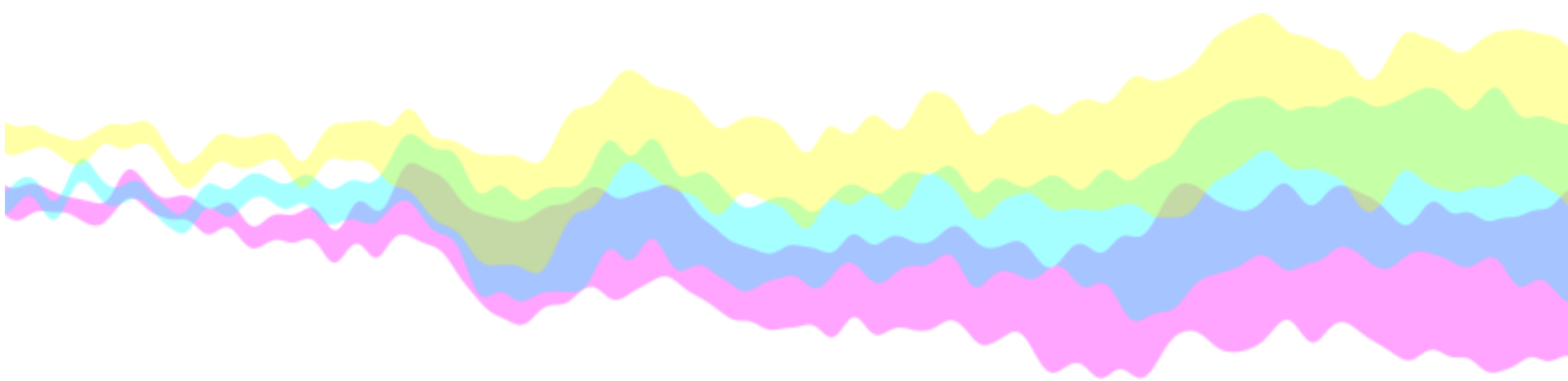
Elementary composition in Spanish: an EEG study



PEDECIBA Biología - Subárea Neurociencias

Tesis de Maestría

Elementary composition in Spanish: an EEG study

Lic. Emilia Fló Rama

Procesamiento del lenguaje
CIBPsi, Facultad de Psicología
Universidad de la República

Advisor Dr. Juan C. Valle Lisboa
Co - advisor: Dr. Álvaro Cabana

Thesis comitee: Dr. Ángel Caputi
Dr. Andrés Pomi
Dr. Jacobo Sitt

Elementary composition in Spanish: an EEG study

Abstract

The ability to combine words in order to represent and convey new meanings is a fundamental operation in the comprehension and production of language, and one of the objectives in language research is to understand the neural computations that allow us to construct complex conceptual representations. Nevertheless, the study of meaning composition is not straightforward as semantic and syntactic computations co-occur; and typical experimental designs have not been able to properly disentangle these processes. Therefore, it is convenient to develop minimal examples of linguistic composition that allows to study the basic operations that take place during language composition.

Recently, a study by Bemis and Pylkkänen (2011) proposed a simple experimental paradigm to evaluate composition at its minimum by restraining composition to pairs of nouns and adjectives. Given the complexity of the experimental tasks usually implemented in language research, this minimal paradigm came as an interesting addition to the field. By reducing stimuli to pairs of words, variables such as memory demands would not intervene and the basic syntactic-semantic computation theoretically elicited at every step in sentence processing could be addressed. These authors report that the left anterior temporal lobe (LATL) showed an increased response to nouns in combinatorial contexts (red boat) over conditions in which composition was not possible or discouraged (xwrt boat, cup boat), 184 - 255 ms after noun onset, as measured by MEG. A replication of this experiment was conducted using EEG (Neufeld, et al. 2016) but the results are to our understanding inconclusive.

In the present study we set out to use EEG to find the composition-related activity described in the MEG literature by applying a robust methodological analysis and a more objective data driven approach than the one carried out in Neufeld et al.'s replication. To this aim, we adapted the basic composition paradigm to Spanish by inverting pairs of stimuli and introducing a second control task.

We successfully reproduced the results reported in the literature. The binding of two items into a single concept yielded a processing advantage, as shown by faster reaction times and accuracy. Moreover, a cluster based permutation analysis revealed an activity specific to the composition condition 260 - 550 ms after second word onset for one of the controls, and at 410 - 600 ms for the second control task. Furthermore, the analyses carried out allowed us to explore the time course of this activity, which was indicative of a potential confound in the experimental design. We propose that this paradigm allows for an expectancy processes to develop only for words in combinatorial contexts. During these trials, the processing of first word was conditional to the second word, and therefore, the contingency between stimuli could account for the reported composition-related activity. We relate this result to the contingent negative variation (CNV) (Walter et al., 1964), an anticipatory potential elicited during tasks in which an initial stimulus announces the appearance of a second stimulus, and a relation between them is gradually established. Following CNV behavior, we show an effect of task progression on the composition-related activity obtained. Moreover, we designed a task in which expectancy contributions were controlled for, and show that this composition-related activity is not elicited when participants are unaware of whether an incoming trial requires them to compose or not.

In this work we provide evidence of a confounding effect of expectancy introduced by the block design of the experimental paradigm proposed by Bemis & Pylkkänen. We discuss the implications of our study to the MEG literature on basic composition and to the EEG replication previously conducted.

Composición elemental en Español: un estudio de EEG

Resumen

La habilidad combinar palabras de modo de representar y comunicar nuevos significados es una operación fundamental en la comprensión y producción del lenguaje, y uno de los objetivos de la investigación en lenguaje es entender las computaciones neuronales que nos permiten construir representaciones conceptuales complejas. Sin embargo, el estudio de la composición de significado no es sencilla, operaciones semánticas y sintácticas ocurren conjuntamente; y los experimentos típicos no han logrado separar exitosamente estos procesos. Por lo tanto, es de interés desarrollar ejemplos mínimos de composición lingüística que permitan estudiar las operaciones básicas que tienen lugar durante el procesamiento del lenguaje.

Recientemente, un estudio por Bemis & Pylkkänen (2011) propuso un paradigma experimental simple para evaluar composición elemental restringiendo los estímulos a pares de sustantivos y adjetivos. De este modo, variables como la atención o memoria de trabajo serían más fácilmente controladas, y las computaciones semánticas-sintácticas básicas producidas en cada paso del procesamiento podrían ser estudiadas. Estos autores reportan que el lóbulo temporal anterior (LATL) muestra una respuesta aumentada en registros de magnetoencefalografía (MEG) frente a sustantivos en contextos combinatorios 200 ms luego de presentado el estímulo, en comparación con sustantivos para los cuales la composición no es requerida o es desestimulada. Una réplica de este estudio fue realizada utilizando electroencefalografía (EEG) (Neufeld, et al. 2016) pero los resultados son a nuestro entender inconcluyentes.

En este trabajo nos propusimos encontrar la actividad de composición reportada en la literatura de MEG utilizando EEG y aplicando un análisis estadístico más robusto. Con este fin adaptamos la tarea original al Español e introdujimos una segunda tarea como control. Reprodujimos los resultados reportados en la literatura. La unión de dos elementos en un único concepto produjo una ventaja de procesamiento, reflejado en tiempos de reacción menores y un mayor porcentaje de aciertos. Asimismo, un análisis de permutación de clusters reveló una actividad específica de composición 260-550 ms luego de presentada la segunda palabra para uno de los controles, y a 410-600 ms para la segunda tarea control. Además, el análisis utilizado permitió explorar el transcurso temporal de esta actividad, que sugirió un posible confounding en el diseño experimental. Proponemos que esta paradigma permite el desarrollo de un proceso de expectativa solo para palabras en contextos combinatorios. El procesamiento de la primer palabra es condicional a la segunda palabra, por lo que una contingencia entre estímulos podría dar cuenta de la actividad de composición obtenida. Relacionamos este resultado con el potencial denominado variación contingente negativa (CNV) (Walter et al., 1964), un potencial anticipatorio evocado en tareas en las que un estímulo inicial anuncia la aparición de un segundo estímulo, y esta relación es gradualmente aprendida. Siguiendo el comportamiento descrito del CNV, mostramos un efecto de la progresión de la tarea sobre la actividad obtenida. Asimismo, diseñamos una tarea en la que controlamos por expectativa, y mostramos que la actividad previamente relacionada con un proceso de composición no se produce cuando los participantes no pueden predecir si van a tener que realizar un acto de composición.

En este trabajo presentamos evidencia de un “confounding effect” producido por la expectativa de los participantes en el paradigma propuesto por Bemis & Pylkkänen, y discutimos las implicancias de nuestro estudio para la literatura de composición usando MEG y sobre la replica conducida previamente utilizando EEG.

Contents

List of Abbreviations

ACC Anterior cingulata cortex

ANOVA Analysis of variance

aSTG Anterior superior temporal gyrus

CNV Contingent negative variation

EEG Electroencephalography

ERP event-related potential

FWER Family-wise error rate

IQR Interquartile range

LA List of adjectives

LN List of nouns

MCP Multiple comparisons problem

MEG Magnetoencephalography

MTG Middle temporal gyrus

PSP Postsynaptic potentials

RN Reticular nucelus

ROIs Regions of interest

SD Standard deviation

SMA Supplementary motor area

SNR-CI Signal-to-noise ratio confidence interval

SNRLB Signal-to-noise ratio confidence interval lower bound

Chapter 1

The neurocognitive basis of language processing

1.1 Introduction

Language productivity lies in the ability to combine a finite number of elements in an infinite way allowing us to create novel conceptual structures to convey original and specific meaning. It is clear that although the combinations are infinite there are legitimate constructions such as *she poured wine into the glass* and sentences that do not "feel" right as is the case of *she poured the glass with wine*. Therefore, there are specific principles that guide and limit the possible expressions enabling an efficient communication. The relationship between these rules and the units of meaning stored in long term memory is complex, expressions not only have to satisfy syntactic and semantic constrains, but also match people's world knowledge (Jackendoff, 2000).

Words are the building blocks we use to communicate with others, and they are typically employed to refer to objects or situations encountered in the world. Therefore, it is fairly intuitive that words are related to concepts. One view in psychology is that word meanings are mentally represented by mapping words onto conceptual structures. In this view, a word gets its significance by being connected to a concept or a coherent structure in our conceptual representation of the world (Murphy, 2002). When hearing a sentence, people are able to retrieve the mental representations associated to these words and combine them into a proposition that evokes the meaning of the expression as a whole, allowing a proper interaction with the environment. Therefore, for language comprehension to occur, complex meaning has to be rapidly constructed from discrete units, and the inverse process has to develop during language production. How the meaning of a complex expression is composed from the meanings of its constituent parts is a central concern for the study of language processing and production.

The principle of compositionality is usually evoked in formal semantics to refer to simple composition: *The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined* (Partee, 1984, 2007). This fits easily for simple constructions, such as *red boat* which requires us to construct a representation of a *boat* that is *red*. In this sense, meaning seems to be constructed bottom up from the lexical units. Nevertheless, semantic composition usually is not as straightforward, as some words can have multiple meanings. Take adjective-noun constructions in which the meaning of the modifier changes according to the head noun it modifies. A clear example proposed by Keenan (1979), is the adjective *flat*. In some combinations, flat means 'without marked projections or depressions', as is the case in *flat road* and other nouns related to surfaces. However, when flat refers to something that typically is inflated, like a tire, the meaning of flat changes to 'deflated', and can even refer to a lack of expected taste as in *flat beer*. Common adjectives share this property, allowing quite different interpretations in a strongly context dependent manner. Therefore, words can refer to multiple conceptual representations showing that the relation between words and concepts is far from one to one. This is of special relevance in most combinatorial contexts, where word content changes as different representations become active and as world knowledge shapes the final mental image.

One of the challenges in language research is to account for the functional and neuroanatomical basis of online meaning composition, specifically what are the neural computations that allow us to construct complex conceptual representations. Addressing this interesting question does not come without problems. There is an intrinsic methodological difficulty as semantic complexity usually correlates with syntactic complexity, making it hard to disentangle semantic composition from syntactical combinatorial processes. Furthermore, there is no consensus in linguistic theories on the interplay or coupling of both operations (Pylkkänen et al., 2011) from where to develop an empirical approach, nor if it is theoretically possible to separate these processes (Pylkkänen, 2008).

Recently, in a series of magnetoencephalography studies an experimental paradigm has been proposed to study the basis of meaning composition. In this work we explore this paradigm using electroencephalography and discuss a potential confounding that could account for the results reported in the literature.

1.2 Classical experimental paradigms in language processing

Research on language has been characterized by the use of elaborate stimuli to study syntactic and semantic composition during sentence processing. To this end, three basic methodological approaches have been proposed. A common line of experiments have compared subjects' responses to normal expressions, to responses to sentences showing some type of violation. Another approach has been to vary the complexity of sentences' syntactic structure or the degree of difficulty of the semantic interpretation. Finally, experiments in which normal sentences are compared to unstructured list of words, and the presence of syntactic or semantic information is varied, have been implemented (Friederici, 2011). These procedures have brought some insights into the complexity of language processing, here we discussed some of the classical paradigms and evidence.

1.2.1 Anomalous sentences: the N400 and P600 event-related potentials

A common approach in the study of semantic composition is to evaluate the characteristics of the event-related potential (ERP) N400, the first ERP related to language processing. It was described by Kutas and Hillyard (1980) as a negative potential peaking at 400 ms after a semantically incongruent word is presented in the context of a highly constrained sentence (e.g, He spread his warm bread with *socks*). The amplitude of this potential was initially correlated with the degree of semantic incongruency of the incoming word, as moderately incongruent sentences such as *He took a sip from the waterfall* elicited a smaller N400 than more semantically incongruent sentences like *He took a sip from the transmitter*. In terms of linguistic theory, this type of violation is considered a world knowledge violation, as semantically all sentences are correct but are in conflict with what we usually encounter in the world (Hagoort et al., 2004; Pylkkänen et al., 2011). This component has been extensively studied in comprehension and production tasks across different modalities (see Kutas and Van Petten (1994) for a review). Research has shown that all sentences' final words elicit a N400, and its amplitude is highly correlated with the cloze probability to that word (the percentage of individuals who would use that word to continue an incomplete sentence). Moreover, a sentential context is not necessary to elicit N400, as presenting pairs of word varying the degree of semantic association between them shows the same modulation in amplitude at this latency (Holcomb, 1993; Kutas, 1988). The gradual nature of this component was initially dually interpreted. One view proposed that N400 amplitude was an index of semantic integration of a word to the working context. The second interpretation, described this component as a reflection of the preactivation of features of the stored representation of the lexical item in long-term memory (Lau et al., 2008). Conciliating this two views, and the fact that N400 is not only elicited by linguistic stimuli (Sitnikova et al., 2008), Kutas and Federmeier (2011) proposed that "it reflects the activity in a multimodal long-term memory system that is induced by a given input stimulus during a delimited time window as meaning is dynamically constructed".

Following N400 discovery, a temporally posterior ERP named P600 was found to be related to language processing. The P600 was initially described as a reflection of the brain's response to syntactic anomalies. This positive potential was elicited around 600 ms after a word violating a syntactically constrained sentence is presented (Osterhout and Holcomb, 1992). P600 is also elicited by syntactic ambiguous sentences and as response to syntactic complexity (see Kuperberg (2007) for a review). Furthermore, it has been shown that this ERP can be modulated by semantic violations in syntactically correct expressions (Kim and Osterhout, 2005), consequently it is postulated that this component is sensitive to morphosyntactic as well as to thematic semantic constraints (Kuperberg, 2007).

