# Is second best good enough? An EEG study on the effects of word expectancy in sentence comprehension

Sofia Frade[a,b], Ana P. Pinheiro[a], Andrea Santi[c], Ana Raposo[a]

[a]*Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal*

[b]*Escola de Ciências Sociais e Humanas, Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal*

[c]*Research Department of Linguistics, University College London, London, UK*

Corresponding author at: Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal.

E-mail address: alraposo@psicologia.ulisboa.pt (A. Raposo).

# Is second best good enough? An EEG study on the effects of word expectancy in sentence comprehension

Sentence comprehension can be facilitated when readers anticipate the upcoming word. Notwithstanding, it remains uncertain if only the most expected word is anticipated, as postulated by the serial graded hypothesis, or if all probable words are pre-activated, as proposed by the parallel probabilistic hypothesis. To test these contrasting accounts, we compared the processing of expected and unexpected words with second-best words, i.e., the second most expected word in a sentence. The results, from 30 participants, revealed a graded facilitation effect for the expected words, indexed by the N400 mean amplitude, which was the least negative for the most expected words, intermediate for second-best words, and most negative for unexpected words. The Post-N400 Positivity analysis did not reveal any significant effects. The facilitation effect found for the most expected and second-best words suggests that readers can pre-activate multiple candidates during sentence comprehension.

**Key-words**: Lexical Prediction, N400, Event-Related Potentials, Sentence Comprehension

## 1. Introduction

When hearing a joke, we are often surprised by the punch-line, whereas when listening to a story we can on many occasions guess what is coming next. These examples show that during language comprehension the initial sentence information guides the processing of upcoming verbal information. In some situations it can mislead individuals, such as in jokes or semantic illusions (Coulson & Kutas, 2001; Raposo & Marques, 2013), yet in most cases it facilitates comprehension. A robust body of evidence reports facilitated processing of expected (e.g., "She cleaned the dirt from her SHOES") relative to unexpected words, i.e., words that provide a syntactically and semantically well-formed completion to the sentence but are unlikely to occur (e.g., "She cleaned the dirt from her BOAT"). Notably, naming is faster (e.g., Duffy, Henderson, & Morris, 1989; Hintz, Meyer, & Huettig, 2016; Simpson, Peterson, Casteel, & Burgess, 1989), reading time is shorter (e.g., Hintz, Meyer, & Huettig, 2016; Matsuki et al., 2011; Ng, Payne, Steen, Stine-Morrow, & Federmeier, 2017), and response times are quicker (e.g., Fischler & Bloom, 1979; Forster, 1981; Schwanenflugel & Shoben, 1985) for expected than unexpected words. Extensive event-related brain potential (ERP) research have also demonstrated that when sentence contexts are strongly constraining, the processing of expected target words is facilitated, which is reflected in a decreased N400 amplitude when compared to unexpected or invalid words (e.g., Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Kutas, 1993; van Petten, 1993; van Petten, Rubin, Parks, Plante, & Coulson, 2000). The N400 is a centro-parietal negativity peaking between 300 and 500 ms after word onset and is sensitive to the relationship of that word to its preceding context (e.g., Federmeier, McLennan, Ochoa, & Kutas, 2002; Kutas & Hillyard, 1980; Pinheiro et al., 2013; Van Petten & Kutas, 1990). This facilitation effect has been argued to reflect predictive mechanisms occurring before word onset and that optimize sentence comprehension and cognitive resource allocation (Huettig & Mani, 2016; Kuperberg & Jaeger, 2016, for a review). Notwithstanding, how these anticipatory processes are implemented is still a matter of debate.

Consistent evidence has supported a role for predictive mechanisms during language processing. Several studies revealed that, besides the facilitated processing of expected words, unexpected words that are semantically related to the expected candidate elicit a reduced (i.e., less negative) N400 compared with unrelated unexpected words (DeLong, Chan, & Kutas, 2019; Federmeier & Kutas, 1999; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Pinheiro et al., 2013; Thornhill & Van Petten, 2012). For example, Federmeier and Kutas (1999) manipulated the semantic relationship between the unexpected and expected words in sentences such as "They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of PALMS/PINES/TULIPS." The results revealed a reduced N400 amplitude for unexpected words that belong to the same category of the expected words [e.g., PINES - PALMS (trees)] compared with unexpected words that belong to a different category [e.g., TULIPS - PALMS (flowers vs. trees)]. Similar findings were reported for unexpected words that share some orthographic and phonologic features with the expected words (e.g., HOOK vs. BOOK) relative to unexpected words with no overlap (e.g., SOFA vs. BOOK; DeLong et al., 2019; Ito et al., 2016; Laszlo & Federmeier, 2009). These findings may be explained by the overlap between pre-activated features of the expected word and the actual features of the unexpected, yet similar, word (Federmeier & Kutas, 1999) or by spreading of activation from the predicted word to related words, leading to a facilitated processing of the latter (Forster, 1979). It should be noted that these facilitation effects may also reveal the eased integration of expected information (e.g., Ferreira & Chantavarin, 2018; Hagoort, 2005; Lau, Phillips, & Poeppel, 2008). Thereby, in addition to predictive mechanisms, lexical-semantic integration processes may also be at play.

To address the question of prediction directly, DeLong and collaborators (2005) examined the processing of indefinite articles - “a” or “an” - when the subsequent expected noun started with either a vowel or a consonant. The authors found more negative N400 amplitudes in response to indefinite articles that mismatched the expected upcoming noun (e.g., “The day was breezy, so the boy went outside to fly an…” where the word “kite” is the most expected word). Yet, a large-scale replication study only found a consistent effect for the noun and failed to find a consistent facilitation effect for the article, which suggests that pre-activation may occur at the semantic but not phonological level (Nieuwland et al., 2018). Additional studies probed facilitation in advance of the predicted noun at the processing of gender-marked articles, gender-marked adjectives, and classifiers (e.g., Fjaellingsdal et al., 2020; Kwon, Sturt, & Liu, 2017; Martin et al., 2013; Szewczyk & Schriefers, 2013; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004). These studies also reported that the N400 amplitude varies as a function of the congruency between the expected target word and its preceding words. These findings support predictive mechanisms in sentence comprehension, as they demonstrate that readers anticipate critical features of the expected word (e.g., gender) before it is presented. An exclusive integration account, which attributes the facilitated processing of expected words to the ease of lexical-semantic integration of those words (Hagoort, 2005; Zhu et al., 2012), cannot accommodate these facilitation effects.

