Introduction

While reading or listening, comprehenders have to deal with sequences of words that incrementally activate the meaning intended by writers or speakers: meanings associated with incoming words are combined with meanings expressed by the previous set of words. These processes are defined as ‘compositional’. As explained by Pykkänen (2008, page 712; Frege, 1892), “natural language interpretation is generally thought to be compositional, that is, the meanings of expressions are a function of the meanings of their parts and of the way the parts are syntactically combined”. Composition is assumed to be incremental, i.e. the meaning of an expression is built word by word as the message unfolds. However, while composing the meaning of an expression, comprehenders actively pre-activate semantic features that are likely to be expressed in the next parts of the sentence/discourse (meaning pre-activation).

Despite composition and meaning pre-activation contributing to the activation of the same kind of semantic information, these two cognitive routines are supposed to be qualitatively different in nature. On the one hand, meaning composition could be defined as the process that links up the semantic information of the currently processed word with contextual information encompassing multiple words (at the sentence or discourse level), which are active in working memory. On the other hand, meaning pre-activation would reflect the facilitation of processing incoming information based on the amount of related semantic information that either has been previously processed or was already available in semantic long-term memory. A critical difference resides in the fact that while semantic composition reflects the processing cost of merging meanings to access the (unknown) message intended by the source (possibly creating new semantic representations that were not available in long-term memory), pre-activation phenomena reflect facilitation of accessing to meaning due to information that is stored in long-term memory (i.e. something that we already know). In terms of processing, these two routines should differ in the fact that while meaning composition would occur after the critical word has been recognized (hence, after its meaning is activated), meaning pre-activation would imply facilitation in the recognition of the critical word based on higher-level semantic information.

Most of the scientific literature focusing on the neurophysiological correlates of semantic processing has claimed that meaning composition and pre-activation interact, affecting brain activity starting around 200 ms after reading a target word in a sentence context (for reviews Hagoort et al., 2009; Kutas and Federmeier, 2011; Lau et al., 2008). Kutas and Hillyard (1980) initially discovered that event-related EEG (electroencephalographic) activity around 400 ms was sensitive to...
semantic manipulations: semantically inappropriate compared to semantically appropriate words elicited in fact more negative amplitudes in this time interval (Kutas and Hillyard, 1980; see also Federmeier and Kutas, 1999; for world-knowledge manipulations, see Hagoort et al., 2004; Van Berkum et al., 2008). In the Event Related Potential (ERP) literature this effect has been associated with the component named N400 (Kutas and Hillyard, 1980, 1984). The term N400 is a “heuristic label for stimulus-related brain activity in the 200–600 ms post-stimulus window with a characteristic morphology and, critically, a pattern of sensitivity to experimental variables” (Kutas and Federmeier, 2011, p. 14; for reviews, see Federmeier, 2007; Lau et al., 2008).

These findings initially indicated that compositional processes could be reflected in the modulation of the N400, since semantically unusual or bizarre messages, to be interpreted, would require additional integration costs (resulting in more negative amplitudes around 400 ms). However, some years later, Kutas and Hillyard (1984) showed that, given the sentence context, different degrees of expectancy for a specific target word (pre-activation effects), elicited the same effect. These authors reported significant correlations (ranging from 0.80 to 0.90) between the ERP amplitudes in the 300–500 ms post-stimulus interval and cloze-probability ratings, i.e. the proportion of people that continued a sentence fragment with a target word in an off-line paper and pencil questionnaire (that is, the more expected the target word, the less negative was the electrophysiological activity around 400 ms).

Thus, despite the fact that semantic compositional difficulties could be associated with electrical brain activity around 400 ms, the level of lexical pre-activation of the target word given the previous context should be carefully taken into account (Federmeier, 2007; Federmeier et al., 2007; Lau et al., 2008). For example, in a sentence like ‘He spread the warm bread with butter/socks’ (Kutas and Hillyard, 1980), where ‘socks’ elicited a more negative N400 compared to ‘butter’ it is hard to dissociate the processing costs elicited by the difficulty of combining the semantic information elicited by socks with the previous sentence context, from the facilitation in the recognition of butter due to the pre-activation of semantic features during the previous semantic context. In other words, it is relatively difficult to distinguish between ‘N400 reductions’ due to semantic pre-activation and ‘N400 increments’ due to impaired semantic compositional processes (facilitation vs. inhibition debate). Nowadays, there is robust evidence that pre-activation effects correlate well with brain related activity in a time window between 200 and 500 ms after the critical word onset (Kutas and Federmeier, 2011). Nonetheless, many authors still emphasize the constructive nature of semantic composition. Hagoort et al. (2005), for example, has suggested that ERP modulations around 400 ms post-stimulus represent a unification process that allows the integration of ideas and concepts that are not associated in semantic memory to already stored patterns (from single words to more general semantic knowledge).

Lau et al. (2008) recently reviewed the available fMRI literature focusing on lexical/semantic processing, trying to (de)construct the brain correlates of the N400 effect and identify a cortical network for semantics. They indicate the main source of this effect in the left posterior middle temporal gyrus (and the neighboring superior temporal sulcus and inferior temporal cortex, see also Halgren et al., 2002; Helenius et al., 1998). In their model, this brain area would be part of a larger cortical network for semantics involving additional areas of the left hemisphere (anterior temporal cortex, angular gyrus and left inferior frontal areas). The most interesting aspect of their proposal resides in the fact that “lexical representations would be stored and activated in the MTG […] and are accessed by other parts of the semantic network” (p. 989). Thus, activity in this area (correlating with the N400 negative amplitude) would not represent semantic composition processes per se, but it would be more related to lexical retrieval and pre-activation phenomena during language processing. Basic combinatorics and semantic integration with context, controlled retrieval of lexical information and selection of lexical candidates would be handled by different parts of the cortical network (that, in turn, could modulate activity in the middle temporal regions). Activity in those areas could still contribute to modulating EEG event-related activity around 400 ms, however Lau et al. suggest that a potential ERP effect also reflecting semantic combinatorial analysis could be the ‘post-N400 positivity.’

Recently, the research on semantic integration has focused also on late stages of processing. Recent empirical evidence has in fact brought increasing attention to EEG evoked activity that develops in a time interval after 500 ms post-stimulus (i.e. in a time interval following the N400 one). DeLong et al. (2011) and Federmeier et al. (2007), among others, reported modulations of a late positive component (LPC) with a frontal distribution over the scalp. Apparently, the integration of an unexpected (but still probable) word in a strongly constraining context requires additional neurocognitive resources to the ones available in the 300–500 ms time interval. Also this post-N400 positivity has been related to pre-activation phenomena: Kutas et al. (2011), for example, conclude that ‘late positive ERPs to unexpected nouns that increase with the degree of constraint violation strongly support the idea that when highly pre-activated input is not received, some form of additional processing may be called for’ (p. 200).

In this last quote, while the focus is still on pre-activation, the real nature of this later effect remains underspecified. For example, Van de Meerendonk et al. (2009, 2010) proposed that late positive effects would reflect monitoring processes triggered by a conflict between an active representation in working memory and unexpected incoming information that renders the expression implausible. However, two pieces of evidence suggest that the LPC modulation is not so tightly related to pre-activation effects: two studies in fact reported the effect for target words in sentence contexts that did not produce a strong semantic expectation. Increased late (after 500 ms) positive ERP activity was in fact observed in pre-frontal areas of the scalp for the processing of nominal metaphors (Unemployment is a plague, De Graauwe et al., 2010) and quantified sentences (Few farmers grow worms…, Urbach and Kutas, 2010). These findings suggest that the processing of natural (non-anomalous) expressions requires additional composition processing routines (reflected in the late EEG evoked activity) when the sentence message is composable but not immediately available in semantic memory.

Additional motivation for better focusing on composition phenomena derives from studies that have investigated the MEG correlates of semantic composition. Bemis and Pykkänen (2011) focused on the composition of minimal adjective–noun constructions in English. In their study, participants read adjective–noun pairs (‘red-boat’), word pairs (‘cup-boats’) or control pairs (‘xkg-boat’). The adjective–noun condition elicited increased evoked activity in the left anterior temporal lobe around 225 ms, followed by an increase of activity in the ventro-medial pre-frontal cortex with later onset around 400 ms. It should be noted that active expectations could have determined the effects recorded at the noun, since the previous adjective could create expectations concerning the following noun (its syntactic head); for example, it has been shown that word-class expectations can influence biomagnetic activity in the early occipital cortex around 100 ms (Dikker et al., 2010).