The ERPs described above have provided evidence that language comprehension is a predictive process, in which context preactivates features to facilitate integration. The relation between N400 and the P600 effects might provide insights of the interaction between semantic and syntactic unification. Nevertheless, these paradigms do not shed light on the computations that occur at every step as words are combined into a unified concept.

1.2.2 Structured sentences and unstructured lists of words

A reasonable procedure to study the neural basis that allow us to generate complex meanings is to vary the amount and complexity of the computations involved when producing and understanding language. In a number of studies, mostly using brain imaging techniques, a paradigm in which structured sentences are compared to unstructured lists of words has been implemented based on the hypothesis that only the former would engage circuits responsible for composition (Friederici et al., 2000; Humphries et al., 2005; Mazoyer et al., 1993; Snijders et al., 2009). In these experiments, stimuli are designed to vary the presence of syntactic and semantic information in an attempt to disentangle their contribution in compositional processes. Subjects are typically presented with normal sentences (e.g, *the man on a vacation lost a bag and a wallet*) which carry both syntactic and semantic information, or sentences comprised of function words and pseudowords (e.g, *the solims on a sonting grilloted a yome and a sovir*), in which only syntactic computations would supposedly be elicited. The response to this sentences can be compared to unstructured word lists where semantic content can also be varied by presenting subjects with lists of pseudowords (e.g, *rooned the sif into lilf the and the foig aurene to*) or lists of words (e.g, *a ball the a the spilled librarian in sign through fire*) (These examples are taken from Humphries et al. (2006)).

The results of these studies indicate that Broca's area and surrounding regions in the inferior frontal cortex support syntactic processing (Friederici, 2011). Furthermore, it has been shown, that the frontal operculum supports local structure building (Friederici et al., 2000). In addition, the anterior superior temporal gyrus (aSTG) appears to be involved in syntactic processing (Humphries et al., 2006; Mazoyer et al., 1993), but the presence/absence of word semantics does not modulate this region (Friederici et al., 2000). Nevertheless, certain tasks and stimuli involving semantic aspects can activate the anterior temporal lobe. Pallier et al. (2011) showed that the anterior superior temporal sulcus and temporo-parietal junction elicited an increase in activity correlated with the size of linguistic structures only when phrases were comprised of words and therefore carried semantic information. A meta analysis on neuroimaging data carried by Hagoort and Indefrey (2014) indicates that ventral regions of the posterior left interior frontal gyrus (BA 45/47) respond to semantic demands whether they are due to violations, complexity or ambiguity, whereas the anterior region (BA 44) activates in relation to syntactic demands. A similar pattern is shown in the left posterior temporal lobe, syntactic factors activate the STG and the middle temporal gyrus (MTG), whereas semantic demands elicit activity in the MTG and inferior temporal gyrus.

Critics of the sentence versus word list paradigm argue that the stimuli used are typically not entirely void of structure, therefore the conclusions should be framed as responses to more or less complex structures. Furthermore, sentences made of pseudowords could be free of semantic information at the word level but have nevertheless meaning, take the phrase *the solims on a sonting grilloted a yome*, even though comprised of pseudowords it still conveys the notion of an event that occurred in the past where an agent performs an action on something (Pylkkänen, 2016). Nevertheless, it is probable that the cognitive operations involved in sentence processing differ considerably from those engaged by lists of words. Another valid criticism is the fact that these paradigms pose memory load demands that can be responsible for some of the observed effects.

The results of this approach show a network of regions involved in semantic and syntactic computations during online sentence comprehension. This paradigm has been used mainly while measuring brain activity with fMRI and PET, therefore the temporal aspect of these processes are not assessed. When trying to understand language processing, timing is of essential importance, as lexical access, syntactic parsing and integration with semantic compositional processes all take place very rapidly, and some probably in parallel (Friederici, 2011). In addition, it is possible that the same brain regions are involved in various aspects of language at different times. A study using this paradigm with magnetoencelography (MEG), evaluated the difference between normal sentences and word lists on a word to word basis. The electrophysiologial results were consistent with the hemodynamic results, observing a time-locked effect in the left anterior superior temporal gyrus (aSTG) and in the anterior middle temporal gyrus (aMTG), occurring between 290 and 340 ms after target word onset. Furthermore, they showed activation of a network of regions in this time window: increased activity was observed in anterior temporal regions bilaterally, posterior temporal regions, the left ventromedial prefrontal cortex, and the left pars opercularis (Brennan and Pylkkänen, 2012).

1.3 Composition and neural oscillations

Composition at the level of syntax and semantics has also been related to power changes induced in the electroencephalogram (EEG). Semantic violations have been show to elicit an increase in power in the theta band in temporal and midfrontal electrodes (Hald et al., 2006). In addition, an increase in theta and beta-band power as syntactically correct sentences unfold has been reported (Bastiaansen et al., 2010). Moreover, Weiss et al. (2005) found increased low-beta coherence between left-lateralised frontal and temporal regions for the processing of complex syntactic structures. Changes in alpha-band have been reported during auditory sentence processing (Krause et al., 1994; Segaert et al., 2018) and an increase of coherence in alpha-band has been shown in a reading task (Kujala et al., 2007). Most of the studies that measure power and coherence changes in language processing are carried out in complex tasks, therefore further research is necessary to understand the underlying processes that contribute to these results.

1.4 A paradigm to study basic composition

As discussed, trying to isolate the different operations that take place during language processing avoiding confounding effects of memory load and attention, has been challenging. Recently, a minimal experimental paradigm has been proposed to address this issue. Aiming to find the basic operation that takes place during language processing, one line of research has focused on modifier noun phrases, restricting stimuli to pairs of words (Bemis and Pylkkänen, 2011, 2013a,b,c; Pylkkänen et al., 2014; Westerlund and Pylkkänen, 2014). These authors propose that by minimizing stimuli complexity, a more precise study of semantic and syntactic computations can be pursued.

Inspired by the list paradigm, the initial study of this series developed a simple design to evaluate composition at its minimum by reducing it to pairs of nouns and adjectives (Bemis and Pylkkänen, 2011). Brain response was measured with MEG in two tasks, a composition task and a list task (Figure 1.1). The composition task consisted of two conditions, a one-word condition in which subjects were presented with a non-word followed by a word denoting a noun (*xkq boat*), and a two-word condition in which subjects saw two words, an adjective followed by a noun (*red boat*). After each condition subjects had to match the presented verbal stimulus to an image. In order to accomplish the task, participants had to unify the adjective and noun to respond correctly in a two-word trial, and only consider the adjective in a non-word adjective trial. For example, when presented with the words *blue star* a trial was congruent if it was followed by an image of a star that was blue, whereas in a *xqrt star* trial any colored image showing a star would be congruent to the verbal stimuli. In the list task, subjects were presented with two conditions, a one-word condition exactly the same as in the composition task, and a two-word condition in which two consecutive nouns were presented. Subjects had to indicate whether a subsequent picture matched any preceding word. The rational behind the experimental design and the data analysis was to find brain regions in which there was a difference in activity between the two-word condition and the one-word condition in the composition task, and no difference between conditions in the list task. By combining task and number of words in this design, they argue that any difference between conditions in the composition task that is not present in the list task can not be due to an effect of word number. They report an increased response in the left anterior temporal lobe (LATL) and ventromedial prefrontal cortex (vmPFC) in response to nouns in combinatorial contexts (*red boat*) over conditions in which composition was not possible or discouraged (*xkq boat, cup boat*), with the LATL response earlier (184 - 255 ms poststimulus), and the vmPFC response later (331 - 480 ms). In this study, LATL was defined as an area encompassing approximately the area around BA 38 and the anterior portions of BA 20 and BA 21; and approximately all of BA 11 was defined as the vmPFC. The authors interpret LATL and vmPFC activities as a reflection of the combinatorial processes elicited by the adjective-noun stimuli.

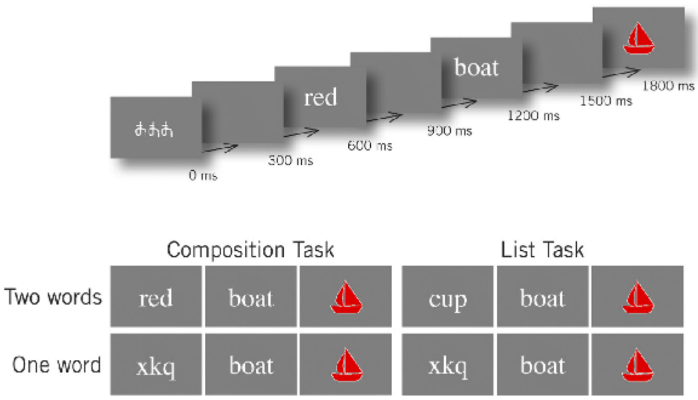


Figure 1.1: Experimental design taken from Bemis and Pylkkänen (2011). In each trial, participants indicated whether the target picture matched the preceding words. In the composition task, all preceding words were required to match, whereas in the list task, any matching word was enough

.

As discussed in their article this design does not allow to isolate semantic from syntactic composition as both processes would be engaged when reading pairs of adjective-noun words. However, this paradigm has been adapted to study semantic and syntactic operations. In this line, Westerlund and Pylkkänen (2014) explored the effects of noun specificity in combinatorial operations, showing a lower activity in the LATL when combining more specific nouns with an adjective (*blue canoe*), than less specific nouns (*blue boat*), indicating a semantic modulation of the activity in this region. Moreover, the inverse effect was shown for modifier specificity (Zhang and Pylkkänen, 2015) and this activity appears to only be elicited by intersective adjectives (Ziegler and Pylkkänen, 2016).

Although, other regions have shown activation for nouns in minimal combinatorial context (Bemis and Pylkkänen, 2011, 2013a; Pylkkänen et al., 2014), the most consistent brain area eliciting activation across studies is the LATL. This region's involvement in composition is supported by a series of fMRI studies on conceptual combination. Baron et al. (2010) and Baron and Osherson (2011) showed that the activation of the LATL for single word concepts such as *girl*, was correlated with the product of the activations of the concepts *child* and *female*. Suggesting that this region is involved in computations related to the act of unifying representations of meaning. It has been postulated that the LATL is involved not in syntactic or semantic computations per se but in constructing complex concepts by adding conceptual features into a unified representation (Poortman and Pylkkänen, 2016).

1.4.1 Adaptations of this paradigm to study basic syntax

This program of research focusing on composition at its minimum brought interesting information on temporal aspects of the combinatorial process, and enabled subsequent research that implemented versions of this paradigm to further explore language composition. In this context, Zaccarella and Friederici (2015), introduced a task to evaluate minimal syntactic composition with fMRI, reporting an activity circumscribed to a subregion of BA 44. Similarly, Segaert et al. (2018) used EEG to measure oscillatory changes in brain activity elicited by syntactic binding of pairs of pseudowords in auditory modality showing a power increase in left frontal-temporal electrodes in the alpha frequency band (8 - 12 Hz) following target word onset.

1.4.2 EEG replication and potential confound

Recently an EEG replication of Bemis & Pylkkänen's study was carried out in Neufeld et al. (2016). In this article the statistical analysis conducted involved a classical ERP analysis by targeting the time window reported by the original study and averaging over anterior and posterior electrodes. Electrode grouping was motivated by trying to relate the combinatorial activity with classical N400 context effects. Although in this work they report having found a composition-related activity in EEG recordings, we consider their results inconclusive as no interaction is found between task (composition, list) and number of words (two-word, one-word). Nevertheless, an interesting contrast to the original MEG study comes from the decision to epoch data to include first and second word, and to baseline the ERP to a time window preceding first word onset. This preprocessing choice enabled to test for differences in the time window preceding the critical noun. By visually inspecting the data, these authors selected a time window 50 ms before and 100 ms after second word onset and performed a t-test between the composition task conditions. The significant difference obtained was reported as a precombinatorial activity which would reflect the building of a syntactic template for the incoming word. Although this analysis should have only been carried out after showing an interaction between task and condition, the proposal of a process taking place before second word presentation only in a combinatorial context points to an alternative interpretation of the results obtained under this paradigm.

Given that Neufeld et al. (2016) and Bemis and Pylkkänen (2011) results were obtained with a block design it is possible that these authors introduced a confound that to our knowledge has not been addressed. In a two-word trial during the composition task, the first word indicated with certainty to subjects that they would have to combine both items. Therefore, it could be questioned whether the observed activity is composition-related or it is a more general activity reflecting expectancy. During the development of this work we came across a voluminous literature analyzing the signatures of expectancy-based processes. In particular an ERP called Contingent Negative Variation (Walter et al., 1964) was brought to our attention. As we will propose and evaluate an alternative hypothesis relating this component to the results obtained, we review here the literature on this topic.

1.5 Contingent negative variation

Organisms tend to find regularities in the world in order to establish causalities and elicit behaviors according to expected results. The ability to predict the occurrence of an event facilitates processing, resulting in a faster and more appropriate response to a specific task (Doya, 2008; J. Valone, 2006; Ridderinkhof et al., 2004). In this section we review one of the first neural correlates of this type of mechanism, the contingent negative variation (CNV).

CNV is a slow negative event-related potential first reported by Walter et al. (1964). In their work, Walter and collaborators presented subjects with pairs of stimuli, S1 and S2 in order to study brain electrical response to conditional and unconditional stimuli. The experiments consisted of a warning signal followed by an imperative stimulus after a variable foreperiod in the range of a few seconds. In the first experiment subjects were presented with a click sound (S1) followed by a flickering light (S2). When subjects had no task associated to the stimuli, a potential was elicited by S1 and a second potential was elicited by S2, disappearing after 50 presentations. However, when subjects were instructed to press a button to terminate S2, a large negative voltage was elicited at frontocentral electrode sites soon after the offset of S1 peaking before S2 onset, and returning abruptly to baseline after subject's response. In this situation, the negative potential was maintained as long as subjects were attentively performing the task. The activity decreased by diminishing the probability of S2 occurrence after S1, and could be recovered by returning to the original pairing of stimuli. Moreover, if subjects where told that they would no longer see S2, the potential disappeared instantly, and falsely stating to them that the light would be restored after the sound, caused again a negative response. Furthermore, subject's intention modulated this negative potential as it only appeared on trials in which participants decided to respond to S2. Finally, once the contingency between stimuli was learned, it was possible to elicit this activity in the absence of S2 by asking subjects to estimate the time-interval and perform a task, such as pressing a button. The authors proposed that CNV reflects a representation of causality, a contingency between two events (a conditional and an imperative stimuli) which is learned after a few dozen presentations and in some cases even less (Cohen, 1969; Hillyard, 1969). The following decades since CNV discovery numerous studies were conducted in order to determine which are the processes underlying this potential. Here we describe some of the evidence, for a review see McCallum (1993), Birbaumer et al. (1990) and Tecce (1972).

