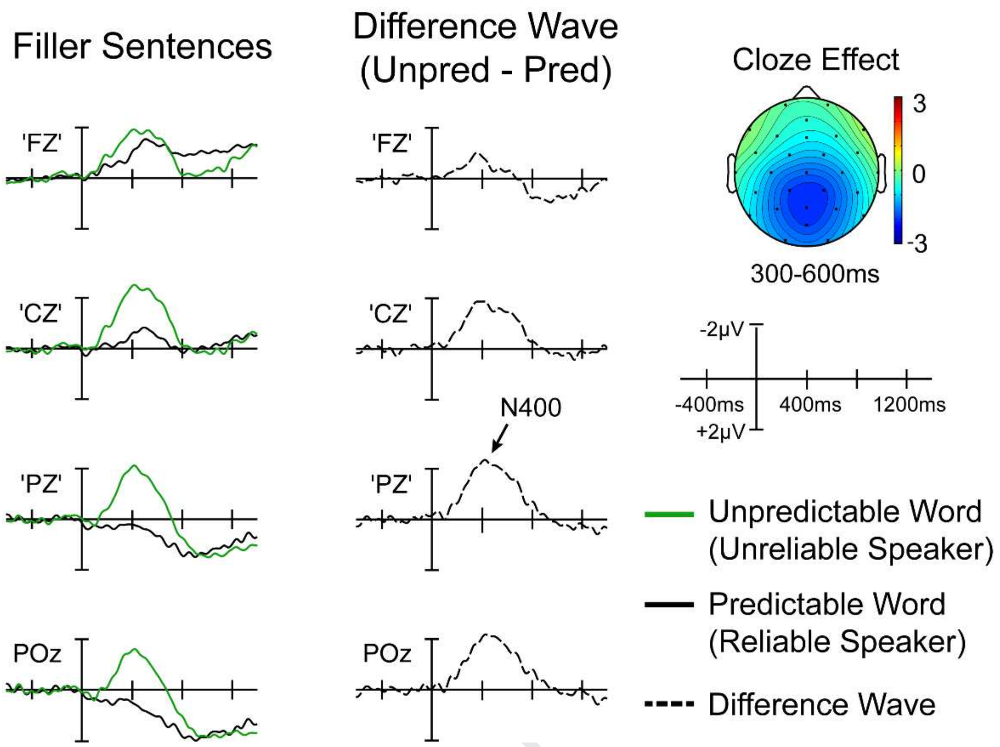


Figure 1. Grand-average event-related potentials, time-locked to final word of the filler sentences (*The volleyball shot barely made it over the car/net*). Waveforms were averaged separately for predictable words spoken by the reliable speaker and unpredictable words spoken by the unreliable speaker. This and all subsequent ERP figures were low-pass filtered at 10Hz for presentation purposes.

