Semantic involvement of initial and final lexical embeddings during sense-making: the advantage of starting late

Petra M. van Alphen[1,2] * and Jos J. A. van Berkum[1,2,3]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[2] Department of Psychology, University of Amsterdam, Amsterdam, Netherlands
[3] Utrecht Institute for Linguistics OTS, Utrecht University, Utrecht, Netherlands

During spoken language interpretation, listeners rapidly relate the meaning of each individual word to what has been said before. However, spoken words often contain spurious other words, like *day* in *daisy*, or *dean* in *sardine*. Do listeners also relate the meaning of such unintended, spurious words to the prior context? We used ERPs to look for transient meaning-based N400 effects in sentences that were completely plausible at the level of words intended by the speaker, but contained an embedded word whose meaning clashed with the context. Although carrier words with an initial embedding (*day* in *daisy*) did not elicit an embedding-related N400 effect relative to matched control words without embedding, carrier words with a final embedding (*dean* in *sardine*) did elicit such an effect. Together with prior work from our lab and the results of a Shortlist B simulation, our findings suggest that listeners do semantically interpret embedded words, albeit not under all conditions. We explain the latter by assuming that the sense-making system adjusts its hypothesis for how to interpret the external input at every new syllable, in line with recent ideas of active sampling in perception.

Keywords: spoken language processing, interpretation, embedded words, lexical competition, N400, syllables

INTRODUCTION

Speakers use only a handful of phonemes to express thousands of different words. As a result, many words share phonemes, and therefore sound similar, while at the same time they mean something completely different. Furthermore, words in a spoken utterance are not separated by periods of silence but are produced in a quasi-continuous fashion. Because of this phonemic overlap and the absence of clear segmentation cues, many spoken words contain other shorter words. For example the word *daisy* starts with the initial-embedded word *day* and at the offset of the word *sardine* there is *dean*. Lexical embeddings are not a rare phenomenon; according to a count by McQueen et al. (1995), no less than 84% of all polysyllabic words in English have shorter words embedded within them. These embedded words are not intended by the speaker, but are nevertheless present in the signal. What do listeners do with these words? Embedded words are unintended, but the listener does not necessarily know the speaker's intention – in fact, that intention is usually what the listener is trying to reconstruct. In the experiment reported here, we ask to what extent listeners consider the meaning of spuriously embedded words when making sense of the utterance, as it unfolds.

Making sense of a spoken utterance involves several different processes. One crucial step is to identify the words that are present in the input stream. After all, words are the main building blocks of linguistically conveyed meaning. Sense-making is therefore tightly linked to the mapping between speech and word-form representations stored in the mental lexicon ("lexical candidates").

As people listen to spoken words, several lexical candidates that match the input to some extent will be activated in parallel (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989; McQueen et al., 1994; Allopenna et al., 1998). As more acoustic information becomes available the set of matching candidates is narrowed down until only one candidate is left. Furthermore, in the literature on spoken word recognition, there is almost universal acceptance that the recognition of spoken words involves a process of competition between lexical candidates (e.g., McClelland and Elman, 1986; Norris, 1994). Strongly activated candidates can suppress the activation of weaker competitors and can help find word boundaries, as such allowing the system to more rapidly settle on the best interpretation for a given stretch of speech.

In line with current theories of word recognition, there is evidence that a spoken word with an initial embedding like *daisy* not only activates the correct lexical candidate *daisy*, but also briefly activates the shorter lexical candidate *day* (e.g., Salverda et al., 2003). Furthermore, as soon as more information in favor of *daisy* becomes available, the activation of *day* is quickly suppressed again (Marslen-Wilson et al., 1994; Isel and Bacri, 1999). What is interesting to note here is that the suboptimal acoustic realization of an embedded word does not prevent its lexical representation from being briefly activated. Initial embeddings are suboptimal acoustic realizations of their real word counterparts because the duration of the first syllable of a multi-syllabic word (e.g., *day-* produced as part of *daisy*) is shorter than that of the same syllable produced as monosyllabic word (e.g., *day* produced on its own). Although

listeners can use these durational differences to bias their lexical interpretation to some extent, these acoustic cues are not strong enough to fully prevent activation of the embedded word's lexical representation in memory (Davis et al., 2002; Salverda et al., 2003).

Whereas for initial embeddings the evidence consistently points toward brief activation, for final embeddings such as *dean* in *sardine*, the story is more complicated. Final embeddings start later in time than their carrier words, so that when the acoustic information starts to match with the shorter lexical candidate (*dean*), the longer candidate (*sardine*) has already gained considerable support. This may make it harder for the final embedding to be activated because the activation of the longer lexical candidate could suppress any activation of the shorter lexical candidate right from the beginning. Some priming studies have indeed not observed priming of final embeddings (Gow and Gordon, 1995; Norris et al., 2006), and some even reported inhibitory priming effects (Marslen-Wilson et al., 1994; Shatzman, 2006). However, in contrast to these findings, there are also studies that found facilitatory priming effects (Shillcock, 1990; Vroomen and de Gelder, 1997; Luce and Cluff, 1998; Isel and Bacri, 1999), suggesting that upon hearing a carrier word such as *sardine*, the final embedding *dean* is lexically activated, at least to some degree[1].

In the current experiment, we use event-related brain potentials (ERPs) to examine the activation of embedded words. However, we go beyond framing the issue as one of lexical activation or word recognition *only*. Listeners are not just detecting words, they are incrementally combining the meanings of those words into a higher-level semantic representation of what is said. This implies that we can ask a further question, namely: do listeners actually take into account the meaning of an embedded word when they incrementally construct the meaning of an incoming sentence? In a previous study, we exploited the N400 to address exactly this question (Van Alphen and Van Berkum, 2010). The N400 is highly sensitive to the degree of fit between incoming spoken words and the preceding sentence-semantic context (see Kutas et al., 2006, for review), with words that fit less well eliciting larger N400 components. Because of this, and because of the continuous and temporally precise measurement that the ERP signal provides, we can use the N400 to assess the semantic involvement of lexical embeddings during sense-making.

In the previous experiment (Van Alphen and Van Berkum, 2010), participants heard carrier words with an initial or final embedding in sentences in which the meaning of the embedding was either supported by the preceding sentence frame or not (while the semantic fit of the carrier word remained the same). For example, a word like *pirate* was presented in a sentence which initial part (preceding the critical word) supported the meaning of the initial embedding *pie*, as in *While Clare was waiting at the bakery she eagerly looked at the pirate on the film poster*, or in a sentence that did not support the embedding, as in *While Clare was waiting at the pharmacy she eagerly looked at the pirate on the film poster*. Similarly, a word like *champagne* with the final embedding *pain* was presented in a sentence supporting the meaning of *pain*, as in *The*

patient asked the nurse when the champagne would be cold enough to be served, or in an unsupporting sentence as *The tourist asked the driver when the champagne would be cold enough to be served*.

Because the N400 is highly sensitive to the relative ease with which the meaning of a word is retrieved and related to the preceding context, we reasoned that if listeners take into account the meaning of spurious words embedded in the carrier words, the N400 elicited by the carrier words should be modulated by the semantic fit of the embedded word to that context. This is exactly what was found: words with a lexical embedding elicited a smaller N400 in the embedding-supporting context than in the embedding-unsupporting context. Interestingly, this N400 modulation was observed for carrier words with initial as well as final embeddings. Thus, as they incrementally construct an interpretation for the unfolding sentence, listeners briefly take into account the meaning of *both* types of embeddings.

