



Event-related brain potentials index cue-based retrieval interference during sentence comprehension

Andrea E. Martin^{a,*}, Mante S. Nieuwland^a, Manuel Carreiras^{a,b,c}

^a Basque Center on Cognition, Brain, and Language, Donostia-San Sebastián, Spain

^b Ikerbasque Foundation, Basque Foundation for Science, Bilbao, Spain

^c Department of Basque Philology, Universidad de País Vasco, Bilbao, Spain

ARTICLE INFO

Article history:

Received 31 January 2011

Revised 15 August 2011

Accepted 20 August 2011

Available online 5 September 2011

Keywords:

Memory

Retrieval cues

ERPs

Sustained anterior negativity

Sentence processing

Ellipsis

Agreement attraction

ABSTRACT

Successful language use requires access to products of past processing within an evolving discourse. A central issue for any neurocognitive theory of language then concerns the role of memory variables during language processing. Under a cue-based retrieval account of language comprehension, linguistic dependency resolution (e.g., retrieving antecedents) is subject to interference from other information in the sentence, especially information that occurs between the words that form the dependency (e.g., between the antecedent and the retrieval site). Retrieval interference may then shape processing complexity as a function of the match of the information at retrieval with the antecedent versus other recent or similar items in memory. To address these issues, we studied the online processing of ellipsis in Castilian Spanish, a language with morphological gender agreement. We recorded event-related brain potentials while participants read sentences containing noun-phrase ellipsis indicated by the determiner *otro/a* ('another'). These determiners had a grammatically correct or incorrect gender with respect to their antecedent nouns that occurred earlier in the sentence. Moreover, between each antecedent and determiner, another noun phrase occurred that was structurally unavailable as an antecedent and that matched or mismatched the gender of the antecedent (i.e., a local agreement attractor). In contrast to extant P600 results on agreement violation processing, and inconsistent with predictions from neurocognitive models of sentence processing, grammatically incorrect determiners evoked a sustained, broadly distributed negativity compared to correct ones between 400 and 1000 ms after word onset, possibly related to sustained negativities as observed for referential processing difficulties. Crucially, this effect was modulated by the attractor: an increased negativity was observed for grammatically correct determiners that did not match the gender of the attractor, suggesting that structurally unavailable noun phrases were at least temporarily considered for grammatically correct ellipsis. These results constitute the first ERP evidence for cue-based retrieval interference during comprehension of grammatical sentences.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Most of the time, the ease with which language is used belies the computational complexity of language processing. The correspondence between linguistic form and meaning must be processed by the brain in real-time, even when the exact relationship between the overt linguistic signal (e.g., speech, sign, or written text) and its meaning or message (what the producer intends or what the comprehender understands) is not straightforward. To make matters worse, often enough some of the components of the understood meaning go missing—sometimes they go unpronounced where they are interpreted in a sentence. Thus, language users must often compute meanings from representations that are no longer in mind, or have passed

out of the current focus of attention. In this sense, memory processes subservise online language comprehension; necessarily then, the language processing system becomes subject to properties of the human memory system. General memory variables, such as interference from other representations in memory that weaken the link between linguistically dependent items, may then determine successful retrieval (Nairne, 2002a, 2002b; Öztekin and McElree, 2007; Watkins and Watkins, 1975). Consequently, aspects of the human memory system are relevant for any neurocognitive theory of the language faculty, yet remain largely underspecified in some extant processing models (e.g., Friederici, 2002; Hagoort, 2005; *but see* Bornkessel-Schlesewsky and Schlewsky, 2009b; Lewis et al., 2006; McElree, 2006). The main aim of the current paper is to identify the source of interference during language comprehension, as indicated by event-related brain potentials (ERPs).

Vulnerability to interference presents itself especially during the formation of grammatical dependencies between words or phrases

* Corresponding author at: Basque Center of Cognition, Brain, and Language, Paseo Mikeletegi 69, 2º, 20009 Donostia-San Sebastián, Spain. Fax: +34 943 309 052.

E-mail address: a.martin@bcbi.eu (A.E. Martin).

that are separated in time by other words and phrases, a frequent state of affairs in natural language (such as English or Spanish, which is used in the current experiment):

(a) Marta se compró la camiseta y Miren cogió otra _.
Marta bought the t-shirt and Miren took another _.

To understand this sentence, the noun *t-shirt* (*camiseta* in Spanish) must be interpreted after *another* (*otra* in Spanish), so that the sentence means that *Miren took another t-shirt*. Processing the linguistic dependency between *another* and the antecedent *t-shirt* requires, at minimum, the retrieval of *t-shirt* from memory given the temporal nature of language. Words at the retrieval site may contain grammatical information, such as grammatical gender, that could help point to possible referents in the prior context. In languages like Spanish, words in certain relationships with one another must agree in grammatical gender. Thus, recovery of the antecedent is likely to be guided by gender cues at the retrieval site that point to the relevant memory representations¹ (e.g., female gender-marking on *otra* that agrees with the gender of the antecedent *camiseta*). However, such a system also entails the possibility that these cues are inadvertently mapped onto other, perhaps more recent or similar, entries in memory.

(b) Marta se compró la camiseta que estaba al lado del vestido y Miren cogió otra _.
Marta bought the t-shirt that was next to the dress and Miren took another _.

To understand this sentence, the same linguistic dependency as in (a) needs to be resolved: *another* must be understood as *another t-shirt*, not as *another dress*. In the English sentence, this occurs because *dress* is structurally unavailable² as antecedent; structural position alone inherently prevents *dress* to be the antecedent for *another*. Crucially, this is different in the Spanish sentence, where the intervening noun (i.e. the local attractor *vestido*) is not only structurally unavailable as antecedent; it also has the wrong gender compared to *another*. This is the grammatical fact that we will exploit in the current study. We compare such sentences with sentences where the intervening structurally unavailable noun does in fact match the gender of *otra*, which offers the opportunity to directly investigate effects of gender cue-based interference during retrieval. An effect of the gender of the structurally unavailable noun on the retrieval of the antecedent of *otra* would constitute evidence for cue-based interference during sentence comprehension. This would be consistent with results from basic memory research that show retrieval interference as a major determinant of retrieval failure (Anderson and Neely, 1996; Keppel and Underwood, 1962; see Nairne, 2002a for a review; Waugh and Norman, 1965), and provide further insights to the major determinants of processing complexity during language comprehension (e.g., Gordon et al., 2001; Lewis et al., 2006; McElree, 2006; McElree et al., 2003; Van Dyke, 2007).

1.1. Cue-based retrieval interference during processing of ellipsis

The linguistic construction *otro/a* in (a/b) is representative of the type of challenge language users routinely face, establishing basic sentential relations despite a missing argument noun phrase, or

¹ Extant research has consistently shown that the retrieval process recruited during sentence processing is a *direct-access* operation, without a search through irrelevant representations (Lewis et al., 2006; Martin and McElree, 2008, 2009, 2011; McElree et al., 2003). This mechanism is possible only under a *content-addressable* system, where antecedent representations are elicited from memory via their content, and are directly accessible via the cues provided at the retrieval site. In this architecture, representations with varying degrees of distinctiveness are recovered in equal time, without a search (McElree and Doshier, 1989).

² Because *dress* is in a relative clause, and nouns in relative clauses cannot be bound outside of them (Chomsky, 1981).

Table 1
Example sentences.

Condition	Sentence
1. Correct attractor-same	Marta se compró la camiseta que estaba al lado de la falda y Miren cogió otra [...] para salir de fiesta. 'Marta bought the t-shirt (fem.) that was next to the skirt (fem.) and Miren took another (fem.) to go to the party.'
2. Correct attractor-different	Marta se compró la camiseta que estaba al lado del vestido y Miren cogió otra [...] para salir de fiesta. 'Marta bought the t-shirt (fem.) that was next to the dress (fem.) and Miren took another (fem.) to go to the party.'
3.*Incorrect attractor-same	*Marta se compró la camiseta que estaba al lado de la falda y Miren cogió otro [...] para salir de fiesta. 'Marta bought the t-shirt (fem.) that was next to the skirt (fem.) and Miren took another (fem.) to go to the party.'
4.*Incorrect attractor-different	*Marta se compró la camiseta que estaba al lado del vestido y Miren cogió otro [...] para salir de fiesta. 'Marta bought the t-shirt (fem.) that was next to the dress (fem.) and Miren took another (fem.) to go to the party.'

interpreting “elided” information. Nowhere else in language does the correspondence between linguistic form and understood meaning fail as dramatically as during *ellipsis*, or omission of a word or phrase that is understood *without being pronounced*. Ellipsis functions as a “natural compression algorithm,” where silent meaning appears to function in the same complex way that overt structure does (Merchant, 2001). Importantly, its antecedents bear no special marking that might engage special memory strategies upon encountering the antecedent. Behavioral studies of ellipsis processing have shown that interpreting more material within the dependency decreased the likelihood of successful retrieval and interpretation (Martin and McElree, 2008, 2009, 2011), consistent with a cue-based interference account.