Nonetheless, Bemis and Pykkänen (2011) did not use complete sentence contexts (which normally induce stronger contextual expectations) so this study is particularly relevant, since it reports activation in brain areas that have not been classicallly identified as the neural sources of lexical pre-activation phenomena (i.e. the N400). Brain regions eliciting the N400 are in fact supposed to involve the middle temporal cortex (Halgren et al., 2002; Helenius et al., 1998; Lau et al., 2008; Pykkänen and Marantz, 2003). The evidence for a qualitatively different brain network involved in semantic composition
(ventro-medial pre-frontal cortex; Bemis and Pyllkänen, 2011; see also Pyllkänen and McElree, 2007) supports the hypothesis that additional processing routines to the ones associated with the N400 component are involved in semantic composition.

In sum, the available findings show that compositional phenomena until now have not been examined with sufficient specificity in the neurocognitive literature. In fact, the neurophysiological correlates of semantic composition have been mainly discussed for semantic manipulations involving different degrees of contextual semantic pre-activation (for a review of the neurocognitive literature, Lau et al., 2008). It appears, then, that further studies are needed, minimizing pre-activation effects, so that compositional mechanisms can be examined with more specificity.

The present study has been designed to evaluate the neurocognitive correlates of semantic composition when ‘cleaned’ of possible contextual pre-activation effects. More specifically, in the present research we reduce as much as possible effects due to contextual pre-activation, so that we can isolate the neurocognitive dynamics involved in semantic composition. To do this, we focus our analysis on the comprehension of minimal Spanish noun–adjective pairs presented in non-constraining sentence contexts. In Spanish, these noun phrases are constructed with the adjective following the noun it modifies (although adjectives can also precede nouns in Spanish). One example could be un monstruo solitario (a lonely monster). In this construction the meaning of the whole noun phrase is obtained, combining the concept expressed by the noun with the specific semantic feature expressed by the adjective, and it is available only after reading the adjective. Differently from English (where adjectives are in pre-nominal position), the noun does not require the following adjective, the use of which is optional. In this context, there is no active expectation for the following (target) adjective, as the noun could be followed by many different stimuli, both closed-class (articles or prepositions) and open-class (adjectives or verbs) words. When reading or listening, the comprehender can thus ‘close’ the noun phrase after reading the noun. This linguistic construction thus permits reduction of semantic pre-activation effects and more focus on semantic combinatorial processes.

This minimal construction presents some interesting properties. Reading of the noun activates a concept that can be defined through a set of semantic features. These features are further expressed by adjectives that can be associated with the noun, and this association leads to a better definition of the concept, from a more general (prototypical) to a more specific representation. Nonetheless, only a limited set of features can be applied to a concept. For example, a monster could be ‘lonely’ but it cannot be ‘geographic’ (in Spanish: un monstruo geográfico). This last phrase violates rules of selectional restriction: each word imposes semantic restrictions on the environment in which it occurs. For example, a verb like eat in a literal context requires that its subject refers to an animate entity and its object to something concrete. A violation of the selectional restrictions of a word results in an illegal anomalous pair. In our experiment, each noun poses a restriction on the type of adjectives that can be applied to it. Thus, a basketball player can be tall or short, and employing a feature varying in the dimension of height to describe him would respect selectional restriction rules.

Neutral-anomalous comparisons have been used very frequently in the neurocognitive literature focusing on semantic integration processes. However, semantically anomalous constructions are not very frequent in natural language, unless they are supported by a specific discourse context. For this reason we extended our analysis to more natural semantic structures that are presumed to tax the semantic system during everyday experience. Among the set of legal adjectives that can be applied to a noun, there is a lot of variability in the typicality of features classically related to the prototypical representation of the noun. Thus, some features can have low typicality, creating an apparent conflict with the prototypical representation expressed by the noun (for example, the prototypical representation of a monster assumes it is not ‘lovely’), while others are highly typical and are frequently used to define the concept (for example, a monster is ‘horrible’ by definition). As compared to the use of neutral adjectives (un monstruo solitario, a lonely monster), in the present study we crucially focus on a fine-grained combination of nouns followed by either low typicality adjectives (un monstruo hermoso, a lovely monster) or high typicality ones (un monstruo horrible, a horrible monster). In the former case, the meaning expressed by the adjective presents a property in an anomalous relation with the more typical one that could be associated with the noun. In the latter case, the adjective redundantly expresses one of the most typical features that could be associated with the previous noun.

Here we present two experiments. In the first experiment, Spanish speakers read sentences containing noun–adjective pairs in which the feature expressed by the adjective could be semantically Neutral (monstruo solitario), Anomalous (monstruo geográfico), or Contrasting (monstruo hermoso). In the second experiment, a similar group of participants read the same sentences in which we replaced the Contrasting condition with the Redundant one (monstruo horrible, see examples in Table 1). We presented the noun–adjective pairs in sentence contexts to induce a more natural linguistic analysis to study meaning composition processes and we used a reading comprehension task to avoid unnatural processing strategies.

**EEG evoked activity around 400 ms**

Assuming that the initial evoked activity (the so-called N400) could reflect initial difficulties in the semantic composition of noun and adjective (as suggested by Hagoort et al., 2009), we expect an increased negativity in this time window only for the Anomalous condition: composition difficulties should arise in earlier stages mainly when the combination of noun and adjective is not legal (as in the Anomalous condition that violates selectional restriction rules). On the other hand, when this relation is semantically legal (Neutral, Contrasting and Redundant conditions), and the combination possible, there should be no cost in associating the adjective with the previous noun. In principle, composition could be more difficult for the Contrasting compared to the Neutral, compared, in turn, to the Redundant condition, based on typicality (as shown by Urbach and Kutas, 2010), but differences for the earlier N400 effects are potentially smaller compared to the effect elicited by the Anomalous condition.

**Late evoked activity and plausibility**

Some studies have reported additional ERP effects extending beyond 500 ms that correlated with semantic manipulations also in weakly constraining contexts (for example Urbach and Kutas, 2010): thus, we do not exclude that comprehenders will pursue additional semantic processing to compose the meaning of noun–adjective pairs in the Contrasting condition. In this case, the meaning expressed by the adjective can be composed with the noun in an initial step (selectional restriction rules are not violated); however, in a later stage, the contrasting relation can be re-evaluated to activate an idiosyncratic meaning (similar to what the monitoring approach states, Van de Meerendonk et al., 2009). For example, if el monstruo (monster) and hermoso (lovely) apparently

---

1 This type of construction, considered figures of speech (oxymora and pleonasms, respectively), is fairly frequent, since many speakers use them quite unconsciously as part of everyday communication. Some of them have become so common that can be considered as conventionalized constructions with their own meaning (e.g., the conventionalized oxymoron living dead for zombie). In this study however, we did not focus our attention on conventional (or very frequent) constructions, but constructed noun-adjective pairs in which the feature expressed by the adjective could be either Contrasting or Redundant, in relation to the prototypical features associated with the previous noun. Since our interest was in semantic composition, the use of conventional pairs that do not require such processes would have been outside the scope of the present research.
constraining contexts, but this is not the case for the present study. De Grauwe et al. (2010) reported a late positive effect for middle-
neural processing (or reduced) for the Redundant condition compared to the Neutral condition (less typical than the Redundant condition).

Late evoked activity and naturality

Meta-linguistic factors, nonetheless, could be the relevant trigger for eliciting the neurocognitive routines correlating with the LPC effect. In fact, Contrasting and Redundant pairs would be less expected in natural language compared, for example, to the Neutral condition; in general, speakers tend to be not too repetitive or too conflicting when communicating, based on cooperative principles proposed by Grice (1975).

These constructions (that can be considered as figures of speech) offer a natural linguistic device for attracting listener attention through the use of creative combinations of words; usually, the ability to generate rhetorical expressions (such as iron curtain or cold war) is highly respected and commonly used, for example, by politicians. In our experiment, the fact that the more (Redundant condition) and less (Contrast-
ing condition) typical feature associated with a concept is emphasized could attract the attention of the reader, who would spend additional neurocognitive resources (correlating with the LPC amplitude) to better interpret the composed meaning of the noun–adjective pairs.