Recent studies have additionally revealed that when predictions are violated, a frontal late positivity is observed after the N400, occurring around 600 to 900ms after word onset (e.g., Brothers, Wlotko, Warnke, Kuperberg, & Watkins, 2020; DeLong & Kutas, 2020; DeLong, Quante, & Kutas, 2014; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Van Petten & Luka, 2012). This ERP component, known as Post-N400 Positivity or PNP, is typically elicited in response to unexpected words appearing in high constraint sentences (e.g., “He bought her a pearl necklace for her COLLECTION”, instead of the most expected word “BIRTHDAY”). This component is thought to reflect additional operations that take place when strong predictions are violated, such as suppression or inhibition of the predicted word (Federmeier et al., 2007; Van Petten & Luka, 2012). An alternative explanation is that the PNP is sensitive to word integration difficulty and reflects the assimilation of new unexpected information into a higher-level representation of sentence meaning (Brothers, Swaab, & Traxler, 2015; DeLong et al., 2014). Notwithstanding, in contrast to the N400, much less is known about the functional significance of the PNP and about the conditions that elicit it. For instance, some studies have documented a PNP effect in response to weakly or moderately constraining sentence contexts (Brothers et al., 2015), whereas others have reported this effect only in response to strongly constraining sentences (Federmeier et al., 2007). Additionally, some studies have reported a PNP effect only for unexpected words (DeLong & Kutas, 2020; Federmeier et al., 2007), whereas others have shown an enhanced PNP in response not only to unexpected words but also to weakly expected words (Ng et al., 2017; Thornhill & Van Petten, 2012).

Current psycholinguistic theories emphasize the importance of prediction as a mechanism to facilitate sentence comprehension, yet how these predictive mechanisms unfold during language processing is still elusive. Some authors have proposed that prediction occurs by a serial graded process, i.e., readers initially predict the most expected word and only when this prediction is violated the system can update the predictions for other probable candidates (Thornhill & Van Petten, 2012). The findings that support the specificity of prediction, such as the facilitation effects for gender-marked articles and adjectives, point towards a highly specific pre-activation (Martin et al., 2013; Szewczyk & Schriefers, 2013; Van Berkum et al, 2005; Wicha et al., 2004). For example, in the sentence context “As it is rainy it is better to go out with” the most expected noun phrase is “an umbrella” (example from Ito, Martin, &

Nieuwland, 2016). Yet, other words related to that sentence context could also be expected to some extent, such as “raincoat”, “parka”, or “wool cap”. Studies showing a reduced N400 amplitude only in response to the article that is congruent with the most expected word suggest that other probable words, which have a weaker expectancy, are not anticipated, or at least are not anticipated to a considerable extent. Alternatively, it has been postulated that prediction is a probabilistic parallel process in which multiple possibilities are considered at the same time, i.e., readers compute and pre-activate all probable candidates at any given time. The level of activation of each of these candidates will reflect their degree of expectancy in a specific sentence context (e.g., DeLong, Urbach, & Kutas, 2005). The degree of expectancy of each final word in a sentence completion task is frequently operationalized as a word’s cloze probability (e.g., Bloom & Fischler, 1980; Pinheiro, Soares, Comesaña, Niznikiewicz, & Gonçalves, 2010; Taylor, 1953). The higher the proportion a given word is used to complete a sentence fragment, the greater the expectancy of that word. Alternatively, two additional measures have been used to assess the fitness of a word in a given task: the word’s *surprisal* – the negative log probability of a word given its preceding context (Frank, Otten, Galli, & Vigliocco, 2015; Smith & Levy, 2013; Willems, Frank, Nijhof, Hagoort, & Van Den Bosch, 2016) and *entropy* – the distribution of next-word probabilities (Frank et al., 2015; Willems et al., 2016). In the present study, the word’s expectancy was defined on the basis of cloze probability, since it is the measure used in most studies testing predictive mechanisms in language comprehension (e.g., DeLong et al, 2005, Kutas & Hillyard, 1984; Federmeier et al., 2007; Thornhill & Van Petten, 2012). The probabilistic parallel hypothesis has been grounded on data showing that a word’s expectancy modulates the magnitude of the facilitation effect. In particular, previous studies have demonstrated a graded facilitation effect for expected words according to their level of expectancy, measured by their cloze probability, as the amplitude of N400 was enhanced for expected words with lower relative to higher cloze probability (Federmeier et al., 2007; Kutas & Hillyard, 1984).

The present study aims to clarify how the predictive mechanisms unfold during sentence comprehension. More specifically, it directly compares the serial graded hypothesis, which postulates that only the most expected word is initially anticipated, with the probabilistic parallel hypothesis that proposes that all probable word candidates are pre-activated. To test these two alternative hypotheses, we probed not only the processing of expected and unexpected words, as it has been done in previous studies (e.g., Federmeier et al., 2007; Pinheiro et al., 2013; Thornhill & Van Petten, 2012), but also the processing of the second-best candidate, i.e., the second most expected word in a sentence. For instance, in the sentence “The dog spent the afternoon chewing the”, the most expected word is “bones”, the second most expected word (i.e., second-best) is “shoes”, whereas the word “glasses” is an unexpected word. Critically, the three words are plausible completions of the sentence but are associated with different expectancy levels (Expected > Second-Best > Unexpected). The most expected word is overall moderately expected (as the sentence is not strongly constrained, otherwise there would not be more than one expected candidate), whereas the second-best word is weakly expected. Previous studies demonstrated that N400 amplitude modulations index word expectancy, hence moderately and weakly expected words elicit a larger N400 amplitude relative to highly expected words (DeLong et al., 2005; Federmeier et al., 2007; Kutas & Hillyard, 1984; Thornhill & Van Petten, 2012; Wlotko & Federmeier, 2012). Critically, in these studies, both moderately and weakly expected words were the most expected targets in a given sentence (e.g., “George could not believe his son stole a CAR” vs. “There was nothing wrong with the CAR”). The expectancy effects were driven by the sentence context that could be either more or less constrained. However, it remains to be clarified if and when an expected word that is not the most expected one will also lead to a facilitation effect.

Specifying how a second-best word is processed allows us to disentangle whether expected words are predicted in a serial or in a parallel way. According to the serial graded proposal there should be no facilitation effect for the second-best word, as only the most expected word is pre-activated. Therefore, the N400 amplitude should be reduced (i.e., less negative) in response to the most expected words relative to the second-best words, with similar N400 amplitudes for second-best and unexpected words, since both have not been pre-activated.