Although both observations were new, in view of the word recognition literature, the observed semantic involvement of final embeddings was particularly surprising. As explained before, final embeddings (such as *pain* in *champagne*) are in a less favorable position, because the initial parts of these words already match the longer candidates (e.g., *champagne*). What our ERP results were now suggesting is that listeners nevertheless briefly consider the meaning of a final embedded word like *pain*, even though this means that the initial part *cham* will be left over. This not only indicates that the sense-making system allows for interpretations that require passing over portions of the input, but it also reveals that the competition mechanism that helps to choose between lexical candidates is not strong enough to prevent final embeddings from briefly taking part in the sense-making process.

However, before relating this to models of word and sentence comprehension, we must address an important question. In the Van Alphen and Van Berkum (2010) study, carrier words always received minimal support from the sentential context, while we varied support for the embedded words across condition. In the embedding-supportive condition, therefore, sentential contexts always also favored the embedded word over the carrier word. For example, in *The patient asked the nurse when the champagne would be cold enough to be served*, the sentential context favored the embedded word *pain* over the carrier word *champagne*. Even though the resulting sentence is semantically well-formed and comprehensible, such bias toward the embedding and away from the carrier word may well matter. So what if things are more favorable for the carrier word, and less for the embedding? For example, while hearing the sentence *Overfishing and lower sea surface temperatures caused a big drop in the sardine population*, do listeners also briefly relate the meaning of the final embedding *dean* to their incrementally unfolding interpretation? Note that this is how we usually encounter unintended embeddings – evidence for their semantic involvement here would thus substantially widen the scope of our observations.

The issue touches on the role of context as a source of top-down prediction, rather than as one of the ingredients for word-driven sentence-semantic integration. There is considerable evidence that strongly constraining sentential and wider context can allow listeners to predict, i.e., preactivate specific upcoming words (see, e.g., Wicha et al., 2004; DeLong et al., 2005; Van Berkum et al., 2005;

---

[1]Note that final embeddings that are not aligned with a syllable boundary or do not carry primary stress are less likely to be activated (Vroomen and de Gelder, 1997).

Federmeier, 2007; Otten and Van Berkum, 2009). Although we do not use strongly predictive contexts in our research on embedded words, such findings do point to a potentially important role of top-down bias in lexical processing (see also Connolly and Phillips, 1994; Van Petten et al., 1999; Van den Brink et al., 2001, 2006; Van Berkum et al., 2003; Dahan and Tanenhaus, 2004). As such, they point to the importance of testing for the semantic involvement of embedded words not just in favorable, but also in somewhat less favorable situations.

We explored this in an ERP experiment where carrier words with an initial or final embedding were presented in sentences that supported the meaning of the carrier word, rather than that of the embedding. The ERPs evoked by these carrier words were compared to those evoked by matched control words containing no embeddings. Importantly, the sentence-semantic fit of these control words was carefully equated to that of the critical carrier words, such that any difference in the ERPs evoked by the carrier and control words could only be attributed to the presence or absence of a lexical embedding. If listeners simply ignore the meaning of the contextually unsupported embedding, carrier, and control words should elicit similar N400 components. However, if listener do briefly activate the meaning of the embeddings and try to relate them to the context, the carrier words with embedded words should show an increased N400.

MATERIALS AND METHODS

PARTICIPANTS
Forty volunteers (24 women and 16 men, mean age 21.9 years), all right-handed psychology students, participated in the experiment for course credits.

MATERIALS
Words
For the experiment 50 Dutch bisyllabic words with an initial embedding and 50 Dutch bisyllabic words with a final embedding were selected, e.g., the word *schilder* (painter) with the initial embedding *schil* (skin/peel) or the word *trofee* (trophy) with the final embedding *fee* (fairy). All embedded words coincided with the stressed syllables of the carrier words. For each of the carrier words with embedding a control word was selected that matched the carrier word in number of phonemes, stress pattern, and frequency. Critically, the syllables of the control words did not coincide with an existing Dutch word. For example, the matching control word for *schilder* was *vlinder* (butterfly), in which the first syllable *vlin* does not correspond to an existing word in Dutch, and the control word for *trophy* was *karaf* (carafe), in which *raf* is not a Dutch word. None of the carrier words or control words were (pseudo-) compounds. The mean durations and frequencies of the embedded words, carrier words, and control words are given in **Table 1**.

Sentences
The sentences were constructed such that the content supported the meaning of the carrier word and the control word to the same extent, while the meaning of the embedded word was odd given the context. For carrier-control word pair, the sentences had a similar syntactic construction and consisted of the same number of words (with the critical word being in the same position within the sentence). Examples are given in **Table 2**.

A paper-and-pencil pretest was conducted to determine the semantic fit of the carrier words, control words, and embedded words (presented as monosyllabic words). The sentences were presented up to and including the critical word to 22 participants (who did not take part in the EEG experiment). They were told that the list contained sentences that were truncated at a particular point and were asked to rate on a six-point scale how plausible they found the critical word (one corresponding to "not at all plausible" and six to "very plausible"). The mean ratings for the embedded words, their carrier words, and the control words are given in **Table 1**, indicating that the embedded words had a lower

Table 1 | Durations (ms), word frequencies (10 log of the token counts per 9 million according to the Corpus Spoken Dutch) and plausibility rating for the critical words.

	Duration (min-max)	Frequency (sd)	Plausibility rating
INITIAL EMBEDDINGS			
Embedded word	269 (152–395)	1.77 (0.90)	2.07 (0.80)
Carrier word (+embed)	430 (320–585)	1.18 (0.74)	4.80 (0.73)
Control word (−embed)	429 (296–564)	1.22 (0.70)	4.72 (0.67)
FINAL EMBEDDINGS			
Embedded word	318 (190–457)	1.77 (0.97)	2.35 (0.89)
Carrier word (+embed)	482 (352–602)	1.09 (0.79)	4.92 (0.64)
Control word (−embed)	477 (301–671)	1.11 (0.75)	5.04 (0.46)

Table 2 | Example materials for initial and final embeddings.

INITIAL EMBEDDINGS

Carrier words with embedding
Omdat Elsa haar huis wilde opknappen, ging ze op zoek naar een **schilder** om haar kozijnen te verven
Because Elsa wanted to refurnish her house, she went looking for a **[peel]/painter** *to paint her window frames.*

Control words without embedding
Omdat Emma het liefst insecten tekende, ging ze op zoek naar een **vlinder** om als voorbeeld te gebruiken
Because Emma preferred to draw insects, she went looking for a **butterfly** *to use as example.*

FINAL EMBEDDINGS

Carrier words with embedding
Omdat Jeroen de marathon opnieuw had gewonnen, kreeg hij een **trofee** met zijn naam erin gegraveerd
Because Jeroen had again won the marathon, he got a **trophy[fairy]** *with his name engraved in it.*

Control words without embedding
Omdat Maarten de huiswijn bijzonder lekker vond, bestelde hij een **karaf** met twee lege glazen erbij
Because Maarten liked the house wine very much, he ordered a **carafe** *with two glasses.*

Critical words are bold and embeddings are underlined. English translations are given in italics.

semantic fit than the carrier words and that the semantic fit was matched between carrier and control words.