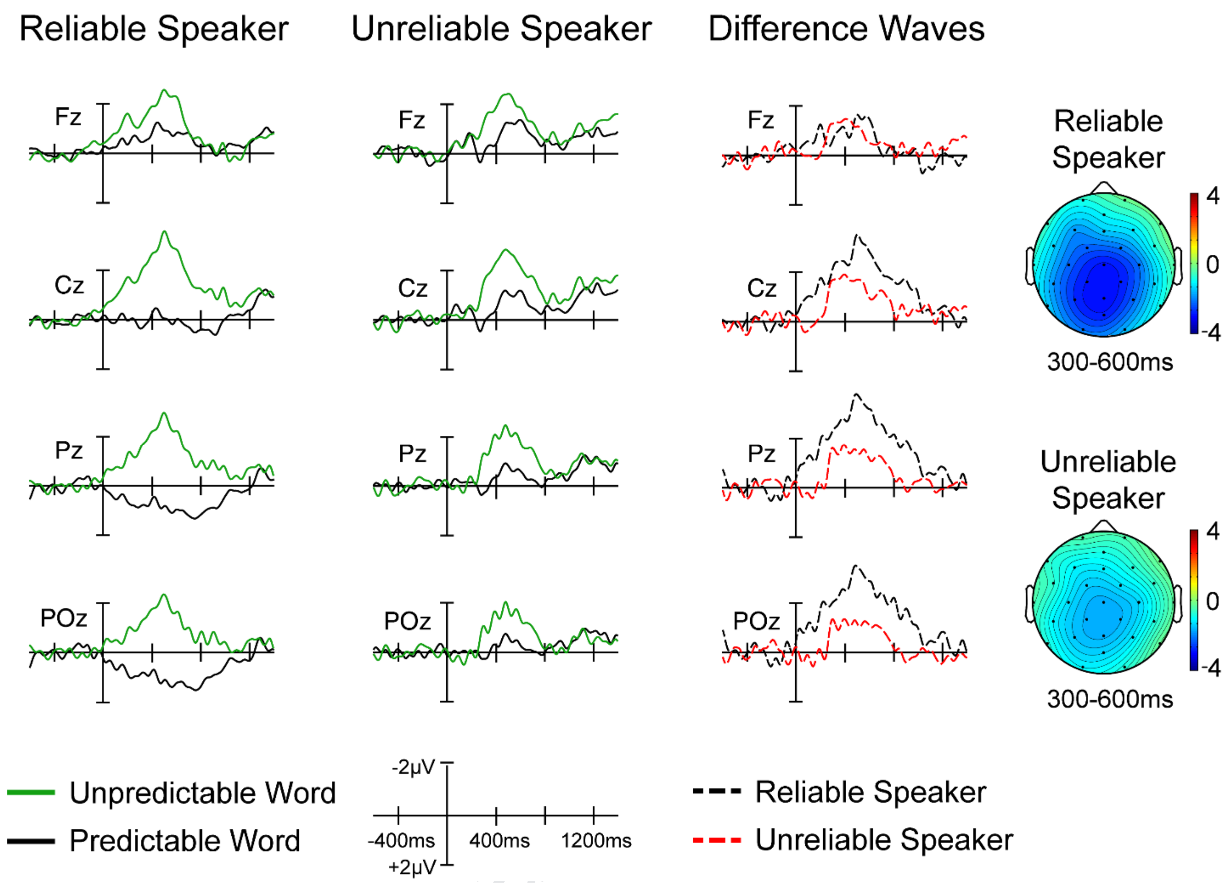


Figure 2. Grand-average event-related potentials, time-locked to the onset of the predictable or unpredictable critical word in the critical sentences (*Eric sued the taxi driver and took him to* court… vs. *Eric picked up his friend and took him to* court…). Waveforms are plotted separately for sentences spoken by the reliable and unreliable speaker (left). Predictability differences waves (Unpredictable minus Predictable) and topographic distributions are also plotted for the two speaker conditions (right). Although these critical sentences were acoustically identical across speaker contexts, participants produced larger and earlier effects of contextual predictability in the reliable speaker condition.

To investigate these differences statistically, we performed a 2 x 2 (*Predictability* x *Speaker Reliability*) ANOVA on mean ERP amplitudes from 300-600ms. In this N400 time-window, we observed robust effects of predictability (Midline: $F(1,39) = 44.9$, $p < .001$, Lateral: $F(1,39) = 34.4$, $p < .001$), particularly over the back of the head (Predictability x Anteriority, Midline: $F(4, 156) = 17.7$, $p < .001$, Lateral: $F(2, 78) = 12.6$, $p < .001$), and no main effects of speaker reliability (*F*s < 1). Critically, we observed a significant interaction between Predictability and Speaker (Midline: $F(1,39) = 4.3$, $p = .046$) that was maximal at central-parietal sites (Predictability x Speaker x Anteriority: Midline: $F(4, 156) = 6.85$, $p = .005$, Lateral: $F(2, 78) = 5.00$, $p = .02$).

In a cluster of central-parietal electrodes, we observed significant effects of predictability for both reliable ($t(39) = -8.35$, $p < .001$, d = 1.32) and unreliable speakers ($t(39) = -3.74$, $p < .001$, d = 0.59). Critically, these effects of sentence constraint were approximately twice as large in the context of a reliable speaker (Reliable N400 effect: 2.67µV, Unreliable N400 effect: 1.33µV, $t(39) = 3.12$, $p = .003$). Hence, local contextual constraint and speaker reliability jointly affected the neurophysiological response to the target words.[1]

To understand the nature of the interaction on the N400, it was important to determine whether the reliable speaker condition resulted in an increase in the *benefits* of predictive context (i.e. stronger word pre-activation), or an increase in the *costs* for encountering an unexpected word (i.e. increased reanalysis). To answer this question, we examined effects of