Alternatively, following the parallel probabilistic account the processing of the second-best word should be immediately facilitated since all probable candidates are pre-activated. Yet, the effect should be of smaller magnitude relative to the expected word, since second-best words are less likely. Specifically, the N400 amplitude should reflect the level of expectancy of the critical words, manifesting as a graded N400 amplitude that increases from expected to second-best to unexpected words.

Finally, we conducted an exploratory analysis of the PNP. Specifically, we probed whether moderately constrained sentences elicit this increased positivity. Most studies that reported this effect relied on strongly constrained contexts (e.g., Brothers et al., 2020; DeLong & Kutas, 2020; DeLong et al., 2014; Ness & Meltzer-Asscher, 2018), and some failed to observe it in response to less constrained sentences (Delong, Urbach, Groppe, & Kutas, 2011; Federmeier et al., 2007). Yet, few studies have observed a PNP effect for sentences with weak and moderately constraining contexts (Brothers et al., 2015; Kutas, 1993). Therefore, it is still uncertain if the degree of sentence constraint affects the emergence of this component.

## 2. Method

### *2.1. Participants*

Thirty college students (18 females, *M* = 22.6 years, *SD*= 6.12) took part in this study. Data from six additional participants were removed from the analysis – four had a high number of trials with artifacts (more than 50% in at least one condition) and the other two were due to technical problems during the EEG recording. Participants were all native speakers of European Portuguese, right-handed, and had no history of neurological impairment or reading disorder. They provided written consent to the experimental procedure, which was approved by the ethics committee of Faculdade de Psicologia da Universidade de Lisboa. All participants received a compensation for their participation (either a 10 € voucher or course credit).

### *2.2. Material*

Two hundred and seventy moderately constrained sentences (see Table 1 for examples) were selected from a pool of 807 sentences that were pre-tested in a cloze probability procedure (Bloom & Fischler, 1980; Pinheiro et al., 2010; Taylor, 1953). In this pre-test, an independent group of participants read a sentence presented without the last word (e.g., “The dog spent the afternoon chewing the”) and had to write down the first word that came to their mind (each sentence was completed by approximately 20 participants). The Cloze Probability (CP) of the word was computed based on the proportion of times each word was used to complete the sentence (Bloom & Fischler, 1980).

Ninety of those sentences were presented with their most expected word (CP = .61, *SD* = .12, range: .41 - .85), their second most expected word (i.e., second-best; CP = .19, *SD* = .05, range: .12 - .29) or an unexpected word (CP = 0, *SD* = .01, range: 0 - .04). The unexpected words were not produced by the participants in the pre-test, yet were semantically congruent with the sentences. To avoid sentence repetition, each sentence frame (e.g., “The dog spent the afternoon chewing the”) was presented only once to each participant, in the expected, second-best or unexpected condition, depending on the final target word (“bones”, “shoes” or “glasses”, respectively). All target words were nouns and matched for various psycholinguistic parameters (see Table 2) obtained from the P-Pal database (Soares et al., 2018), including word frequency ($F < .1$), length ($F < 1.5$), orthographic, and phonological neighbours ($F < 1$ in both cases).

-------------- Table 1 --------------

To ensure that the same target words were presented in the three experimental conditions, the ninety second-best words were also presented in other sentence contexts where they were the most expected word (CP = .61, *SD* = .15, range: .32 - .90) and an unexpected word (CP = 0, *SD* = .01, range: 0 - .05). For instance, the target word “shoes” which was the second-best in the example above, appeared as the most expected word in the sentence “Tiago, when he got home, took off his”, and as the unexpected word in the sentence frame “Grandmother always goes to the market to buy the”. In this way, we guaranteed that, between participants, the same sentence frame appeared in the three experimental conditions, and that the same target word was the most expected, second-best or unexpected word, depending on the previous sentence context. Importantly, to ensure that each participant saw each sentence and target word only once, we created four lists containing 45 sentences from each condition. Each participant saw one experimental list, with the four lists evenly distributed across participants.

-------------- Table 2 --------------

### *2.3. Procedure*

Each trial began with a fixation cross presented for 500 ms in the centre of the screen. The sentence was then presented word by word, with a duration of 200 ms per word and a 300 ms inter-word interval. The order of sentence presentation was pseudo-randomized, ensuring that there were always less than 3 sentences of each condition appearing consecutively. To ensure attentive reading, participants were asked, 1000 ms after the target word of the sentence was shown, to judge whether a probe word had appeared in the preceding sentence. Probes were displayed for 1500 ms and during this period participants had to press a key with the right index finger to decide if that word was presented (i.e., old) and another key with the left index finger to decide if that word did not appear in the sentence (i.e., new). Half of the probe words were selected from the immediately preceding sentence and half were randomly selected from other sentences. Probes were content words (noun, verbs, adjectives, adverbs) of the presented sentences, but were never the final word. A new trial started after a blinking period that lasted 1500ms. Presentation software (version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) was used for stimulus presentation and behavioural response recording.

Participants were instructed to avoid eye blinks and body movements during the presentation of the sentences. Before the experimental session, they were presented with nine practice trials to get familiarised with the task. The main experimental session was divided into five blocks, including brief resting periods between blocks; recording time was approximately fifty minutes. The order of block presentation was counterbalanced between participants.

After the EEG recording session ended, participants provided plausibility ratings for each of the previously presented sentences. They were instructed to rate the plausibility of each sentence using a 5-point scale (1= completely implausible to 5= completely plausible).

### *2.4. Behavioural analysis*

The results from the recognition task were analysed in two mixed effects models. Proportion of correct responses was analysed using a generalized linear model and response times (RTs) for correct trials with a linear mixed effects model, using the lme4 package (1.1-2.1) in RStudio (https://www.rstudio.com/). The RTs above or below 2.5 the standard deviation from the mean (by subject) were excluded, this resulted in 2.5% of the original data being removed. The fixed factor of each model was Target Word (expected, second-best, unexpected) and participants and items were included as random effects (intercepts only; the model would not converge with Target Word included in the participants random effect structure). Backward difference contrasts were used for the Target Word factor, where expected was compared to unexpected [unexpected -2/3, expected 1/3, second-best 1/3], and second-best to expected [unexpected -1/3, expected -1/3, second-best 2/3]. *P*-values were determined through treating the *t*-value as a *z*-statistic (Barr, Levy, Scheepers, & Tily, 2013).