Furthermore, to have a sensitivity check in case of a null result for our critical words, we included 40 filler sentences with a highly predictable word that matched the control words in all other aspects (number of syllables and phonemes, stress pattern, and frequency). The average cloze-probability (as obtained from a cloze-test with 22 participants) for these high-cloze words was 0.63 (in comparison to 0.11 for the control words). If listeners were indeed paying attention and processing the sentences (as we asked them to do), then these highly predictable words should elicit a smaller N400 than the control words. All sentences were recorded by a female native speaker of Dutch who was unaware of the purpose of the experiment.

PROCEDURE

The experiment consisted of all 240 sentences, presented in four blocks of 60 sentences. The different conditions (initial-carrier, initial-control, final-carrier, final-control, initial-high-cloze, final-high-cloze) were evenly distributed among the blocks [such that each block always contained 12 items of the first four conditions, plus one extra item of two of these conditions ($50/4 = 12 + 12 + 13 + 13$), plus five items of both high-cloze conditions]. One block on average took 12 min and was followed by a break. Two different randomized versions of the experiment were used (which were created by reversing the order of the sentences). At the beginning of the experiment there were 10 practice sentences. After the experiment there was a short questionnaire.

Participants were informed that the experiment consisted of a large number of unrelated sentences that would be played over two loudspeakers in front of them. They were instructed to alertly listen the sentences and for each sentence imagine the situation that was described. They could start the next sentence by pressing the button that was attached to the chair. They were asked to sit still as soon as they had pressed the button and blink as little as possible. Each trial started with a silent interval of 1000 ms after which the sentence was played. At 1000 ms after the offset of each sentence a plus sign appeared in the middle of the screen for 2000 ms, indicating that the participant could again press the button to initiate the next sentence.

EEG-RECORDING

The EEG was recorded from 30 silver-chloride electrodes mounted in an elastic cap at standard 10–20 locations (Fz, Cz, Pz, Oz, Fp1/2, F3/4, F7/8, F9/10, FC1/2, FC5/6, FT9/10, C3/4, T7/8, CP1/2, CP5/6, P3/4, P7/8), all referenced to the left mastoid, and with impedances below 5 kΩ. Signals were amplified with BrainAmps DC amplifiers (0.03–100 Hz band-pass), digitized at 500 Hz, and re-referenced off-line to the mastoid average. Additional HEOG and VEOG signals were computed from F9–F10 and from Fp1-V1 (an electrode below the left eye) respectively. Then, EEG segments ranging from 500 ms before to 1000 ms after critical word onset were extracted and baseline corrected (by subtraction) to a 200-ms pre-onset baseline. Segments with potentials exceeding ±75 μV were rejected. If the total rejection rate exceeded 50% in any of the experimental conditions, all data of the participant were

excluded. Across the remaining 28 participants (17 women and 11 men, mean age 22.2) the average segment loss was 19%, with no asymmetry across conditions. EEG segments were averaged per participant and condition. Because our hypotheses specifically involved the N400, repeated-measures analyses of variance (ANOVAs) were conducted over mean amplitudes in the standard 300–500 ms latency range across all 16 posterior electrodes.

RESULTS

WORD-INITIAL EMBEDDINGS

Figure 1 displays the grand average waveforms for nine electrodes for the carrier words with initial embeddings, for the matching control words without embedding and for the high-cloze words. All waveforms were time-locked to the onset of these critical words (which for the carrier words also corresponds to the onset of the embeddings).
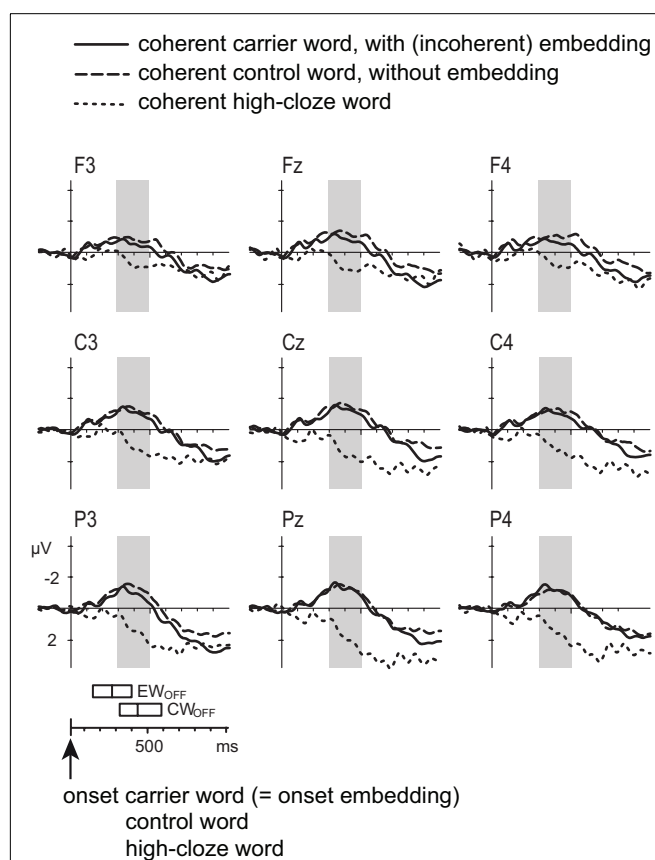
FIGURE 1 | Grand average event-related potentials from nine scalp sites to coherent carrier words with initial embeddings (solid line), coherent control words without embeddings (dashed line), and to coherent high-cloze words (dotted line), after baseline correction in the 200-ms prestimulus interval, time-locked to the onset of the carrier/control words, which corresponds to the onset of the initial embeddings. In these and all other figures, the time axis is in milliseconds, negative polarity is plotted upward, and waveforms are filtered (5 Hz high cut-off, 12 dB/oct) for presentation purposes. The bars in the lower left corner show the offset of the embedded words (EW$_{OFF}$), and the offset of the carrier words (CW$_{OFF}$). The start of the bar corresponds to the minimal value, the end to the maximal value and the middle to the mean.