If this were the case, additional routines would be similarly used to analyze both contrasting and redundant semantic information. In other words, composition in both cases would be more difficult given the pragmatic status of the expressions: in the Redundant condition, readers’ attention would be attracted by the fact that a very typical feature of the noun (that has already been activated at the noun’s reading) is repeated and this can cause an integration cost, since the two pieces of information (largely overlapping in meaning) have to be finely analyzed and composed (this is a reason why speakers in general tend not to be too repetitive; Grice, 1975).

In sum, in this study we focus on semantic composition processing, trying to prevent possible semantic pre-activation routines before reading the target adjective. We perform this analysis using Spanish noun–adjective pairs focusing on the semantic relation between noun and adjective. Manipulating both the composability and the typicality of the adjective (in relation to the noun) in a non-constraining con-
struction we can thus evaluate the modulation of previous ERP components associated with semantic neurocognitive analyses.

Experiment 1: Contrasting pairs

Method

Participants

Twenty-one young adult Spanish native speakers (13 females, age range: 18–31) took part voluntarily in the study. They were paid 15€ for their collaboration: all of them were right hand dominant with no history of neurological disease and normal or corrected to normal vision. One participant was excluded from the analyses given the high number of artifacts evident in the EEG recordings.

Materials

135 Spanish noun–adjective pairs were created (e.g., monstruo solitario). The aim was to construct neutral pairs in which there was no strong semantic relation between the two words. These pairs were inserted in sentences, avoiding strongly constraining contexts in which the two target words (both the noun and the adjective) could be predictable. The position of the critical pair was constant across the sentence (always in sixth–seventh position). Additional words were always included after the critical noun phrases, in order to avoid end-of-sentence wrap-up effects.

We then created the Anomalous condition by substituting the neutral adjective (solitario) with an anomalous one (geográfico). This adjective referred to a feature that could not be composed with the previous noun. Finally, the Contrasting condition was constructed in the following way: we first identified a high typically feature associated to the target noun (like horrible) and then we identified its antonymous word (hermoso). On average the Neutral adjectives were 8 (SD = 1.97) letters long, the Anomalous adjectives 7.81 (SD = 1.96) and the Contrasting 7.86 (SD = 1.82). The frequen-
ty of use (calculated from the LEXESP database, Sebastian-Gallés et al., 2000) was also similar across conditions (Neutral: M = 23.91, SD = 37.43; Anomalous: M = 21.41, SD = 25.09; Contrasting: M = 21.81, SD = 36.19). Finally, the number of neighbors did not differ significantly (Neutral: M = 1.49, SD = 2.12; Anomalous: M = 1.38, SD = 1.95; Con-
trasting: M = 1.52, SD = 1.81).

The resulting pairs in the three conditions were distributed across three experimental lists balancing the lexical parameters associated to the target adjectives. To evaluate possible differences in the number of letters, frequency of use and number of neighbors of the critical adjective, we ran an ANOVA with two factors: List (three levels: List_1, List_2 or List_3) and Condition (three levels: Neutral, Contrasting and

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Example sentences used in Experiment 1 and Experiment 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1: Contrasting pairs</strong></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>Esa revista fotografió a un monstruo solitario que resulta muy interesante. That magazine photographed a lonely monster that turned out to be very interesting.</td>
</tr>
<tr>
<td>Contrasting</td>
<td>Esa revista fotografió a un monstruo hermoso que resulta muy interesante. That magazine photographed a lonely monster that turned out to be very interesting.</td>
</tr>
<tr>
<td>Anomalous</td>
<td>Esa revista fotografió a un monstruo geográfico que resulta muy interesante. That magazine photographed a geographic monster that turned out to be very interesting.</td>
</tr>
<tr>
<td><strong>Experiment 2: Redundant pairs</strong></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>Esa revista fotografió a un monstruo solitario que resulta muy interesante. That magazine photographed a lonely monster that turned out to be very interesting.</td>
</tr>
<tr>
<td>Redundant</td>
<td>Esa revista fotografió a un monstruo horrible que resulta muy interesante. That magazine photographed a geographic monster that turned out to be very interesting.</td>
</tr>
<tr>
<td>Anomalous</td>
<td>Esa revista fotografió a un monstruo geográfico que resulta muy interesante. That magazine photographed a geographic monster that turned out to be very interesting.</td>
</tr>
</tbody>
</table>
Anomalous). In all the cases the statistics did not reveal any main effects or interactions between the critical factors.

Since we were interested in triggering processes of semantic integration between noun and adjective we checked the frequency of the noun–adjective pairs in the Corpus of the Real Academia Española (http://corpus.rae.es/creanet.html) to exclude the frequent pairs that had idiosyncratic meaning. We then excluded six of the Contrasting pairs (such as muerto vivo, living dead, or noche blanca, white night), thus resulting in an overall number of 129 items, 43 per condition.

The resulting sentences were then further normed through two additional questionnaires that were administered to two independent groups of 20 Spanish speakers who did not take part in the following experiment. In the first questionnaire we tested the cloze-probability of the target adjective, i.e. the proportion of participants that continued the sentence presented until the target noun with the target adjective. In the three conditions the cloze-probability was almost null (Neutral: M = 0.03, SD = 0.01; Anomalous: M = 0, SD = 0; Contrasting: M = 0.02, SD = 0.01; see Table 2). Interestingly, only in 5.2% of the cases were the sentence fragments completed with an adjective.

In the second questionnaire we tested the semantic composability of all the sentences on a 7-point scale [Q: Does the following sentence make sense to you?; 1: not meaningful; 7: very meaningful]. In this norming the Neutral condition was the one rated as most composable (M = 5.54, SD = 1.31), while the Anomalous condition was the least composable (M = 2.18, SD = 1.82); the Contrasting condition was rated as less composable than the Neutral, but the ratings still remained above the mid value of 4 (M = 4.82, SD = 1.17; see Table 2).

The stimuli were counterbalanced across lists so that each participant did not see the same item in two different conditions. In addition to the experimental material, 71 filler sentences were presented to the participants, 40 of them containing semantic violations of various types. The semantic violations were presented half of the time in the initial position of the sentence (such as La niña llueve rápida delante de..., The girl rains fast in front of…) and half at the end of the sentence (such as En el bosque profundo de África yo vi una gacela que tromaba despacio. In the deep forest of Africa I saw a gazelle that was thudding slowly). In addition, noun–adjective pairs were presented in initial and final position of the other sentences. The fillers were constructed to avoid participants developing expectations for some particular noun–adjective construction in the middle of the sentence. Indeed, when debriefed at the end of the ERP experiment, participants did not report noticing any particular syntactic structure across the whole material; quite the opposite, all of them recognized that there were many sentences containing semantic anomalies in different sentence positions.

**Procedure**

Sentences were visually presented word by word in the center of a computer screen. The participants were instructed to silently read the sentences for comprehension. The instructions were given in written form and then orally repeated after a brief training (10 practice sentences). Each trial began when a participant pressed a keypad button. A fixation point presented for 500 ms on the center of the screen was followed by single words presented for 300 ms and separated by a blank of 300 ms. The 600 ms SOA is well suited for this experiment since it allows the N400 elicited by the noun to come back to baseline before the adjective presentation; this could be interpreted as reflecting a complete semantic analysis of the noun before the adjective is presented for composition. The presentation of each sentence was followed by a variable blank between 1000 and 3000 ms. Every 5 sentences on average, the participants were randomly asked to answer a YES–NO question about the content of the sentence just read. The experiment lasted approximately 45 min.

**EEE recording and signal extraction**

The electroencephalogram (EEG) was amplified and recorded with the BrainAmp system from 27 active electrodes placed on the scalp (Fp1, Fp2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, C3, C4, T7, T8, CP1, CP2, CP5, CP6, P3, P4, P7, P8, O1, O2, Fz, Cz, Pz) and referred to the left mastoid. One additional electrode was placed on the right mastoid plus four electrodes placed around the eyes for eye movement monitoring (two at the external ocular canthi and two below the eyes). EEG and EOG signals were amplified and digitized continuously with a sampling rate of 500 Hz. During recording, impedance values were kept below 5 kΩ for each scalp electrode and below 10 kΩ for the external ones. Adequate trigger signals time-locked to the target adjectives were generated and recorded for synchronization.