To explore the sentence plausibility ratings, a cumulative link mixed effects model was computed with fixed factor Target Word (expected, second-best, unexpected) and participants and items were included as random effects (intercepts only; the model would not converge with Target Word included in the participants random effect structure). Backward difference contrasts were used for the Target Word factor, where second-best was compared to unexpected [unexpected -2/3, expected 1/3, second-best 1/3], and expected to second-best [unexpected -1/3, expected -1/3, second-best 2/3]. *P*-values were determined through treating the *t*-value as a *z*-statistic (Barr et al., 2013).

The code used for data analysis and the dataset with behavioural responses, can be found here: https://osf.io/utfjc.

### *2.5. EEG recording and analysis*

The electroencephalogram (EEG) was recorded with an ActiveTwo Biosemi electrode system with 64 Ag/AgCl active scalp electrodes, mounted on an elastic cap (for further details see http://www.biosemi.com; BioSemi, Amsterdam, The Netherlands). EEG was continuously sampled at 2048Hz, with a bandpass of 0.01-100Hz, and stored for later analysis. Additionally, two electrodes were placed at the left and right temples (horizontal electrooculogram: EOG) and one below the left eye (vertical EOG) to monitor horizontal and vertical eye movements, and two electrodes were placed on left and right mastoids for offline referencing.

EEG data were pre-processed using EEGLAB v14.1.1 (Delorme & Makeig, 2004), with supplementary plugins: ERPLab (Lopez-Calderon & Luck, 2014), ADJUST (Mognon, Jovicich, & Bruzzone, 2011), and CleanLine. Data were downsampled to 512Hz, referenced offline to the average of the left and right mastoids, and high-pass filtered with a 0.1 Hz filter. Data from individual channels that were consistently noisy for a given subject were replaced using a nearest-neighbour interpolation. The average number of replaced channels was 0.67 (range = 0 – 2). We ran an independent component analysis (ICA) and used ADJUST plugin to identify and correct artefacts (e.g., blinks) in the raw EEG data. Individual epochs were created for each target word from 200ms before word onset to 1000ms after onset, and a baseline correction from -200ms to 0ms preceding word onset was applied. A final round of semi-automatic artifact rejection with a threshold of ±100 μV was used to remove any remaining artifacts. Following artefact rejection, ERP averages were based on at least 70% of the trials of each condition per participant. The number of trials did not differ between conditions [$F < 1$; Expected = 42.37 (3.0), Second-best = 41.93 (3.26), Unexpected = 42.23 (3.0)].

Separate linear mixed effects models were calculated with the lme4 package (1.1-2.1) in RStudio (https://www.rstudio.com/) for each component at specific regions. For the N400 component, the dependent variable was the single-trial amplitude between 300-500ms after word onset averaged across three regions of interest (ROIs): Frontal: Fz, F1, F3, F2, F4, FCz, FC1, FC3, FC2, FC4; Central: Cz, C1, C3, C2, C4, CPz, CP1, CP3, CP2, CP4; Parietal: Pz, P1, P3, P2, P4, POZ, PO3, PO4 (see Figure 1A). The selection of the time-window and the electrodes to compute the average amplitude in the ROI was created considering the electrodes usually used in the literature (e.g., Brothers, Swaab, & Traxler, 2017; Comesaña et al., 2012; Pinheiro et al., 2013; Thornhill & Van Petten, 2012). The fixed factors in the model were Target Word (expected *vs*. second-best *vs*. unexpected) and ROI (frontal *vs.* central *vs*. parietal) along with their interaction. Both participants and items were included as random effects (intercepts only; the model would not converge with Target Word included in the participants random effect structure). Backward difference contrasts were used for the Target Word factor, where second-best was compared to unexpected [unexpected -2/3, expected 1/3, second-best 1/3], and expected to second-best [unexpected -1/3, expected -1/3, second-best 2/3]. Sum contrasts were used for the ROI factor. *P*-values were determined through treating the *t*-value as a *z*-statistic (Barr et al., 2013).

For the PNP component, the dependent variable was the single-trial amplitude between 600-900ms across three regions of interest (ROIs): Anterior Left: AF3, AF7, F3, F5, F7; Anterior Midline: AFz, Fz; Anterior Right: AF4, AF8, F4, F6, F8 (see Figure 1B). The time window and the electrodes used for the analysis were based on prior studies (e.g., Brothers et al., 2017; Ness & Meltzer-Asscher, 2018). The fixed factors in the model were Target Word (expected *vs*. second-best *vs*. unexpected), ROI (anterior left *vs.* anterior midline *vs.* anterior right), and their interaction. Both participants and items were included as random effects (intercepts only; the model would not converge with Target Word included in the

Participant random effect structure). Backward difference contrasts were used for the Target Word factor, where expected was compared to unexpected [unexpected -2/3, expected 1/3, second-best 1/3], and second-best to expected [unexpected -1/3, expected -1/3, second-best 2/3]. Sum contrasts were used for the ROI fixed effect. *P*-values were determined through treating *t*-value as a *z*-statistic (Barr et al., 2013).

The code used for data analysis, the dataset for N400 and PNP components, can be found here: https://osf.io/utfjc.

--------------- Figure 1 ---------------

# 3. Results

## *3.1. Behavioural Results*

### *3.1.1 Probe Recognition*

Overall accuracy in the post-sentence probe recognition task was .98 (*SD* = .03), indicating that participants were reading the sentences attentively (see Table 3). In both models, the model predicting accuracy [*ACC ~ Target Word + (1 | Participant) + (1 | Item)*] and the model predicting response time [*RT ~ Target Word + (1 | Participant) + (1 | Item)*], neither of the two contrasts were significant.

--------------- Table 3 ---------------

### *3.1.2. Plausibility Ratings*

The sentence plausibility questionnaire showed that in general all sentences were plausible, with a mean score above 3 in the 5-point scale. Sentences completed with an unexpected word had a lower plausibility score ($M$ = 3.67, $SD$ = 1.34) than sentences completed with the most expected ($M$ = 4.62, $SD$ = 0.75) and sentences completed with the second-best words ($M$ = 4.42, $SD$ = 1.00). In the model predicting plausibility scores [PlausibilityScore ~ *Target Word + (1 | Participant) + (1 | Item)*], a significant difference between expected and unexpected condition was observed ($\beta$ = 2.74, $t$ = 9.35, $p$ < .001), showing that the plausibility scores were lower in sentences with an unexpected target word than an expected target word. Additionally, the difference between expected and second-best word was also significant ($\beta$ = -0.73, $t$ = -4,67, $p$ < .001), showing that the plausibility scores were slightly higher in sentences with an expected target word than a second-best target word.