As for initial embeddings, we did not observe an embedding-dependent N400 modulation in this experiment. An ANOVA with presence/absence of embedding (2) and electrodes (16) as within-subject factors indeed showed no main effect for our critical manipulation [$F(1, 27) = 0.10$, $p = 0.755$] in the 300- to 500-ms latency range[2]. Critically, the high-cloze words elicited a substantially smaller N400 than the control words [$F(1, 27) = 28.04$, $p < 0.001$], indicating that our participants were interpreting the sentences. The absence of a difference between the two critical conditions (carrier versus control) can therefore not been explained by a lack of attention. Thus, these results suggest that when listeners encounter a word (e.g., *daisy*) that is consistent with the context, they are not taking into account the meaning of spurious *onset*-embedded words (e.g., *day*).
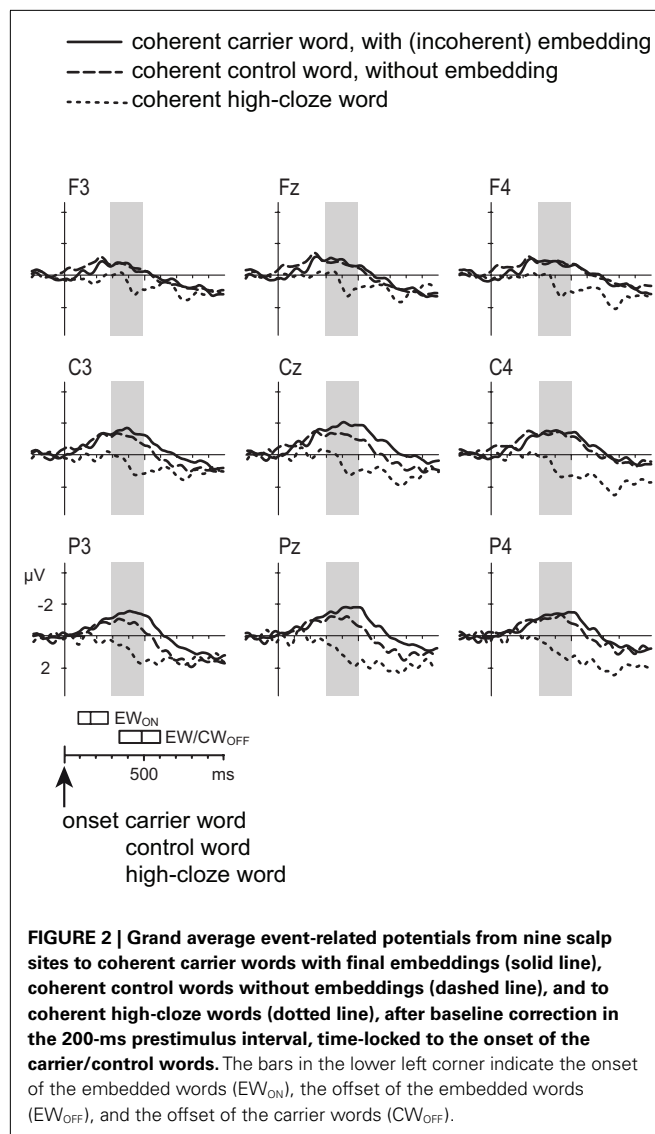
WORD-FINAL EMBEDDINGS

Figure 2 shows the grand average waveforms for the carrier words with final embeddings, for the matching control words without embedding and for the high-cloze words. Again, all waveforms were time-locked to the onset of the critical words (note that now this alignment does not correspond to the onset of the final embeddings, which start at the second syllable). As before, we found a significant reduction of the N400 for the high-cloze words in comparison to the control words [$F(1, 27) = 15.61$, $p = 0.001$].

Remarkably, there is now a noticeable difference between the carrier words with final embeddings and the control words without embeddings (when time-locked to the onset of the carrier and control words). To statistically test if this difference indeed reflected an N400 effect caused by the presence of a final embedding, we time-locked the waveforms to the onset of the second syllable (for the carrier word this corresponds to the onset of the embedded word, e.g., *fee*; for the control words it corresponds to a nonsense syllable, e.g., *raf*). The new waveforms are displayed in **Figure 3**. The results show a significant larger negativity between 300 and 500 ms for the carrier words with final embeddings relative to the control words [$F(1, 27) = 15.06$, $p = 0.001$].

To verify that this embedding-dependent effect is indeed an N400 effect, reflecting the same neural generator(s) as the standard N400 effect that we observed for the (low-cloze) control words compared to their highly predictable (high-cloze) counterparts, we plotted the topographies for both differential effects in **Figure 4**. Although the size of the embedding-dependent ERP effect is smaller than the size of the cloze-probability effect, the scalp distributions of these effects are virtually identical, and typical for the N400. This supports our assumption that the embedding-dependent effect is indeed a modulation of the N400.

In addition, an omnibus ANOVA with position (initial/final), presence/absence of embedding and electrodes (16) as within-subject factors showed a significant interaction between position and presence/absence of embedding [$F(1, 27) = 9.17$, $p < 0.01$]. Thus, in contrast to initial embeddings, the presence of a final embedded word does modulate the N400, suggesting that listeners do take into account the meaning of final embeddings when making sense of the incoming speech.

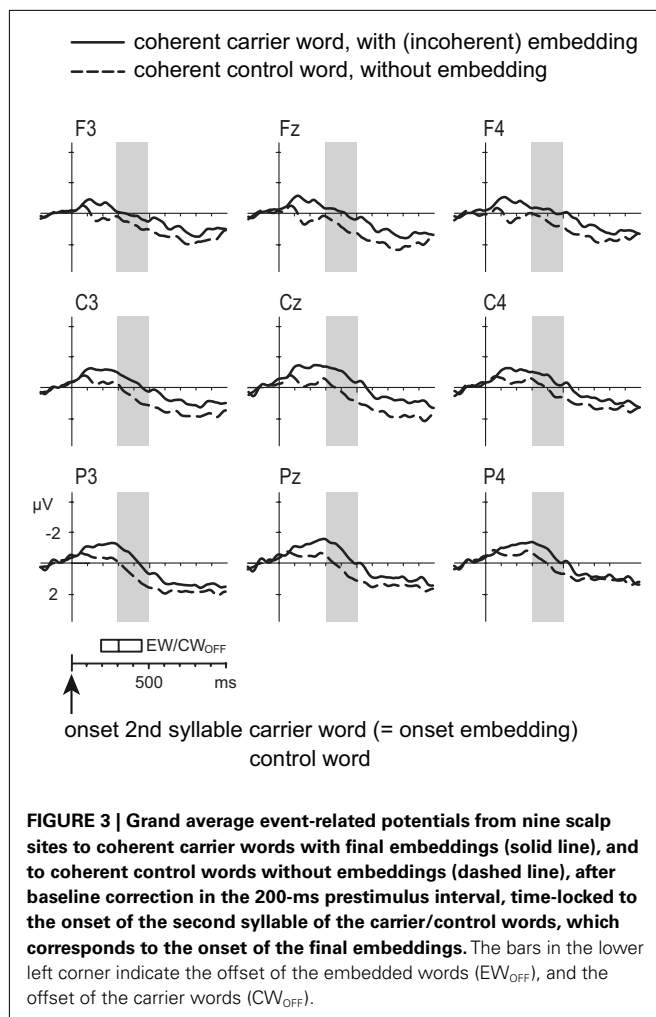[2]There was also no significant difference between the carrier and control condition in any later time window.

FIGURE 2 | Grand average event-related potentials from nine scalp sites to coherent carrier words with final embeddings (solid line), coherent control words without embeddings (dashed line), and to coherent high-cloze words (dotted line), after baseline correction in the 200-ms prestimulus interval, time-locked to the onset of the carrier/control words. The bars in the lower left corner indicate the onset of the embedded words (EW$_{ON}$), the offset of the embedded words (EW$_{OFF}$), and the offset of the carrier words (CW$_{OFF}$).

DISCUSSION

The purpose of the study was to examine the semantic involvement of initial and final lexical embeddings during sentence comprehension. In contrast to prior work (Van Alphen and Van Berkum, 2010) we focused on the most commonly occurring situation, where the sentence supports the carrier word but not the embedded word. The results show that in such a "carrier-favoring" context, the presence of a word-initial embedding (e.g., *day* in *daisy*) does not affect the N400 to the carrier words containing them. At the same time, the presence of a word-*final* embedding (e.g., *dean* in *sardine*) does considerably modulate the N400 elicited by its carrier word.