For the purpose of testing the role of cue-based retrieval interference during language comprehension, we constructed an ellipsis paradigm in Spanish based on the earlier example sentence (b), as illustrated in Table 1. In (1) and (2), the determiner *otra* can only mean the phrase *otra camiseta*, and thus introduces a new discourse referent of the same type as *camiseta*. It should be noted that *otra* does not refer to the same instance of *camiseta* (Eguren, 2010), and is therefore not a pronoun or anaphor (which, except for cataphoric pronouns, refer back to an already given referent). Eguren and Sanchez (2004) and Eguren (2010) describe *otra* in this instance as a determiner³ that allows nominal ellipsis.⁴ This determiner has no grammatical gender of its own; it agrees with the gender of the elided noun as a result of the semantic entailment of this noun, and this gender morpheme information serves in resolving the dependency (Badecker and Straub, 2002; Sturt, 2003).

One essential prediction from the cue-based retrieval interference hypothesis is that retrieval success can be modulated by the presence of other, perhaps more recent or similar representations in memory (from here on ‘attractors’; Lewis et al., 2006; McElree, 2006; Wagers

³ Eguren and Sanchez (2004) argue that *otro/a* is a determiner, not an adjective. *Otro/a* differs from French *autre*, which is demonstrably an adjective, in two significant ways. First, *otro/a* cannot appear in predicate adjective position. For example, “Elle l’a fait autrement” (‘She did it differently’) is grammatical, whereas “*Elle lo hizo autrement” is ungrammatical. Second, *otro/a* cannot be used to form an adverb, as is true of other adjectives. For example, “Elle est autre que ce qu’elle paraît” (‘She is other than what she seems’) is grammatical, but “*El es otro de lo que parece” is not.

⁴ Eguren (2010, page 440) describe evidence that suggests that our construction is elided, with examples that illustrate that, in Spanish, deleted nominals and their antecedents need not agree in number, but must in gender. For example, “Los (pl.m.) estudiantes de París y el (sg.m.) de Madrid” (‘The students from Paris and the one from Madrid’) is possible, whereas “Los (pl.m.) estudiantes de París y *la (sg.f.) de Madrid” (‘The male students from Paris and the female one from Madrid’) is not possible. Eguren (2010) notes that although agreeing gender morphology on a remnant of ellipsis is not a crucial licensing condition for empty nominals (because some remnants, due to their particular lexical features, do not bare gender, e.g., *grande*), “nominal ellipsis allows a mismatch in number between the elided element and its antecedent, but requires a match in gender features.”

et al., 2009; Van Dyke, 2007; Van Dyke and Lewis, 2003; Van Dyke and McElree, 2006). To illustrate this paradigm from the example in Table 1, the intervening attractors the feminine *falda* “skirt” in (1) and the masculine *vestido* “dress” in (2) have recently been processed when *otra* is encountered. The gender of these attractors matches the feminine *otra* in gender (*falda* in 1) or not (*vestido* in 2). In either case, *otra* refers to its intended antecedent *camiseta*, because the attractor is not in a structurally licensed position to be its antecedent. The gender of the attractor noun is thus irrelevant if the parsing process only takes into account the correct antecedent. Under a cue-based retrieval interference account, however, the gender of the attractor, as well as structural information, might nevertheless impact processing of ellipsis. More specifically, the mismatching gender of the attractor noun in (2) might incur processing costs in this grammatically completely correct sentence.

To investigate the effect of retrieval failure during ellipsis processing, thereby establishing a benchmark for the abovementioned comparison, we also constructed two ungrammatical conditions where *otra* cannot refer to the antecedent *camiseta* at all because of gender-mismatch (3 and 4). In this way, we crossed the match of the gender of *otra/a* to the antecedent with a match in gender to the intervening attractor noun. The manipulation of these two factors, grammaticality and the gender of the attractor noun, allows us to observe two things: first, the ERP signature of retrieval failure due to grammatically incorrect ellipsis, and, second, whether this signature is modulated by a structurally unavailable but similar attractor. In the following section, we outline how these processes might play out in the observed ERP responses.

1.2. Predictions for event-related brain potentials

Previous work on cue-based retrieval interference has solely relied on behavioral measures (Lewis et al., 2006; McElree, 2006; Van Dyke, 2007; Van Dyke and McElree, 2011; but see Xiang et al., 2009, for a related ERP study). In our study, we used ERPs to index retrieval operations while participants read sentences for comprehension. The basic advantage of ERPs is that their multidimensional nature provides potentially relevant information that cannot be obtained with behavioral response tasks. ERPs can reflect whether the brain is differentially sensitive to certain linguistic manipulations and provide clues to the identity of the cognitive event at hand (by virtue of possible differences in their polarity, morphology and scalp-distribution), and can do so on the timescale needed to track retrieval during sentence comprehension (for review, see Kutas et al., 2006; Van Berkum, 2004).

First and foremost, based on the existing ERP literature we predicted that, compared to the case where there is correct gender agreement, grammatically incorrect ellipsis should elicit a P600 effect, an enhanced broadly distributed centroparietal positivity that is most commonly associated with syntactic computations (e.g., Hagoort et al., 1999). P600 effects are not only reliably elicited by outright syntactic violations (e.g., determiner-noun gender agreement violation; Barber and Carreiras, 2005; Wicha et al., 2004; or subject-verb person agreement violations; Mancini et al., 2011; Molinaro et al., 2011) but also by denial of syntactic preference or expectation, i.e., by constructions that are well-formed but whose syntactic properties do not fit the analysis currently being pursued or that was previously expected (e.g., Carreiras et al., 2004; Van Berkum, 2009). Given that the gender of the determiner (*-o/-a*) was grammatically correct or incorrect given that of the antecedent (*camiseta*), grammatically incorrect determiners in our study may thus initially be perceived as a morphosyntactic “dead end”, triggering re-analysis (Gouvea et al., 2010; Kaan and Swaab, 2003; Osterhout and Holcomb, 1992). Perhaps for similar reasons, pronouns with unexpected or incorrect gender-marking will also evoke P600 effects (e.g., Nieuwland and Van Berkum, 2006; Osterhout and Mobley, 1995; Van Berkum et al., 2007; Xiang et al., 2009).

However, the electrophysiological literature on resolving sentence- and discourse-level referential dependencies points to an alternative possibility. Referential ambiguity, the difficulty in interpretation

which arises whenever language users are unable to converge on a unique referent (e.g., noun phrase anaphor or pronoun in the extent literature) from multiple candidates in the discourse, is known to evoke a sustained anterior negativity (the Nref; Van Berkum et al., 1999). Possibly related sustained negativities have been reported when participants interpret a pronoun with an unexpected gender as referring to an unmentioned person (see Osterhout and Mobley, 1995; King and Kutas, 1997; see also Kutas et al., 2000; Nieuwland et al., 2007b). Importantly, it stands to argue that, while an elided constituent is not a pro-form or a trace (Johnson, 2001), grammatically incorrect determiners during ellipsis may similarly constitute a referential (rather than only a strictly morphosyntactic) problem, when it is insufficient to retrieve a unique antecedent. Such is also assumed in memory-based processing models of anaphor resolution (e.g., Gerrig and O'Brien, 2005; McKoon and Ratcliff, 1998; Myers and O'Brien, 1998): antecedents that have sufficient features in common with an anaphor are automatically activated or elicited from memory. However, when no antecedent resonates to sufficient degree (e.g., when there is little featural overlap between anaphor and antecedent, when the antecedent is out of the focus of attention, or when there are strong attractors), this state of affairs is said to trigger additional “episodic” retrieval processes, either to resolve interference or possibly to initiate the recovery of additional information that might help to infer the most plausible referent. Sustained negativities that have been reported in the literature to referentially problematic expressions might be an electrophysiological correlate of these additional retrieval processes (Van Berkum, 2009), and may be related to sustained frontal shifts that are evoked by linguistically complex structures that are thought to tax working memory (see Kutas et al., 2006 for a review). Because grammatically incorrect ellipsis could elicit similar retrieval processes when dependencies over relatively long distances are compromised, they might elicit a sustained anterior negativity in addition to, or instead of, a P600 modulation.