---

[1] We also examined this critical interaction separately as a function of the gender of the unreliable speaker. The magnitude of the Predictability x Speaker interaction was highly similar in each case (unreliable male speaker: 1.52µV ± 1.2; unreliable female speaker: 1.28µV ± 1.1)

speaker reliability on the N400 within each level of cloze probability (see Figure 3). In this analysis, unpredictable words produced equally large N400s across speaker contexts (Reliable: -2.08 µV, Unreliable: -1.83 µV, $t(39) = 0.79$, $p = .43$), but predictable words showed more N400 facilitation when spoken by a reliable speaker (Reliable: 0.59 µV, Unreliable: -0.50 µV, $t(39) = 3.44$, $p = .001$). This finding suggests that, in the N400 time-window, the effects of speaker reliability were primarily driven by an enhancement in the benefits of informative semantic constraint.[2]

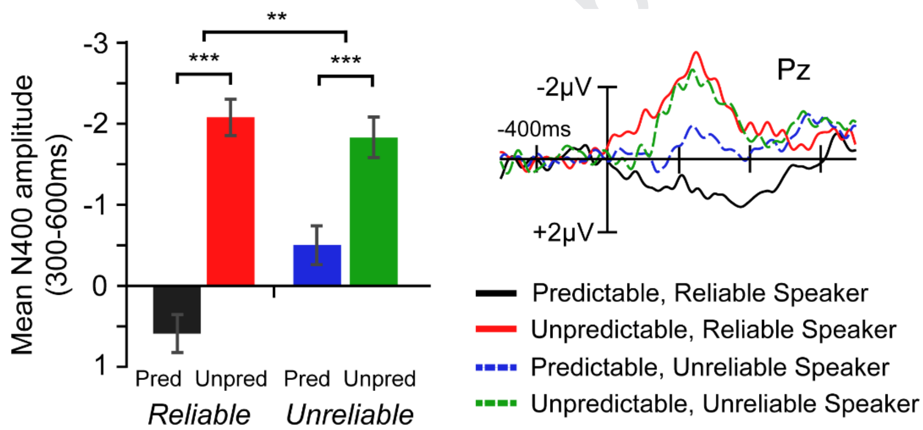


Figure 3. The bar graph shows mean ERP amplitudes (in microvolts) over central-posterior electrode sites (Pz, POz, CP1/2, P3/4) as a function of critical word predictability (Predictable vs. Unpredictable) and speaker reliability (Reliable vs. Unreliable). Error bars represent within-subject standard errors. Note the significant interaction between predictability and speaker reliability, driven by the enhancement in N400 facilitation in the reliable speaker context. The right panel shows grand-average event-related potentials for these conditions, plotted at electrode site Pz. *** $p < .001$, ** $p < .01$

3.4 Onset of the predictability effect

[2] For the set of critical words, we observed no clear differences in the amplitude of the frontal positivity (800-1200ms). Over frontal electrode sites (AFz, Fz, F3/4) we observed no effect of predictability (F < 1), or speaker reliability ($F(1,39 = 2.32$, $p = 0.14$), and no interaction (F < 1).

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

In addition to examining the magnitude of the word predictability effect, we also tested whether contextual constraints had an *earlier* influence on lexical processing in the reliable speaker context. In this time-course analysis, we analyzed mean ERP amplitudes in sequential, 50ms time bins, testing for effects of lexical predictability separately in the two speaker conditions. For the reliable speaker, significant effects of lexical predictability first emerged in the 100-150ms time window. For the unreliable speaker, the same contextual constraints did not affect listener's neural responses until 250-300ms post word-onset (see Table 1).[3]

We also directly compared the peak latency and onset latency of the N400 context effect across speakers.[4] Overall, no significant differences were found in the peak latency of the N400 effect across contexts ($t(39) = -1.08$, $p = .29$), but the 25% peak latency measure showed an earlier onset for the N400 effect in the reliable speaker context (63ms) than the unreliable speaker context (235ms), $t(39) = 2.00$, $p = .05$). Together, these findings suggest that constraining sentence contexts can produce facilitation during the early stages of word identification (100-250ms), but that this early facilitation is only observed when the global utility of lexical prediction is high.