EEE signal was off-line filtered with a Butterworth low-pass filter at 30 Hz (48 dB/oct) and then re-referenced to the average activity of the two mastoids. Based on the critical triggers, we segmented the EEG recording starting 200 ms before the target until 800 ms after target word presentation. A semi-automatic artifact rejection procedure was then pursued: epochs in which the electrophysiological activity exceeded ± 100 μV were visually inspected and, in the case of artifact detection, rejected from the following analyses. These resulted in a similar amount of rejections across conditions (Neutral: 6.2%, Anomalous: 5.6%, Contrasting: 4.3%) as also revealed by the one-way ANOVA with Condition as a critical factor: F(2,38) = 1.11 (n.s.). Activity was baseline-corrected depending on the average amplitude in the 200 ms pre-stimulus interval. Individual epochs were averaged separately for each condition. Individual ERPs were employed for statistical analyses in specific time windows of interest and for grand-averaging across participants for visualizing group-level effects.

**Data analysis**

ERP effects were statistically evaluated focusing on the amplitude activity in the different conditions in subsequent time windows of interest. Based on our expectations we ran our analyses on average amplitudes in 5 time intervals of 100 ms each, starting from 250 ms to 750 ms. In each time interval we ran two distinct analyses of variance; in all cases the Greenhouse–Geisser correction was applied in the case of lack of sphericity in the data. The first *Midline ANOVA* was conducted on the midline electrodes, which are highly representative of the ERP effects of interest: an initial two-way ANOVA was run across the Electrode factor (three levels: Fz, Cz, Pz) with the Condition factor (three levels: Neutral, Anomalous, Contrasting). In the case of a significant main effect of Condition or an interaction between the two factors, additional ANOVAs comparing the conditions in

<table>
<thead>
<tr>
<th>Condition</th>
<th>Cloze-probability</th>
<th>Composability</th>
<th>Naturality</th>
<th>increased EEG evoked activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.03 (±0.01)</td>
<td>5.54 (1.31)</td>
<td>5.44 (1.30)</td>
<td>Baseline</td>
</tr>
<tr>
<td>Redundant</td>
<td>0.04 (0.02)</td>
<td>5.32 (1.98)</td>
<td>4.59 (1.20)</td>
<td>Short LPC (positivity: 550–650 ms)</td>
</tr>
<tr>
<td>Contrasting</td>
<td>0.02 (±0.01)</td>
<td>4.82 (1.17)</td>
<td>3.21 (1.60)</td>
<td>Long LPC (positivity: 550–750 ms)</td>
</tr>
<tr>
<td>Anomalous</td>
<td>0 (0)</td>
<td>2.18 (1.82)</td>
<td>2.87 (1.29)</td>
<td>N400 (negativity: 350–550 ms)</td>
</tr>
</tbody>
</table>
a pairwise manner were performed to determine the source of the effect. In the case of interaction between Electrode and Condition in the pairwise ANOVAs, post-hoc comparisons were run between conditions for each electrode, correcting the p-values through the False Discovery Rate procedure.

In order to evaluate possible lateralized ERP effects we also conducted a second ANOVA on the remaining electrodes. We grouped the activity elicited by contiguous electrodes calculating the mean values in six homogeneously distributed Clusters: Left Anterior (LA: mean activity of F3, F7, FC1), Right Anterior (RA: F4, F8, FC2), Left Central (LC: C3, CP1, CP5), Right Central (RC: C4, CP2, CP6), Left Posterior (LP: P3, P7, O1) and Right Posterior (RP: P4, P8, O2). Mean amplitudes in each time window entered a three-way overall ANOVA with two topographical factors, Longitude (three levels: Frontal, Central and Posterior) and Hemisphere (two levels: Left and Right), and the three level Condition factor. Relevant main effects or interactions involving the Condition factor were further evaluated by ANOVAs with the same design, comparing the three conditions in a pairwise manner. Possible interactions of the Condition factor with the topographical factors were followed by post-hoc comparisons in each Cluster between condition pairs, FDR corrected.

Result

Comprehension accuracy

The twenty participants in the ERP study showed a very good level of comprehension. Accuracy varied between 86% and 98%, with an average accuracy of 94%.

Event Related Potentials

Visual inspection of the ERP waveforms (Fig. 1) for the three critical conditions shows increased negativity for the ANOMALOUS condition compared to the other two conditions around 400 ms. This effect (that is mainly posterior) reflects a modulation of the negative peak that could be identified as the N400 component. No relevant difference could be identified between the NEUTRAL and the CONTRASTING conditions. After the N400 effect a later frontal positive effect is evident for the CONTRASTING condition compared to the NEUTRAL condition: this effect starts around 500 ms after word onset and lasts for several milliseconds. This last effect is identified as an increased Late Positive Component, the frontal distribution of which is in line with previous positive shifts reported in the ERP studies focusing on semantic manipulations.

The earliest time window of interest (250–350 ms) did not show any relevant effects (all Fs < 2). In the following time window (350–450 ms, classically associated to the N400 component), a main effect of Condition emerged in both the Midline (F(2,38) = 4.502, p < 0.05) and the Clusters (F(2,38) = 6.322, p < 0.05) overall ANOVAs. This main effect was accompanied by an interaction between condition and Electrode in the Midline analysis (F(4,76) = 4.323, p < 0.01). Pairwise comparisons showed critical differences between the NEUTRAL and the ANOMALOUS conditions: a main effect of Condition [emerging mainly in the Cluster analysis: Midline: F(1,19) = 3.594, p = 0.073; Clusters: F(1,19) = 6.053, p < 0.05] was evident for this contrast; in addition, the interaction between Condition and Electrode in the Midline analysis (F(2,38) = 8.054, p < 0.01) and the marginal interaction between Condition and Longitude (F(2,38) = 4.014, p = 0.051) were evident. These later interactions suggested that the N400 effect is mainly posterior, as confirmed by FDR corrected post-hoc comparisons [Midline post-hoc analyses: Pz: t(19) = 3.323, p < 0.01; Clusters post-hoc analyses: RC cluster: t(19) = 2.845, p < 0.01; LP: t(19) = 3.124, p < 0.01; RP: t(19) = 4.487, p < 0.001]. Also the ANOMALOUS–CONTRASTING contrast showed a main effect of Condition [Midline: F(1,19) = 17.933, p < 0.001; Clusters: F(1,19) = 22.363, p < 0.001]. In the Midline analysis an additional interaction between Condition and Electrode (F(2,38) = 5.425, p < 0.05) indicates that the N400 effect is not homogeneously distributed over the scalp, but shows its maximum in the posterior electrodes, as evidenced by subsequent FDR corrected post-hoc

Fig. 1. Event Related Potentials (ERPs) elicited by the target adjective in the three conditions in Experiment 1 in a set of representative homogeneously distributed electrodes. The Neutral condition is represented by the solid thin line, the Anomalous condition by the dashed thin line and the Contrasting condition by the solid thick line. Negative voltages are plotted up.
comparisons [Cz: t(19) = 3.995, p < 0.001; Pz: t(19) = 5.213, p < 0.001]. The Clusters analysis for this contrast showed a marginal interaction between Condition and Longitude [F(2,38) = 3.275, p = 0.071]: post-hoc comparisons confirmed the central-posterior distribution of the effect [RC: t(19) = 5.067, p < 0.001; Lp: t(19) = 4.307, p < 0.001; Rp: t(19) = 5.367, p < 0.001]. The comparison between the neutral and the contrasting conditions did not show any reliable index of an effect [all Fs < 1].

The third time window of interest (450–550 ms) showed critical effects in the overall ANOVAs: main effects of Condition [Midline: F(2,38) = 4.921, p < 0.05; Clusters: F(2,38) = 7.086, p < 0.01] and critical interactions in the anterior–posterior dimension [that were marginally significant in the Cluster analysis: Midline: Condition Electrode: F(4,76) = 1.757, p < 0.001; Clusters: Condition Longitude: F(4,76) = 2.585, p = 0.053]. Again the critical differences were due to the larger N400 effect elicited by the anomalous condition. In fact the comparison between the neutral and the anomalous conditions showed a main effect of Condition in the statistics [Midline: F(1,19) = 5.057, p < 0.05; Clusters: F(1,19) = 7.216, p < 0.05] that was accompanied by the interaction between Condition and Electrode in the Midline analysis [F(2,38) = 12.077, p < 0.001], and between Condition and Longitude in the Clusters analysis [F(2,38) = 5.865, p < 0.05]. Both interactions were due to the posterior distribution of the N400 effect, as showed by the FDR corrected post-hoc contrasts [Midline: Cz: t(19) = 3.207, p < 0.01; Pz: t(19) = 4.049, p < 0.001; Clusters: Rc: t(19) = 3.138, p < 0.01; Lp: t(19) = 3.701, p < 0.01]. The neutral–contrasting conditions contrast did not reveal any relevant effects, but only a marginal interaction between Condition and Longitude in the Midline ANOVA [F(2,38) = 2.919, p = 0.096], and between Condition and Longitude in the Clusters one [F(2,38) = 2.711, p = 0.079]. Finally, the comparison between anomalous and contrasting conditions showed main effects of Condition both in the Midline [F(1,19) = 12.258, p < 0.01] and in the Clusters analyses [F(1,19) = 19.076, p < 0.001], thus supporting the observation of a larger N400 for the anomalous condition. No interaction with the topographical factors emerged in this last contrast.