1.5.1 CNV consists of two different components

The results presented in Walter et al. (1964) were obtained by showing subjects pairs of stimuli separated by 0.5 and 1 seconds. Subsequent research in which S1 and S2 time interval was increased to 3 - 4 seconds, revealed a biphasic nature of the CNV (Loveless and Sanford, 1974; Weerts and Lang, 1973). An early component, starting after S1 onset and persisting for about 1.2 s with its maximum over frontal electrodes, was related to the orientation of attention and initially called *O*. The later wave, starting 1s prior to S2 and terminating around 0.2 s after, had a central distribution and was named *E* as it was considered an expectancy-related wave (Gaillard, 1976; Rohrbaugh and Gaillard, 1983; Rohrbaugh et al., 1976). Furthermore, it was hypothesized that this two components reflected different processes (Figure 1.2). The *O* component is sensitive to S1 properties. It has been shown that this component amplitude varies as a function of S1 intensity, having greater amplitude for more intense stimuli (Loveless, 1975). Moreover the *O* wave topography is in part determined by S1 sensory modality (Gaillard, 1976) and S2 properties do not affect this early component (Ritter et al., 1980). The terminal CNV or *E* wave has been consistently observed in paradigms in which S2 required a motor response but also in experiments in which S2 conveyed task-relevant information to the subject but did not require a response (Cohen and Walter, 1966; Klorman and Ryan, 1980; Simons et al., 1979). Therefore it was proposed that the early CNV was more related to the alerting properties of the stimuli and the later component reflected expectancy and preparation for motor response (Loveless and Sanford, 1974; Rohrbaugh and Gaillard, 1983).

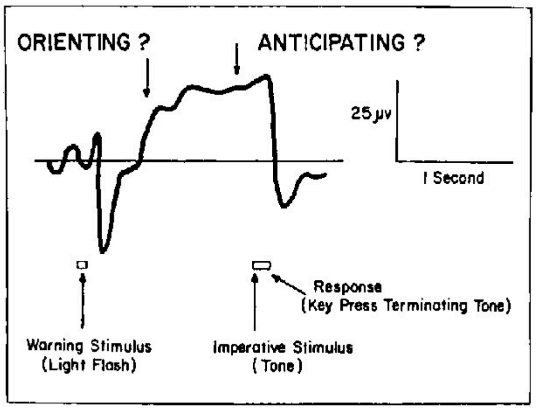


Figure 1.2: Prototipical CNV elicited in a foreperiod paradigm. Callaway (1975)

1.5.2 CNV and motor response

Various studies aimed at eliciting CNV without a motor response to isolate the contribution of this component to the CNV wave. This research was further motivated by the intention to distinguish the CNV from other slow negative potentials reported contemporary in the literature, such as the readiness potential (BP) (Jahanshahi and Hallet, 2003; Kornhuber and Deecke, 1965) and more recently the stimulus preceding negativity (SNP) (Brunia et al., 2011; Damen and Brunia, 1987). In Walter et al. (1964) original paper, it had already been shown how CNV could be elicited without motor response by asking participants to estimate the time until S2 appearance. This potential was also elicited in posterior experiments with longer time intervals and no motor response (Jarvilehto and Fruhstorfer, 1970; Ruchkin et al., 1986). Interestingly, CNV elicited without motor response is typically of lower amplitude, supporting the idea that there is a contribution of the readiness potential to the CNV wave (Irwin et al., 1966). This was further evidenced by a paradigm developed by Brunia and Vingerhoets (1981). In this experiment, subjects had to initiate a voluntary movement which triggered the presentation of a visual stimulus 4 s later, at that moment participants had to repeat the movement. Data shows that a BP was elicited for the first movement but a CNV arose only prior to the second movement. Additionally, both neural activities had similar topographies but CNV potential was bigger in amplitude.

1.5.3 CNV and attention

Attentional processes have been found to modulate CNV. Low et al. (1967) showed using a light and click sounds as S1 and S2, respectively, how decreasing the clicking sound intensity resulted in a higher amplitude CNV, suggesting that this potential is related to the attentiveness involved in detecting the imperative stimuli. In this line, Rebert et al. (1967) presented participants two types of trials in a random manner, a light that was paired with an easily detected sound, and a second type of trial in which the conditional stimuli was followed by a barely detectable auditory stimulus. CNV was more negative when the warning stimuli announced the low intensity sound. The same pattern was found in a speech discrimination task (Jacobson and Gans, 1981), and in a sound image pairing task (Simons et al., 1979). Furthermore, Tecce and Scheff (1969) presented participants with competing stimuli to S2 showing that distractors decrease CNV amplitude and result in a poorer performance. These experiments support the hypothesis that the more effective attention is given to a task, the bigger is the elicited CNV, whereas a state of impaired attention, gives rise to a smaller CNV (Tecce et al., 1976).

1.5.4 CNV as a temporal estimator

This negative potential has been also related to temporal estimation. In a temporal production task, CNV time course correlates with the interval length estimated by participants (Macar et al., 1999). This was seen in a temporal discrimination task in which subjects have to report if a probe interval is shorter or longer than a previously memorized duration (Pfeuty et al., 2005). In trying to distinguish CNV as a reflection of response preparation from a temporal accumulator (van Rijn et al., 2011), studies in which a passive CNV is elicited have been undertaken. In a passive temporal oddball task in which no decision nor motor response was required, participants were presented with pairs of stimuli separated by a variable ISI, manipulated according to an oddball probabilistic distribution. A CNV was elicited with a time course reflecting the most probable ISI (Mento et al., 2013). This evidence supports the fact that CNV reflects in part the temporal structure of the events and can be dissociated from other processes that contribute to this negative potential.

1.5.5 Learning effect on CNV

Different studies have reported an effect of task progression in CNV development. Firstly, enough trials have to be presented to subjects in order for them to extract the pairing pattern. Once the pairing is established there is conflicting evidence on the effects of time on CNV amplitude (Donald, 1980). For example in Walter et al's light-click task, the CNV is consistently elicited. However, during a classic defense-reflex conditioning an habituation occurs and CNV extinguishes after 60 trials, nonetheless subjects response is maintained. This effect of CNV diminishing over time has been also reported during a learning interval task (McAdam, 1966) and during a complex pair associate learning task (Peters et al., 1977). Although these studies show that the CNV tends to diminish or disappear with time while performance improves, there are some paradigms which sustain the CNV for longer periods of time. During a difficult pattern discrimination task, Poon et al. (1974) found that CNV amplitude developed until the pattern had been mastered, and no diminution was shown in over-learned trials. Furthermore, a learning-contingent CNV has been demonstrated in a reaction time paradigm only for attended pairs of stimuli in a sequence, showing no habituation (Proulx and Picton, 1980). Additionally, in a passive temporal oddball task (Mento et al., 2013), the CNV amplitude increases over time as subjects extract the most probable interval between stimuli. In short, there is evidence to support that some training is needed to elicit CNV as most studies show an increase of CNV amplitude during learning, and this development seems to vary according to task complexity. On the other hand, once the contingency is learned the effect of later trials on CNV amplitude is not as clear. Some tasks show a decrease or even a total disappearance of CNV, while others show this component to be constant. It is possible that CNV amplitude is dependent upon the type of task. Specifically, larger cognitive demands appear to sustain the CNV for a longest period after initial acquisition (Donald, 1980). However, there is evidence showing that CNV amplitude is inversely proportional to task difficulty or even unrelated (Tecce, 1972; Tecce et al., 1976). Donald (1980) proposes a motivational interpretation of CNV, where a subject's interest in a task might be unrelated to its complexity at the start of learning. Thus CNV would be initially similar across a wide range of tasks but would tend to decrease in later trials as the task becomes less engaging or demanding for the subject. This interpretation could account for the differences reported in the literature.

1.5.6 Neural correlates of CNV

The CNV is a complex negative potential that reflects multiple cognitive processes. It has been shown to be sensitive to attentional demands, sensory processing and motor preparation.

EEG studies have measured CNV in central parietal regions using a temporal estimation paradigm with visual and auditory stimuli (Macar and Vidal, 2002, 2003), this has been further explored by PET suggesting the supplementary motor areas (SMA) and the anterior cingulate cortex (ACC) (Macar et al., 2002) as possible CNV origins. This is supported by evidence from MEG studies which associate the primary motor cortex, the ACC, and the SMA with CNV generation (Ioannides et al., 1994). Similar results are shown with fMRI (Lee et al., 1999), where they further circumscribe SMA involvement to its anterior portion. Furthermore, (Nagai et al., 2004) showed that thalamic, cingulate, and insular activity are related to the early CNV component during a forewarning reaction time task. Additionally, basal ganglia has been implicated in CNV generation (Bares and Rektor, 2001; Ikeda et al., 1997). The involvement of SMA in the generation of CNV is expected as this area is related to motor preparation and execution of movements (Lee et al., 1999). As for the ACC, it has been shown that activity in this region is modulated by uncertainty to an event, detection of negative rewards and is involved in the generation of states of autonomic arousal (Critchley et al., 2001, 2003; Martin et al., 2009). Furthermore, the ACC is implicated in tasks that require working memory and orienting of attention (Luks et al., 2002).

Slow cortical negative potentials are considered to result from an increase of excitatory postsynaptic potentials in the upper layers of the cortex. Early studies indicate that this potentials are produced by the depolarization of apical dendrites of cortical pyramidal cells by thalamic afferents (Birbaumer et al., 1990). Thus, the origin of these slow potentials come from sub cortical levels, specifically both the mesencephalic reticular formation and the thalamic reticular nucleus (RN) play a role in this process. In this line, subcortical structures have been related to CNV generation both in humans and non human primates (McCallum et al., 1973; Rebert, 1972; Yingling and Skinner, 1976). It has been proposed that CNV modulation is given by the thalamic RN acting as a filter of sensory information between the thalamus and the cortex (Birbaumer et al., 1990; Brunia et al., 2011; Elbert, 1993; Nagai et al., 2004; Skinner and Yingling, 1976). Increasing attention to a specific event requires filtering sensory information that is not relevant, thus the CNV increase observed when subjects are asked to sustain attention during a time interval, could be given by an increase in thalamocortical sensory information flow.

Chapter 2

Methods

2.1 Measuring brain activity with MEG and EEG

The movement of ions across a neuronal cell produces electrical currents. Under certain conditions, non invasive methods such as MEG and EEG are sensitive to the electric and magnetic fields associated with this current flux. For this to occur, it is necessary that a sufficiently large spatially organized group of neurons respond in a coordinated way. Pyramidal neurons in the cortex often satisfy these conditions.

A prototypical pyramidal cells has a conic-shaped cell body with two dendritic domains. A long apical dendrite emerges from the soma and bifurcates, giving rise to a dense tree shaped dendritic pole, and a basal dendritic domain comprised of short dendrites. The soma and the axon receive inhibitory inputs, whereas excitatory synaptic input takes place at the level of the dendrites. Typically, proximal dendrites receive excitatory inputs from local sources while the distal apical tree receives inputs from more distant cortical and thalamic locations (Spruston, 2008). Dendritic trees of pyramidal cells are arranged parallel to each other and perpendicular to the cortical surface. Given their morphology, activation of these cells can generate strong dipoles along the somatodendritic axis. When an excitatory input arrives to the apical dendritic tree, positive ions will flow from the extracellular to the intracellular space, producing a local negative imbalance of charges outside the cell. Current will flow out from the cell body and basal dendrites to achieve electroneutrality, resulting in a positive potential in these areas. Due to the spatial separation between the source and the return currents, an open field is generated which can be detected at a distance from the neuronal sources (Buzsáki et al., 2012). As current flows through a conductor it is accompanied by a magnetic field that flows around it, thus in pyramidal cells magnetic field lines are created around the neuronal main axis (Hämäläinen et al., 1993). When neural activity occurs in a synchronous manner across a great number of neurons, the coherent field potentials and local magnetic fields produced become big enough to be picked up by EEG and MEG, respectively. Different sources are responsible for this activity but postsynaptic potentials (PSP) are the main contributors. These are small and slow neuronal responses which can last between tens or hundreds of milliseconds and occur locally in the dendritic trees. The orientation of pyramidal cells allows a summation of the small dipoles of many neurons, and the resulting voltage can be measured at a distance. Usually, due to its temporal and spatial properties, typical action potentials are not detected by these methods.

Even though there is great overlap between the information provided by MEG and EEG, the techniques have some important differences. Firstly, the electrical and geometrical properties of the layers of tissue that separate the cortical sources from the surface, distort and undermine the signal. This has greater impact on EEG as electrical conductivities are more variable across structures than magnetic permeability (Okada et al., 1999). Furthermore, because of the folded nature of the cortex, pyramidal neurons have their dendrites parallel or perpendicular to the cortical surface resulting in dipoles that are tangential and radial to the skull. Whereas EEG is sensitive to both, MEG can only measure magnetic fields that are perpendicular to the skull, as neuronal currents oriented radially do not generate a magnetic field outside the head. Estimating the brain sources that are responsible for a given potential distribution as measured from the scalp surface, is called the inverse problem. It is possible to offer an infinite number of sets of dipoles that can explain a given voltage distribution (Von Helmholtz, 2004), therefore to solve this it is necessary to apply certain assumptions. A common approach corresponds to model the source as a current dipole (equivalent to the net postsynaptic electrophysiological activity of local groups of neurons), or as distributed patches of activation on the cortex, typically minimum norm estimates (Hansen et al., 2010). Even though EEG is sensitive to radial and tangential dipoles, it is typically consider that MEG is better to resolve source localization. EEG source reconstruction requires numerous electrodes, knowing their positions with precision, having head shapes measurements and accurate estimates of tissue conductivities. Thus, in practice, MEG is often better to resolve brain areas' location as magnetic fields are not distorted by the different brain structures. As for temporal resolution, both techniques have a millisecond resolution, therefore are useful when there is an interest in the temporal course of a neural process (Luck, 2005).

2.2 Event-related potentials and time-frequency analysis

The recorded cortical responses using EEG can be characterized in the time domain to examine event-related potentials (ERPs), or in the time-frequency domain to study phase and power of oscillatory activity.

Neural oscillations are rhythmic fluctuations in the excitability of neurons or populations of neurons, and this activity has been related to a number of cognitive processes (Buzsáki, 2006). It has been proposed that oscillations originate from the interaction between inhibitory interneurons and excitatory pyramidal cells. The activation of pyramidal cells increases as they excite each other, the concurrent activation of interneurons within this population produces the inhibition of pyramidal cells. As the activity of the interneurons decline, pyramidal cells return to an excitatory state. The fluctuation between an excitatory and an inhibitory state is considered to be the basis of neural oscillations. Furthermore, oscillations can be present in exclusively inhibitory or excitatory circuits due to the specific properties of voltage gated channels expressed in neurons (Buzsáki, 2006).

ERPs are a measure of brain activity corresponding to a stereotyped electrophysiological response to a given stimulus. They are obtained by averaging subject's response to a given stimulus, and typically present peaks of negative and positive voltage at different latencies according to the type of stimuli or cognitive process involved (Luck, 2005).

Brain activity measured by EEG can be classified into ongoing, evoked or induced oscillations. Evoked and induced oscillations differ in their phase-relationship to the stimulus (Tallon-Baudry and Bertrand, 1999), thus they are also referred as phase-locked and non-phased-locked respectively. Phase-locked activity is phase-aligned in time to a specific event, therefore it can be obtained when averaging in the time domain (ERPs) and when averaging in the frequency domain. Contrarily, non-phased-locked activity occurs when the responses associated to the presented stimuli are not in phase, averaging over time results in their cancellation, rendering it undetectable by means of ERPs. This activity can be observed in the frequency domain as a difference in power. Finally, ongoing or background activity is unrelated to the task or stimuli presented (Cohen, 2014).