[3]Consistent with this time-course analysis, we observed significant interactions between Predictability and Speaker Reliability in an earlier 100-200ms time-window, Midline: $F(1,39) = 10.6$, $p = .002$, Lateral: $F(1,39) = 8.2$, $p = .007$. Over posterior electrode sites, mean amplitudes in the Reliable-Unpredictable condition were more negative than both the Reliable-Predictable condition, $t(39) = -3.78$, $p < .001$, and the Unreliable-Unpredictable condition, $t(39) = -2.62$, $p = .012$, although only this first comparison remained significant after correction for multiple comparisons.

[4] For this analysis, we first calculated difference waveforms for each participant by performing a point-by-point subtraction between the unpredictable and predictable conditions, separately for the reliable and unreliable speakers (see Figure 2). After applying an 8Hz lowpass filter to each difference wave, we then calculated jackknife sub-averages within each speaker condition to test for differences in the peak or onset latency of the N400 predictability effect.

Table 1. Time-course of the word predictability effect (μV) for reliable and unreliable speakers

Latency	Reliable Speaker			Unreliable Speaker		
Window (ms)	Mean (SD)	t-value	*p*-value	Mean (SD)	t-value	*p*-value
0-50	-0.32 (1.2)	-1.67	*.10*	-0.07 (1.3)	-0.1	*.75*
50-100	-0.49 (1.7)	-1.80	*.08*	0.08 (1.4)	0.34	*.73*
100-150	**-0.56 (1.4)**	**-2.60**	**.01**	0.14 (1.3)	0.67	*.51*
150-200	**-0.64 (1.4)**	**-2.88**	**.006**	0.05 (1.7)	0.18	*.86*
200-250	**-0.96 (1.6)**	**-3.76**	**<.001**	-0.12 (1.6)	-0.48	*.64*
250-300	**-0.73 (1.7)**	**-2.78**	**.008**	**-0.81 (1.4)**	**-3.66**	**<.001**
300-350	**-1.16 (1.4)**	**-5.18**	**<.001**	**-0.96 (1.7)**	**-3.48**	**<.001**
350-400	**-1.25 (1.7)**	**-4.58**	**<.001**	**-1.01 (1.7)**	**-3.77**	**<.001**

Note. *Analyses were conducted on mean ERP amplitudes averaged across the scalp. Latency windows represent milliseconds following the onset of the predictable/unpredictable critical word.*

4. General Discussion

In the present study, we examined how the global reliability of predictive cues would influence the use of semantic context during natural listening comprehension. Participants listened to a series of sentences produced by a male or female speaker while brain activity was recorded at the scalp. Throughout the experiment, one of the speakers was *reliable* and would continue high constraint sentence contexts with a highly predictable word (*The dairy cow produced a lot of milk*). The other speaker was *unreliable* and would routinely violate the listeners' contextual expectations (*The dairy cow produced a lot of noise*). Our goal was to determine whether listeners would detect differences in the reliability of predictive cues across speakers and use this information to modulate their online comprehension of the two speakers' utterances.[5]

As expected, violations of predictive constraints (*milk* vs. *noise*) produced clear differences in the amplitude of the N400 across speakers, as well as a late positivity over frontal electrodes sites (DeLong, Quante, & Kutas, 2014; Brothers et al, 2015). More importantly, these confirmed/disconfirmed lexical predictions also influenced the relative strength of predictive processing across speakers. In a separate set of critical sentences, we observed an interaction between speaker reliability and lexical predictability on the N400, with larger effects of contextual constraint when the speaker was reliable. We believe that these results have important implications for the role of prediction error in anticipatory processing and the nature of the N400, as well as more general implications for understanding the locus of sentence context effects during lexical processing.