These results suggest that if listeners hear a word like *daisy* in a supporting context, they simply ignore the meaning of the initial embedding *day*, whereas if they hear a word like *sardine* they also relate the meaning of the embedded word *dean* to the context. They do the latter even though the longer candidate *sardine* fits the sentence-semantic context (which *dean* does not) *and* also explains some of the immediately prior acoustic input

coherent carrier word, with (incoherent) embedding
coherent control word, without embedding

F3   Fz   F4

C3   Cz   C4

μV
-2
2

P3   Pz   P4

EW/CW$_{OFF}$

500 ms

onset 2nd syllable carrier word (= onset embedding)
control word

FIGURE 3 | Grand average event-related potentials from nine scalp sites to coherent carrier words with final embeddings (solid line), and to coherent control words without embeddings (dashed line), after baseline correction in the 200-ms prestimulus interval, time-locked to the onset of the second syllable of the carrier/control words, which corresponds to the onset of the final embeddings. The bars in the lower left corner indicate the offset of the embedded words (EW$_{OFF}$), and the offset of the carrier words (CW$_{OFF}$).

(*sar-*). Recall that if the sentence-semantic context did not support the carrier words (Van Alphen and Van Berkum, 2010), both initial and final embeddings were briefly affecting the sense-making process. Taken together, it appears that listeners *always* relate a final embedding to the sentence-semantic context, regardless of whether there is a good context-supported alternative, but *selectively* take an initial embedding into account, doing so only when there is no reasonable context-supported alternative at hand[3]. We return to this asymmetry later in the Discussion.

WORD-FINAL EMBEDDINGS

Let us first focus on the semantic involvement of the final embeddings, such as *dean* in *sardine*. The finding that this type of embedding is taken into account during sense-making, even when the
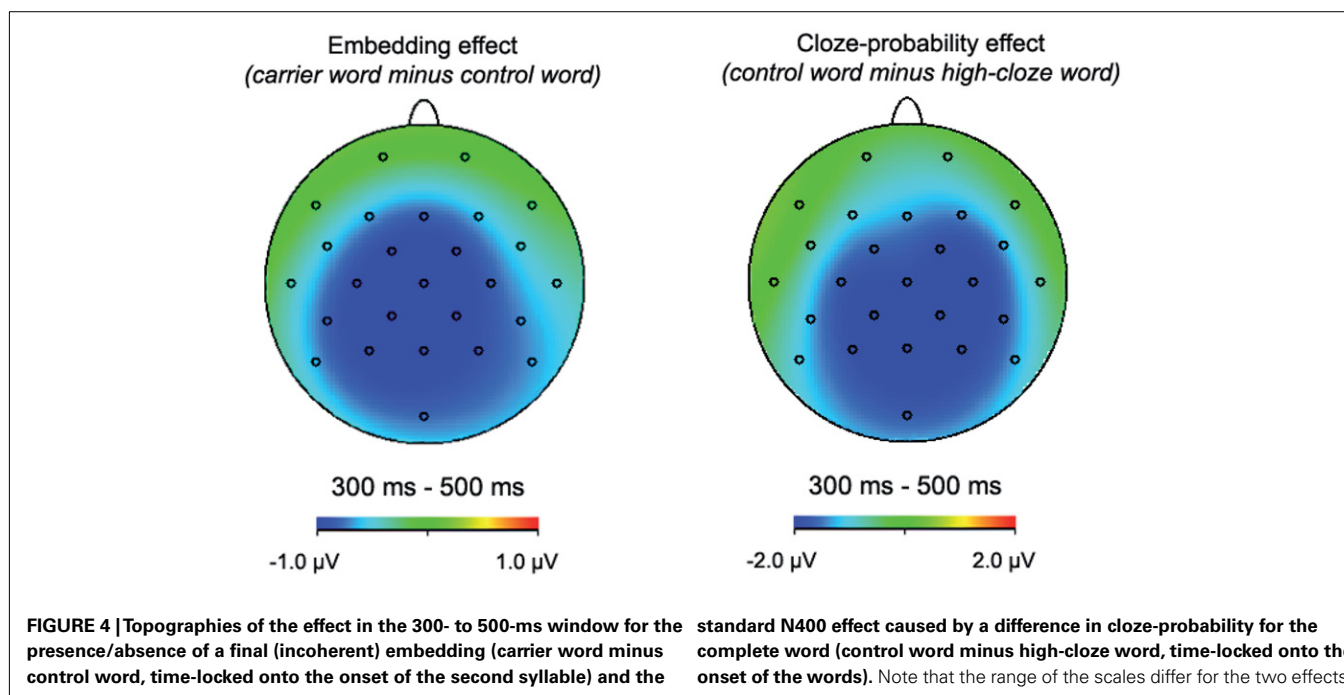
[3]That is, for embedded words that are aligned with the onset of a stressed syllable (in a language in which a strong syllable is an important segmentation cue) and that are morphologically and semantically unrelated to the carrier words. It would be beneficial to conduct more experiments using the same paradigm with carrier words with embeddings that vary along these properties to further explore the dynamics of the sense-making system and its link to lexical activation.

context favors the interpretation of the longer carrier word, has at least two interesting implications. First, it implies that at the lexical level, final embeddings words are not such poor candidates after all, in that the preceding acoustic information in favor of the longer carrier word is not sufficient to block the activation of the final embedding. Apparently, the position of final embeddings is not as unfavorable as most current models of word recognition would lead one to expect (e.g., McClelland and Elman, 1986; Norris, 1994; Norris and McQueen, 2008). What may well have contributed to the strength of final embeddings here is that all embedded word syllables in the present and previous experiment had primary stress. A strong syllable is not only a better acoustic match to the monosyllabic lexical candidate than an unstressed syllable (in terms of duration and vowel quality), it is also an important cue for segmentation. The majority of words in Dutch (87%) have lexical stress on the first syllable (Schreuder and Baayen, 1994), and a strong syllable is thus most likely the onset of a new word. Listeners indeed assume that word boundaries occur before a strong syllable (Cutler and Norris, 1988, for English; Vroomen et al., 1996, for Dutch). It is therefore more likely that a final syllable of word that corresponds to another existing word is recognized as a separate monosyllabic word when it carries primary stress.

A second implication of the observed semantic involvement of final embeddings like *dean* in *sardine* is that this indicates that, even when the system has already found a suitable candidate, this does not simply prevent it from considering other serious lexical candidates. Even though *sardine* is clearly the best acoustic match to the input and makes perfect sense in the sentence-semantic context, the meaning of *dean* is also briefly taken into consideration. This seemingly counterproductive aspect of our sense-making system may come as a surprise. But note that it does make sense, in that it indicates that the system is not blindly relying on its current hypothesis, and is prepared to reexamine that hypothesis when the next acoustically salient bit of input – a final stressed embedding – comes along.