The origins of ERPs are not well understood but different hypotheses have been proposed (Burgess, 2012). One view is that ERPs are a reflection of the evoked activity by an external stimuli, or an internal process, which is added to the ongoing oscillations. When averaging the responses in time, only the ERP remain as it is the only activity time-locked to the stimuli. Another proposal is the phase reset model, in which ERPs are considered the result of a phase alignment of the ongoing oscillations in response to a stimulus. Although ERPs origin is not clear, they have been related to motor, sensory and cognitive processing, and given EEG temporal resolution they are a good tool to study the temporal course of brain's response to a specific stimulus.

2.3 Analyzing EEG data

When measuring EEG data one is typically interested in testing if at least two experimental conditions are different. To this end one records from a number of channels and may want to evaluate differences in a given time window. If one decides to test each channel - data point pair by means of a t-test there is a probability of finding a positive result as the number of comparisons increase merely because of random fluctuations in the data. This gives rise to the multiple comparisons problem (MCP). The MCP entails that given the number of statistical comparisons it is not possible to control for the family-wise error rate (FWER), specifically for type I errors, which is the probability of obtaining a significant difference between conditions given that the null hypothesis si true (see Goodman (1998)). There are different ways to solve this problem. For parametric tests there are various methods to correct for multiple comparisons (Shaffer, 1995) such as Bonferroni correction (Dunnett, 1955), which decreases the critical alpha value in order to compensate for the amount of comparisons tested. This usually results in type II errors, as effects have to be very big in order to surpass such a strict threshold. For this reason if there is a clear hypothesis on electrode sites and data points, one should record from a few electrodes and reduce the number of comparisons. However, if one has no a priori hypotheses as where to expect the effect and decides to register data from numerous channels, one should avoid implicitly performing multiple comparisons by exploring the obtained ERP waveforms and based on this visual inspection decide on the latencies, amplitudes and electrodes to test (Luck, 2005).
Another approach to EEG data analysis is the use of non-parametric methods which are described in detail in the following section.

2.3.1 A non-parametric test for EEG - permutation analysis

Non-parametric permutation analysis for EEG data was first introduced by Blair and Karniski (1993) and is illustrated clearly in Maris and Oostenveld (2007). This method accounts for the MCP, admits flexibility as to which test statistic is used to evaluate differences between conditions and allows the introduction of biophysical constraints that improve their sensitivity. Furthermore, non-parametric tests take into account our ignorance when spatial and temporal distribution of the pursued effect are unknown.

This tests consists on calculating a stastistic for the original data, and then obtaining the same statistic for many permutations of the data. For each permutation the condition labels are randomly exchanged between conditions and then the statistic is calculated, finally a distribution of the statistic is constructed. The p-value for the original data is computed relative to this generated distribution and is set equal to the proportion of permuted datasets that produced a test statistic more extreme than the actual data. A step by step for a two condition experiment based on Maris and Oostenveld (2007):

1. Collect the condition average of the two experimental conditions for all subjects in a single set.

2. Randomly draw as many condition averages from this combined data set as there are subjects and place them in subset 1. Place the remaining in subset 2.

3. Calculate the test statistic on this permuted dataset.

4. Repeat steps 2 and 3 a large number of times and construct a histogram of the test statistics.

5. From the test statistic obtained for the real data and the histogram in step 4, calculate the p-value as the proportion of permuted datasets that resulted in a larger test statistic than the observed one.

6. If the p-value is smaller than the critical alpha-level, then it is possible to conclude that the data in the two experimental conditions are significantly different.

Moreover, this test allows to design a specific statistic by incorporating prior knowledge and enhancing the sensitivity of the statistical test. For example when measuring EEG it is expected that closely placed electrodes will measure similar neuronal activity, therefore it makes sense to include neighboring information. In addition, we know that cognitive processes develop over time. Accordingly, a process measured at one time point probably will still be occurring in subsequent time points. A type of analysis that incorporates this knowledge is the cluster permutation analysis.

Cluster permutation analysis

Cluster permutation analysis is a non-parametric method that follows the steps previously described but in which the test statistic defined is based on clustering of adjacent time-samples and electrodes that exhibit a similar difference. An example of a cluster permutation analysis using the maximum sum cluster statistic:

1. Perform a t-test between conditions for every data point.

2. Select data points that have a t-value larger than a given threshold.

3. Cluster the selected samples based on temporal, spatial and/or spectral adjacency.

4. Calculate cluster-level statistics by taking the sum of the t-values within a cluster.

5. Select the largest-value.

6. Repeat the previous steps for a large number of permutations and use the largest-value for each permutation to build the cluster statistic distribution

7. Finally compare the clusters statistics obtained for the real data with the generated cluster statistic distribution.

With this procedure instead of evaluating the difference between the experimental conditions for each data point and electrode, it is evaluated by means of a single test statistic for the entire data and therefore, the MCP is avoided.

2.4 Bemis & Pylkkänen's cluster permutation analysis

The original experiment by Bemis and Pylkkänen (2011) uses a highly hypothesis-driven cluster permutation analysis. Specifically, they expected the composition-related activity to be greater within the two-word condition compared with the one-word condition, and no difference between conditions in the list task. Initially they identify clusters of continuous data points that show an interaction between task (composition, list) and number of words (two-word, one-word) by means of a 2×2 repeated measures ANOVA. Clustering was done by imposing that at least ten adjacent data points had a p-value for the interaction of less than 0.30. Secondly, for each data point of each cluster a t-test was performed within each task. In this way, for each data point they obtained two t values (t_{comp}, t_{list}), corresponding to the difference between conditions in each task (see Figure 2.1). The statistic for a cluster was calculated as:

$$s = \sum_{i}^{N} t_{comp} - \sum_{i}^{N} t_{list} \tag{2.1}$$

Furthermore, they select regions of interest (ROIs) previously described in the literature as having show relation to the binding of items, or well established regions that support language processing, and performed the above analysis for each ROI.

In an initial attempt to follow this procedure in our study we coded an implementation in MATLAB (code can be found in this repository on github). However, selecting the parameters such as the initial threshold, number of time points, and ROIs for EEG in an objective way was an issue we did not manage to resolve and therefore decided to follow a more general analysis.

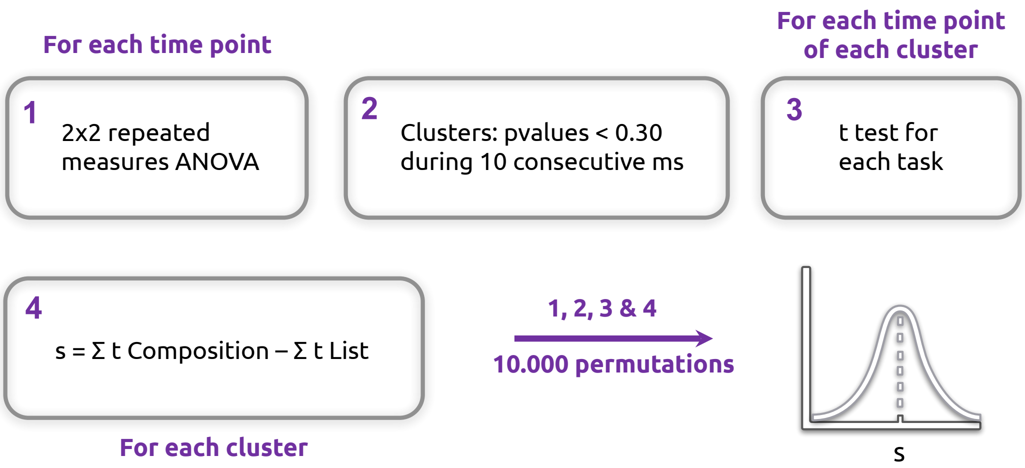


Figure 2.1: Outline showing Bemis and Pylkkänen (2011) cluster permutation analysis statistic

2.5 Cluster permutation analysis on event-related potentials

After recording from numerous channels, the selection of a group of electrodes to proceed the analysis conflicts with the amount of resources destined to collect the data, and can incorporate researcher bias as well. Furthermore, when analyzing a specific time window it is possible to obviate unexpected effects. Finally, it is important to show the boundaries of an effect, by displaying other channels and data points that do not reach significance (Picton et al., 2000). Accordingly, in our study we use this method with a data-driven approach, evaluating all recorded channels and data points from epoch onset until the moment subjects were asked to elicit a response. In each analysis, a t-test was performed on every sample and t values were clustered depending on if they exceeded a dependent samples t-test threshold of $p < 0.05$ (two-tailed). t values for each data point within each cluster were summed in order to obtain a value per cluster. The maximum negative and positive cluster statistic values were kept. This was done for 5000 permutations of the data resulting in a null hypothesis distribution of the statistic, against which we tested the real data. We considered the critical α level here to be 0.025. For all the analyses using this method on event-related potentials we set to three the minimum number of electrode neighbors that had to be significant for a given time point in order to be part of a cluster. All cluster permutation analyses were conducted on MATLAB using Fieldtrip toolbox (Oostenveld et al., 2011) and neighbors were defined following Biosemi 64 neighbors template.

2.6 Threshold-free cluster permutation analysis

When applying a cluster permutation analysis there are various parameters that can be defined, such as the critical value that will be used for thresholding the sample specific t statistics, the minimum number of continuous data points that have to surpass this threshold and the minimum number of neighbors. Specifically, the initial threshold can have big consequences: a low one can result in a single large cluster, whereas an extremely high threshold may result in no clusters being found at all. The set of parameters will influence the final result and while doing exploratory research there are no clear reasons that justify selecting a set of parameters above others. Another limitation is the interpretation of the result, as there is no information on the contribution of each element that takes part in the cluster, only inferences of the cluster as a whole can be made.

In order to address these issues and support the results obtained with the cluster permutation analysis we tested our data with a threshold-free cluster enhancement analysis (TFCE). This method is implemented for EEG data in Matlab (Mensen and Khatami, 2013) and is based on a similar solution for fMRI (Smith and Nichols, 2009). The idea behind this method is to find signals of big magnitude but which show a very focal distribution and at the same time enhance low signals that are more widely distributed across channels. The initial statistic is enhanced at every data point by integrating the height of the point of the statistic (h) and the extent of the neighborhood surrounding that point (e), computed for all possible cluster forming thresholds.

The difference between two ERPs at a given channel and time point (p) is some combination of this parameters. It is possible to give more relevance to each parameter by modifying their weights, as denoted by E and H. The TFCE score can be aproximated for every p integrating the height and extent obtained for all possible cluster thresholds (Figure 2.2), following equation 2.2.

$$TFCE(p) = \int_{h=0}^{hp} e(h)^E \times h^H dh \tag{2.2}$$

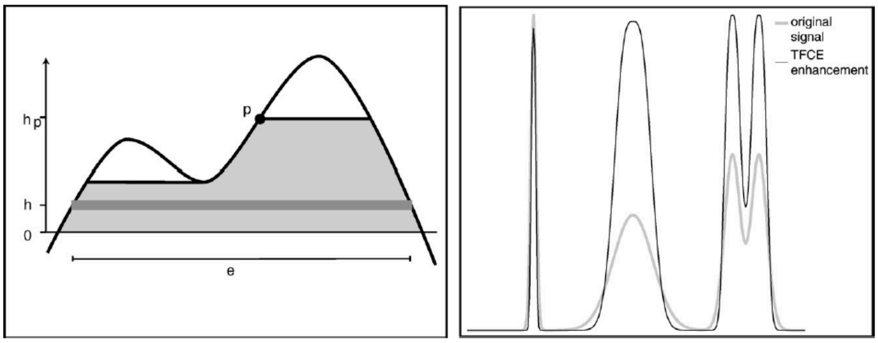


Figure 2.2: Illustration taken from Smith and Nichols (2009) Left: the TFCE score at time-location p is given by the sum of the scores of all incremental supporting sections (such as the one shown with a dark-grey band) within the area of "support" of p (light grey). The score for each section is a function of its height *h* and extent *e*. Right: example of input data and TFCE-enhanced output. The input contains a localized high signal, a spatially extended signal but of lower intensity and overlapping signals of intermediate extent and height.The TFCE output shows the same values for all three cases maintaining local maximums

The integral is estimated as a sum using finite dh by applying the thresholding in evenly spaced steps (s), between statistical values of 0 and the maximum value found in the data (eq. 2.3).

$$TFCE(p) = \sum_{h=0,s,2s...hmax} e_p^E \times h_p^H dh \quad (2.3)$$

Importantly, information about peaks and troughs will be maintained as the signal is enhanced at every point and then analyzed.

An implementation of TFCE and the set of parameters for EEG that control for type I FWER, was shown by Mensen and Khatami (2013) (see also Pernet et al. (2014)). Simulations used in this article suggest setting E=2/3 and H=1, which are the parameters followed in the present study. The drawback of this analysis is that at the moment there are few studies applying it to EEG data, and therefore the choice of parameters can only be guided by the mentioned publication.

Chapter 3

Simple composition with EEG

An discussed in the previous chapters an experimental paradigm to study the fundamentals of language composition has been proposed in Bemis and Pylkkänen (2011). Showing participants modifiers and nouns, specifically adjective-noun pairs, elicited an activity around 250 ms in the LATL as measured by MEG. This response is interpreted as a reflection of the successful act of joining two pairs of words into a unified concept. Although, it is not exactly clear if this activity corresponds to a purely semantic operation, this paradigm presented itself as an interesting starting point to study how meaning is constructed. An EEG replication of the original study was carried out by Neufeld et al. (2016). In this article the analysis conducted is questionable and the results are inconclusive. Therefore, in this work we carry out a version of this paradigm using a more objective and robust statistical analysis to explore EEG's sensitivity to this reported composition marker, and explore alternative interpretations of Bemis & Pylkkänen and Neufeld's results.

3.1 Hypothesis

Language composition is associated with a specific neural activity peaking around 200 ms after the presentation of a word in a combinatorial context, as measured by MEG. Given the similarities between EEG and MEG, we expect that the composition-related activity reported in the literature can also be observed in EEG recordings.

3.2 Objectives

General

Detect and study the neural signatures of language composition using EEG.

Specific

1. Adapt the paradigm proposed in Bemis and Pylkkänen (2011) to Spanish by creating two control conditions to account for the different linguistic properties between languages.
2. Use a rigorous statistical method to analyze the results avoiding the MCP and biased selection of samples.

3.3 Materials and methods

3.3.1 Participants

Twenty-nine non-colorblind undergraduate Uruguayan students participated in the experiment (22 female, average age, 25.31 $\pm$ 0.56). Participants were recruited through the Centro de Investigación Básica en Psicología. All subjects were native Rioplatense Spanish speakers, right handed, and had normal or corrected vision. The experiments were approved by the ethics committee of the Facultad de Psicología, UdelaR. All participants gave informed consent and were not awarded any economic or academic retribution, according to the national established guidelines (Decree N°379/008 - Investigación en seres humanos).

3.3.2 Stimuli and procedure

Experimental design

Bemis and Pylkkänen (2011)'s study was designed for English speakers. This language follows very rigid rules as for the type of constructions allowed for nominal phrases. Specifically, English noun modification is pre-nominal meaning that nouns are always preceded by the adjective. In contrast, in Spanish we can find different constructions depending on the type of adjective. Interestingly, intersective adjectives such as color, shape or evaluation tend to be used in a post-nominal way, and their pre-nominal uses are mainly restricted to literary resources as is the use of epithets (Bosque and Demonte, 1999).