4.2 Active Prediction or Facilitated Integration

In the language comprehension literature, there has been a long-standing debate over the locus of contextual facilitation effects during word recognition (Marslen-Wilson, 1987; Zwitserlood, 1989; Lau, Phillips & Poeppel, 2008). At a basic level, researchers disagree over whether these effects are predictive or integrative in nature – i.e. whether sentential

[5] Previous ERP studies investigating the role of speaker identity have often focused on so-called "semantic-pragmatic mismatches", in which pragmatic information about the identity of the speaker conflicts with the meaning of their utterance; for example, a young child saying "*Every evening I drink some wine…*" (Lattner & Friederici, 2003). In the present study we utilized speaker identity in a very different way to examine the flexibility of predictive processing across speaker contexts.

constraints exert their influence before or after a word has been encountered in the bottom-up input (Brothers et al., 2015; Elman, 1991; Foss & Ross, 1983; Hess, Foss, & Carroll, 1995; Schwanenflugel & Shoben, 1983; van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). According to *prediction accounts*, lexical processing is facilitated in congruent contexts because listeners use context-based semantic and syntactic constraints to pre-activate features of upcoming words. When lexico-semantic features of a word have been pre-activated, the time and neural effort required for lexical access is reduced. While different predictive processing accounts make different assumptions about the exact nature of this pre-activation (parallel or serial, lexical or semantic), all of these accounts agree that contextual word probability is one of the primary determinants of the amount of bottom-up "work" necessary during lexical identification (see Schwanenflugel & LaCount, 1988 for an early version of this account; and Kuperberg & Jaeger, 2016 for a recent review).

Opposed to this view are *facilitated integration accounts*, which claim that the benefits of congruent sentence contexts do *not* involve the prediction of upcoming content (Foss, 1982, Hess et al., 1995; Morris, 1994). Instead, these accounts claim that contextually supported words are easier to process because they can be integrated more easily into the listener's ongoing representation of the discourse. According to these integration accounts, the early stages of lexical access proceed entirely automatically, without any influence from broader contextual constraints (Forster, 1981; Marslen-Wilson, 1987; Swinney, 1979). It is only once a word or a distributed cohort of words has been identified that context can exert an effect (Marslen-Wilson & Zwitserlood, 1989; Zwitserlood, 1989; see Ferreira & Chantavarin, 2018 for a recent review).

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

It is difficult to see how integration difficulty alone could account for the present pattern of ERP results. In the current paradigm, we compared N400 responses to the same sentences while manipulating the global validity of predictive cues across speakers. Although these sentences were identical across speakers in terms of plausibility and semantic content, we nonetheless observed greater N400 facilitation for predictable words in a global context that was more conducive to lexical predictions. While it is difficult to imagine integration processes unfolding differently across speakers for the same set of simple congruent sentences (e.g. *The web was spun by the large spider…*), one can easily explain this pattern of ERP results by assuming that listeners were engaged in lexical prediction to different degrees in environments with differing levels of prediction error.

4.3 Flexible models of anticipatory processing

In our view, the boost in N400 facilitation that we observed in the reliable speaker condition can only be accounted for by a particular subset of predictive processing models. First, normal sentence comprehension must recruit an *active expectancy mechanism* that produces feed-forward activation for relevant lexico-semantic features that have not yet appeared in the bottom-up input (Brothers et al., 2015; Schwanenflugel & Shoben, 1983; Schwanenflugel & LaCount, 1988). Secondly, this expectancy mechanism must show dynamic sensitivity to prediction error beyond the level of the individual sentence. This would involve implicitly tracking the success or failure of anticipatory processing over time and using this information to regulate the strength of top-down anticipation in the future.

According to this account, in a language environment in which prediction error is high, the comprehension system will engage in less anticipatory processing and instead devote more resources to bottom-up stimulus evaluation. Conversely, in environments in which contextual predictions had been particularly successful, the comprehender will rely more strongly on top-down predictions, resulting in faster and more efficient word recognition when these predictions are correct. While many language theorists have proposed this type of flexible architecture for predictive processing (Huettig, 2015; Lupyan & Clark, 2015; Kuperberg & Jaeger, 2016), we believe that the current experiment provides the most direct evidence for adaptive prediction mechanisms in the context of natural sentence comprehension (for a discussion of adaptive expectancy mechanisms in the visual system, see Summerfield & Lange, 2014; for a recent Bayesian formalization in semantic priming, see Delaney-Busch, Morgan, Lau & Kuperberg, 2019).

4.4 Time-course of the speaker reliability effects

The early time-course of these ERP differences places important constraints on the neurocognitive mechanisms underlying the speaker reliability effect. Previous ERP studies have demonstrated very early effects of lexical predictability on ERP responses in natural, connected speech, with effects sometimes appearing within 50 to 100ms after word onset (Holcomb & Neville, 1991; Romero-Rivas, Martin & Costa, 2016). This is especially surprising considering that spoken word input unfolds gradually over time. Within 100ms, participants have heard only 1-2 phonemes of the critical word, which is generally insufficient to identify any single lexical candidate from the bottom-up acoustic input (Van Petten et al., 1999; van den Brink, Brown & Hagoort, 2006).