Not all final embeddings will be equally "salient" in the signal, though. For example, a relatively long and high-frequent monosyllabic word embedded in a low-frequent carrier word that is only a few phonemes longer will be more strongly activated than a relatively short and low frequency word embedded in a much longer high-frequent carrier. All else equal, these more salient embeddings should be more likely to be taken into account during sense-making. Hence, if the N400 effect to carrier words with contextually unsupported final embeddings indeed reflects semantic involvement of the embedding, then it follows that this N400 effect should be largest for carrier words containing the most salient embeddings, with local salience determined along some relevant metric. To assess the viability of our current account, we examined this *post hoc* prediction in a model-based reanalysis of our data.

To estimate the salience of our final embeddings regardless of a given sentence context, i.e., in carrier words presented in isolation, we ran a simulation in Shortlist B (Norris and McQueen, 2008). Shortlist B is a Bayesian model of continuous speech recognition that evaluates multiple lexical hypotheses in a parallel and competitive fashion, taking into account several "stimulus-tied" factors that may well affect the probability of a final embedded

FIGURE 4 | Topographies of the effect in the 300- to 500-ms window for the presence/absence of a final (incoherent) embedding (carrier word minus control word, time-locked onto the onset of the second syllable) and the standard N400 effect caused by a difference in cloze-probability for the complete word (control word minus high-cloze word, time-locked onto the onset of the words). Note that the range of the scales differ for the two effects.

word, such as the degree of phonemic overlap, the frequency of the embedding, neighborhood density, and the neighborhood frequency (including the frequency of the carrier word). We split our set of carrier words with final embeddings into two groups based on the highest probability value in any time slice starting from the onset of the embedded word (see Appendix for more details) and computed ERPs for the more salient and the less salient embeddings separately.

In line with our prediction, the results in **Figure 5** confirm that, the embedding-dependent N400 effect is largest for the most salient embeddings. Thus, the final embeddings with the highest probability to be recognized in their local carrier word context – as quantified by Shortlist B – are showing the strongest involvement in higher-level sense-making. This analysis strengthens our claim that the N400 effects we find are indeed related to the semantic involvement of the final embeddings. Along the way, it also illustrates how a Bayesian model of continuous speech recognition can be used to link probabilities of lexical hypotheses to electrophysiological measures.
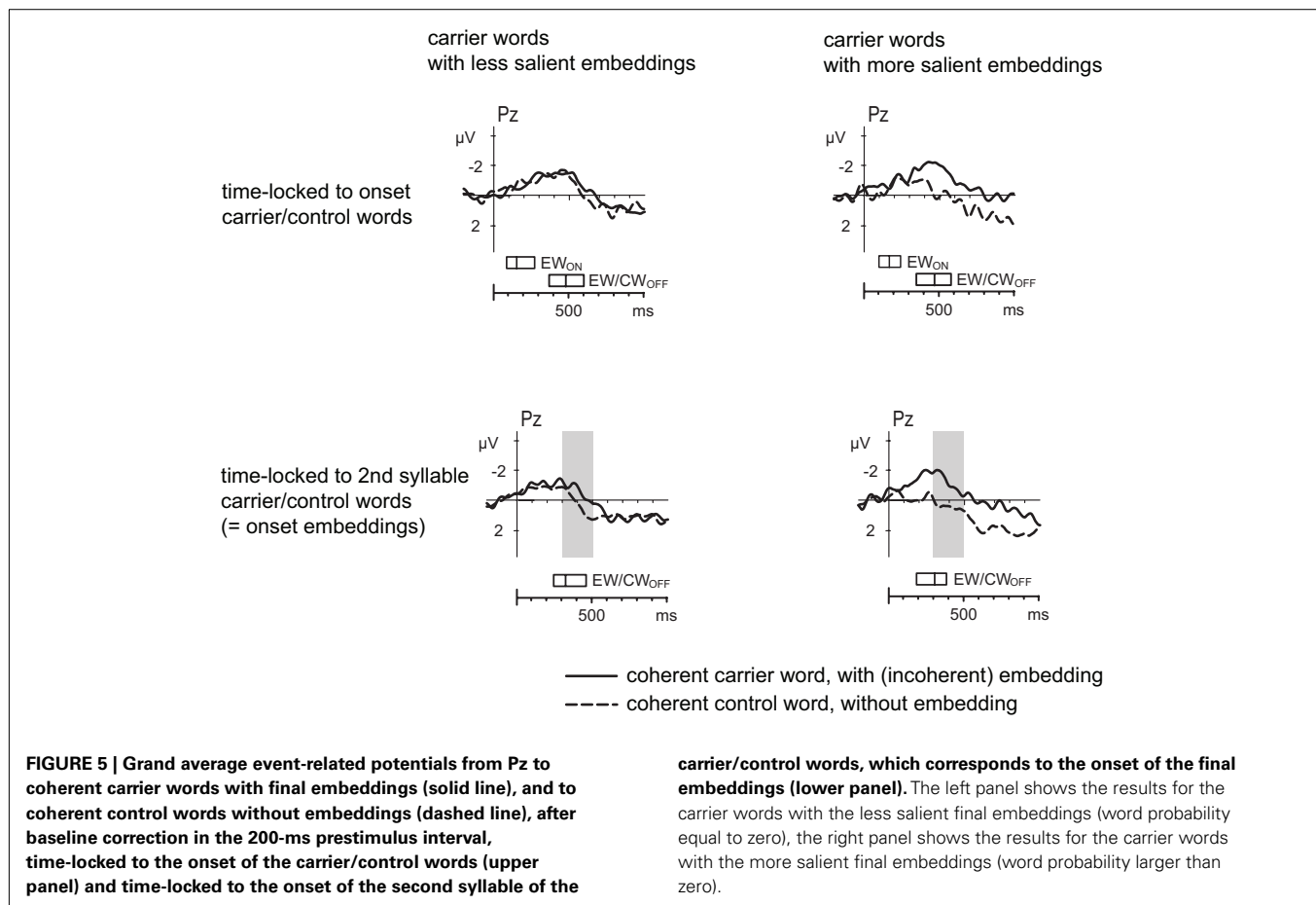
WORD-INITIAL EMBEDDINGS

Now let us return to the findings for the initial embeddings. In our previous study (Van Alphen and Van Berkum, 2010), initial embeddings were momentarily related to the prior semantic context. So why is there no sign of semantic involvement of initial embeddings in the current experiment? Note that for final embeddings, both experiments reveal an N400 effect indicative of semantic involvement, so it is not that the current study is insensitive to embedded words in general. Below, we will provide a tentative interpretation in terms of a differential interaction between signal-driven lexical competition and context-dependent sentence-level sense-making, an interpretation that is consistent with our results as well as with several other findings in the literature.

We start from the observation that in the case of carrier words with initial embeddings there are two very strongly activated lexical candidates (e.g., *day* and *daisy*) competing *for the same (strong) syllable as their onset*[4]. At the level of word recognition, this competition for the same syllable does not seem to be a bottleneck, since there is ample evidence that lexical candidates that start with the same onset are activated in parallel (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989; McQueen et al., 1994; Allopenna et al., 1998), and good evidence that initial embeddings are at least temporarily activated (e.g., Salverda et al., 2003). This fits with the idea that the main task of the word recognition system is to calculate which lexical candidates best match the acoustic input (hence "recognition"), a task that allows for the presence of multiple viable options.