Taking this into account we adapted the composition task so that subjects were presented with *noun adjective* trials in one of the conditions and *non-word adjective* trials in the other condition. For the list task we were not entirely certain in using a list of nouns as we would be presenting subjects with a noun as second word in the two-word condition, and would differ with the type of word presented to subjects in the same condition for the composition task (*bote rojo VS. bote tren*). Nevertheless, using a list of adjectives could have its own drawback as color adjectives are typically more abstract than regular nouns. We took a conservative approach and decided to have subjects do both types of list tasks. Accordingly, subjects did a list of nouns task (LN) in which they were presented with *noun noun* trials and *non-word noun* trials, and a list of adjectives task (LA) in which they saw *adjective adjective* trials and *non-word adjective* trials (Figure 3.1).

During the composition task participants were instructed to answer whether the image matched the preceding words. In contrast, in the list tasks subjects had to answer whether the image matched any of the preceding words. Half of the participants started the experiment with the composition task, and half started with the list tasks. The list task order and the yes/no response hand was counterbalanced across subjects. Although participants knew they would be doing three tasks, they were not given any specific instructions of the incoming task until the previous one was completed. Subjects were encouraged to answer as quickly and accurately as possible. Each task consisted of 200 trials that were preceded by a 40 trial practice to ensure that participants understood the task and learned the response keys. During each trial participants saw a fixation cross for 1 s and all stimuli except for the images were presented for 300 ms and were followed by a 300 ms blank screen. The images presented at the end of each trial remained on screen until subjects pressed a key or after 3s (see Figure 3.1). Following the end of each trial a sound was elicited to encourage participants to blink at that time. The inter-trial interval was randomly varied between 0.8 - 1.5 s. Stimuli presentation was coded in Psychopy (Peirce, 2007) and displayed on a CRT monitor with a 60 Hz refresh rate.

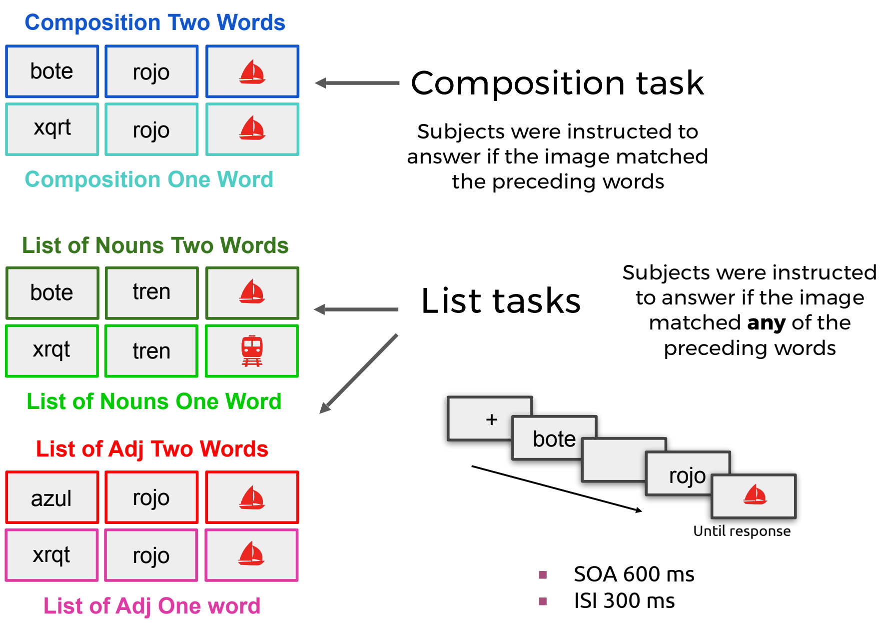


Figure 3.1: Spanish adaptation to Bemis and Pylkkänen (2011) experiment. Subjects were presented with three tasks. A composition task and two list tasks with two-word and one-word conditions. The list tasks differed in word type, nouns or adjectives

Stimuli construction

Following Bemis and Pylkkänen (2011), 11 nouns (*tren, bote, lápiz, reloj, globo, gorro, cable, cepillo, tenedor, teléfono, zapato*) and 11 color adjectives (*amarillo, azul, blanco, celeste, marrón, naranja, negro, rojo, rosado, verde, violeta*) were selected to generate the stimuli. Nouns and adjectives were matched for frequency ($p = 0.059$), number of letters ($p = 0.51$), number of substitution neighbors ($p = 0.92$), number of phonemes ($p = 0.50$), number of syllabus ($p = 0.58$), number of homophones ($p = 0.22$), and number of phonological neighbors ($p=0.95$). It was not possible to compare words on familiarity, imageneability and concreteness as there was no available information for 7 out of the 11 color adjectives. p-values correspond to a two tail t-test and properties were taken from Duchon et al. (2013), a lexical database of psycholinguistic variables for Castillian Spanish. This database was used as there is none available for Rioplatense Spanish which is Uruguayans' population main language. Non-words (*brnlqs, slgrl, grsd, vrpng, jlcrfsmt, cxgnff, drbcw, tphn, dpjzb, pkrdt, vqdfnsm*) were constructed using Wuggy software (Keuleers and Brysbaert, 2010), and did not differ in number of letters from the adjectives ($p=0.40$) nor from the nouns ($p=0.87$).

For the composition task a python script was created in Spyder 3.3.1 to automatically select 10 nouns, and 10 adjectives, and to combine them in order to generate 100 congruent and 100 incongruent trials. This was done independently for every subject. Half of the trials for each condition were incongruent. In the two-word condition a trial was incongruent if the noun (25 trials) or the adjective (25 trials) did not match the image. In the one-word condition a trial was incongruent if the adjective did not match the image (50 trials).

A second python script was created to generate the stimuli for the list tasks. For each subject for each list task, permutations of the 11 words were obtained and 10 items were randomly selected and discarded. For the one-word condition the 11 words and the 11 non words were combined, 21 items were randomly selected and discarded. For the one-word condition of every task, a trial was incongruent if the image did not match the word (50 trials). For the two-word condition a trial was incongruent if the image did not match any of the preceding words (50 trials).

3.3.3 Behavioral data analysis

Reaction times

Response times were measured from image onset until subjects pressed no/yes key. For the behavioral data we used R (R Core Team, 2013) and lme4 package (Bates et al., 2015) in Rstudio 1.1.456 (RStudio Team, 2016), to implement a linear mixed-effects analysis of the relationship between reaction times and the interaction between task and number of words.

For each participant and for each task, reaction times were analyzed by removing all incorrect and missing responses and trials in which response times were over or under two standard deviations were discarded (4.58% of all trials). A model with task as a three level factor (*composition - list of nouns - list of adjectives*) and number of words as a two-level factor (*two-word - one-word*) was fitted. Random intercepts were also fitted for subject, first and second word, as words differed across tasks and were slightly different across participants (eq. 3.1). Visual inspection of the residual plots revealed some deviation from normality therefore a logarithmic transformation of the reaction times was conducted.

$$log(RT) \sim Task \times NumberWords + (1|Subject) + (1|Word1) + (1|Word2) \tag{3.1}$$

The model was tested with an ANOVA-like Chi Square-test and pairwise comparisons were carried out using emmeans R package (Lenth, 2018). Bonferroni method was used to correct for multiple comparisons.

Accuracy

In order to test the relationship between accuracy and the interaction between task and number of words, we implemented a generalized linear model with task as a three level factor (*composition - list of nouns - list of adjectives*), number of words as a two-level factor (*two-word - one-word*) and image congruency as a two-level factor (*congruent - incongruent*). We incorporated subject as random effect to the model (eq. 3.2).

$$Accuracy \sim Task \times NumberWords + Congruency + (1|Subject) \tag{3.2}$$

Following the analysis for reaction times, the general linear model for accuracy was tested with an ANOVA-like Chi-squared test. Pairwise comparisons were performed using emmeans R package (Lenth, 2018) and Bonferroni method was used to correct for multiple comparisons.

3.3.4 EEG recording and preprocessing

EEG signal was recorded using a Biosemi Active-Two system (Biosemi, B.V., Amsterdam, Netherlands), which has a DC coupled amplifier. Sixty-four Ag-AgCl scalp electrodes where placed on an elastic head cap following the location and label of the 10-20 system (Jasper, 1958). Ocular movements were monitored by 4 electrooculographic (EOG) electrodes (above, below the left eye, and on the outer canthi). Electrodes were placed on mastoids as a potential reference which we finally decided not to use. The activity recorded was referenced online to the common mode sense (CMS; active electrode) and grounded to a passive electrode (Driven Right Leg, DRL), creating a feedback loop that drives the average potential of the participant to the AD-box reference potential. Data was digitized with a sample rate of 512 Hz with a fifth-order low-pass sinc filter with a -3 dB cutoff at 102 Hz.

Epochs time-locked to first word onset

Data was preprocessed in MATLAB using fieldtrip toolbox (Oostenveld et al., 2011). Continuous data was two-pass filtered with a second-order high-pass Butterworth filter at 0.1 Hz and a fourth-order low-pass Butterworth filter at 30 Hz. Data was epoched 0.2 s prior to the onset of the first word until 1.5 s (0.3 s after image presentation). Epochs were baselined to activity 200 ms preceding first word onset. Noisy trials and channels were rejected using an adaptation of Junghöfer et al. (2000). Channels were rejected if they were above a threshold given by equation 3.3.

$$th = mstd + 3 \times \sqrt{\frac{\sum_{i}^{N}(std_i - mstd)^2}{N}} \tag{3.3}$$

where $mstd$ is the median of all channels standard deviation, std_i is the standard deviation for channel i and N is the number of channels. Trials were then referenced to the averaged of the remaining electrodes. Trials were rejected if for any electrode they were above both thresholds given by equation 3.4 and equation 3.5.

$$thtrialStd = mstdE + 4 \times \sqrt{\frac{\sum_{i}^{t}(stdtrial_i - mstdE)^2}{t}} \tag{3.4}$$

$$thtrialMax = mMaxE + 4 \times \sqrt{\frac{\sum_{i}^{t}(Mxtrial_i - mMaxE)^2}{t}} \tag{3.5}$$

Where $mstdE$ is the median standard deviation of all trials for electrode E, $stdtrial_i$ is the standard deviation of trial i for electrode E, and t is the number of trials. As for equation 3.5, $mMaxE$ corresponds to the median absolute maximum for all trials of electrode E, and $Mxtrial$ is the maximum value for trial i for electrode E. Furthermore, trials in which participants responded incorrectly were discarded. The remaining trials for each condition where averaged in order to obtain one ERP per subject per condition. For the composition task the number of rejected electrodes and trials was 1.52 ± 1.16 and 23.19 ± 6.97, respectively. A t-test showed no difference in the number of trials rejected between conditions ($p = 0.16$). For the LA task the average of discarded electrodes was 2.17 ± 1.97 and the number of trials rejected was $21, 76 \pm 6.80$, there was no difference in number of trial rejected between conditions ($p = 1.22$). For the LN task the average of electrodes rejected corresponded to 1.55 ± 1.22 and the average for trials was 23.30 ± 7.22. No difference between conditions for the number of rejected trials was found ($p = 0.96$). A one-way ANOVA was conducted to compare the effect of task on number of channels rejected, no difference was found $F(2, 78) = 1.56, p = 0.22$. Accordingly, a second one-way ANOVA showed no difference between task and number or discarded trials $F(2, 78) = 0.51, p = 0.60$.

Subjects' ERP signal-to-noise ratio was evaluated following Parks et al. (2016). The approach uses bootstrap resampling of ERP waveforms from each subjects' trials to compute a signal-to-noise ratio confidence interval (SNR-CI) for the average waveform. This article suggests a lower bound of SNR-CI (SNRLB) of 3.0 dB as threshold to ensure signal quality, which is the criterion followed here. For the composition task the SNRLB ranged from 3.54 to 21.87 dB (mean = 11.25, median = 10.26 dB, SD = 5.23 dB, IQR = 7.33 dB). The SNRLB for the list of adjectives task ranged from 0.15 to 16.10 dB (mean = 8.31 dB, median = 8.34 dB, SD = 3.41 dB, IQR = 3.65 dB). Subjects 13 and 19 failed to meet the SNRLB criterion. Finally for the list of nouns task the SNRLB was between 4.37 and 22.14 dB (mean = 11.76 dB, median = 10.46 dB, SD = 3.66 dB, IQR = 5.39 dB). In order to compare the data between tasks we excluded from the analysis the two participants that did not met the critical value in the list of adjectives task.

3.3.5 Event-related potentials analysis

Cluster permutation analysis

Since EEG does not have good spatial resolution we could not use information on the regions tested with MEG in the original experiment. Consequently we did not have an hypothesis on possible ROIs in which to expect a composition-related activity and thus decided to use a cluster permutation analysis to evaluate all electrodes and epoch data points from -0.2 s to 1.2 s (until image presentation). We used two-tailed t-tests to contrast the difference between the composition conditions (two-word - one-word) with the difference between the LN conditions. Independently, we used the same approach to contrast differences between the composition conditions with the difference between the LA conditions. Additionally, as we decided to use two controls, conditions in the LN and in the LA were also contrasted to test for possible differences. Furthermore, we tested our data with the TFCE method in order to validate the results obtained by the classical cluster permutation analysis.

3.4 Behavioral results

Reaction times

In this experiment participants completed three separated tasks, a composition task, and two list tasks which differed in word category (nouns and adjectives). As described in the materials and methods section, a linear mixed model was fitted to the logarithm of reaction times, with task and number of words as fixed factor, and subject and words as random effects.

A main effect of task was found ($\chi^2(2) = 54.68, p < 0.001$) as well as an effect for number of words ($\chi^2(1) = 729.03, p < 0.001$). More interestingly, an interaction between task and number of words was found to be significant ($\chi^2(2) = 471.35, p < 0.001$). Multiple comparisons were tested between number of words for each task. There was no difference between the two-word condition and the one-word condition for the composition task ($Z = -0.81, p = 0.42$). Contrarily, reaction times were lower for the one-word condition for the list of nouns ($Z = -27.37, p < 0.001$) and for the list of adjectives ($Z = -22.89, p < 0.001$) (Figure 3.2).

Accuracy

Participant's accuracy was in line with the reaction time results (Figure 3.3). There was a main effect of task ($\chi^2(2) = 13.53, p = 0.001$) and number of words ($\chi^2(1) = 45.45, p < 0.001$), as well as a congruency effect ($\chi^2(1) = 7.27, p = 0.007$) which was not further explored. The analysis shows a significant interaction between task and number of words ($\chi^2(2) = 11.06, p = 0.004$). Pairwise comparisons revealed no difference between conditions for the composition task ($Z = 1.31, p = 0.58$) and a significant difference between conditions in the list of nouns ($Z = 5.76, p < 0.001$), and list of adjectives task ($Z = 4.63, p < 0.001$).

Participants answered as fast and accurate to the one-word condition (*xqrt rojo*) than to the two-word (*bote rojo)* when both words had to be unified into one concept. When subjects where presented with two-word which had to be maintained in memory as independent elements, reaction times were higher and more mistakes where elicited. These results suggest that there is a processing advantage to the composition of elements. This is in line with the behavioral results reported by Bemis and Pylkkänen (2011) and Neufeld et al. (2016), as well as in previous research (Potter and Faulconer, 1979).