What then should we make of the early interactions between lexical predictability and speaker reliability in the present paradigm? At a basic level, this finding is inconsistent with the claim that expectancy effects can only influence late, post-lexical processing stages (Seidenberg, Waters, Sanders & Langer, 1984). Instead these findings suggest that the anticipatory mechanisms - which were preferentially active in the reliable speaker context - were operating at a lexical, or even pre-lexical processing stage (Staub, 2015). It may be the case that early ERP differences of contextual constraint (100-250ms) reflect facilitation during the early perceptual decoding of a word, as reflected in modulations of ERP components such as the N2 or N250 (Lau, Holcomb & Kuperberg, 2013; Brothers, Swaab & Traxler, 2015; Ng, Payne, Stine-Morrow & Federmeier, 2018, although see Nieuwland, 2019). In the present study, we did not see clear differences in scalp topography for facilitation effects occurring in early (100-250ms) and late (300-600ms) time-windows. Nonetheless, the clear temporal dissociation observed for the reliable and unreliable speaker contexts is consistent with the idea that form level prediction is more strongly engaged in environments where the utility of predictive cues is high.

4.5 Do prediction violations drive adaptation?

In the current experiment, we observed speaker-specific adaptation following the violation/confirmation of strong lexical predictions. While it is clear that readers respond differently to predictable and unpredictable words, there is an ongoing debate whether readers are particularly sensitive to the *violation* of high certainty predictions during natural language comprehension (Luke & Christianson, 2016). There is some ERP evidence suggesting that readers show differential neural responses to prediction violations (*The pilot was flying the kite*) compared to unexpected words within non-constraining contexts (*The old man saw the kite*) (Federmeier, et al., 2007, but see Thornhill & van Petten, 2012). In contrast, there is evidence from reading time studies suggesting that prediction violations do *not* produce robust differences in online reading behavior (Luke & Christianson, 2016; Frisson, Harvey & Staub, 2017; although see Payne & Federmeier, 2017).

In the current ERP study, it is unclear whether adaptation across speaker contexts was driven by overall differences in lexical predictability, or specifically by the presence of prediction violations in the unreliable speaker context (*The dairy cow produced a lot of noise*). Across participants, we observed no correlation between the amplitude of the frontal positivity in expected vs. unexpected filler sentences and the size of the prediction validity effect on the N400 ($r(39) = 0.07$, $p = .67$). Nonetheless, to more closely examine the relationship between adaptation and prediction error, it may be informative to replace prediction violations in the unreliable speaker context with neutral, non-constraining sentences (e.g. *None of the workers had noticed the noise*). This experiment would provide a clearer answer to whether disconfirmed predictions *per se* are necessary for driving adaptation across speaker contexts.

4.6 Residual Priming for the Unreliable Speaker

Finally, we believe it is important to address some small differences between the present set of results and the results of prior, related work (Brothers, Swaab & Traxler, 2017). In this prior reading-time study, the proportion of prediction violations across environments also interacted with the effect of lexical predictability. Critically, participants in the lowest validity group (12.5%) showed no significant reading time differences between predictable and unpredictable critical words. This result differs from the present experiment, which showed diminished but still highly significant effects of lexical predictability on the N400 in the unreliable speaker context.

It is unclear what factors might have contributed to this difference. It is possible that differences in validity proportions (12.5% vs. 20% validity) or the use of a within-subjects design may have weakened the prediction validity manipulation in the current experiment. It is also possible that the two dependent measures – behavioral reading times and N400 amplitudes – are differentially sensitive to effects of constraining sentence contexts. For example, while the N400 and behavioral reading times are both sensitive to message-level predictability, the N400 may be more sensitive to low-level semantic overlap between a critical word and its preceding context (e.g. *book* – *library*, Federmeier, Van Petten, Schwartz & Kutas, 2003). This factor might explain the residual N400 facilitation observed in the unreliable speaker context, despite the low overall validity of predictive cues.