Our critical manipulation was designed to tap into retrieval interference from a recently occurring attractor. To remind the reader, *Correct Attractor-same* is the case where the gender of *otra* matches the gender of the to-be retrieved antecedent, and the potentially interfering noun has the same gender as the antecedent. *Incorrect Attractor-same* is where the potentially inferring noun again has the same gender as the antecedent, but this time the gender of *otra* mismatches that of the antecedent. The remaining two conditions in the 2x2 design are where the potentially interfering noun does not share gender with the antecedent. We therefore predicted an interaction between grammaticality (gender match between antecedent and determiner) and the gender of the attractor noun, as reflected in a modulation of the neural signature of ongoing retrieval operations: grammatically incorrect determiners would elicit a smaller P600 effect and/or Nref effect in the presence of an attractor that mismatches the gender of the determiner, i.e., ((*Incorrect Attractor-same* “camiseta-falda-otra” > *Correct Attractor-same* “camiseta-falda-otra”) > (*Incorrect Attractor-different* “camiseta-vestido-otra” > *Correct Attractor-different* “camiseta-vestido-otra”)). This pattern of results may manifest in two different ways. The first possibility is that because the attractor in the *Incorrect Attractor-different* condition might, at least temporarily, function as a valid target, interference may lead to a reduced effect to this condition compared to the correct conditions. This is prediction is predicated on finding of a main effect of grammaticality which indexes successful retrieval and interpretation of the antecedent. However, another possibility, albeit it perhaps less plausible than the first one because agreement attraction effects have not been observed for comprehension of grammatically correct sentences (see Phillips et al., 2010), is that the interaction comes about from an increased P600 response to the *Correct Attractor-different* condition. This would provide unique evidence that retrieval interference can ‘cut both ways’, that mismatching gender information disrupts processing even in a completely grammatically licensed situation—and not only through local coherence in an ungrammatical situation (Tabor et al., 2004).

2. Method

2.1. Materials and pretests

One hundred and twenty sentence quadruplets were constructed. Each quadruplet featured the grammaticality manipulation and the presence of an attractor noun in a relative clause. Thus, the sentences per quadruplet differed only in two ways: (1) in the gender morphology of the critical word (*otro* or *otra*; the remaining determiner of the elided noun phrase) and (2) the presence of an antecedent gender-matching or -mismatching noun phrase that served as an agreement attractor (see Table 1). The gender of the determiner (-o/-a) was grammatically correct or incorrect given that of the antecedent (*camiseta*). The ellipsis site occurred in the context of an intervening noun (*la falda/el vestido*) that either matched or mismatched the antecedent gender (the *Attractor-same* and *Attractor-different* condition respectively, indexing whether the gender of the attractor was the same as that of the antecedent). Neither the critical word nor the next word was ever sentence-final, and the mean log frequency of the attractor nouns and antecedent nouns was matched across conditions with no reliable differences between them (mean log frequency, antecedent = 1.12, same gender attractor = 1, different gender attractor = 1.03; Davis and Perea, 2005). Grammatical gender of the antecedent was balanced across items.

We collected verb–noun plausibility norms to examine whether the antecedent and attractor nouns were equally plausible as objects of the verb in the second clause (e.g., of *cogió*). Twenty students from the University of the Basque Country community completed a questionnaire in which they read verb–noun pairs (featuring the verb which was used in the second clause of a given experimental item and the corresponding antecedent and attractor nouns for that same item). Each of the three pairs for an item was pseudo-randomly mixed with those from other items, such that no item's pairs appeared in succession. The participants were instructed to rate the plausibility of the verb–noun pairs on a five-point scale (1 = Implausible, 5 = Quite Plausible). The mean plausibility rating for the verb plus antecedent noun was 4.37 (SD = .44) on this scale, while the mean plausibility rating for the same gender attractor was 4.22 (SD = .42), and the mean plausibility rating for the different gender attractor was 4.20 (SD = .46). Pairwise comparisons showed that both the attractors were reliably less plausible as an argument of the second clause verb than the antecedent noun (mean differences were .15 and .17 scale units, respectively; same gender as antecedent: $t = 4.72, p < .001$; different gender as antecedent: $t = 4.60, p < .001$), but not different from each other in plausibility ($t = .64, p < .47$). Importantly, this result excludes the possibility that observed differences are due to verb argument plausibility differences between gender-matching and -mismatching nouns, and assures that antecedents were always more plausible than either attractor.

Additionally, since no Spanish corpora allow the extraction of relative lexical frequencies concerning verbs, we estimated the relative frequency of the verb–noun pairs by collecting Google hits to several strings representing lemmata with and without wildcards in between verb and noun; paired *t*-tests between the hits for antecedents and attractors nouns in the context of the embedded verb found no significant differences between antecedent and attractor nouns, or between attractor nouns for any of the strings searched. (antecedent vs. same gender attractor: $t = 3.63, p < .18$; antecedent vs. different gender attractor: $t = 1.35, p < .41$; same vs. different gender attractors: $t = .36, p < .79$).

We created four counterbalanced lists so that each item appeared in only one condition per list, but in all conditions equally often across lists. Within each list, items were pseudo-randomly mixed with 60 filler sentences to limit the succession of identical sentence types while matching trial types on average list position. The filler sentences were of identical initial structure except they resolved without ellipsis, so that

participants could not predict ellipsis based on initial sentence form (“Elena cortó una margarita que estaba junto a una rosa y Ana compró una orquidea para tenerla en su casa.”—‘Elena picked a daisy that was next to a rose and Ana bought an orchid to have it in her house.’). The fillers were all grammatical, contained no agreement errors, nor *otro/a*.

2.2. ERP experiment

2.2.1. Participants

Twenty-two right-handed students (5 males; mean age = 22.3 years) from the University of the Basque Country gave written informed consent. All were native Castilian Spanish speakers, and none had neurological or psychiatric disorders or had participated in the pretests. All had normal or corrected-to-normal vision.

2.2.2. Procedure

Participants were instructed to silently read sentences that were centered on a computer monitor in black letters on a light gray background, while minimizing movements, eye movements and blinks. A trial began with a 1 second fixation cross. Then the sentence was presented word-by-word, with variable word-length scaled presentation except for pre-critical, critical and post-critical words, which were presented for 300 ms with 200 ms inter-stimulus-interval. For the variable length presentation, we used number of letters*30 + 190 ms, with a maximum of 430 ms per word (e.g., Otten and Van Berkum, 2008). The interstimulus interval was also 200 ms for these words. Sentence-final words were presented for 800 ms. The sentence-final word was either followed by a fixation cross upon which participants could blink as needed and then proceed on to the next sentence, or by a yes/no comprehension question that probed their comprehension related to the events and referents in the sentence (e.g., Sentence: “Marta se compró la camiseta que estaba al lado del vestido y Miren cogió otro igual para salir de fiesta.” Question: “¿Cogió algo Miren para salir de fiesta?”—‘Did Miren take something to go to the party?’). There were 60 comprehension questions of which 50% required a ‘yes’-response by a left- or right-hand button-press, and they were evenly distributed to occur after fillers and experimental sentences. Participants had very good performance on these comprehension questions (mean percent correct across all conditions and fillers = 88%, participant range = 70–96%; mean percent correct by condition ($n = 20$ questions per condition over subjects): *Correct Attractor-same* condition (camiseta/falda/otra) = 83%, *Correct Attractor-different* condition = 91%, *Incorrect Attractor-same* condition = 86%, *Incorrect Attractor-different* condition = 91%). We performed a 2(Grammaticality: Correct, Incorrect) × 2(Attractor: Antecedent-matching, Antecedent-mismatching) repeated measures ANOVA on the mean accuracy over subjects and found no reliable differences. Participants were given a practice session and six experimental sessions separated by short breaks. Total time on task was approximately 50 min.

2.2.3. Electroencephalogram recording

The electroencephalogram (EEG) was recorded from 29 standard scalp electrodes (referenced to the left mastoid; one additional right mastoid electrode and four electro-oculogram electrodes), amplified (band-pass filtered at 0.01–30 Hz), and digitized at 250 Hz. Impedance was kept below 5 kOhm for EEG electrodes. The EEG was re-referenced to the average of the left and right mastoid electrode offline. We corrected for ocular artifacts using a procedure based on independent component analysis (see Jung et al., 2000), and the single-trial waveforms were automatically screened for amplifier blocking and muscle artifacts over 1400 ms epochs (starting 200 ms before critical word onset). Two participants were excluded due to excessive artifacts (mean trial loss > 30%). For the remaining 20 participants, average ERPs (normalized by subtraction to a 200 ms pre-stimulus baseline) were computed over artifact-free trials for critical words in all conditions (mean trial loss across conditions = 9.5%, range = 9–10%).