3.5 Electrophysiological results

3.5.1 Event-related potentials time-locked to first word onset

Event-related potentials were obtained for each condition for each task (Figure 3.4). As described in the methods section, data was epoched to first word onset, therefore 0 s is the time at which first word is

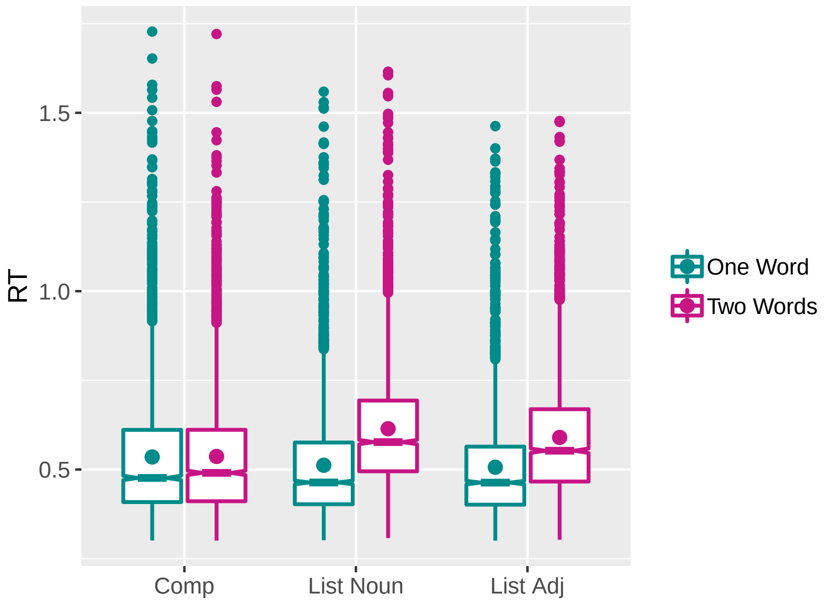


Figure 3.2: Subject's reaction times for the three tasks. Left: composition task. Middle: list of nouns task. Right: list of adjectives task. one-word conditions in green and two-word conditions in magenta

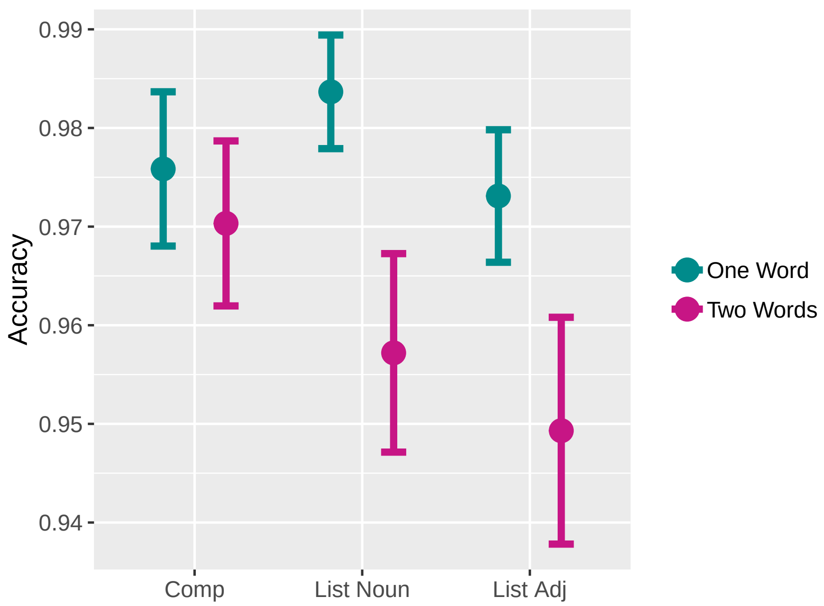


Figure 3.3: Participant's accuracy for the three tasks. Left: composition. Middle: list of nouns. Right: list of adjectives. one-word conditions in green and two-word conditions in magenta. Bars indicate 95% confidence interval.

presented and 0.6 s corresponds to the moment that subjects saw the second word. Figure 3.5 shows a more informative graphical representation of the data.

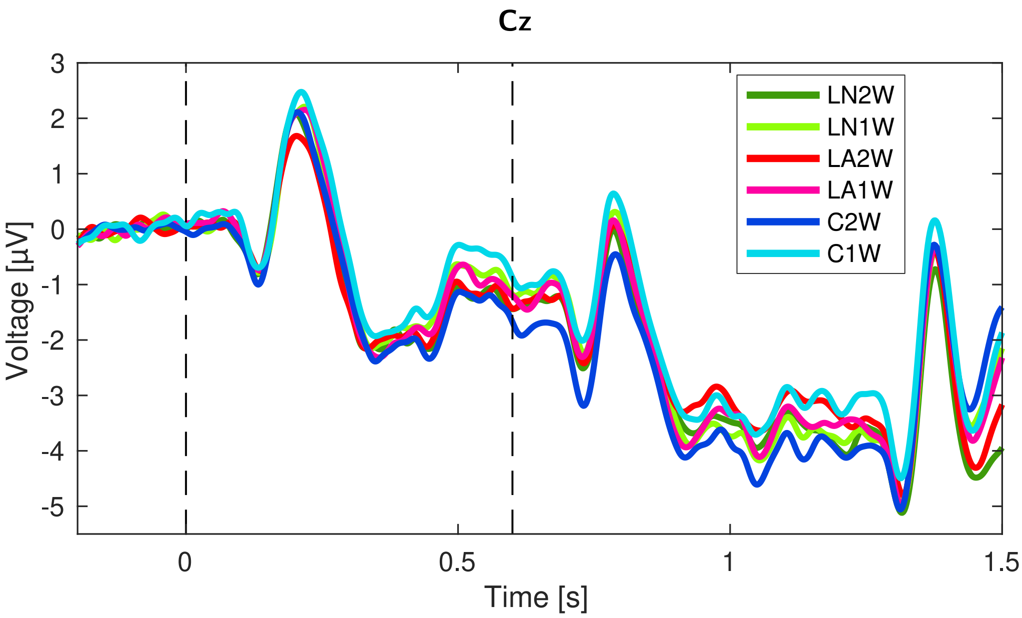


Figure 3.4: ERP grand average for each condition of each task superimposed. C1W: composition one-word condition, C2W: composition two-word condition, LN1W: list of nouns one-word condition, LN2W: list of nouns two-word condition, LA1W: list of adjectives one-word condition and LA2W: list of adjectives two-word condition. Vertical dotted lines indicate first and second word onset.

Cluster permutation analysis results

A cluster permutation analysis was carried out to evaluate an interaction between number of words (two-word - one-word) and task (composition - list). For this purpose we tested differences between the composition task conditions and the list of nouns conditions, as well as differences between the composition task conditions and the list of adjectives conditions. Finally we contrasted the list tasks to look for possible differences.

Interaction between composition and list of nouns

The permutation cluster analysis revealed a significant negative cluster for the interaction between task and number of words (p = 0.006, t = 0.86 - 1.15 s) composed of 26 electrodes with a central distribution (Figure 3.6a and Figure 3.6b).

A post-hoc cluster permutation analysis was carried out for each task on the electrodes and time points that participated in the cluster that resulted from the interaction. The post-hoc analysis showed no difference between conditions for the LN task (Figure 3.7a) and two significant negative clusters were obtained for the composition task ($p = 8.0 \times 10^{-4}$, t = 0.95 - 1.15 s and $p = 6.6 \times 10^{-3}$, t = 0.86 - 094 s) (Figure 3.7b).

Interaction between composition and list of adjectives

The permutation cluster analysis to test the interaction between the LA and composition task, and number of words showed similar results. A significant negative cluster comprised of 16 electrodes with a

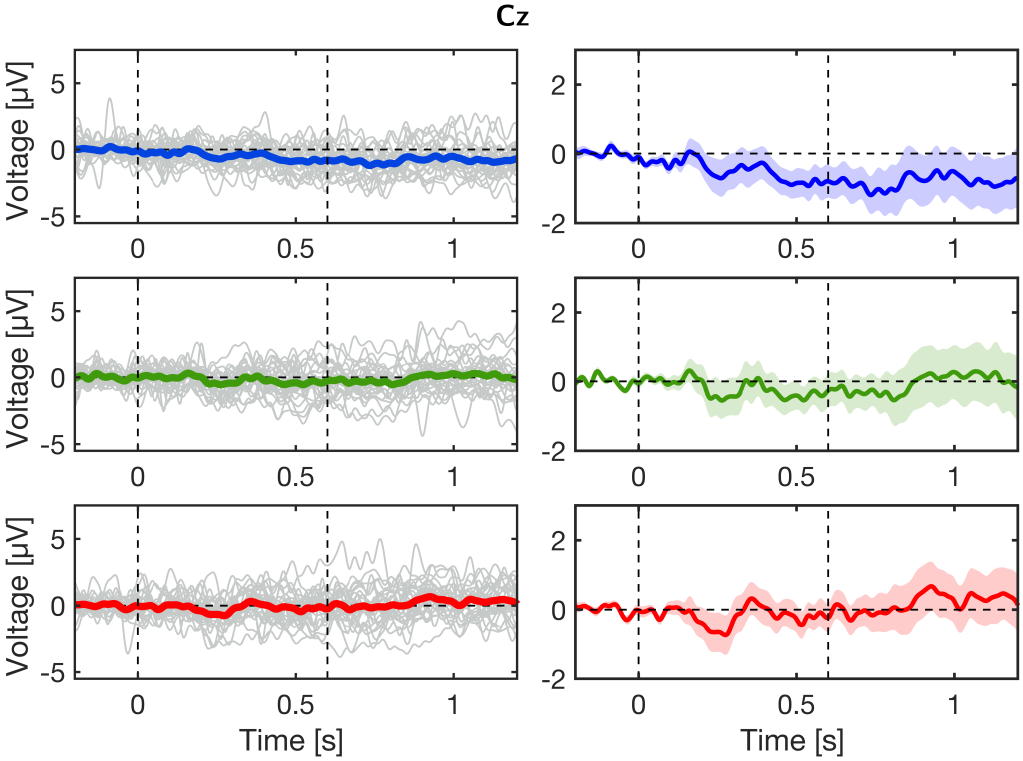


Figure 3.5: Left: Each gray stroke is the average difference between two-word condition and one-word condition for each participant for Cz electrode. Top graph corresponds to the composition task, middle graph shows the list of nouns task and bottom graph corresponds to the list of adjectives task. The thick color line indicates the mean difference. **Right**: Mean differences between conditions with 95% confidence interval. Vertical dotted lines indicate first and second word onset

central-frontal distribution was obtained ($p = 0.007$, t = 1.01 - 1.20s), (Figure 3.8a and Figure 3.8b). We carried out a post-hoc analysis for each task taking only the electrodes and data points that participate in the interaction cluster. No cluster was obtained for the list of adjectives. There was a negative significant cluster for the composition task ($p = 5.0 \times 10^{-4}$, t = 1.01 - 1.20 s) (Figure 3.9).

Interaction between list of nouns and list of adjectives

The same analysis was done between both list tasks. No significant clusters were obtained (Figure 3.10).

Threshold-free cluster enhancement results

In order to evaluate the consistency of our results, and its independence to the parameters arbitrarily selected we applied a threshold-free cluster permutation analysis as implemented in Mensen and Khatami (2013). The interaction between the composition task and the noun list revealed three significant clusters with a central distribution: (1) consisting of one channel, peak channel P2; t = 0.97 - 1.02 s, $p = 0.033$, (2) comprised of nine unique channels, peak channel Cz; t = 0.95 - 1.15 s, $p = 0.015$ and (3) consisting of 8 unique electrodes, peak channel FC1; t = 0.88 - 0.95, $p = 0.032$.

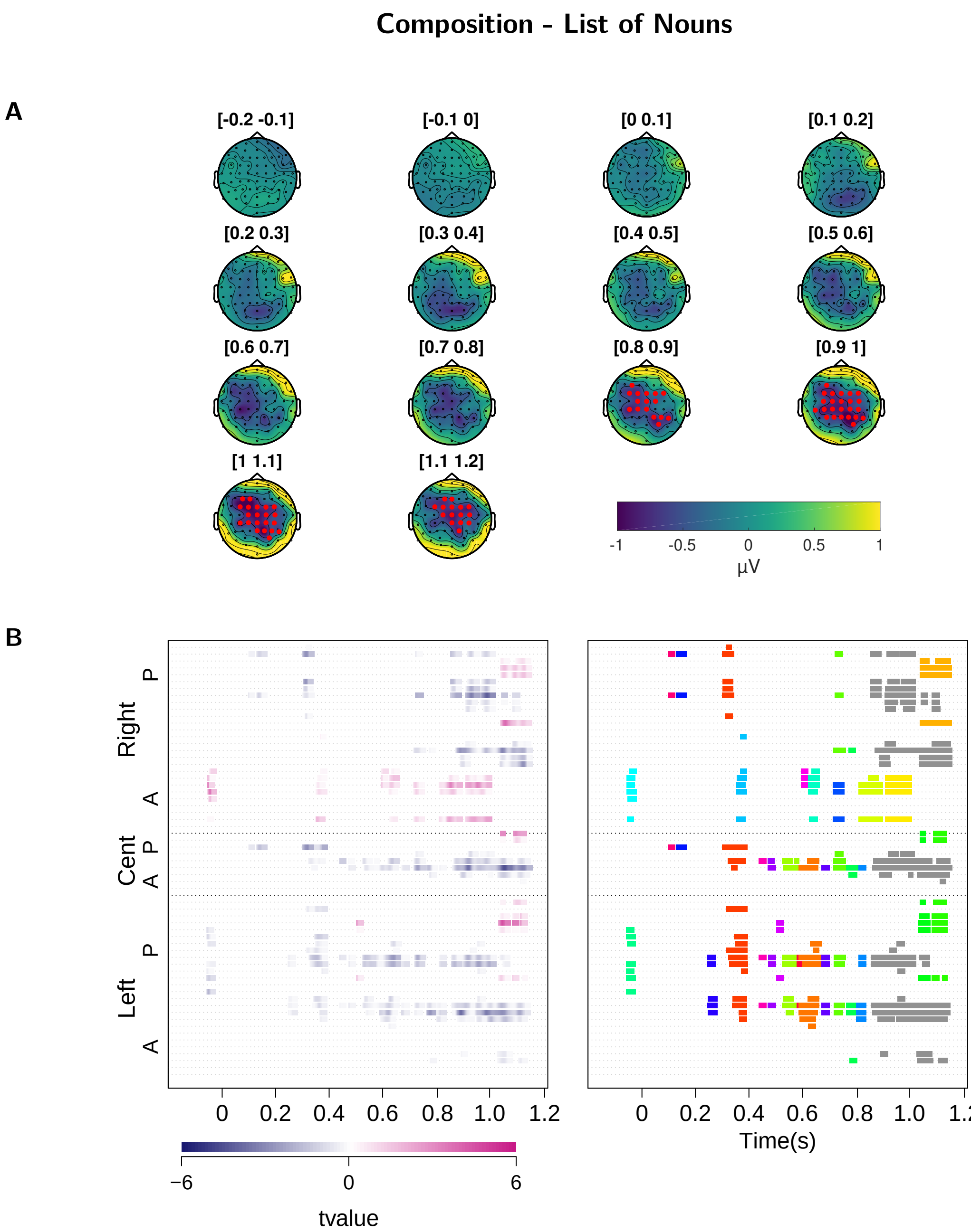


Figure 3.6: **A**. Topography of the differences between the composition task conditions and the list of nouns conditions. Red points indicate electrodes that are part of the significant cluster. **B**. **Left**: t-values of the clusters obtained for the permutation cluster analysis of the interaction between composition and list of nouns, A refers to anterior and P to posterior electrodes. **Right**: The color code represents data points that participate in a given cluster. Grey colored cluster corresponds to the only significant cluster.