However, the ultimate goal of higher-level sense-making is to converge on a single most likely interpretation for some piece of signal – after all, interlocutors often need to respond to what is said, and such behavior simply requires convergence. At this level, therefore, a workable balance must be struck between committing to a single unfolding interpretation and "opening up" to new, and potentially ambiguous information in the acoustic signal. What is important here is that to handle the temporal constraints involved in speech comprehension and conversational turn-taking (e.g., Stivers et al., 2009), the balance between commitment and uncertainty must be struck incrementally, as an *unfolding* utterance is being processed. In line with recent ideas of top-down processing in the brain (Friston, 2010; Bar, 2011), a viable strategy for the sense-making system would therefore be to always incrementally pursue a single interpretation of what is said based on prior words

---

[4]Of course there may well be other additional strongly activated lexical candidates that also briefly take part in the recognition process, but for the sake of simplicity we only take into account the two strongest lexical candidates (the carrier word and its embedding) in this Discussion.

FIGURE 5 | Grand average event-related potentials from Pz to coherent carrier words with final embeddings (solid line), and to coherent control words without embeddings (dashed line), after baseline correction in the 200-ms prestimulus interval, time-locked to the onset of the carrier/control words (upper panel) and time-locked to the onset of the second syllable of the carrier/control words, which corresponds to the onset of the final embeddings (lower panel). The left panel shows the results for the carrier words with the less salient final embeddings (word probability equal to zero), the right panel shows the results for the carrier words with the more salient final embeddings (word probability larger than zero).

and the expectations raised by those words (and by wider context, if available), but to frequently check and adjust the most likely interpretation in the light of the incoming data (i.e., the outcome of the lexical activation process), at a "sampling rate" or "grain size" that makes sense.

If this is indeed the strategy employed by our comprehension system, what would be a sensible sampling rate? For many languages – including Dutch, the language used in this study – the relevant sampling unit might well be the syllable. The syllable has been suggested as a fundamental unit for speech perception and production, serving as the interface between sound and meaning (Greenberg et al., 2003; Greenberg and Ainsworth, 2006). This makes it an obvious candidate for the strategy under discussion. Hence, at least for Dutch and comparably syllable-timed languages, we suggest that the sense-making system continually pursues a single interpretive "Gestalt" but adjust its hypothesis to the external input at every new syllable.

How can this account for the current findings, as well as those of our previous study? We need to make just one additional assumption. If our sense-making system assesses, at every syllable, the most likely chunk of meaning to be added to the partial interpretation constructed so far, words with salient initial embeddings, like *daisy*, should be particularly problematic, as there are multiple strongly activated lexical candidates (in this case *day* and *daisy*) that are *competing for the same syllable as their onset.*

How should a sense-making system aiming to converge on a single interpretation at each syllable handle this? We assume that it will provisionally commit itself to the meaning that is most likely given the preceding interpretive context. If the sentential context is entirely supportive of the carrier word (e.g., *She walked in the meadow and picked a daisy*), the onset-embedded word is not even considered – this explains the absence of an embedding-dependent N400 effect for onset embeddings in the current study. However, if the sentential context is not at all supportive of the carrier word but does support the meaning of the embedding (e.g., *She looked in her agenda for a suitable daisy*), the onset-embedded word *is* considered and provisionally committed to – this explains the presence of an embedding-dependent N400 effect in the prior Van Alphen and Van Berkum (2010) experiment.

DIFFERENTIAL ROLE OF CONTEXT IN WORD-FINAL AND WORD-INITIAL EMBEDDINGS

What about carrier words with stressed *final* embeddings, such as *sardine*? Note that with such embeddings, we observed embedding-dependent N400 effects when the prior sentential context was not at all supportive of the carrier word but did support the embedding (e.g., *The university faculty was led by a very inspiring sardine*; Van Alphen and Van Berkum, 2010), but also when the prior sentential context *was* entirely supportive of the carrier

word (e.g., *Overfishing caused a big drop in the sardine population*; this study). Why would listeners still consider *dean* if the context strongly supports *sardine*? We suspect that in this case, context plays less of a decisive role, because the two competing candidates (*sardine* and *dean*) are not aligned in time, and thus do not compete for the same stressed syllable as their onset.

To summarize, we argue that listeners *always* relate a final embedding to the sentence-semantic context regardless of whether there is a good context-supported alternative, but *selectively* take an initial embedding into account, doing so only when there is no reasonable context-supported alternative at hand. The more decisive role of context in dealing with word-initial embeddings makes sense, because both candidate words compete for the same syllable, which *immediately* suggests that only *one* of them can be correct. With word-final embeddings, the situation is different: *dean* in *sardine* is simply the next syllable, and a strong one at that – two good reasons for the system to see if it can be added to the interpretation constructed so far.

The idea that the sense-making system continually pursues a single interpretive "Gestalt" but adjust its hypothesis to the external input at every new syllable (at least in syllable-timed languages) is speculative, and in need of further testing. So is our additional assumption that context factors in different ways, for different types of embeddings. However, we note that the general idea is in line with the established central role of syllables, both in perception and production, serving as the interface between sound and meaning (Greenberg et al., 2003; Greenberg and Ainsworth, 2006). Related to this, recent work suggests that oscillatory brain dynamics in the theta frequency range is functionally related to the retrieval of lexical-semantic information (Bastiaansen et al., 2008). This is of relevance here because, as pointed out by Luo and Poeppel (2007), theta band activity corresponds to a temporal window of about 125–200 ms, which matches the mean syllable length across languages (Greenberg et al., 2003; Greenberg and Ainsworth, 2006). Thus, evidence from brain oscillations, although circumstantial, is consistent with our idea that the language interpretation system "samples" the lexical-semantic input at a syllable-based rate, at least in syllable-timed languages. Finally, the idea of a comprehension system that is actively sampling the spoken language environment, and that is prepared to reconsider its interpretation at every syllable, also resonates with emerging ideas about active sensing in the perceptual brain (Schroeder and Lakatos, 2009; Friston, 2010; Bar, 2011).

CONCLUSION

To conclude, the research reported here has generated both some challenging findings that call for a modification in theories of spoken language interpretation, and some tentative ideas on the shape such modifications might take. As for the first, our current results unequivocally show that when listeners encounter a word with a lexical embedding in a sentence that supports the meaning of the intended carrier word (e.g., *sardine* in a sentence about fishing), their interpretive system also briefly takes into account the meaning of the embedding (*dean*), at least for the word-final case illustrated here. This surprising observation runs counter to what might be expected under standard models of lexical competition and word segmentation, while at the same time testifying to the central role of strong syllables in word recognition. Furthermore, by using the Shortlist model of continuous speech recognition to reanalyze our EEG data, we have also demonstrated that the involvement of a final embedding covaries with its "perceptual salience" along other relatively signal-bound dimensions, such as its frequency and duration relative to that of the carrier word. Next and moving beyond relatively signal-bound word recognition phenomena, our findings point to an intimate yet complex connection between sentence-level sense-making and lexical activation. In particular, the combined results of our prior and current study reveal that the semantic involvement of embedded words depends on the interplay between the sentence-level contextual fit and the exact position of the embedding within its carrier word. To make sense of the latter, we propose to consider the idea that, at least in syllable-timed languages, the sense-making system continually pursues a single interpretive "Gestalt" while adjusting its hypothesis to the external input at every new syllable, in line with recent ideas of active sampling in perception.