2.2.4. Statistical analysis

We performed a 2(Grammaticality: Correct, Incorrect) × 2(Attractor: Same, Different) repeated measures analysis of variance (ANOVA) using average amplitude repeated across all EEG electrodes in three time windows (100–400, 400–1000, and 1000–1200 ms windows) after critical word onset. We choose the 400–1000 ms time window because this window is suitable to examine modulations of the P600 induced by gender mismatch of referring expressions as well as of the Nref, which are both relatively extended ERP effects (see [Nieuwland and Van Berkum, 2006](#)). We choose the 100–400 ms time window to explore relatively early effects of grammatically incorrect cues although we did not predict to find effects in this time window. We examined the 1000–1200 ms time window because it was the latest time window in which the ERP waveforms would be free from potential sentence wrap-up effects.

To test the distribution of observed effects, we divided the electrodes into those posterior or anterior to the central cross-line, and followed-up by performing a 2(Grammaticality: Correct, Incorrect) × 2(Attractor: Same, Different) × 2(Distribution: Anterior, Posterior) repeated measures ANOVA, and additional 2(Grammaticality: Correct, Incorrect) × 2(Attractor: Same, Different) × 2(Hemisphere: Left, Right) repeated measures ANOVA.

3. Results

Critical words elicited a sustained negativity in the Incorrect conditions compared to the Correct conditions (see [Figs. 1 and 2](#)). This negativity started at about 300–400 ms after critical word onset, dissipated around 1100–1200 ms, and had a broad central distribution. Statistical analysis revealed that Incorrect sentences elicited reliably more negative voltage compared to Correct sentences in the 400–1000 ms window (all but four participants showed a more negative voltage for Incorrect sentences), resulting in a significant main effect of Grammaticality (see [Table 2](#)). Consistent with the broadly distributed nature of the sustained negativity, this effect did not differ between hemispheres nor was it different for anterior versus posterior electrodes. No main effect of Grammaticality was found in the other time windows. No main effect of Attractor was found in any of the time windows (see the Appendix, for grand average ERPs collapsed across the Attractor levels, and grand average ERPs collapsed across the Grammaticality levels), but there was a reliable Grammaticality × Attractor interaction in the 400–1000 ms window. Following up on this interaction, pairwise comparisons revealed that the mean amplitude difference between the Correct conditions was reliable, with the Attractor-different condition being more negative (see [Table 3](#)). The mean amplitude difference between the Attractor-same conditions was also reliable, with the Incorrect condition being more negative (see [Table 3](#)). No other comparisons in this time window were reliable. We wish to refer to the reader to the Appendices for additional figures and complementary analyses.

In summary, we found a sustained negativity to Incorrect conditions compared to Correct conditions, and this negativity was modulated by the gender of a local attractor in the Correct conditions. Both the sustained negativity to Incorrect conditions and the modulation of this effect by the agreement attractor were observable in the ERP waveform at about 400–600 ms window after critical word onset, hence before the onset of the next word (we refer the reader to Appendix 2 that includes more focused statistical analyses on a selection of central electrodes using 200 ms time windows). When the determiner was grammatically correct but there was a locally mismatching noun, this elicited a reliably more negative going ERP response than in the context of a locally matching attractor noun, suggesting that when the attractor does not match the retrieval cue, this also impacts processing, even if the antecedent is retrievable and the sentence grammatical. This strongly suggests that the gender of the intervening nouns was impacting retrieval processes. Concomitantly, the negativity

to incorrect determiners was slightly reduced in the presence of a matching attractor, although there was no fully significant difference between Incorrect conditions (but see Appendix 2) or between Antecedent-different conditions.

4. Discussion

Our results present two novel contributions: (1) grammatically incorrect gender-marking during ellipsis elicits a broadly distributed, sustained negativity in the ERP waveform compared to correct gender-marking, and, (2) this neural response was modulated by the gender of the attractor noun that intervened between the ellipsis site and antecedent in the grammatical conditions. Regarding (1), the predicted main effect of Grammaticality supports the conclusion that participants are reliably dissociating between grammatical and ungrammatical ellipsis. That this reliable difference is due to Grammaticality strongly implies that, in the grammatical cases, ellipsis antecedents have been successfully retrieved and interpreted, but in ungrammatical cases, retrieval failure or difficulty has occurred. Importantly, the main effect shows that participants are not simply interpreting the attractor noun as the antecedent. Our second finding, (2) above, presents evidence for an asymmetry between cue-based interference in grammatical and ungrammatical cases. Our results have not borne out our first, and perhaps most intuitively plausible prediction regarding the interaction between grammaticality and the attractor—that the *Incorrect Attractor-different* (*camiseta-vestido-otro*) condition might temporarily function as an antecedent and lead to a reduced effect in this condition compared to the correct conditions or the *Incorrect Attractor-same* (*camiseta-falda-otro*) condition. Instead, our results confirmed our second interaction prediction—that the interaction comes about from a change in response to the *Correct Attractor-different* condition. Mismatching gender information from the attractor noun disrupts processing in a completely grammatically licensed situation—not in an ungrammatical one—and stems from a noun that is structurally unavailable. We take these ERP results as evidence for cue-based retrieval interference during sentence comprehension. After outlining the current results and their implications for cue-based retrieval in more detail below, we will briefly discuss how our findings could lend some insight to neurocognitive accounts of sentence comprehension.

In our study, determiners could be grammatically incorrect with regard to the gender of the antecedents. Critically, under any neurocognitive account of sentence comprehension ([Bornkessel-Schlesewsky and Schlewsky, 2009a](#); [Friederici, 2002](#); [Hagoort, 2005](#)), this outright syntactic violation should have evoked a P600 effect, which is the ERP effect most commonly associated with syntactically problematic or dispreferred linguistic expressions ([Hagoort et al., 1999](#); [Van Berkum et al., 2007](#); [Kutas et al., 2006](#)). Moreover, a multitude of electrophysiological studies have reported P600 effects for related gender mismatch violations across different languages ([Barber and Carreiras, 2005](#); [Molinero et al., 2011](#); [Wicha et al., 2004](#); [Van Berkum et al., 2007](#); [Xiang et al., 2009](#)). The fact that the main manipulation in the current study elicited a sustained negativity rather than a P600 effect suggests that the processing consequences of incorrectly gender-marked ellipsis can be qualitatively different from the repair or reanalysis processes as assumed to be indexed by P600 modulations, and perhaps that participants did not treat failure to retrieve an antecedent as a consequence of the gender agreement violation. That is, it is possible that they simply failed to retrieve an antecedent, at least temporarily, rather than directly correct or repair the gender of the retrieval cue.

As we see it, this result may be retraced to at least two important differences between the current design and those that have reported P600 effects to morphosyntactic gender violations. First, just as pronouns typically refer to things given in the discourse, ellipsis can only refer to concepts entailed by the discourse (*ipso facto* given), yet *otra* necessarily introduces a new discourse referent, crucially one of the same type as its antecedent *camiseta*. Thus, referential context comes into play in our design. Second, the dependency between