In the 550–650 ms time interval the overall ANOVA revealed both main effects of Condition [Midline: F(2,38) = 3.896, p < 0.05; Clusters: F(2,38) = 5.836, p < 0.05] and the interactions between Electrode and Condition in the Midline analysis [F(4,76) = 3.9, p < 0.05] and, marginally, between Longitude and Condition in the Clusters analysis [F(4,76) = 2.551, p = 0.072]. The ANOVA considering the neutral and the contrasting conditions revealed a marginal effect of Condition [Midline: F(1,19) = 3.24, p = 0.088], and the critical interactions between Condition and Electrode in the Midline [F(2,38) = 4.717, p < 0.05] and between Condition and Longitude in the Clusters analysis [F(2,38) = 5.033, p < 0.05]. Here, post-hoc comparisons for the midline electrodes confirm the observation of a larger positive effect for the contrasting condition that shows its maximum in the frontal electrodes [Fz: t(19) = −2.344, p < 0.05]; this was also confirmed by the significant effect emerging in the FDR-corrected post-hoc comparison for the left anterior cluster [L.a: t(19) = −2.624, p < 0.05]. The comparison between neutral and anomalous in this time window showed an interaction between Condition and Electrode in the Midline analysis [F(2,38) = 7.297, p < 0.01; FDR corrected post-hoc: Pz: t(19) = 2.158, p < 0.05] and the marginal interaction between Condition and Longitude in the Clusters analysis [F(2,38) = 2.917, p = 0.09]. As evident from the grand-average (Fig. 1) this is the last part of the N400 effect for the anomalous vs. neutral condition. Finally, the anomalous–contrasting comparison revealed main effects of Condition in both analyses [Midline: F(1,19) = 10.311, p < 0.01; Clusters: F(1,19) = 19.416, p < 0.001]. Since we consider the neutral condition as the control condition, this difference mainly indicates that two overlapping effects are affecting ERP’s in this time interval: the last part of the N400 for the anomalous condition (maximum in posterior areas of the scalp) and the initial part of the frontal positivity for the contrasting condition (maximum in frontal areas of the scalp).

In the 650–750 ms time window the overall ANOVA showed a main effect of Condition both in the Midline [F(2,38) = 3.931, p < 0.05] and in the Clusters [F(2,38) = 5.491, p < 0.05] analysis. In this last time window no effect emerged in the statistics that contrasted the neutral and the anomalous conditions. The neutral–contrasting comparison, on the other side, revealed a main effect of Condition [F(1,19) = 5.891, p < 0.05] and the marginal interaction between Condition and Electrode [F(2,38) = 3.592, p < 0.068] in the Midline analysis. In the Clusters analysis a main effect of Condition [F(1,19) = 6.737, p < 0.05] and the interaction between Condition and Longitude [F(2,38) = 5.141, p < 0.05] emerged. Post-hoc comparisons showed a relevant effect in the left anterior [La: t(19) = −3.488, p < 0.01], right anterior [Ra: t(19) = −2.298, p < 0.05] and left central [Lc: t(19) = −2.905, p < 0.01] clusters. These results support the fact that an increased positive effect is present for the contrasting condition in the frontal areas of the scalp (Fig. 1). Similarly, when comparing the anomalous and the contrasting conditions, a main effect of Condition emerged both in the midline [F(1,19) = 9.777, p < 0.01] and in the clusters [F(1,19) = 17.711, p < 0.001] analysis: this supports the evidence for a long-lasting frontal positive effect for the contrasting condition.

Summary of results of experiment 1 and interim conclusions

The first experiment focusing on the processing of contrasting noun-adjective pairs showed interesting effects in two time windows. Around 400 ms, the evoked EEG activity was more negative for the anomalous condition compared to the neutral and the contrasting conditions, which did not differ between them (see Cz electrode amplitudes in Fig. 2). After this time window, a long-lasting frontal positive effect was elicited by the contrasting condition compared to the neutral one (see Fz electrode amplitudes in Fig. 2).

A clear-cut dissociation emerged in this experiment, with the anomalous condition eliciting increased N400, while the contrasting condition elicited the increased LPC. These findings suggest that qualitatively distinct neurocognitive routines are operating in the two cases: evaluating if the semantic combination of noun and adjective is legal (Anomalous condition) affects the brain evoked activity in the early time interval (around 400 ms), while the contrast between the more typical features associated with the noun and the low-typicality (but composable) feature expressed by the adjective (Contrasting condition) triggers a later processing cost.

The N400 modulation suggests that the time interval around 400 ms could represent initial evaluation of the combination of the noun–adjective pair, even when there is no active expectation for the target word: if the adjective represents a feature that cannot be used to describe the previous noun, this cost emerges (correlating with the larger N400 effect). However, when the noun is followed by an adjective representing a feature that can be applied to the previous noun (semantically legal), there seems to be no explicit cost of composition: the lack of difference between the neutral and the contrasting condition in this time interval shows that the neurocognitive system only evaluates if the combination between noun and adjective is semantically possible.

The effect of typicality emerges in a later time interval (after 500 ms), correlating with the later frontal positive effect: compared to the neutral condition, an adjective that specifies a feature in contrast with the more typical ones associated with the previous noun (Contrasting condition) triggers this increased processing. To better evaluate this later effect we ran a second ERP experiment in which we substituted the contrasting condition with the redundant condition.

We expect the same early N400 effect between neutral and anomalous condition as in Experiment 1; on the other hand, the redundant condition in this time interval should not differ from
the Neutral, given the fact that the noun–adjective combination in this time window is legal.

The main interest after the findings of Experiment 1 now shifts to the later time interval: the late positive component, triggered by the Contrasting condition in Experiment 1, could modulate differently in Experiment 2 for the Redundant condition, in which the adjective expresses a highly typical feature that can be assigned by the concept expressed by the previous noun. Assuming that the increased late positive effect reflects composition processes triggered by low plausible but semantically legal information (as in the Contrasting condition), no LPC should be found. However, a larger LPC for the Redundant compared to the Neutral condition could also be predicted, assuming that the LPC reflects meta-linguistic analysis of infrequent expressions that are used in everyday language to attract the comprehender’s attention.

Experiment 2: Redundant pairs

Method

Participants

Twenty Spanish native speakers (14 females, age range: 19–26) took part in the study and were paid 15€ for their participation: all of them were right hand dominant with no history of neurological disease and normal or corrected to normal vision.

Material

We used the same set of sentences as in Experiment 1, substituting the Contrasting adjectives in the 129 items with a Redundant adjective. The adjective was selected as being a highly typical one used to describe the previous noun. The Redundant adjectives (horrible) did not have critical lexical differences compared to the other conditions; they were on average 7.86 letters long (SD = 1.96) with a similar frequency of use as extracted by the LEXESP database (M = 23.08, SD = 40.28) and a similar number of neighbors (M = 1.53, SD = 2) compared to the other conditions: the ANOVA with List (three levels: List_1, List_2 or List_3) and Condition (three levels: Neutral, Redundant and Anomalous) factors did not show any relevant difference among conditions. None of the noun–adjective pairs was associated with a high number of occurrences in the CREA database.

Cloze-probability was checked evaluating the participant responses in the cloze-probability test described for Experiment 1: cloze-probability of the Redundant adjectives was still very low (M = 0.04, SD = 0.02), indicating low level of expectation induced by the sentence context. Composability of the Redundant sentences was evaluated including the 129 item of the Redundant condition as filler sentences of questionnaires testing composability of a different set of sentences. A group of 20 Spanish young adults evaluated the Redundant condition as highly composable (M = 5.32, SD = 1.98), similarly to the rating reported for the Neutral condition. A summary of the linguistic norming for all the relevant conditions is reported in Table 2.