### *3.2. ERP Results*

Visual inspection of the grand average ERP waveforms revealed similar patterns to those observed in previous sentence comprehension studies using a word-by-word visual presentation paradigm (e.g., DeLong et al., 2019; Federmeier et al., 2007; Thornhill & Van Petten, 2012). Target words in all conditions elicited an initial positive going peak (P1), a negative going peak (N1), followed by a positivity (P2) peaking around 200ms that was broadly distributed across the scalp. These components were followed by a negativity, peaking between 300 and 500ms (N400) that was largest at centro-parietal sites (Figure 2). After the N400, the ERPs in all conditions became more positive. Visual inspection showed a slight increased positivity for unexpected words, specially over the left anterior electrodes (between 700-900ms).

-------------- Figure 2 --------------

*3.2.1. N400*

In this model [*MeanAmplitude ~ Target Word * ROI + (1 | Participant) + (1 | Item)*], there was a significant difference between second-best words and expected words ($\beta = 0.92$, $t = 3.55$, $p < .001$), showing that the second-best words (Mean: -0.33 μV; SE = 0.17) elicited a more negative N400 than expected words (Mean: 0.50 μV, SE = 0.17). Additionally, the difference between second-best words and unexpected words was also significant ($\beta = .66$, $t = 2.53$, $p = .01$), with unexpected words (Mean: -0.88 μV, SE = 0.16) having a more negative N400 than second-best words (see Figures 2 and 3). There were no other significant effects ($p > .05$; see Table 4).

----------- Figures 3 ---------------

--------------- Table 4 ---------------

*3.2.2. PNP*

In this model [*MeanAmplitude ~ Target Word * ROI + (1 | Participant) + (1 | Item)*], none of the contrasts were significant: neither expected words *vs.* second-best words nor expected words *vs.* unexpected words or any other effects reached significance ($p > .1$, for both contrasts; see Figure 2 and Table 4).

## 4. Discussion

In the present study, we investigated predictive mechanisms in sentence comprehension by examining whether a facilitation effect is extended to second-best words, which have an intermediate level of cloze probability between the most expected and unexpected words, while controlling for contextual constraint. Accordingly, we tested two competing hypotheses accounting for how predictive processes are implemented – through a serial graded cascade *vs.* probabilistic parallel activation.

The processing of second-best words in moderately constrained sentences was associated with an N400 response that was distinguished from both expected and unexpected words. The second-best words showed a reduced N400 amplitude compared with the unexpected words. This result shows that the facilitation effect extends to expected words with an intermediate level of cloze probability, even when those words appear in a context where there is another, more expected word. Thus, the reduced N400 amplitude is not exclusive to the most expected word in a given sentence context. However, when the word is not supported by the previous sentence context, which is the case of unexpected words, there is an increased effort to process that word, as indexed by a more negative N400 amplitude (e.g., Federmeier et al., 2007; Marta Kutas & Hillyard, 1984; Pinheiro et al., 2013).

In the same time-window (300-500ms), the second-best words showed an enhanced N400 amplitude compared with the most expected words. This result reveals that the processing of second-best words is not facilitated to the same extent as expected words, even though second-best words are also expected candidates. This difference is consistent with prior studies showing that word expectancy modulates the N400 amplitude: the more a given word is expected in a sentence context, the smaller the N400 amplitude (e.g., DeLong et al., 2005; Federmeier, 2007; Wlotko & Federmeier, 2012). Yet, in those studies the words with lower cloze probability were presented in weakly constraining sentence contexts. Critically, our results show that in moderate constraint sentences, there is a facilitation effect that extends to other expectable words.

The reduced N400 amplitudes found for both the second-best words (when compared with unexpected words) and the most expected words might be associated, at least in part, with the pre-activation of those words during sentence reading. Although our data analysis was focused on EEG effects occurring after the word onset, prior studies have consistently demonstrated that specific words or word features are pre-activated in sentence comprehension (e.g., DeLong et al., 2005; Kamide, Altmann, & Haywood, 2003; Szewczyk & Schriefers, 2013; Van Berkum et al., 2005).

Some authors have claimed that the N400 effect observed for the target word might only index a word's ease of integration into a sentence context, which in turn may reflect a more general unification process involved in generating a coherent interpretation of the sentence meaning (Hagoort, 2005; Zhu et al., 2012). The subjective sentence plausibility score is often used to measure how well a word fits into the sentential context. Previous studies have shown that sentence plausibility affects the processing of equally probable words. Unexpected words that are less plausible (e.g., "It was difficult to understand the visiting professor. Like many foreigners he spoke with an APRON…", where the expected word was "ACCENT") elicited an enhanced N400 effect (i.e., more negative) relative to more plausible unexpected words (e.g., "It was difficult to understand the visiting professor. Like many foreigners he spoke with a LISP…"; e.g., Brothers et al., 2015; DeLong, Quante, & Kutas, 2014). In our study, the mean plausibility score confirmed that all sentences were considered plausible (mean score above 3 in a 5 point-scale), although the sentences ending with unexpected words had a lower level of plausibility compared to sentences ending with the most expected and second-best words. This difference could explain the enhanced N400 found for the unexpected words compared with the second-best words and with the most expected words[1]. Critically, the plausibility score for sentences completed with the most

[1] Note that in the linear mixed effects model presented we did not directly compare the mean amplitude of the N400 between unexpected and the most expected words. Yet, in an additional model we compared the processing of the most expected words with the second-best words and with the unexpected words and both the effects were statistically significant.

expected and the second-best word was very similar (4.62 *vs.* 4.42). Even though the plausibility scores difference was significant, since both conditions showed a higher score in plausibility it seems unlikely that the difference in N400 amplitude is only related with the sentence plausibility. Thus, this ERP component does not seem to exclusively reflect difficulty in integrating words for sentence unification, so the facilitation effect found has to be associated, exclusively or to some degree, to the pre-activation of the expected words.