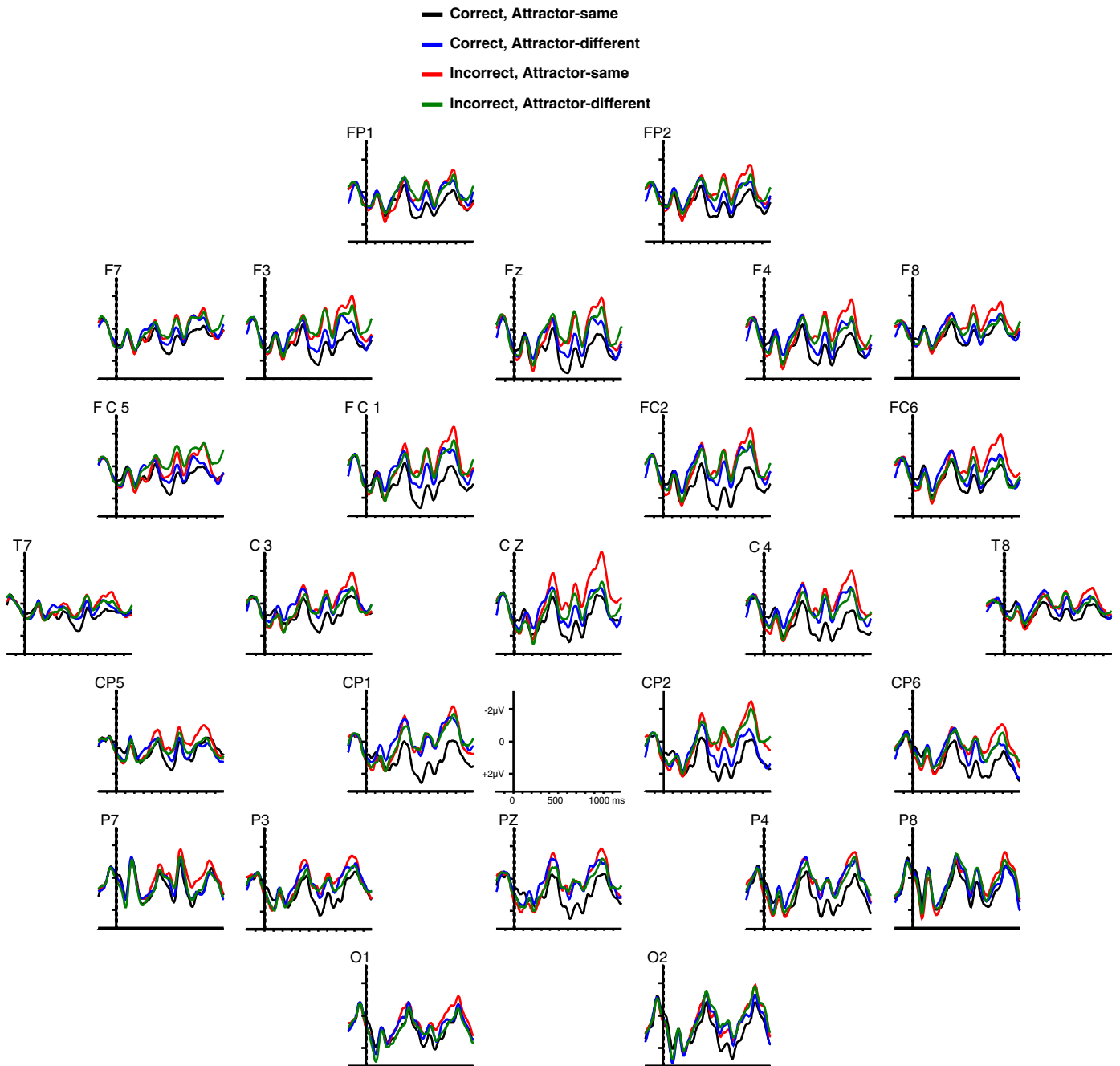


Fig. 1. Grand averages at all electrodes elicited by critical words in the four conditions: Correct attractor-same (*camiseta-falda-otra*; black lines); Correct attractor-different (*camiseta-vestido-otra*; blue lines); Incorrect attractor-same (*camiseta-falda-otro*; red lines); Incorrect attractor-different (*camiseta-vestido-otro*; green lines). Note that in this and the following figure, negativity is plotted upwards and that the waveforms are filtered (5 Hz high cut-off, 12 dB/oct) for presentation purpose only.

agreeing elements in our study is probably longer than in most examples of agreement computation in the processing literature. With respect to the first difference, an incorrectly gender-marked determiner in our study perhaps constitutes not simply a morphosyntactic error but also a referentially problematic situation. Problems with referential interpretation, in particular referential ambiguity when one anaphor maps on to two possible referents, are known to elicit sustained negativities (the Nref, see Nieuwland et al., 2007a; Nieuwland and Van Berkum, 2006; Van Berkum et al., 1999). However, sustained negativities have also been reported for pronouns whose gender does not match the gender of the only structurally available referent(s) in the sentence, but that are resolved by invoking a new entity. For example, Osterhout and Mobley (1995) reported that whereas

participants who judged such sentences to be grammatically incorrect (i.e., 'he' should have been 'she') showed a P600 effect, a small number of participants who judged "The aunt heard that he..." to be grammatically correct (by invoking a new male referent for 'he', e.g., the aunt's nephew) showed a sustained negativity over Wernicke's area and at frontal electrodes (see also King and Kutas, 1997). It must be noted that the currently observed sustained negativity does not have a scalp distribution that is fully consistent with the Nref or with the effects reported by Osterhout and Mobley (1995), suggesting at least partially different underlying neural generators, and therefore the comparison between these effects comes with an important caveat. If, however, these effects nevertheless have some amount of functional overlap, this could mean that when grammatically

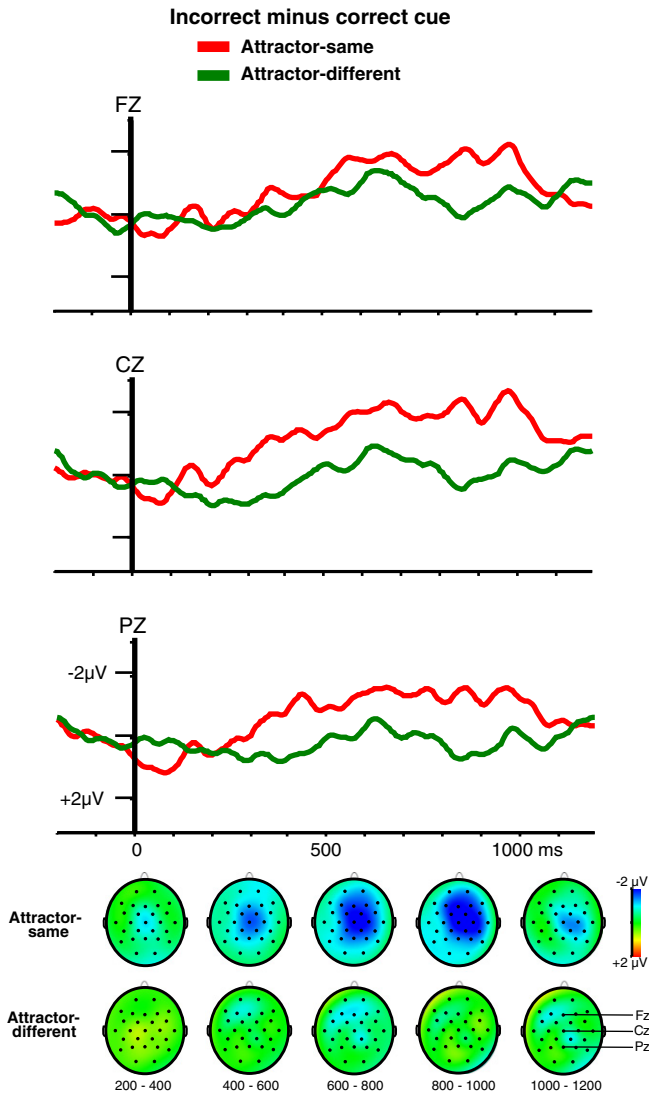


Fig. 2. Difference waves at Fz, Cz and Pz for incorrect minus correct in the context of a same-gender attractor (red lines) or a different-gender attractor (green lines), and corresponding scalp distributions for five adjacent 200 ms time windows between 200 and 1200 ms after critical word onset.

incorrect determiners in our paradigm did not readily elicit the intended referent, participants take this lack of an elicited antecedent as uncertainty about the previous referential context, perhaps trying to recollect a matching antecedent that has been forgotten or that occurred somewhere else in the discourse. However, we must note that such an account seems hard to reconcile that we

Table 2
F-values and MSE from ANOVA on average over all electrodes between 100–400 ms, 400–1000 ms, and 1000–1200 ms (df = 1, 19).

		Time (ms)		
		100–400	400–1000	1000–1200
Grammaticality	F	0.28	10.01**	2.78
	MSE	1.15	1.15	1.61
Attractor	F	1.89	1.21	0.10
	MSE	0.91	1.01	1.47
Grammaticality × Attractor	F	0.67	4.63*	0.20
	MSE	1.24	1.18	2.58

** $p < .01$.
* $p < .05$.

Table 3
Pairwise comparisons on average voltage (μV) over all electrodes between 400 and 1000 ms (df = 1, 19).

	Incorrect	Correct	Incorrect–Correct (95% CI)
Attractor-same	–0.35	0.93	–1.28** (–0.53 to –2.04)
Attractor-different	–0.08	0.16	–0.24 (0.43 to –0.91)
Attractor-different– attractor-same (95% CI)	0.28 (0.93 to –0.38)	–0.77* (–0.04 to –1.50)	

** $p < .01$.
* $p < .05$.

found stronger effects of the attractor for grammatically correct determiners than for incorrect determiners.

The second difference between the current study and other agreement computation studies is the distance between agreeing elements. This factor of greater distance could be somewhat related to the referential difference discussed above—the uncertainty about the referential context could also be increasing as distance between the antecedent and ellipsis site increases. The reliable match of the retrieval cue to the antecedent may decrease as distance increases, due to the processing of potentially interfering representations, which in turn render the cue less diagnostic to a unique antecedent. These two factors may give rise to a situation similar to when there is no unique referent available. Controlled retrieval processes to overcome such a referentially problematic situation may be what are reflected by the referentially-induced sustained negativity (e.g., Van Berkum et al., 1999). The fact that a superficially similar, though smaller negativity was elicited by correct determiners when preceded by a mismatching attractor suggests at least momentary interference during the comprehension process, and the intervening attractor possibly may have been considered as a potential antecedent (see also Phillips et al., 2010, for discussion, and cf. Xiang et al., 2009). Thus, in this case, resolving ellipsis may have relied more on overcoming retrieval interference from inconsistent but retrieval-relevant information. In the section below we relate our findings to cue-based retrieval accounts in more detail.