The stimuli were counterbalanced across lists so that each participant did not see the same item in two different conditions. In addition to the experimental material, 71 filler sentences were presented to the participants, 60 of them containing semantic violations of various types. Compared to the previous experiment we used a larger number of semantic violations: Redundant noun–adjective pairs are in fact more plausible and natural than Contrasting pairs, so that the overall number of semantically inappropriate sentences could appear (to the new group of participants) smaller in Experiment 2 compared to the conditions of Experiment 1. For this reason we increased the number of sentences containing semantic violations; as in Experiment 1, half of the sentences had a semantic violation in the initial part of the
sentence and half at the end. Post-experiment participant debriefing confirmed that they did not notice any specific constructions (apart from the semantic anomalies) across the set of sentences.

Procedure
The same as in Experiment 1.

EEG recording and signal extraction
The whole procedure was exactly the same as in Experiment 1. Visual inspection for artifact rejection resulted in a slightly higher number of rejections in Experiment 2 (Neutral: 7.4%, Anomalous: 6.1%, Redundant: 6.3%) that was not, however, different among conditions (F(2,38) = 0.69, n.s.).

Data analysis
The same statistical analyses as in Experiment 1 were employed in Experiment 2, substituting the Contrasting condition in the ANOVAs with the Redundant condition.

Results
Comprehension accuracy
In this second experiment, accuracy in the comprehension questions was still very high, with an average value of 96% (varying between 89% and 100%).

Event Related Potentials
Visual inspection of the ERP waveforms (Fig. 3) for the three conditions clearly indicates that the ANOMALOUS condition triggered more negative evoked activity compared to the other two conditions. The NEUTRAL and the REDUNDANT conditions dissociate in a later time interval, i.e. after 500 ms, with a larger positive effect for the REDUNDANT condition. This increased positivity is mainly evident in the frontal and central areas of the scalp, similar to the LPC effect reported in Experiment 1 for the CONTRASTING condition; however, the late positivity emerging in this last experiment is shorter in time, with an onset around 550 ms up to 650 ms. The ANOMALOUS condition in this later time window did not differ from the NEUTRAL one.

Statistical evaluation of the ERP effects in the 250–350 ms time interval did not reveal any significant differences in the overall ANOVAs. In the following time window (i.e., 350–450 ms) a main effect of Condition emerged mainly in the Midline analysis [Midline: F(2,38) = 3.776, p < 0.05; Clusters: F(2,38) = 3.168, p = 0.054] concerning the overall evaluation. Pairwise comparisons showed significant difference between the NEUTRAL and the ANOMALOUS conditions as shown by a main effect of Condition [Midline: F(1,19) = 5.149, p < 0.05; Clusters: F(1,19) = 4.512, p < 0.05]. This contrast also revealed marginally significant interactions between Condition and Longitude [F(2,38) = 3.473, p = 0.059] and between Condition and Hemisphere [F(1,19) = 3.056, p = 0.097]. Post-hoc FDR corrected contrasts showed that the larger N400 effect for the ANOMALOUS condition mainly emerges in the Right–Central [RC: t(19) = 2.939, p < 0.01] and the Right–Posterior clusters [RP: t(19) = 2.766, p < 0.01]. The larger N400 for the ANOMALOUS condition was confirmed by the main effect of Condition emerging in the REDUNDANT–ANOMALOUS contrast [Midline: F(1,19) = 6.996, p < 0.05; Clusters: F(1,19) = 5.577, p < 0.05]. On the other hand, no relevant effects emerged between the NEUTRAL and REDUNDANT conditions in this time interval.

In the following time interval (450–550 ms) the overall Midline analysis revealed an interaction between Condition and Electrode [F(4,76) = 3.327, p < 0.05], while the Clusters analysis revealed an interaction between Condition and Longitude [F(4,76) = 3.345, p < 0.05]. These interactions were further explored in the pairwise analyses that showed the interaction between Condition and Electrode in the Midline statistics of the NEUTRAL–ANOMALOUS contrast [F(2,38) = 5.957, p < 0.05]; post-hoc comparisons showed that the larger N400 for the ANOMALOUS condition was mainly parietal [similar to Experiment 1: Pz: t(19) = 2.551, p < 0.05]. This comparison also showed

![Fig. 3. ERPs elicited by the target adjective in the three conditions in Experiment 2 in a set of representative homogeneously distributed electrodes. The Neutral condition is represented by the solid thin line, the Anomalous condition by the dashed thin line and the Redundant condition by the solid thick line. Negative voltages are plotted up.](image-url)
the interaction between Condition and Longitude in the Clusters analysis [F(2,38) = 5.795, p < 0.05]; following planned comparisons showed that the N400 effect has its maximum in the Right-Central [RC: t(19) = 2.621, p < 0.05] and the Right-Parietal clusters [t(19) = 3.526, p < 0.01]. In the same time window, the comparison between the neutral and the redundant condition did not reveal any effects, while a main effect of Condition emerged for the redundant–anomalous comparison mainly in the Midline analysis [Midline: F(1,19) = 4.472, p < 0.05; Clusters: F(1,19) = 4.141, p = 0.056], probably reflecting the overlapping of the N400 effect for the anomalous condition and the LPC for the redundant condition (both compared to the neutral).

The fourth time interval of interest (550–650 ms) showed in the overall analysis an effect of Condition [Midline: F(2,38) = 3.155, p < 0.05] and the interaction between Condition and Longitude in the Clusters analysis [F(4,76) = 3.175, p < 0.05]. Following pairwise comparisons showed for the neutral–redundant contrast a marginal effect of Condition [F(1,19) = 3.789, p = 0.067] and the interaction between Condition and Electrode [F(2,38) = 3.739, p < 0.05] in the Midline statistics. FDR corrected post-hoc analyses confirmed that there was an increased positive effect for the redundant condition with maximum in the frontal electrode [Fz; t(19) = −2.673, p < 0.05]. The Clusters statistics showed a significant interaction between Condition and Longitude [F(2,38) = 5.142, p < 0.05], with the positivity mainly emerging in the left-anterior cluster [LA: t(19) = −2.237, p < 0.05] for the redundant condition. In this time interval, no significant effects emerged from the analyses comparing the neutral with the redundant condition. The redundant–anomalous contrast on the other hand showed a main effect of Condition [Midline: F(1,19) = 5.87, p < 0.05]. This last result supports the fact that the redundant condition elicits a larger positivity in this time interval compared to the other two conditions.

Finally, the last time interval of interest (650–750 ms) only showed an interaction between Condition and Hemisphere [Clusters: F(2,38) = 4.596, p < 0.05]. No significant effects emerged in the pairwise comparisons among the three conditions; only a marginal triple interaction among Condition, Longitude and Hemisphere emerged in the contrast between neutral and redundant conditions [Clusters: F(2,38) = 3.195, p = 0.062]. FDR corrected post-hoc analyses, however, did not reveal any significant effects between these two conditions in any topographical cluster.

Summary of results of Experiment 2

Experiment 2 revealed a pattern of effects almost parallel to Experiment 1. The anomalous condition elicited a larger negative effect around 400 ms compared to the other two conditions (see Cz electrode amplitudes in Fig. 4), while the redundant condition elicited an increased frontal-central positivity after 500 ms mainly compared to the neutral condition (see Fz electrode amplitudes in Fig. 4). While the N400 recorded for the anomalous condition was similar to the effect recorded in Experiment 1 (thus supporting the across-experiment comparison), the late positive effect recorded for the redundant condition has a shorter duration compared to the one recorded for the contrasting condition in Experiment 1. Even with different duration, the two LPC effects show a similar frontal scalp distribution (compare Figs. 2 and 4) and appear identical in the earlier time interval (550–650 ms, see difference waveforms in Fig. 5).

Across-experiment evaluation

In order to better evaluate the similarity/differences between the two LPCs we ran additional analyses in the 550–650 and the 650–750 ms time windows to compare the two effects. We thus ran a Greenhouse–Geisser corrected ANOVA on the amplitudes recorded at three frontal electrodes (F3, Fz, F4). More specifically, we ran a four-way ANOVA with three within-subject factors and a between-subject factor; the three within factors were Condition (two levels: neutral vs. contrasting for Experiment 1 and vs. redundant for Experiment 2), Time Window (two levels: early time window, 550–650 ms, and late time window, 650–750 ms) and Laterality (three levels: F3, Fz, F4).