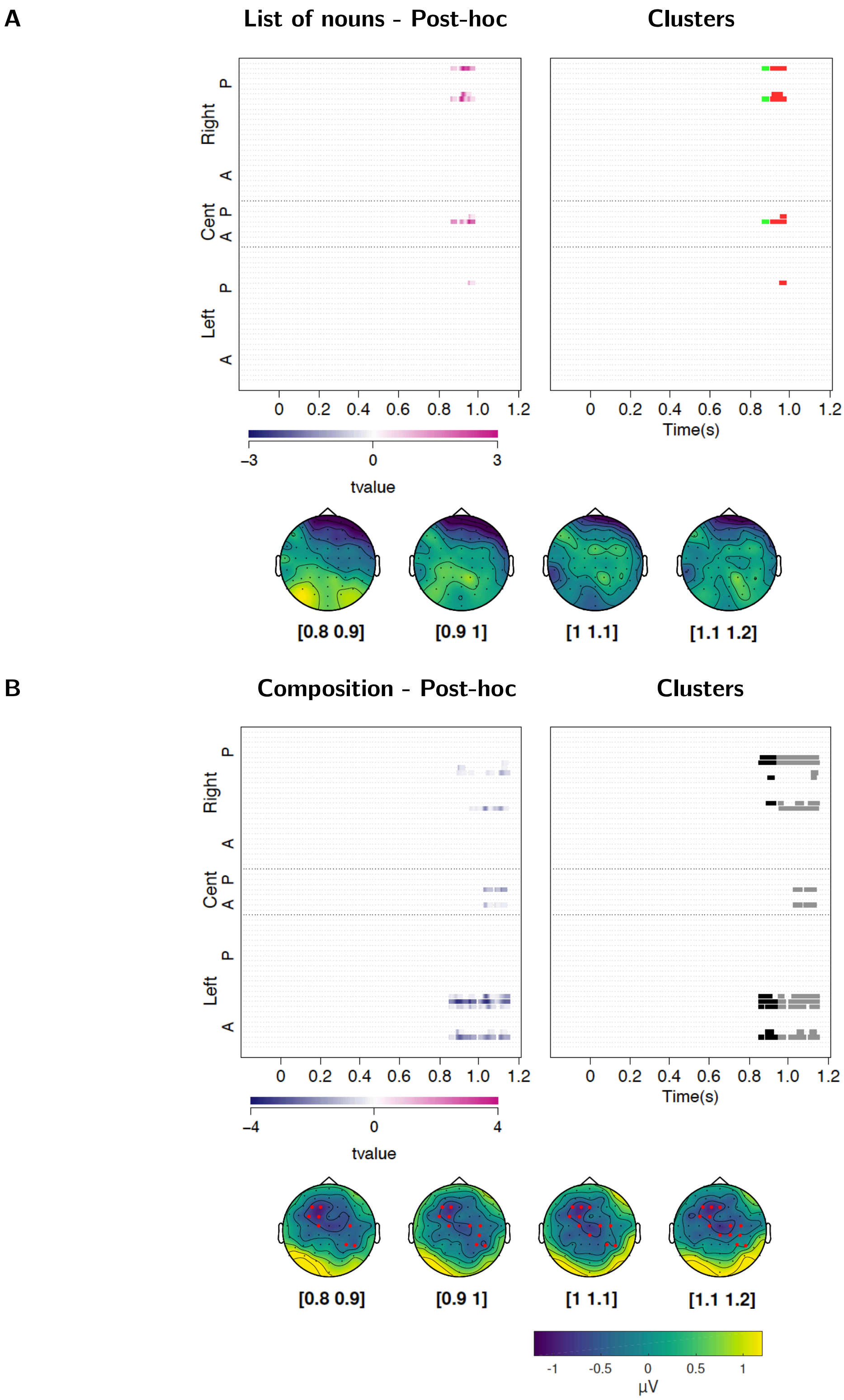


Figure 3.7: **A**: **Left**. t-values of the clusters obtained for the post-hoc permutation cluster analysis between the list of nouns conditions, A refers to anterior and P to posterior electrodes. **Right**: Color code represents data points that participate in a given cluster. There were no significant clusters. **Bottom**: Topography of the difference between the list of noun conditions. **B**: **Left**. t-values of the clusters obtained for the post-hoc permutation cluster analysis between the composition conditions. **Right**: Grey and black indicate significant clusters. **Bottom**: Topography of the difference between the composition conditions. Red points indicate electrodes that are part of the significant clusters.

Composition - List of Adjectives

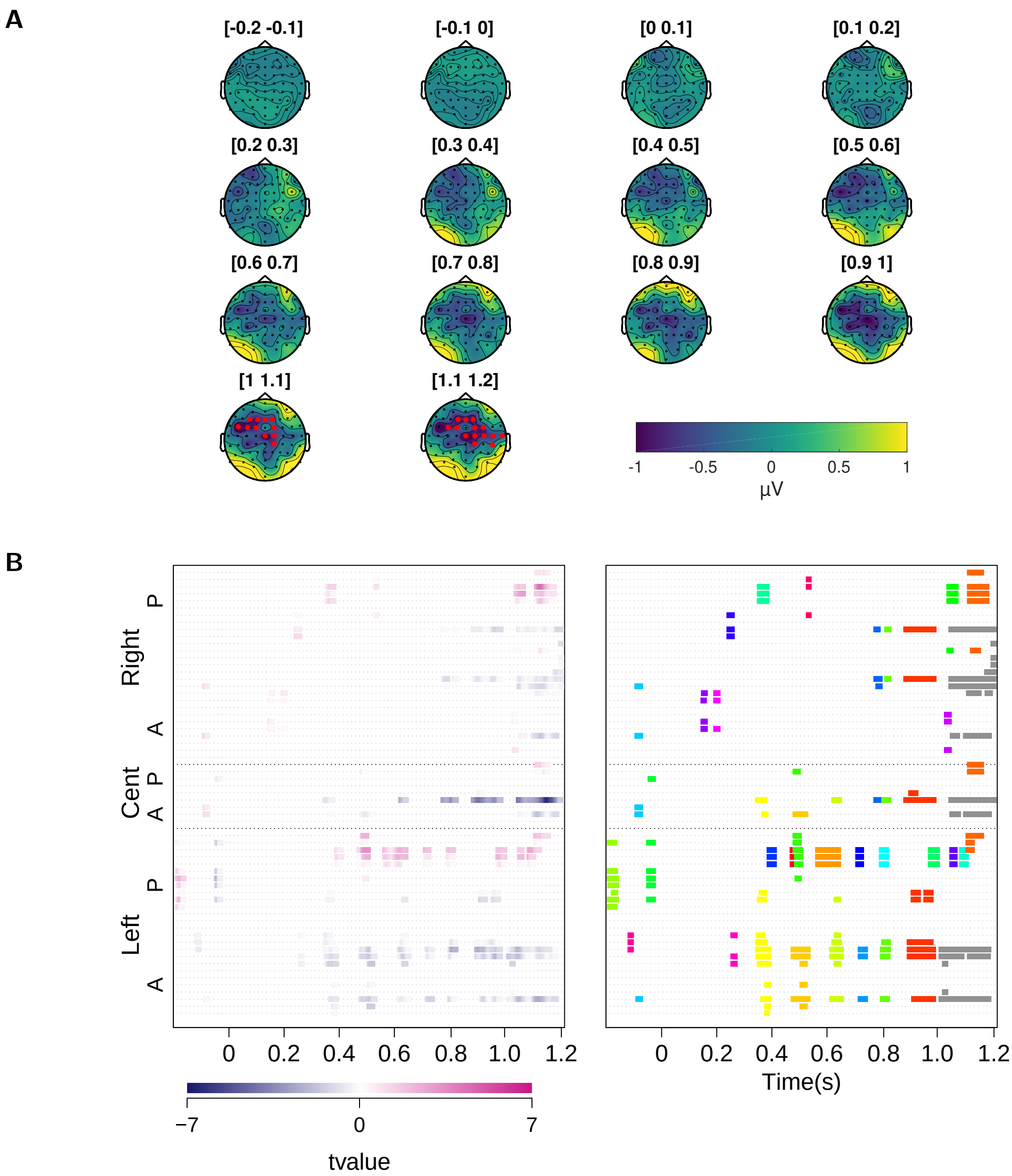


Figure 3.8: A. Topography of the differences between the composition task conditions and the list of adjectives conditions. Red points indicate electrodes that are part of the significant cluster. **B**. **Left**: tvalues of the clusters obtained for the permutation cluster analysis of the interaction between composition and list of adjectives, A refers to anterior and P to posterior electrodes. **Right**: The color code represents data points that participate in a given cluster. Grey colored cluster indicates the only significant cluster.

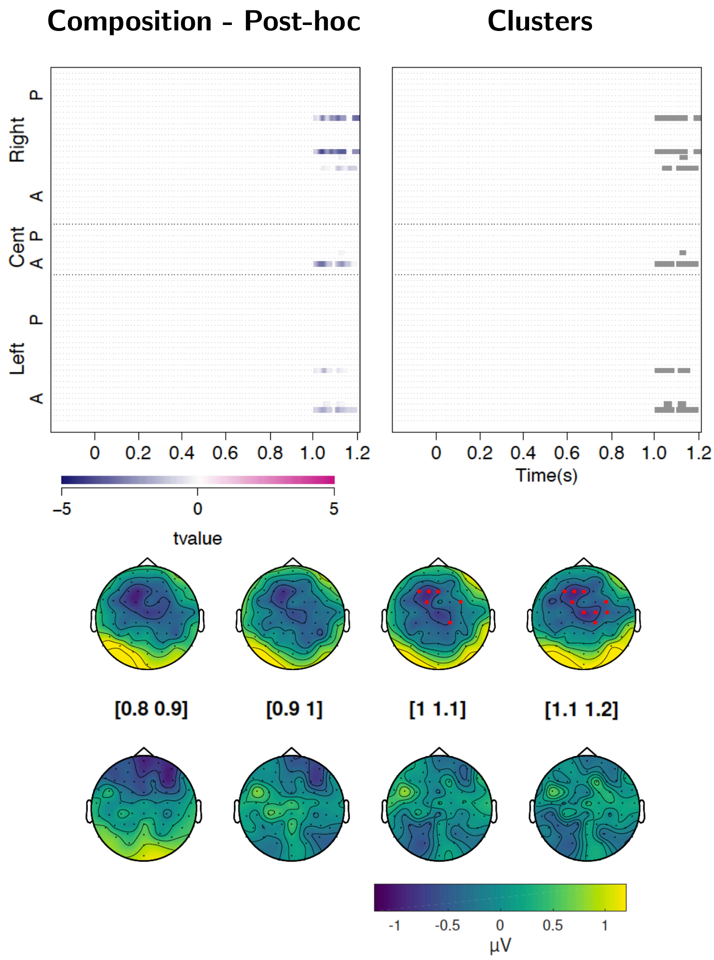


Figure 3.9: Top left: t-values of the clusters obtained for the post-hoc permutation cluster analysis between the composition conditions, A refers to anterior and P to posterior electrodes **Top right**: In grey the significant cluster. **Bottom**: Topography of the difference between the composition conditions (top) and list of adjectives conditions (bottom) for the analyzed time window. Red points indicate electrodes that are part of the significant clusters.

Likewise we applied this method to evaluate an interaction between the composition and list of adjectives task. The analysis resulted in five small clusters comprised of one electrode each: (1) peak channel PO8; t = 1.12 - 1.14 s, $p = 0.041$, (2) peak channel Cz, t = 1.09 - 1.19 s, $p = 0.0036$, (3) peak channel Cz, t = 1.04 - 1.07, $p = 0.032$, (4) peak channel Cz, t = 0.96 - 0.97s, $p = 0.049$ and (5) peak channel Cz, t = 0.89 - 0.93 s, $p = 0.044$.

These results are similar to the ones obtained by the cluster permutation analysis. There is an interaction effect of task and number of words at central electrodes in a time window around 0.86 - 1.20 s, which corresponds to a more negative potential for the two-word condition compared to the one-word condition in the composition task.

3.6 Intermediate discussion

An adaptation of Bemis & Pylkkänen's paradigm was implemented in Spanish in a group of 29 subjects. The behavioral results show a facilitation for the processing of two items that can be merged into one concept. This is shown by lower reaction times and a higher accuracy rate during the two-word

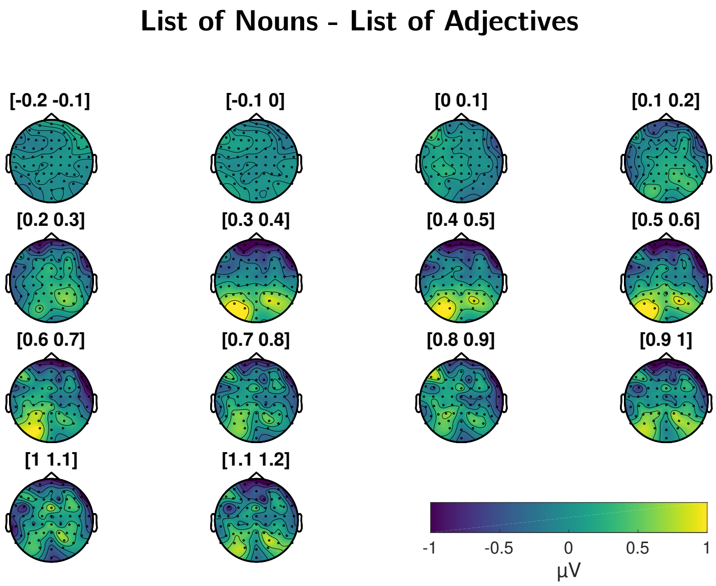


Figure 3.10: Topography of the difference of the two-word condition minus the one-word condition between the list tasks.

composition condition in comparison to the two-word condition for both list tasks.

The electrophysiological results presented in the previous section are in line with the results reported in the literature. Particularly, an interaction between task and number of words for both lists was found. For the composition and list of nouns, an interaction was obtained at 260 ms - 550 ms after second word onset. Similarly, for the list of adjectives task an interaction was found at 410 to 600 ms after second word presentation. Moreover these results were obtained without restricting the analysis to a few data points or electrodes, but using the entire data set instead. Furthermore, a threshold-free permutation cluster analysis gave similar results, thus supporting the chosen parameters. As shown by the post-hoc analyses, these effects corresponds to a more negative potential for the two-word condition than to the one-word condition in the composition task.