4.1. Cue-based retrieval and cue-diagnosticity in sentence comprehension

Our results suggest that the successful retrieval of previously processed words and phrases, a routine event during language processing, is a function of the degree to which cues at the retrieval site match or ‘pick out’ a unique antecedent in memory. This degree of match, or *cue-diagnosticity*, can be defined as the retrieval cue’s resonance the antecedent relative to its match to other items in memory (Nairne, 2002b). More specifically, the probability that a given antecedent will be successfully retrieved by a given cue, is determined by the resonance between cue and antecedent divided by the resonance between that cue and other items in memory, modulated by salience and other properties of the memory trace (Gillund and Shiffrin, 1984; Nairne, 2002b, see Van Dyke and McElree, 2011 for a discussion of cue combinatorics). To illustrate, a clear-cut example of insufficient cue-diagnosticity is referential ambiguity (e.g., ‘the girl’ in a context with two girls), and a perhaps more nuanced case comes from the following sentences, from Van Dyke and Lewis (2003):

- (c) The worker was surprised that the resident who said that the warehouse was dangerous __was complaining about the investigation.
- (d) The worker was surprised that the resident who was living near the dangerous warehouse __was complaining about the investigation.

Here, the syntactic cues at the gap site generate interference in (c) compared to (d) because in (c) both *the resident* and *the warehouse* are subjects, while in (d) *the warehouse* is the object of a preposition. Indeed, example (c) showed reliably slower reading times at the verb following the gap site than (d), suggesting that shared syntactic features of the nouns within the dependency affected processing at the verb (Van Dyke and Lewis, 2003; Van Dyke, 2007).

The idea that similarity of other items in memory affect retrieval makes the prediction that identical gender of *camiseta* and *falda* should have negatively affected processing, as gender morphology on the cue is not diagnostic to a unique antecedent. However, if structural position is also taken into account, arguably *otra* is in fact diagnostic to *camiseta*. Thus, how diagnostic a cue is to an antecedent depends on exactly how the retrieval cues are encoded, ranked, and combined, which is only beginning to be investigated (Van Dyke and McElree, 2011). In our study, it appears that morphosyntactic gender information and structural information combine as retrieval cues, and that given sufficient resonance between the cues and the contents of memory, inconsistent gender information must be overridden. However, in the absence of sufficient resonance, gender information does not intrude. This characterization also seems satisfactory to explain the failure of *otro* as a retrieval cue; there is no entry in memory that *otro* picks out—except in the Incorrect Attractor-different case, where *vestido* occurs prior to *otro*. This condition may induce a local attraction between cue and illicit antecedent that does not occur when there is a matching and also *licit* item in memory, although our results do not bear evidence for this hypothesis (but see Wagers et al., 2009). Cue-diagnostics-based attraction thus offers an explanation for the difference between the two Correct conditions—if the attractor noun does not match the features of the antecedent, then local mismatch between the attractor and the correct determiner also impacts ongoing neural processing—as well as for the reduction of the sustained negativity for the Incorrect condition in the presence of a local *cue-matching* attractor.

To our knowledge, our results form the first observation in the language comprehension literature that grammatically unavailable nouns intrude on the resolution of referring expressions in grammatically correct sentences (for discussion, see Sturt, 2003; Phillips et al., 2010). This pattern seems related in spirit to instances of so-called “local coherence,” where local constraints can sometimes outweigh global ones, producing “illusions” of grammaticality (Phillips et al., 2010; Tabor et al., 2004). A similar phenomenon of attraction errors during subject–verb agreement production has been well-documented in production (e.g., Bock and Miller, 1991; Staub, 2009) and in comprehension in certain cases (Pearlmutter et al., 1999; Wagers et al., 2009). Anaphor resolution has also shown vulnerability to interference that could be attributed to differences in cue-diagnostics (Badecker and Straub, 2002; Sturt, 2003). Accordingly, the existence of cue-based attraction supports a processing account where the resonance between recent neural states or recent information and the cues at retrieval, as well as the resonance of those cues with the antecedent, comprise the probability of successful retrieval, and in turn, successful language comprehension.

4.2. Potential implications for neurocognitive accounts of sentence comprehension

It is important to note that while we believe there is a crucial role for cue-based retrieval and cue-diagnostics in psycholinguistic and neurocognitive theories of language, this role has not been fully spelled out in extant models and thus it is not clear how to adapt the predictions of these models to make contact with our results. However, we will attempt to argue that the architecture of most models implies an important role for cues and thus cue-diagnostics.

Within hierarchical models such as the neurocognitive model of auditory sentence comprehension (Friederici, 2002), retrieval cues

could shape many aspects of processing (from as early as word identification to later stages like structure building and construction of semantic relations), if they are conceived of as multiplicative composites of information maintained in the current focus of attention that retrieve or *point* to representations as processing requires (Gillund and Shiffrin, 1984; Lewis et al., 2006; Martin and McElree, 2008; McElree, 2006). In this way, as general “working memory” appears alongside the online language processing stream in the model, cues could also serve as interfaces between nodes in the different processing streams, perhaps with degree of cue-diagnostics capturing some of the two-stage processing bottlenecks that have been classically observed in the literature. Cues also could play an important role in the diametric Memory Unification Control framework (MUC; Hagoort, 2005). Here retrieval is subsumed in the Memory component, so a role for cues is already implied, though again, if cues are conceived as the contents of the current focus of attention, cue-diagnostics could play an important role not only during retrieval but also during Unification, perhaps being part of the mechanism—conceivably a pointer—that determines which information types resonate and are unified. Lastly, within the more typologically-inspired Extended Argument Dependency Model (eADM; Bornkessel and Schlesewsky, 2006; Bornkessel-Schlesewsky and Schlesewsky, 2009b) cues and cue-diagnostics could again come into play in each processing stage, certainly pertaining to the “centerpiece” operations, Compute Prominence and Compute Linking. Furthermore, the notion that cues and cue-diagnostics shape processing fits in naturally with a typological framework, where the model depends on specification of scales of information types to provide a processing stream that can account for both similarity and variation in the features of human languages (for similar arguments, see Bates and MacWhinney, 1989; Bornkessel-Schlesewsky and Schlesewsky, 2009b; Van Valin, 2005). Though the eADM gives clear, stage-based predictions of ERP components and their processing loci, at the moment neither this model nor any other predicts sustained shifts for dependency-processing where the dependent element is not “maintained.”

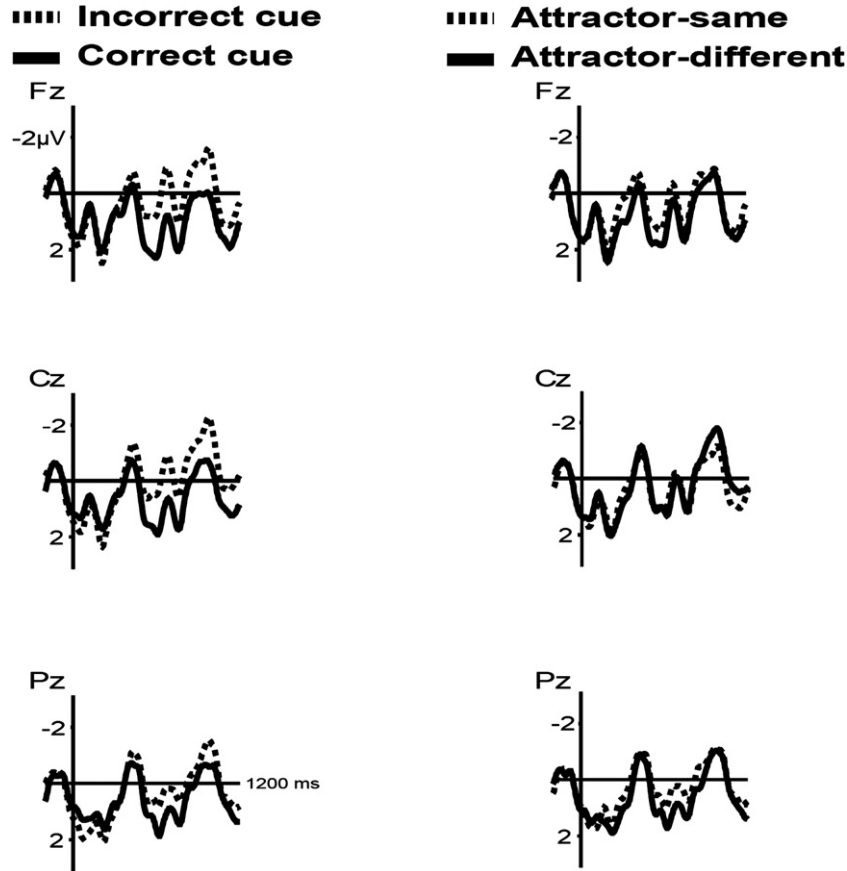
4.3. Conclusions

Successful language use requires access to products of past processing during current processing within an evolving discourse, such as the retrieval of a noun phrase antecedent to resolve ellipsis. In this study, we provided ERP evidence that this retrieval process can be subject to interference from intervening noun phrases, even if the latter are structurally unavailable for ellipsis and the sentence is grammatically correct. Importantly, our results are consistent with cue-dependent direct-access retrieval of content-addressable representations underlying online language comprehension, and they point towards further incorporation of these computational principles and architectures into extant models of language processing.