---

**Experiment 2**

![Fig. 4. Detailed effects in Experiment 2. The left panel represents the Cz electrode at which the N400 effect (in the 350–450 ms time interval) elicited by the Anomalous condition shows its maximum, as evident by both the relative voltage maps on its right and the histogram of the relative absolute amplitudes in the three conditions below. The right panel represents the Fz electrode at which the short-lasting LPC effect (in the 550–650 ms time interval) elicited by the Redundant condition shows its maximum, as evident by both the relative voltage maps on its right and the histogram of the relative absolute amplitudes in the three conditions below.](image-url)
In the Introduction, we proposed that the increased LPC effect for the Contrast and Redundant conditions compared to the Neutral condition could indicate that similar meta-linguistic analyses of the material could be operating in the two situations. The use of either high or low typicality adjectives, related to the previous noun, could attract the comprehender’s attention to pursue additional semantic combinatorial processing of those stimuli. If this is the case, the cognitive system would spend additional composition analyses resources to better integrate the unusual expression in its mental representation of the message. The duration of these additional processes could depend on the typicality of the feature that has to be composed with the noun.

Thus, a good indicator of the LPC duration could be the pragmatic evaluation of the stimuli. For this reason we prepared a second questionnaire in which we asked a group of Spanish speakers to rate on a 7-point scale the naturality of our items, asking the following question: Would you expect a Spanish speaker to produce this expression? Rating the sentence with 1 would indicate a very unnatural expression, while rating the item as 7 would mean that the sentence was very natural. The most natural condition was the Neutral (M = 5.44, SD = 1.30), followed by the Redundant (M = 4.59, SD = 1.20), then the Contrast (M = 3.21, SD = 1.60), and finally the Anomalous (M = 2.87, SD = 1.29), as indicated in the histograms in Fig. 6. This graded pattern was statistically reliable: the Neutral condition was significantly more natural than the Redundant (t(128) = 6.17, p < 0.001), which in turn was more natural than the Contrast (t(128) = 8.93, p < 0.001); finally, also the Contrast–Anomalous comparison was statistically significant (t(128) = 2.04, p < 0.05). Based on the observation that naturality ratings aligned well with the ERP effects (see Table 2), especially for the critical conditions that are semantically composable, we performed additional correlation analyses to see if those ratings correlate with the LPC amplitude.

We thus considered the two critical conditions eliciting the LPC effect (Contrast and Redundant). Then, we reclassified all the items independently for each condition based on the naturality ratings and created eight bins per condition each containing 5 items: after ordering our stimuli from the least natural to the most natural item within each condition, we created the first bin with the 5 least natural items per condition, the second bin with 5 items that were more natural in the scale and so on.

Based on the bin subdivision, we extracted ERPs for each bin based on the same procedure used in the previous analyses. Given the high across-participants variability in the ERPs corresponding to each bin, we averaged single-subject averages for each bin across participants who saw the same experimental list; this procedure is possible since participants who saw the same list were exposed to items with the same levels of naturality. Independently for the Contrast and Redundant conditions, we then extracted average amplitudes in the 600–800 ms time interval for each electrode, each group of subjects and each bin. These values represent the amplitude.

In all ANOVAs, Condition was the within factor and Electrode the between factor. In this ANOVA, a main effect of Condition, while rating the item as 7 would mean that the sentence was very natural. The most natural condition was the Neutral (M = 5.44, SD = 1.30), followed by the Redundant (M = 4.59, SD = 1.20), then the Contrast (M = 3.21, SD = 1.60), and finally the Anomalous (M = 2.87, SD = 1.29), as indicated in the histograms in Fig. 6. This graded pattern was statistically reliable: the Neutral condition was significantly more natural than the Redundant (t(128) = 6.17, p < 0.001), which in turn was more natural than the Contrast (t(128) = 8.93, p < 0.001); finally, also the Contrast–Anomalous comparison was statistically significant (t(128) = 2.04, p < 0.05). Based on the observation that naturality ratings aligned well with the ERP effects (see Table 2), especially for the critical conditions that are semantically composable, we performed additional correlation analyses to see if those ratings correlate with the LPC amplitude.

We thus considered the two critical conditions eliciting the LPC effect (Contrast and Redundant). Then, we reclassified all the items independently for each condition based on the naturality ratings and created eight bins per condition each containing 5 items: after ordering our stimuli from the least natural to the most natural item within each condition, we created the first bin with the 5 least natural items per condition, the second bin with 5 items that were more natural in the scale and so on.

Based on the bin subdivision, we extracted ERPs for each bin based on the same procedure used in the previous analyses. Given the high across-participants variability in the ERPs corresponding to each bin, we averaged single-subject averages for each bin across participants who saw the same experimental list; this procedure is possible since participants who saw the same list were exposed to items with the same levels of naturality. Independently for the Contrast and Redundant conditions, we then extracted average amplitudes in the 600–800 ms time interval for each electrode, each group of subjects and each bin. These values represent the amplitude.
of the LPC in the late time interval where the effect was differently modulated for the Contrasting and the Redundant condition (see Fig. 5). Given the low number of trials per bin entering in the correlation analyses, the SNR value was too high to have a reliable amplitude value within a small 100 ms time interval. For this reason, we considered a slightly longer time window. Adding 50 ms before and after the critical time interval allowed us to enter more reliable amplitude values in the correlation analyses.

The Pearson correlation between the bin amplitudes of the ERP for each electrode and the corresponding naturality rating was calculated separately for the Contrasting and the Redundant conditions. In the Contrasting condition, the electrode that showed the strongest significant inverse correlation ($r = -0.39$, $t(23) = -1.93$, $p=0.05$) was FC1. Interestingly, the Redundant condition showed overall higher inverse correlation levels, with the maximum on the same left frontal-central electrode (FC1: $r = -0.50$, $t(23) = -2.62$, $p=0.01$); the topographical maps of the correlation are reported in Fig. 6 for both conditions.

**General discussion**

In the present study we evaluated the real-time electrophysiological correlates of semantic composition during the comprehension of minimal noun–adjective pairs. The pattern that emerged showed that these combinatorial analyses affect ERPs in a time window starting around 350 ms and continuing at least until 800 ms. The onset of the effects is slightly later compared to the ERP effect that has been shown to correlate with high levels of semantic constraint induced by the context, usually starting 100–150 ms earlier (Kutas and Federmeier, 2011; Lau et al., 2008). It could be that pre-activation phenomena already show their effects in the N400 time interval around 200 ms, while compositional processes emerge later (i.e. around 350 ms; see also Molinaro and Carreiras, 2010).

It is thus reasonable to assume that combinatorial semantic analyses start to show their effects after 350 ms if they are unaffected by contextual pre-activation of partial semantic information. Based on these findings we do not exclude that semantic contextual pre-activation (already evident around 200 ms) could modulate semantic combinatorial processes (starting around 350 ms), since it modulates the N400 amplitude in many studies. This interpretation aligns well with theoretical proposals that assume a complex interaction between semantic pre-activation and composition in the N400 time interval (Federmeier and Laszlo, 2009; Hagoort et al., 2009; Lau et al., 2008), while it does not support pure lexical accounts (for reviews see Kutas and Federmeier, 2011; Lau et al., 2008).

The present findings indicate that an initial evaluation of the possible combination of the target adjective with the previous noun is pursued for semantic composition around 400 ms. In this time interval, typicality does not seem to play a role in affecting the N400 amplitude (as suggested in the study reported by Urbach and Kutas, 2010). When the contextual constraint is lowered and the sentence construction is minimal, as in our study, no typicality effect emerges in the N400 time interval (i.e. between 350 and 500 ms approximately). We report two different sets of data in which no N400 effect emerged either for low- (Experiment 1) or for high-typicality adjectives (Experiment 2). Interestingly however, these types of semantic relations (concept-feature) have been shown to have an effect in behavioral prime-target paradigms (Lucas, 2000, 2001; Smith et al., 1988).