The observed graded facilitation effect (Expected < Second-Best < Unexpected) is consistent with the proposal that the prediction mechanism occurs in a parallel probabilistic way (DeLong et al., 2005; Frisson, Harvey, Drieghe, & Staub, 2017; Luke & Christianson, 2016). The facilitated processing of the most expected and second-best words suggests that readers are likely able to use the sentence information to activate all the probable candidates for that sentential context. In addition, the attenuated facilitation found for the second-best words compared with the most expected words suggests that the degree of activation of each candidate is modulated by the degree of expectancy of that word. In our study, similar to the majority of studies in this field (e.g., DeLong et al., 2005; Federmeier et al., 2007; Ito, et al., 2016; Thornhill & Van Petten, 2012), the degree of expectancy was defined by the cloze probability score. Yet, some authors argued that the cloze probability scores might not be a precise measure of word's expectancy degree, at least in weak constraint sentences or in the case of less expected words (Staub, Grant, Astheimer, & Cohen, 2015; Thornhill & Van Petten, 2012): cloze probability is frequently computed on the basis of single response, with no control over time response or alternative candidates. Future studies comparing the processing of words with intermediate levels of cloze probability could use cloze probabilities obtained in multiple response paradigms (e.g., McDonald & Tamariz, 2002; Schwanenflugel, 1986) or include additional measures such as type-token ratio (e.g., McDonald & Tamariz, 2002; Staub et al., 2015), word's surprisal (e.g., Frank et al., 2015) or transitional probabilities (e.g., Frisson, Rayner, & Pickering, 2005), which may provide a finer measure of the degree of word expectancy.

The graded effect of the N400 across conditions does not support the hypothesis of a serial graded account that postulates that only the most expected word can be pre-activated (Thornhill & Van Petten, 2012). Indeed, we observed a facilitation effect for the second-best words, which elicited a reduced N400 amplitude compared with the unexpected words. Moreover, our results challenge previous findings indicating that representations pre-activated in high constraint sentences are highly specific (e.g., DeLong et al., 2005; Ito et al., 2016; Kwon et al., 2017; Laszlo & Federmeier, 2009; Szewczyk & Schriefers, 2013; Wicha et al., 2004). If the pre-activation was highly specific, facilitation would only occur for words that share semantic features with the most expected word. However, in our study, even though all the critical words were matched in gender and number, their semantic similarity varied across sentences. In some sentences, the words shared some semantic features especially between the most expected and the second-best condition, (e.g., both “milk” and “syrup” are drinkable items); yet in other sentences the critical words were semantically unrelated (e.g., in the sentence “Grandma takes good care of her” the most expected word was “granddaughters” while the second-most expected word was “flowers”). Thus, our data suggests that the pre-activation of expectable candidates is not highly specific, and that it seems to be best accounted for in terms of word’s expectancy.

Although we have argued for a leading role of predictive mechanisms in the N400 effects, some authors have proposed that the N400 is a complex component reflecting both the retrieval of lexical features and the cognitive demands underlying integration of words into a sentence context (Chow et al., 2014; Ferreira & Chantavarin, 2018; Lau et al., 2008). According to this view, the facilitation effect for the most expected words and the second-best words could be due to either a facilitated retrieval or a simplified integration of these words when compared with the unexpected words. This proposal could accommodate the results of the current study with the serial graded account, if we consider that the reduced N400 amplitude in response to the most expected words reflects both the pre-activation of those words and the eased integration, whereas the facilitation effect observed for the second-best words may only reflect eased integration. Yet, we think this is an unlikely hypothesis. On the one hand, all target words should be easy to integrate into the respective sentence contexts since sentences were considered plausible in all conditions. For example, even when completed with unexpected words the sentences had a plausibility score above average (3.7, ranging from 1-5). The study of DeLong et al. (2014) showed an impact of the plausibility score on the N400 amplitude but presented sentences with very low plausibility (1.2 vs. 2.8, ranging from 1-5). On the other hand, the decreased N400 amplitude for the most expected and second-best words, occurring at the same time-window, suggests that the underlying processes are likely to be the same. Thus, we consider that the facilitation effects observed in our study are more likely associated with the pre-activation of the expectable words, in line with the parallel probabilistic account.

In contrast to some studies that have observed a late positivity for unexpected words, we did not find evidence for the PNP component. In part, this result could be a consequence of the constraint degree of our sentence contexts. Previous studies (DeLong et al., 2014; Federmeier et al., 2007) presented strong constraint sentences with a mean cloze probability above .80, whereas our sentences were moderately constrained with a mean cloze probability of .61. As such, the absence of a PNP component for the unexpected words is consistent with the view that this response is only observed when unexpected words are integrated into high constraint contexts that are strongly biased towards a specific word (Brothers et al., 2020; Kuperberg, Brothers, & Wlotko, 2020). Moreover, a recent study suggests that the semantic similarity between the unexpected and the highly predicted words affects the emergence of this component; only unexpected words that were unrelated to expected words elicited a PNP component (DeLong & Kutas, 2020). In our study, neither unexpected nor second-best words were controlled for semantic overlap with the most expected words. As such, in some sentences the critical words belonged to the same semantic category or were semantically related (e.g., when the word “bread” was the most expected, the words “cake” and “tomato” were the second-best and unexpected words, respectively). Importantly, the PNP amplitude can also be influenced by task demands. In tasks that explicitly instructed participants to predict the upcoming words, the PNP elicited by the unexpected words was characterized by a more positive amplitude compared with passive reading tasks (Brothers et al., 2015, 2017). Tasks that instruct participants to guess the upcoming words presumably promote pre-activation processes which may lead to increased demands to suppress those pre-activated words when they are not presented. In our study, participants were asked to read the sentences and perform a memory task at the end of each sentence. There was no mention to predict upcoming words, the unexpected words were not necessarily unrelated to the most expected words, and the sentence context was not strongly biased. As such, these methodological options could have undermined the elicitation of the PNP component. Our findings confirm that this component only emerges in specific experimental conditions, which suggests that the processes associated with this late positivity are not key to sentence comprehension.

Since the mixed effects models did not converge with a maximal random-effect structure (Barr et al., 2013), we could not assess the possible impact of subject variability on the processing difference across target words. Even though the models we used had a simple random effects structure (only including intercepts), this need not give rise to concern. There is evidence that parsimonious models can provide better power than maximal models while maintaining an acceptable Type I error rate, especially in studies where the experimental design has an adequate number of subjects and items (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). As this study included 30 participants and 45 items per condition (i.e., approximately 1350 observations per condition), there is an acceptable number of observations to ensure good power for the analysis (Brysbaert & Stevens, 2018; Matuschek et al., 2017). Therefore, we believe thar our models are suitable and that the results reported are reliable.