Acknowledgments

We are grateful to Saioa Larraza and Javi Miqueleiz for assistance with material creation and data collection. We are also grateful to Eneko Anton and Eri Takahashi for assistance with data collection. We thank Ivan Ortega-Santos and Jason Merchant for their gracious consultations regarding the linguistic analysis of our stimuli and we thank Gerry Altmann, Phil Monahan, and two anonymous reviewers for very helpful feedback on an earlier version of this manuscript. MC was partially supported by Grant CONSOLIDER-INGENIO 2010 CSD2008-00048 from the Spanish Ministry of Science and Innovation. MSN was partially supported by a Plan Nacional research grant from the Spanish Ministry of Science and Innovation (PSI2010-18087).

Appendix 1



Appendix 1. Grand averages at Fz, Cz, and Pz electrodes corresponding to the main effect of grammaticality (left graph; Incorrect = dashed lines, Correct = solid lines) and of Attractor (right graph; Attractor-same = dashed lines, Attractor-different = solid lines).

Appendix 2

In order to better examine the time course of the interaction, we conducted follow-up analysis on the average amplitude of a central electrode selection in 5 consecutive 200 ms time windows starting

at 200 ms after critical word onset (see Table 4). In support for this electrode selection, we compared effects of Grammaticality at central electrodes (Cz, FC₁, FC₂, CP₁, CP₂) with those at peripheral electrodes (FP₁, FP₂, F8, FC₆, T8, CP₆, P8, O₂, O₁, P7, CP₅, T7, FC₅, F7) between 400 and 1000 ms post-stimulus onset, which resulted in a statistically

Table 4

F-values and MSE from moving window ANOVAs on average voltage over Central electrodes [Cz, FC₁, FC₂, CP₁, CP₂] in 200 ms time windows, (df = 1, 19).

		Time (ms)				
		200–400	400–600	600–800	800–1000	1000–1200
Grammaticality	F	0.00	6.06*	12.33**	8.33**	5.11*
	MSE	2.98	2.46	3.02	3.05	2.75
Attractor	F	3.03	0.93	3.15	0.65	0.27
	MSE	1.60	2.17	2.46	1.90	2.42
Grammaticality × Attractor	F	2.29	5.39*	2.93^	8.09**	0.52
	MSE	2.51	1.69	2.37	2.48	3.58
Pairwise comparisons:		200–400 ms	400–600 ms	600–800 ms	800–1000 ms	1000–1200 ms
Mean diff. (95% CI)						
Attractor-same:		−0.53 (−1.54 to 0.48)	−1.54** (−2.52 to −0.55)	−1.95** (−3.08 to −0.83)	−2.13** (−3.38 to −0.88)	−1.14^ (−2.51 to 0.23)
Attractor-different:						
Incorrect–Correct		0.54 (−0.63 to 1.72)	−0.19 (−1.11 to 0.73)	−0.78 (−1.82 to 0.27)	−0.13 (−1.05 to 0.80)	−0.53 (−1.48 to −0.41)
Correct: attractor diff.–attractor-same		−1.03* (−1.92 to −0.13)	−0.99* (−1.86 to −0.13)	−1.21* (−2.28 to −0.14)	−1.25* (−2.34 to −0.16)	−0.48 (−1.77 to −0.8)
Incorrect: attractor-diff.–attractor-same		0.04 (−0.95 to 1.04)	0.36 (−0.61 to 1.33)	−0.03 (−1.02 to .95)	0.75^ (−0.10 to 1.61)	0.12 (−0.87 to 1.11)

** p < .01.
 * p < .05.
 ^ p < .1

significant 2(Grammaticality: Correct, Incorrect) \times 2(Electrode selection: Central, Peripheral) interaction effect ($F(1, 19) = 6.73, p < .05$), reflecting the fact that the effect of Grammaticality was greater at these central electrodes than at peripheral electrodes.

Incorrect sentences elicited a negativity compared to Correct sentences in each of the four time windows between 400 and 1200 ms, resulting in a reliable main effect of Grammaticality. There was a reliable Grammaticality \times Attractor interaction between 400 and 600 ms and between 800 and 1000 ms, whereas this effect was marginally significant in the 600–800 ms time window (we speculate that this smaller effect size may be due to increase signal variance associated with onset potentials of the word following the critical word). Pairwise comparisons in all time windows with a significant interaction showed that the Correct conditions differed reliably from one another, with Attractor-different being more negative than Attractor-same, and that the Attractor-same conditions reliably differed from each other, with Incorrect being more negative than Correct (see Table 4). Again, in the three time windows between 400 and 1000 ms, the Incorrect Attractor-same condition was reliably more negative than the Correct Attractor-same condition and the Correct Attractor-different condition was reliably more negative than the Correct Attractor-same condition. No other comparisons in these time windows were reliable, except for a marginally significant difference between the Incorrect Attractor-same and the Incorrect Attractor-different condition in the 800–1000 ms time window (Incorrect Attractor-same was more negative than Incorrect Attractor-different; In a statistical analysis that involved all electrodes in this time window, this pair-wise comparison reached full statistical significance, $F(1,19) = 4.66, p = .04$). We presented this analysis to demonstrate that depending on electrode selection and on time window, a reliable difference between the Incorrect conditions does emerge.