6. Conclusion

In summary, the present ERP findings suggest that the effects of confirmed and disconfirmed predictions have real consequences during auditory sentence comprehension, even in the absence of artificial task demands or slow, word-by-word presentation rates. These results provide evidence that comprehenders generate implicit predictions for upcoming content during regular comprehension and that the accuracy (or inaccuracy) of these predictions has important downstream consequences for how we predict in the future.

Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes.

REFERENCES

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

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2017). Goals and strategies influence lexical prediction during sentence comprehension. *Journal of Memory and Language*, *93*, 203–216.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34–44.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*, 234–272.

Chomsky, N. (1957). Syntactic structures. Oxford, England: Mouton.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*, 181–204.

Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. R. (2019). Neural evidence for bayesian trial-by-trial adaptation on the N400 during semantic priming. *Cognition*, *187*, 10–20.

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162.

den Heyer, K., Briand, K., & Dannenbring, G. L. (1983). Strategic factors in a lexical-decision task: Evidence for automatic and attention-driven processes. *Memory & Cognition*, *11*, 374–381.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195–225.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychology and Aging*, *18*, 858-872.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, *1146*, 75–84.

Ferreira, F., & Chantavarin, S. (2018). Integration and prediction in language processing: a synthesis of old and new. *Current Directions in Psychological Science*, *27*, 443–448.

Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior*, *18*, 1–20.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*, 465–495.

Foss, D. J., & Ross, J. R. (1983). Great expectations: Context effects during sentence processing. *The Process of Language Understanding*, 169–191.

Frisson, S., Harvey, D. R., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language, 95*, 200–214.

Grodner, D., & Sedivy, J. C. (2011). 10 The Effect of Speaker-Specific Information on Pragmatic Inferences. In: *The processing and acquisition of reference*, vol. 2327, eds. E. Gibson & N. Pearlmutter, pp. 239–272. MIT Press .

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*, 878–887.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*, 62–82.

Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and Language*, *35*, 66–85.

Holcomb, P. J., & Neville, H. J. (1991). Natural speech processing: An analysis using event-related brain potentials. *Psychobiology*, *19*, 286–300.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, *1626*, 118–135.

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 645–662.

Johnson, K (2005). Speaker normalization in speech perception. In D.B. Pisoni & R.E. Remez (Eds.), *The Handbook of Speech Perception*. pp. (363-389). Malden, MA: Blackwell.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*, 66–71.

Kiesel, A., Miller, J., Jolicø eur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*, 250–274.

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

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*, 32–59.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

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

Lattner, S., & Friederici, A. D. (2003). Talker's voice and gender stereotype in human auditory sentence processing–evidence from event-related brain potentials. *Neuroscience Letters*, *339*, 191–194.

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*, 484–502.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*, 920-933.

Liberman, A. M., Delattre, P., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *The American Journal of Psychology*, 497–516.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60.

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*, 279–284.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576-585.

Neely, J. H. (2012). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In *Basic processes in reading* (pp. 272–344). Routledge.

Nieuwland, M. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, *96*, 367-400.

Ng, S., Payne, B. R., Stine-Morrow, E. A., & Federmeier, K. D. (2018). How struggling adult readers use contextual information when comprehending speech: Evidence from event-related potentials. *International Journal of Psychophysiology*, *125*, 1-9.

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

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Payne, B. R., & Federmeier, K. D. (2017). Pace yourself: Intraindividual variability in context use revealed by self-paced event-related brain potentials. *Journal of Cognitive Neuroscience*, *29*, 837–854.

Posner, M. I., Snyder, C. R., & Solso, R. (2004). Attention and cognitive control. *Cognitive Psychology: Key Readings, 205*.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245–255.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 344-354.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 82–102.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232–252.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre-and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, *12*, 315–328.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, *9*, 311–327.

Stolz, J. A., & Neely, J. H. (1995). When target degradation does and does not enhance semantic context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 596–611.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*, 745–756.

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

Traxler, M. J., & Foss, D. J. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1266–1282.

Van Berkum, J. J., 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 Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 364–372.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 394-417.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*, 770–785.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.

Highlights

- Participants listened to sentences for comprehension
- The global validity of predictive cues was manipulated across speakers
- N400 sentence context effects were enhanced in “reliable” speakers
- Only reliable speakers showed early context effects, 100-200ms after word onset
- Prediction mechanisms are flexible and context-dependent