Notwithstanding, the present study was run in a laboratory setting and its experimental design created an artificial reading scenario, in which sentences were shown word by word in a rapid serial visual presentation. Therefore, it could be argued that the predictive mechanisms are biased by the type of sentences we have presented, all being moderately constraining, and/or by the procedure adopted, which may not reflect what would happen in a natural reading scenario. Yet, similar results have been found using co-registered eye-tracking measures and neural measures which allow the whole sentence to be presented (e.g., Kliegl, Dambacher, Dimigen, Jacobs, & Sommer, 2012; Schuster, Hawelka, Hutzler, Kronbichler, & Richlan, 2016). In addition, studies simulating natural conversations have also found evidence supporting the operation of predictive mechanisms (Mandel, Bourguignon, Parkkonen, & Hari, 2016; Pérez, Dumas, Karadag, & Duñabeitia, 2019). Thus, we believe that the effects observed in our study are replicable to a greater extent in a more natural setting.

In conclusion, the N400 findings showed a graded facilitation effect for the expected words, as there was an enhanced reduction of N400 amplitude for the most expected words and a moderate reduction for the second-best words. The facilitation effect found for the most expected words is consistent with prior literature (DeLong et al., 2005; Federmeier et al., 2007; Kutas & Hillyard, 1980; Pinheiro et al., 2013). The facilitation effect observed for the second-best word shows that other expectable words in a given sentence are promptly and easily processed, which suggests that pre-activation might extend to all expectable words in sentence comprehension. Therefore, these findings are consistent with the parallel probabilistic proposal of predictive mechanisms in language processing (Delong, Troyer, & Kutas, 2014; DeLong et al., 2005). The lack of a PNP component suggests there are no additional costs in processing unexpected words, at least in moderately constraining sentences.

Acknowledgments

The authors wish to thank Maria Amorim for assistance with ERP data collection and Susana Araújo for assistance with task programming.

Declaration of interest statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a PhD grant from Fundação para a Ciência e a Tecnologia (FCT, Portugal: PTDC/PSI-PCO/118148/2010) to SF and by the Research Centre for Psychological Science (CICPSI, Faculdade de Psicologia, Universidade de Lisboa).

## References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, *8*(6), 631–642. https://doi.org/10.3758/BF03213783

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. https://doi.org/10.1016/j.cognition.2014.10.017

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. https://doi.org/10.1016/j.jml.2016.10.002

Brothers, T., Wlotko, E. W., Warnke, L., Kuperberg, G. R., & Watkins, K. E. (2020). Going the extra mile: Effects of discourse context on two late positivities during language comprehension. *Neurobiology of Language*, *1*(1), 135–160. https://doi.org/https://doi.org/10.1162/nol_a_00006 Running

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

Calvo, M. G., & Castillo, M. D. (1996). Predictive inferences occur on-line, but with delay: Convergence of naming and reading times. *Discourse Processes*, *22*(1), 57–78. https://doi.org/10.1080/01638539609544966

Chow, W. Y., Lago, S., Barrios, S., Parker, D., Morini, G., & Lau, E. (2014). Additive effects of repetition and predictability during comprehension: Evidence from event-related potentials. *PLoS ONE*, *9*(6). https://doi.org/10.1371/journal.pone.0099199

Christensen, R. H. B. (2011). A Tutorial on fitting Cumulative Link Mixed Models with clmm2 from the ordinal Package. *Analysis*, *1*(1), 1–18.

Comesaña, M., Sánchez-Casas, R., Soares, A. P., Pinheiro, A. P., Rauber, A., Frade, S., & Fraga, I. (2012). The interplay of phonology and orthography in visual cognate word recognition: An ERP study. *Neuroscience Letters*, *529*(1), 75–79. https://doi.org/10.1016/j.neulet.2012.09.010

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters*, *316*(2), 71–74.

DeLong, K. A., Chan, W. H., & Kutas, M. (2019). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*, *56*(4), e13312. https://doi.org/10.1111/psyp.13312

DeLong, K. A., & Kutas, M. (2020). Comprehending surprising sentences: sensitivity of post-N400 positivities to contextual congruity and semantic relatedness. *Language, Cognition and Neuroscience*, 1–20. https://doi.org/10.1080/23273798.2019.1708960

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

Delong, K. A., Troyer, M., & Kutas, M. (2014). Pre-Processing in Sentence Comprehension: Sensitivity to Likely Upcoming Meaning and Structure. *Linguistics and Language Compass*, *8*(12), 631–645. https://doi.org/10.1111/lnc3.12093

Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology*, *48*(9), 1203–1207. https://doi.org/10.1111/j.1469-8986.2011.01199.x

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121. https://doi.org/10.1038/nn1504

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *15*(5), 791–801. https://doi.org/10.3758/BF03198255

Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name : Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language*, *41*, 469–495. https://doi.org/10.1006/jmla.1999.2660

Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133–146. https://doi.org/10.1111/1469-8986.3920133

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. https://doi.org/10.1016/j.brainres.2006.06.101

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science*, *27*(6), 443–448. https://doi.org/10.1177/0963721418794491

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), 1–20. https://doi.org/10.1016/S0022-5371(79)90534-6

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 Section A*, *33*(4), 465–495.

https://doi.org/10.1080/14640748108400804
Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/10.1016/j.bandl.2014.10.006
Frisson, S, Harvey, D., Drieghe, D., & Staub, A. (2017). No prediction error cost in reading: Evidence from eye movements. *Journal of Memory and Language*, *95*, 200–214. https://doi.org/10.1016/j.jml.2017.04.007
Frisson, Steven, Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *31*(5), 862–877. https://doi.org/10.1037/0278-7393.31.5.862
Goregliad Fjaellingsdal, T., Schwenke, D., Scherbaum, S., Kuhlen, A. K., Bögels, S., Meekes, J., & Bleichner, M. G. (2020). Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in German. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-62155-z
Hagoort, P. (2005a). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004
Hagoort, P. (2005b). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423. https://doi.org/10.1016/j.tics.2005.07.004
Hintz, F., Meyer, A. S., & Huettig, F. (2016). Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading. *The Quarterly Journal of Experimental Psychology*, *69*(6), 1056–1063. https://doi.org/10.1080/17470218.2015.1131309
Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience*, *31*(1), 19–31.

https://doi.org/10.1080/23273798.2015.1072223

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171. https://doi.org/10.1016/j.jml.2015.10.007