References

- Anderson, M.C., Neely, J.H., 1996. Interference and inhibition in memory retrieval. In: Bjork, E.L., Bjork, R.A. (Eds.), *Handbook of Perception and Memory: memory*, 10. Academic Press, San Diego, pp. 237–313.
- Badecker, W., Straub, K., 2002. The processing role of structural constraints on the interpretation of pronouns and anaphora. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 748–769.
- Barber, H., Carreiras, M., 2005. Grammatical gender and number agreement in Spanish: an ERP comparison. *J. Cogn. Neurosci.* 17 (1), 137–153.
- Bates, E., MacWhinney, M., 1989. Functionalism and the competitive model. In: MacWhinney, B., Bates, E. (Eds.), *The Crosslinguistic Study of Sentence Processing*. Cambridge University Press, Cambridge.
- Bock, J.K., Miller, C.A., 1991. Broken agreement. *Cogn. Psychol.* 23, 45–93.
- Bornkessel, I., Schlesewsky, M., 2006. The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* 113, 787–821.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2009a. The role of prominence information in the real-time comprehension of transitive constructions: a cross-linguistic approach. *Lang. Ling. Compass* 3, 19–58.
- Bornkessel-Schlesewsky, I., Schlesewsky, M., 2009b. Processing Syntax and Morphology: A neurocognitive perspective. Oxford University Press, Oxford, UK.
- Carreiras, M., Salillas, E., Barber, H., 2004. Event-related potentials elicited during parsing of ambiguous relative clauses in Spanish. *Cognitive Brain Res.* 20, 98–105.
- Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.
- Davis, C.J., Perea, M., 2005. BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behav. Res. Methods* 37, 665–671.
- Eguren, L., 2010. Contrastive focus and nominal ellipsis in Spanish. *Lingua* 120, 435–457.
- Eguren, L., Sanchez, C., 2004. Contrast and addition in Romance: a case study in microvariation. In: Auger, J., et al. (Ed.), *Contemporary Approaches to Romance Linguistics*. John Benjamins, Amsterdam, pp. 159–176.
- Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.* 6, 78–84.
- Gerrig, R.J., O'Brien, E.J., 2005. The scope of memory-based processing. *Discourse Processes* 39 (2–3), 225–242.
- Gillund, G., Shiffrin, R.M., 1984. A retrieval model for both recognition and recall. *Psychol. Rev.* 91, 1–65.
- Gordon, P.C., Hendrick, R., Johnson, M., 2001. Memory interference during language processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 1411–1423.
- Gouvea, A., Phillips, C., Kazanina, N., Poeppel, D., 2010. The linguistic processes underlying the P600. *Lang. Cognitive Processes* 25, 149–188.
- Hagoort, P., 2005. On Broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9 (9), 416–423.
- Hagoort, P., Brown, C.M., Osterhout, L., 1999. The neurocognition of syntactic processing. In: Brown, C.M., Hagoort, P. (Eds.), *Neurocognition of Language*. Oxford University Press, Oxford, UK, pp. 273–316.
- Johnson, K., 2001. What VP ellipsis can do, and what it can't, but not why. In: Mark, Baltin, Chris, Collins (Eds.), *The Handbook of Contemporary Syntactic Theory*. Blackwell Publishers, pp. 439–479.
- Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111 (10), 1745–1758.
- Kaan, E., Swaab, T.Y., 2003. Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *J. Cogn. Neurosci.* 15, 98–110.
- Keppel, G., Underwood, B.J., 1962. Proactive inhibition in short-term retention of single items. *J. Verbal Learn. Verbal Behav.* 1, 153–161.
- King, J.W., Kutas, M., 1997. Is she an engineer? Brain potentials and anaphora. Poster presentation at the 4th Annual Meeting of the Cognitive Neuroscience Society, Boston, MA.
- Kutas, M., Federmeier, K.D., Coulson, S., King, J.W., Münte, T.F., 2000. Language. In: Cacioppo, J.T., Tassinari, L.G., Berntson, G.G. (Eds.), *Cambridge University Press*, pp. 576–601.
- Kutas, M., Van Petten, C., Kluender, R., 2006. Psycholinguistics electrified II: 1994–2005. In: Traxler, M., Gernsbacher, M.A. (Eds.), *Handbook of Psycholinguistics*, 2nd ed. Elsevier, New York, pp. 659–724.
- Lewis, R.L., Vasishth, S., Van Dyke, J., 2006. Computational principles of working memory in sentence comprehension. *Trends Cognitive Sci.* 10, 447–454.
- Mancini, S., Molinaro, N., Rizzi, L., Carreiras, M., 2011. When persons disagree: an ERP study of disagreement in Spanish. *Clin. Psychophysiol.* 48 (10), 1361–1371.
- Martin, A.E., McElree, B., 2008. A content-addressable pointer underlies comprehension of verb-phrase ellipsis. *J. Mem. Lang.* 58, 879–906.
- Martin, A.E., McElree, B., 2009. Memory operations that support language comprehension: evidence from verb-phrase ellipsis. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1231–1239.
- Martin, A.E., McElree, B., 2011. Direct-access retrieval during sentence comprehension: evidence from sluicing. *J. Mem. Lang.* 64, 327–343.
- McElree, B., 2006. Accessing recent events. In: Ross, B.H. (Ed.), *The Psychology of Learning and Motivation*, Vol. 46. Academic Press, San Diego.
- McElree, B., Doshier, B.A., 1989. Serial position and set size in short-term memory: time course of recognition. *J. Exp. Psychol. Gen.* 18, 346–373.
- McElree, B., Foraker, S., Dyer, L., 2003. Memory structures that subservise sentence comprehension. *J. Mem. Lang.* 48, 67–91.
- McKoon, G., Ratcliff, R., 1998. Memory-based language processing: psycholinguistic research in the 1990s. *Annu. Rev. Psychol.* 49, 25–42.
- Merchant, J., 2001. *The Syntax of Silence: Sluicing, Islands and the Theory of Ellipsis*. Oxford University Press.
- Molinaro, N., Barber, H.A., Carreiras, M., 2011. Grammatical agreement processing in reading: ERP findings and future directions. *Cortex* 47, 908–930.
- Myers, J.L., O'Brien, E.J., 1998. Accessing the discourse representation during reading. *Discourse Processes* 26, 131–157.
- Nairne, J.S., 2002a. Remembering over the short-term: the case against the standard model. *Annu. Rev. Psychol.* 53, 53–81.
- Nairne, J.S., 2002b. The myth of the encoding-retrieval match. *Memory* 10, 389–395.
- Nieuwland, M.S., Van Berkum, J.J.A., 2006. Individual differences and contextual bias in pronoun resolution: evidence from ERPs. *Brain Res.* 1118, 155–167.
- Nieuwland, M.S., Otten, M., Van Berkum, J.J.A., 2007a. Who are you talking about? Tracking discourse-level referential processing with ERPs. *J. Cogn. Neurosci.* 19, 228–236.
- Nieuwland, M.S., Petersson, K.M., Van Berkum, J.J.A., 2007b. On sense and reference: examining the functional neuroanatomy of referential processing. *NeuroImage* 37 (3), 993–1004.
- Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806.
- Osterhout, L., Mobley, L.A., 1995. Event-related brain potentials elicited by failure to agree. *J. Mem. Lang.* 34, 739–773.
- Otten, M., Van Berkum, J.J.A., 2008. Discourse-based lexical anticipation: prediction or priming? *Discourse Processes* 45, 464–496.
- Öztekli, I., McElree, B., 2007. Retrieval dynamics of proactive interference: PI slows retrieval by eliminating fast assessments of familiarity. *J. Mem. Lang.* 57, 126–149.
- Pearlmutter, N.J., Garnsey, S.M., Bock, K., 1999. Agreement processes in sentence comprehension. *J. Mem. Lang.* 41, 427–456.
- Phillips, C., Wagers, M., Lau, E., 2010. Grammatical illusions and selective fallibility in real-time language comprehension. To appear. In: Runner, J. (Ed.), *Experiments at the Interfaces, Syntax & Semantics*, 37. Emerald Publications, Bingley, UK.
- Staub, A., 2009. On the interpretation of the number attraction effect: response time evidence. *J. Mem. Lang.* 60, 308–327.
- Sturt, P., 2003. The time-course of the application of binding constraints in reference resolution. *J. Mem. Lang.* 48, 542–562.
- Tabor, W., Galantucci, B., Richardson, D., 2004. Effects of merely local syntactic coherence on sentence processing. *J. Mem. Lang.* 50, 355–370.
- Van Berkum, J.J.A., 2004. Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? In: Carreiras Jr., M., Clifton, C. (Eds.), *The On-Line Study of Sentence Comprehension: Eye-tracking, ERPs and Beyond*. Psychology Press, New York, pp. 229–270.
- Van Berkum, J.J.A., 2009. The neuropragmatics of “simple” utterance comprehension: an ERP review. In: Sauerland, U., Yatsushiro, K. (Eds.), *Semantics and Pragmatics: From Experiment to Theory*. Palgrave Macmillan, Basingstoke, pp. 276–316.
- Van Berkum, J.J.A., Brown, C.M., Hagoort, P.M., 1999. Early referential context effects in sentence processing: evidence from event-related brain potentials. *J. Mem. Lang.* 41, 147–182.

- Van Berkum, J.J.A., Koornneef, A.W., Otten, M., Nieuwland, M.S., 2007. Establishing reference in language comprehension: an electrophysiological perspective. *Brain Res.* 1146, 158–171.
- Van Dyke, J.A., 2007. Interference effects from grammatically unavailable constituents during sentence processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 407–430.
- Van Dyke, J.A., Lewis, R.L., 2003. Distinguishing effects of structure and decay of attachment and repair: a retrieval interference theory of recovery from misanalyzed ambiguities. *J. Mem. Lang.* 49, 285–413.
- Van Dyke, J.A., McElree, B., 2006. Retrieval interference in sentence comprehension. *J. Mem. Lang.* 55, 157–166.
- Van Dyke, J.A., McElree, B., 2011. Cue-dependent interference in comprehension. *Journal of Memory and Language.* 65 (3), 247–263.
- Van Valin Jr., R.D., 2005. *Exploring the Syntax–Semantics Interface*. Cambridge University Press, Cambridge.
- Wagers, M., Lau, E., Phillips, C., 2009. Agreement attraction in comprehension: Representations and processes. *J. Mem. Lang.* 61, 206–237.
- Watkins, O.C., Watkins, M.J., 1975. Build-up of proactive inhibition as a cue overload effect. *J. Exp. Psychol. Hum. Learn. Memory* 104, 442–4152.
- Waugh, N.C., Norman, D.A., 1965. Primary memory. *Psychol. Rev.* 72, 89–104.
- Wicha, N.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. *J. Cogn. Neurosci.* 16, 1272–1288.
- Xiang, M., Dillon, B., Phillips, C., 2009. Illusory licensing across dependency types: ERP evidence. *Brain Lang.* 108, 40–55.