REFERENCES

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *J. Mem. Lang.* 38, 419–439.

Baayen, H. R., Piepenbrock, R., and Gulikers, L. (1995). *The CELEX lexical database. Linguistic Data Consortium.* University of Pennsylvania, Philadelphia, PA.

Bar, M. (ed.). (2011). *Predictions in the Brain: Using Our Past to Generate a*

Future. New York: Oxford University Press Inc.

Bastiaansen, M. C. M., Oostenveld, R., Jensen, O., and Hagoort, P. (2008). I see what you mean: theta power increases are involved in the retrieval of lexical semantic information. *Brain Lang.* 106, 15–28.

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

Cutler, A., and Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *J. Exp. Psychol. Hum. Percept. Perform.* 14, 113–121.

Dahan, D., and Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. *J. Exp. Psychol. Learn Mem. Cogn.* 30, 498–513.

Davis, M. H., Marslen-Wilson, W. D., and Gaskell, M. G. (2002). Leading

up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 218–244.

DeLong, K., Urbach, T., and Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.

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

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138.

Gow, D. W., and Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: evidence from priming. *J. Exp. Psychol. Hum. Percept. Perform.* 21, 344–359.

Greenberg, S., and Ainsworth, W. A. (2006). *Listening to Speech: An Auditory Perspective.* Mahwah, NJ: Erlbaum.

Greenberg, S., Carvey, H., Hitchcock, L., and Chang, S. (2003). Temporal properties of spontaneous speech – a syllable-centric perspective. *J. Phon.* 31, 465–485.

Isel, F., and Bacri, N. (1999). Spoken-word recognition: the access to embedded words. *Brain Lang.* 68, 61–67.

Kutas, M., Van Petten, C., and Kluender, R. (2006). "Psycholinguistics electrified II: 1994–2005," in *Handbook of Psycholinguistics*, 2nd Edn, eds M. Traxler and M. A. Gernsbacher (New York: Elsevier), 659–724.

Luce, P. A., and Cluff, M. S. (1998). Delayed commitment in spoken word recognition: evidence from cross-modal priming. *Percept. Psychophys.* 60, 484–490.

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010.

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

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., and Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychol. Rev.* 101, 3–33.

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86.

McQueen, J. M., Cutler, A., Briscoe, T., and Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Lang. Cogn. Process.* 10, 309–331.

McQueen, J. M., Norris, D., and Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *J. Exp. Psychol. Learn Mem. Cogn.* 20, 621–638.

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

Norris, D., Cutler, A., McQueen, J. M., and Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cogn. Psychol.* 53, 146–193.

Norris, D., and McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychol. Rev.* 115, 357–395.

Otten, M., and Van Berkum, J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Res.* 1291, 92–101.

Salverda, A. P., Dahan, D., and McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90, 51–89.

Schreuder, R., and Baayen, R. H. (1994). Prefix stripping re-revisited. *J. Mem. Lang.* 33, 357–375.

Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18.

Shatzman, K. B. (2006). *Sensitivity to Detailed Acoustic Information in Word Recognition (MPI Series in Psycholinguistics No. 37).* Doctoral dissertation, University of Nijmegen, Nijmegen.

Shillcock, R. C. (1990). "Lexical hypotheses in continuous speech," in *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, ed. G. T. M. Altmann (Cambridge, MA: MIT Press), 24–49.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., Yoon, K.-E., and Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592.

Van Alphen, P. M., and Van Berkum, J. J. A. (2010). Is there pain in champagne? Semantic involvement of words within words during sense-making. *J. Cogn. Neurosci.* 22, 2618–2626.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., and Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn Mem. Cogn.* 31, 443–466.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., and Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Brain Res. Cogn. Brain Res.* 17, 701–718.

Van den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn. Neurosci.* 13, 967–985.

Van den Brink, D., Brown, C. M., and Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *J. Exp. Psychol. Learn Mem. Cogn.* 32, 364–372.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., and Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *J. Exp. Psychol. Learn Mem. Cogn.* 25, 394–417.

Vroomen, J., and de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 710–720.

Vroomen, J., van Zon, M., and de Gelder, B. (1996). Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Mem. Cognit.* 24, 744–755.

Wicha, N. Y., Moreno, E. M., and 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. *J. Cogn. Neurosci.* 16, 1272–1288.

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

APPENDIX

DATA-SPLIT BASED ON SHORTLIST B SIMULATION

Introduction

When their carrier word is presented in isolation, some final embedded words will be more likely to be activated than others. To estimate the salience of our final embeddings regardless of a given sentence context we ran a simulation in Shortlist B (Norris and McQueen, 2008). Shortlist B is a Bayesian model of continuous speech recognition that evaluates multiple lexical hypotheses in a parallel and competitive fashion, taking into account several factors that may well affect the probability of a final embedded word, such as the frequency of the embedding, neighborhood density and the neighborhood frequency (including the frequency of the carrier word).

Methods

The input for the Shortlist simulation consisted of the 50 carrier words with final embeddings of our experiment set. In line with Norris and McQueen (2008) we used a lexicon of the 20,000 most frequent words in Dutch, plus a few additional words that were used in our experiment but that were not in the top 20,000. However, rather than using the dated written language frequency count of CELEX (Baayen et al., 1995), we used the recently developed *Corpus Spoken Dutch* (CGN) database. For each carrier word input (e.g., *trofee*) we tracked the word probability for the embedded word (e.g., *fee*) in each time slice (three per phoneme, see Norris and McQueen, 2008 for further technical details). For each embedded word we selected the highest probability value in any time slice starting from the onset of the embedded word. The mean maximal probability for the embedded words was 0.0065 (ranging from 0 to 0.1585)[1].

We then split the set of carrier items in two groups based on these values: carrier words with embeddings yielding a shortlist probability larger than zero (the "more salient embeddings") and carrier words with embeddings with a probability equal to zero (the "less salient embeddings"). This resulted in 19 carriers with strong embeddings (mean word probability of 0.017), and 31 carriers with weak embeddings (mean probability of zero). Then we computed the average ERP waveforms separately for these two groups.

Results

Figure 5 displays the grand average waveforms for the carrier and control words, time-locked onto the onset of the second syllable of the critical word (corresponding to the onset of the final embedding) for the parietal-central electrode, plotted separately for the carrier words with the less salient and more salient embeddings. Relative to control words, carriers with both types of embeddings elicit a significant N400 effect in the 300- to 500-ms window (less salient embeddings: $F(1, 27) = 4.65$, $p = 0.04$; more salient embeddings: $F(1, 27) = 7.76$, $p = 0.01$. However, as is evident from the graphs, carrier words with a more salient embedding show a larger and more extensive N400 effect than carrier words with a less salient embedding [$F(1, 27) = 4.35$, $p = 0.047$, 200–800 ms range].

---

[1]Note that the magnitude of the probabilities are relatively low, as a result of the presence of a strong lexical competitor that starts earlier in time (average maximal probability for the carrier words was 0.9933). According to Shortlist B final embeddings have therefore little chance to be recognized. However, the values would already be higher if Shortlist model would give words that start at the onset of strong syllables an extra boost over words that do not start at the onset of a strong syllable.