Ito, A., Martin, A. E., & Nieuwland, M. S. (2016). How robust are prediction effects in language comprehension? Failure to replicate article-elicited N400 effects. *Language, Cognition and Neuroscience*, *32*(8), 954–965. https://doi.org/10.1080/23273798.2016.1242761

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, *76*(2), 145–158. https://doi.org/10.1007/s00426-011-0376-x

Kuperberg, G., Brothers, T., & Wlotko, E. (2018). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *BioRxiv*, 404780. https://doi.org/10.1101/404780

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language*, *Cognition and Neuroscience*, *31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M, & Federmeier, K. D. (2000). Electropsysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, Marta. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572. https://doi.org/10.1080/01690969308407587

Kutas, Marta, & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese : Evidence from ERPs. *Cognition*, *166*, 433–446. https://doi.org/10.1016/j.cognition.2017.06.010

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB : an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neurosciencer*, *8*(213), 1–14. https://doi.org/10.3389/fnhum.2014.00213

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. https://doi.org/10.1016/j.cogpsych.2016.06.002

Mandel, A., Bourguignon, M., Parkkonen, L., & Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. *Neuroscience Letters*, *614*, 99–104. https://doi.org/10.1016/j.neulet.2015.12.054

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013).

Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588. https://doi.org/10.1016/j.jml.2013.08.001

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-Based Plausibility Immediately Influences On-Line Language Comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(4), 913–934. https://doi.org/10.1037/a0022964

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McDonald, S. A., & Tamariz, M. (2002). Completion norms for 112 Spanish sentences. *Behavior Research Methods, Instruments, and Computers*, *34*(1), 128–137. https://doi.org/10.3758/BF03195431

Mognon, A., Jovicich, J., & Bruzzone, L. (2011). ADJUST : An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Ness, T., & Meltzer-Asscher, A. (2018). Lexical inhibition due to failed prediction: Behavioral evidence and ERP correlates. *Journal of Experimental Psychology: Learning Memory and Cognition*, *44*(8), 1269–1285. https://doi.org/10.1037/xlm0000525

Ng, S., Payne, B. R., Steen, A. A., Stine-Morrow, E. A. L., & Federmeier, K. D. (2017). Use of Contextual Information and Prediction by Struggling Adult Readers: Evidence From Reading Times and Event-Related Potentials. *Scientific Studies of Reading*, *21*(5), 359–375. https://doi.org/10.1080/10888438.2017.1310213

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468. https://doi.org/10.7554/eLife.33468

Pérez, A., Dumas, G., Karadag, M., & Duñabeitia, J. A. (2019). Differential brain-to-brain entrainment while speaking and listening in native and foreign languages. *Cortex*, *111*, 303–315. https://doi.org/10.1016/j.cortex.2018.11.026

Pinheiro, A. P., del Re, E., Nestor, P. G., McCarley, R. W., Gonçalves, ó scar F., & Niznikiewicz, M. (2013). Interactions between mood and the structure of semantic memory: Event-related potentials evidence. *Social Cognitive and Affective Neuroscience*, *8*(5), 579–594. https://doi.org/10.1093/scan/nss035

Pinheiro, A. P., Soares, A. P., Comesaña, M., Niznikiewicz, M., & Gonçalves, O. F. (2010). Sentence-final word completion norms for European Portuguese children and adolescents. *Behavior Research Methods*, *42*(4), 1022–1029. https://doi.org/10.3758/BRM.42.4.1022

Raposo, A., & Marques, J. F. (2013). The contribution of fronto-parietal regions to sentence comprehension: Insights from the Moses illusion. *NeuroImage*, *83*, 431–437. https://doi.org/10.1016/j.neuroimage.2013.06.052

Schuster, S., Hawelka, S., Hutzler, F., Kronbichler, M., & Richlan, F. (2016). Words in Context: The Effects of Length, Frequency, and Predictability on Brain Responses during Natural Reading. *Cerebral Cortex*, *26*(10), 3889–3904. https://doi.org/10.1093/cercor/bhw184

Schwanenflugel, P. J. (1986). Completion norms for final words of sentences using a multiple production measure. *Behavior Research Methods, Instruments, and Computers*, *18*(4), 363–364.

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*(2), 232–252. https://doi.org/10.1016/0749-596X(85)90026-9

Simpson, G. B., Peterson, R. R., Casteel, M. a, & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 88–97. https://doi.org/10.1037//0278-7393.15.1.88

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Soares, A. P., Iriarte, Á., de Almeida, J. J., Simões, A., Costa, A., Machado, J., … Perea, M. (2018). Procura-PALavras (P-PAL): A Web-based interface for a new European Portuguese lexical database. *Behavior Research Methods*, *50*(4), 1461–1481. https://doi.org/10.3758/s13428-018-1058-z

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (2015). The influence of cloze probability and item constraint on cloze task response time. *Journal of Memory and Language, 82*, 1–17. https://doi.org/10.1016/j.jml.2015.02.004

Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words : An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, *68*(4), 297–314. https://doi.org/10.1016/j.jml.2012.12.002

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Bulletin, 30*(4), 415–433. https://doi.org/10.1177/107769905303000401

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, *83*(3), 382–392. https://doi.org/10.1016/j.ijpsycho.2011.12.007

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*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*(4), 485–531. https://doi.org/10.1080/01690969308407586

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*, *18*(4), 380–393. https://doi.org/10.3758/BF03197127

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Van Petten, C., Rubin, S., Parks, M., Plante, E., & Coulson, S. (2000). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning Memory and Cognition*, *108*(5), 2643–2643. https://doi.org/10.1121/1.4743849

West, R. F., & Stanovich, K. E. (1981). The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(3), 658–672.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating Words and Their Gender: An Event-related Brain Potential Study of Semantic Integration, Gender Expectancy, and Gender Agreement in Spanish Sentence Reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288. https://doi.org/10.1162/0898929041920487

Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & Van Den Bosch, A. (2016). Prediction during Natural Language Comprehension. *Cerebral Cortex*, *26*(6), 2506–2516. https://doi.org/10.1093/cercor/bhv075

Wlotko, E. W., & Federmeier, K. D. (2012). So that’s what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*(1), 356–366. https://doi.org/10.1016/j.neuroimage.2012.04.054

Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H. C., Bastiaansen, M., & Wang, S. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage*, *60*(4), 2230–2237. https://doi.org/10.1016/j.neuroimage.2012.02.036