While these last behavioral findings apparently contrast with the lack of ERP modulation in the N400 interval, EEG evoked activity showed sensitivity to the typicality of the noun–adjective pair in a later time interval. ERPs after 500 ms showed interesting modulations, triggering a positivity with a long duration for the Contrasting condition compared to the Redundant, which in turn showed a short-lasting increased positivity compared to the Neutral condition. Frontal late positive components have received increasing attention in the neurophysiological literature on semantic processing. Some authors interpret late positive effects (independently from their topography) as an index of a conflict-monitoring process that is triggered to reanalyze a stimulus due to conflicting internal representation (De Grauwe et al., 2010; Kuperberg, 2007; Van de Meerendonk et al., 2009, 2010). Kutas et al. (2011) also discussed the frontal late components: in many studies (Coulson and Van Petten, 2007; DeLong et al., 2011; Federmeier et al., 2007) they reported late positive components for low expectancy words in high-constraining contexts. The focus in Kutas et al.’s (2011) discussion is on the neurocognitive routines correlating with the late frontal positivities as triggered by the ‘disconfirmation of strongly pre-activated linguistic trajectories’ (DeLong et al., 2011, p. 5).

However, we report LPCs in a linguistic scenario in which the contextual expectation for the target word is almost absent. It seems then that the processes correlating with the LPC are not only triggered in experimental conditions of strong contextual constraint.

It is interesting that both the Contrasting and the Redundant conditions elicited these late neurocognitive routines. As discussed in the Introduction, cloze-probability of the target adjectives was very low.

---

2 It should be noted, however, that their typicality N400 effect was contaminated by contextual expectations: when reading the fragment ‘farmers grow...’ the completion ‘worms’ could be more expected than ‘worms’ in that minimal context, thus eliciting the N400 modulation they reported.
in all the conditions: only in 5.2% of the cases were our sentence fragments continued with an adjective completion. In addition, correlation analyses in the present study showed that an inverse relation emerged between the duration of the LPC and the naturality of the expression. This piece of evidence suggests that meta-linguistic evaluation of the stimulus could influence the use of additional compositional routines. In other words, it is possible that the stimuli we used (in the Contrastign and the Redundant conditions) attracted the attention of the reader, since they were not completely natural. The Contrasting condition requires the composition of two words that apparently contrast between them, while the Redundant condition requires the composition of information that has been repeated twice. In other words, in the former case the representations related to the two words have to be integrated despite going (semantically) in opposite directions, while in the latter case the two words have to be integrated despite the fact that they are emphasizing a highly similar conceptual representation. In this framework, the less natural the expression is, the more prolonged is the analysis required. We thus propose that meta-linguistic evaluations (e.g., assessing the naturality of the expression) of the linguistic stimuli attract the comprehender’s attention, which is induced to carry out additional compositional analyses to make sense of the novel but possible expression.

The present study reports a clear-cut dissociation between the N400 and the LPC time intervals: only the anomalous adjective elicited the earlier effect, while the other non-anomalous conditions that varied in typically elicited late positive components to different degrees. These findings suggest that we are facing two subsequent and maybe interacting stages of semantic analysis (see also Federman et al., 2007).

The present results could be well integrated in a developing model that identifies a series of stages involved in semantic composition. The recent MEG study by Bemis and Pylkkänen (2011) proposed that around 225 ms initial syntactic composition of the target word with the context is carried out in the left anterior temporal lobe. In our study we did not find evidence for such a stage of processing since we only used syntactically well-formed constructions. After this initial processing stage, the pre-frontal cortex would be involved in a network (also involving lateral fronto-temporal areas) activated around 400 ms that correlates with the N400 ERP component. There is increasing evidence in the literature that the N400 component is not a unitary phenomenon, but the result of the activation of different neural sources (see Halgren et al., 2002; for similar proposals see Molinaro and Carreiras, 2010; Molinaro et al., 2010). This stage would be sensitive to the composability of the relation between the target word and the previous context: when this relation is anomalous, increased EEG evoked activity around 400 ms would be observed.

Unlike Bemis and Pylkkänen (2011) we also found evidence for a late frontal effect the duration of which correlates with the naturality of the expression. One plausible reason for this difference relies on the fact that this positivity mainly emerged due to the embedding of our critical pairs within a sentence context, since Bemis and Pylkkänen employed isolated word pairs. Given the frontal distribution of our late effect, we are tempted to speculate that the late positivity reflects prolonged brain activity in frontal areas of the brain (maybe in the pre-frontal cortex) that operates in a controlled way depending on attentional focusing. The pre-frontal cortex might be engaged during this LPC period in a more refined (top-down) access (select/control) to the representations (Badre and Wagner, 2007), or more generally in performing monitoring operations, meta-linguistic judgments, or integration of the information encountered.

These findings could have interesting implications for brain-based models of semantic processing since the areas that underlie the combination and integration of lexical/semantic representations are still poorly understood. Many studies have, in fact, indicated the left middle temporal regions as the main cortical source of the N400 effect (Halgren et al., 2002; Helenius et al., 1998; Lau et al., 2008); the strong sensitivity of this ERP effect to lexical pre-activation led Lau et al. (2008) to suggest that the MTG (middle temporal gyrus) is involved in the storage of lexico-semantic information and its activity could be modulated by other brain areas in a top-down manner. However, when the semantic composition of the noun–adjective pair is not possible (Anomalous condition) increased activity in the N400 time interval was observed in the present study. We do not think that any lexical processing difficulty should arise in our experiment, thus, it is possible that the additional parts of the network depicted in the Lau et al. (2008) model are contributing to the elicitation of the N400 effect. More specifically, it is possible that the earlier part of the N400 (up to 350 ms) is more related to lexical activity in the middle temporal regions, while the later part of this component (after 350 ms) is reflecting basic semantic combinatorial analyses: according to Lau et al. these processes would be carried out by the anterior part of the temporal lobe and possibly by the angular gyrus.

These authors also discussed the possibility that ‘post-N400’ EEG evoked activity could reflect later brain activity carried out by left inferior frontal regions. The robust late frontal positive effects we report in the present study for the non-anomalous pairs could indicate more complex semantic combinatorial analyses. As discussed above, we suggest that the comprehender’s attention is attracted by the unnatural expression, and for this reason a prolonged semantic combinatorial analysis is carried out, to better interpret the meaning of the expression. This controlled process could be responsible for eliciting the late positive effect and it presents interesting similarities with the activity that would be carried out by the left inferior frontal regions in the Lau et al. model: according to these authors “inferior frontal areas may be involved in relating the resulting compositional meaning to stored knowledge about the world” (p. 923) but they could also contribute to linking the composed meaning to other semantically related stimuli in long term memory. Critically, the contribution of the left frontal cortex in eliciting the late frontal ERP effect has to be empirically verified. However, available MEG evidence (Bemis and Pylkkänen, 2011) seems to point to a relevant role of the frontal lobe in semantic composition (see Badre and Wagner, 2007).

Conclusions

In the present research we evaluated the neurophysiological correlates of semantic composition, trying to dissociate them from semantic pre-activation phenomena. Studies that have manipulated the degree of expectation of the target word usually report modulations of EEG event-related activity starting around 200 ms after target word presentation, followed in some cases by frontal positive components when the expression is non-anomalous. In the present study, only the not composable constructions elicited an increased N400, whose onset is 150 ms later than the N400 effects related to pre-activation (Kutas and Federmeier, 2011); in addition, frontal late positive effects were reported for non-anomalous constructions even when the relation of the noun–adjective pair was highly typical.

These findings attest the complex interaction between the pre-activated contextual information and the incoming semantic input in semantic composition. ERP effects related to pre-activation and composition phenomena largely overlap in time and space, showing interesting modulations in the evoked electrophysiological activity starting after 350 ms: these ERP effects probably represent a complex network of brain areas whose interaction is modulated by a large number of parameters that have to be carefully considered, such as contextual pre-activation, but also the naturality and compositability of the linguistic expression. The present findings pose interesting constraints for future research aiming at uncovering the brain networks underlying semantic combinatorial analyses.
Acknowledgments

This work was partially supported by the Spanish Ministry of Science and Innovation (grants CONSOLIDER-INGENIO 2010 CSD2008-00048 and PSI2009-08889 to M.C.). N.M. was supported by a ‘Juan de la Cierva’ grant from the Spanish Ministry of Science and Innovation. We would like to thank Margaret Gillon-Dowens, Pedro Paz-Alonso, Francesco Vespiignani and two anonymous reviewers for useful comments on the present manuscript.

References