As noted in Neufeld et al. (2016), by visually exploring their data a difference between conditions for the composition task could be observed before second word onset. Exploring our cluster permutation analysis results (Figure 3.6b and Figure 3.8b), it is possible to notice that the same electrodes that comprise the significant clusters, seem to take part in temporally smaller clusters that do not reach significance before second word onset (this is more clear for the anterior and posterior left hemisphere electrodes). During these three tasks participants knew after seeing the first word what type of word would be presented subsequently. Nevertheless, it was only the case for the composition task that subjects had to manipulate the first word in relation to the second word. In this way, participants' processing of the first word was conditional to the second word. Thus, there was a contingency in this task between the first and second stimuli. In this sense, we hypothesize that the activity obtained for the two-word condition in the composition task could in fact be a slow wave that starts before second word onset but reaches actual significance after the presentation of the second stimuli. This possibility is addressed in the following sections.

3.6.1 Evaluating time effects on the composition-related activity

The result obtained in the previous section, which matches with the composition-related activity reported by Neufeld et al. (2016) and Bemis and Pylkkänen (2011), could also be explained by anticipatory effects. Since the two-word condition in the composition task obliged subjects to hold first word processing until second word was shown, it is possible that first word onset oriented subjects' attention to second word eliciting an anticipatory process.

As discussed in the introduction, it has been shown that subject's expectancy can result in a slow negative ERP named CNV. This complex expectancy wave is modulated by task demands, sensory processing and motor preparation. CNV is reported in the literature as a negative activity that increases as the contingency between two stimuli is learned, once the pattern is established the effect of later trials depends on task demands and subjects engagement. Therefore, we wonder if the activity shown to be significant by the cluster permutation analysis, had a time-dependent behavior over the course of the experiment.

In order to test the relationship between voltage and task progression, we implemented a generalized linear model with interaction between condition and trial number as fixed effects, and subject and electrode as random effects (eq. 3.6). We took the average voltage for all electrodes and time points that were part of the previously obtained interaction clusters and evaluated an interaction between condition and trial number in each task.

$$Voltage \sim Condition * Trial + (1|Subject) + (1|Electrode) \tag{3.6}$$

We tested the cluster resulting from the interaction between the composition task and the LN task and number of words for a trial effect (Figure 3.11). For the composition task, a main effect of condition was found ($\chi^2(1) = 16.18, p < 0.001$) and there was no effect for trial ($\chi^2(1) = 3.09, p = 0.079$). Finally, no interaction between condition and trial was obtained ($\chi^2(1) = 0.64, p = 0.42$). For the LN task, there was a main effect of condition ($\chi^2(1) = 14.06, p < 0.001$) and an effect of trial ($\chi^2(1) = 7.19, p = 0.007$). No interaction between condition and trial was found ($\chi^2(1) = 1.67, p = 0.20$).

Correspondingly, we tested the cluster resulting from the interaction between the composition and the LA, and number of words, for a trial effect (Figure 3.12). For the composition task, a main effect of condition was found ($\chi^2(1) = 4.92, p = 0.027$) and no effect for trial was obtained ($\chi^2(1) = 1.50, p = 0.22$). Furthermore, there was a significant interaction between condition and trial ($\chi^2(1) = 4.77, p = 0.029$). For the LA task, there was no effect of condition ($\chi^2(1) = 0.34, p < 0.56$) and no effect of trial ($\chi^2(1) = 0.29, p = 0.59$). Finally, no interaction between condition and trial was found ($\chi^2(1) = 2.72, p < 0.099$).

The interaction cluster between the composition task and the LA task shows a clear effect of task progression. As task develops and subjects are exposed to a greater number of trials, the negative potential elicited increases.

It has been shown that during tasks in which subjects are presented with two contingent stimuli and are asked to elicit a motor response, the readiness potential contributes to CNV waveform. Therefore,

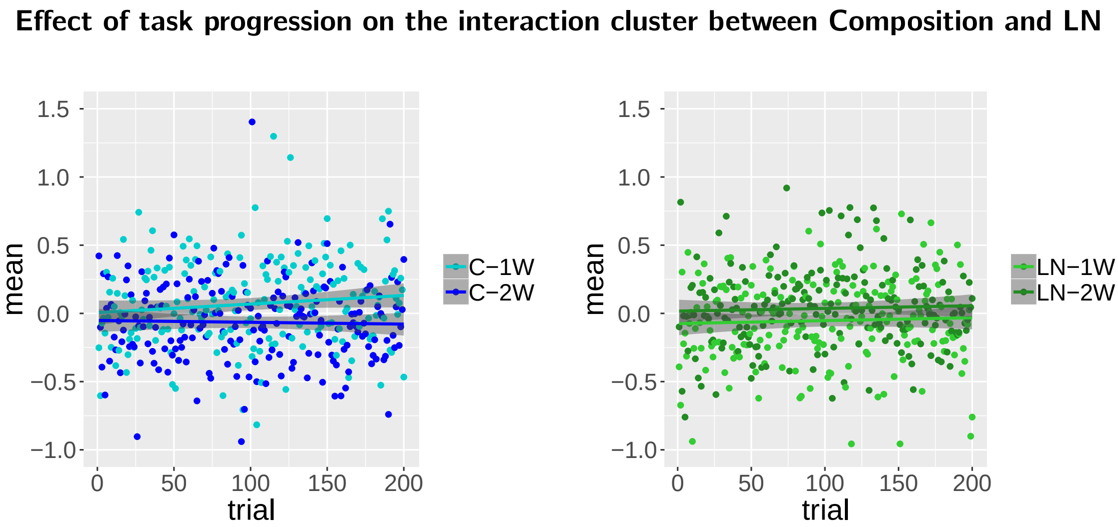


Figure 3.11: Mean voltage for the 26 electrodes that comprise the resulting cluster of the permutation cluster analysis between task (composition, list of nouns) and number of words (two-word, one-word). **Left**: Average for the composition task. In blue, mean voltage for composition two-word (C-2W), in cyan mean voltage for composition one-word (C-1W). **Right**: Average for the list of nouns task, in dark green mean voltage for list of nouns two-word (LN-2W) and in light green mean voltage for list of nouns one-word (LN-1W)

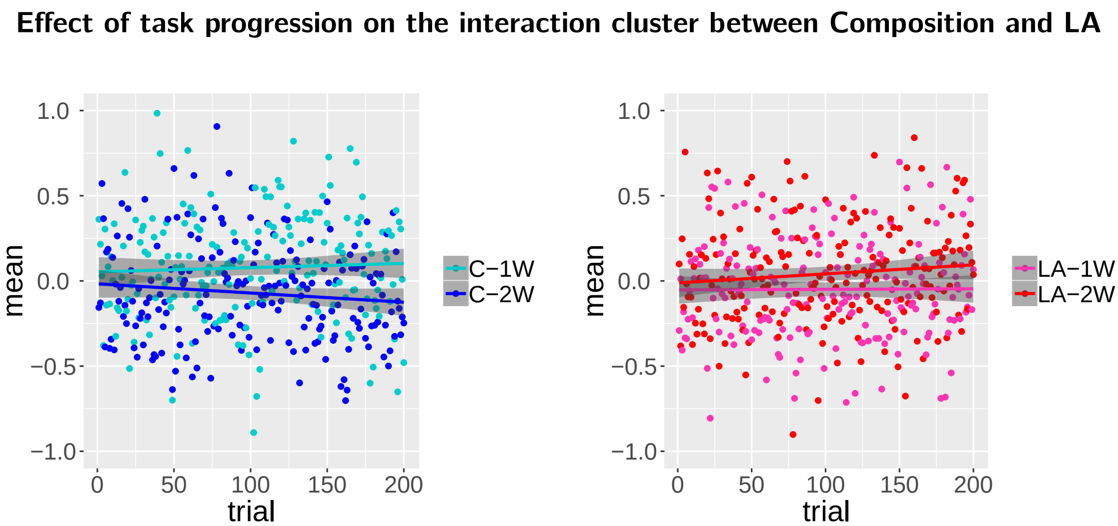


Figure 3.12: Mean voltage for the 16 electrodes that take part in the resulting cluster of the permutation cluster analysis between task (composition, list of adjectives) and number of words (two-word, one-word). **Left**: Average for the composition task. In blue mean voltage for composition two-word (C-2W) and in cyan mean voltage for composition one-word (C-1W). **Right**: Average for the list of adjectives task, in red mean voltage for list of adjectives two-word (LA-2W) and in magenta mean voltage for list of adjectives one-word (LA-1W)

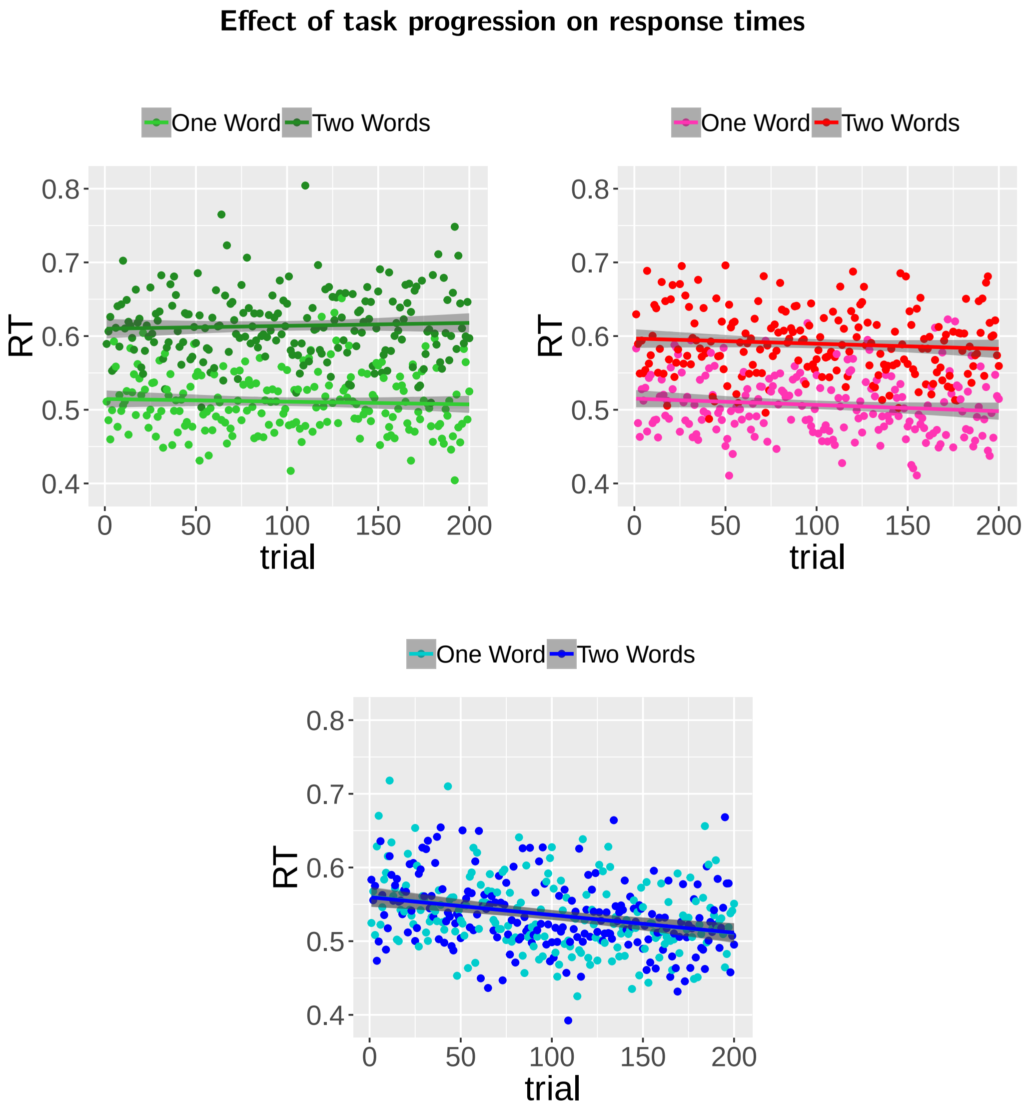


Figure 3.13: Trial effect on reaction times for the list of nouns task (green), the list of adjectives (red) and the composition task (blue)

to discard an effect of motor preparation or execution on the negative potential obtained, we evaluated the interaction effect between condition and trial number on reaction times for the composition task Figure 3.13. Our analysis showed no effect of condition ($\chi^2(1) = 0.20, p = 0.65$), but an effect of trial ($\chi^2(1) = 71.46, p < 0.001$). Finally no interaction between condition and trial was obtained ($\chi^2(1) = 0.67, p = 0.41$). These results indicate that the increase in negativity is not related to motor execution, as reaction times decrease through out the composition task but equally for the two-word and one-word conditions.

The results show that there is a change in the electrophysiological response which is compatible with non-stationary and learning effects. This points to the possibility that the results are due to the presence of a CNV. The question remains, thus, as to whether there is a specific composition activity. We disentangle these effects in the next chapter.

Chapter 4

Removing expectancy effects

The first objective of this work was to find the composition-related activity reported in the literature with EEG by means of adapting Bemis & Pylkkänen design. In the previous section we successfully reproduce Bemis and Pylkkänen (2011) and Neufeld et al. (2016) results. Nevertheless, given the evidence presented for a potential effect of expectancy that could give account for the results, in this chapter we propose a task to evaluate composition in a expectancy-free context.

4.1 Hypothesis

Composition effects have been reported in the MEG and EEG literature using an experimental paradigm in which subjects' expectancy could be acting as a confounding factor that would explain the obtained results. If the activity showed in Chapter 3 is in fact related to composition we would expect it to be also present during a task in which participants can not anticipate if a given word will have to be used to perform composition. Contrarily, if this activity is not elicited in a expectancy-free context it would further support the hypothesis that this brain activity corresponds to an anticipatory potential generated by task demands. Furthermore, without subject preparation to compose, elicited composition could display temporal variation on a trial-to-trial basis, and therefore not be time-locked to the stimulus. Under this condition, time-frequency analysis could be useful to reveal a potential compositional process.

As discussed in the introduction, neural oscillations have been related to language processing, of particular interest to this work are Segaert et al. (2018) results. Studying syntactic composition at theta, alpha, low beta and beta-bands, they found an increase in the alpha frequency band in combinatorial contexts. Although our research is interested in meaning composition, as discussed in the introduction syntactic and semantic computations are difficult to disentangle, and a two-word paradigm should elicit both processes. Here we follow Segaert et al. (2018) analysis to test for a compositional marker in the frequency domain.

4.2 Objectives

General

Evaluate if the results obtained using Bemis & Pylkkänen paradigm persist when subjects expectancy is not involved.

Specific

1. Design a task to evaluate language composition in a expectancy-free context.
2. Evaluate an oscillatory response related to composition following Segaert et al. (2018) analysis.

4.3 Materials and methods

4.3.1 Participants

Thirty-nine non-colorblind subjects participated in the experiment (23 female, average age, 23.13 ± 3.65). All subjects were native Rioplatense Spanish speakers, right handed, and had normal or corrected vision and were not aware of the previous experiment. Three of the subjects' recordings had to be discarded due to an error in stimuli timing presentation. All participants gave inform consent and were not awarded any economic or academic retribution.

4.3.2 Stimuli and procedure

Experimental design

In this experiment we tried to maintain the design similar to the one used in Chapter 3 while ensuring that participant expectancy would not intervene. During this task participants were presented with four different conditions: *noun adjective: NA*, *non-word adjective: XA*, *noun noun: NN* and *non-word noun: XN*. In all conditions, each word pair was followed by an image. Subjects were instructed to answer if the image matched the preceding verbal material and were encouraged to answer as quickly and accurately as possible. Hence, there were two-word conditions (NA and NN) and one-word conditions (XA and XN), and composition was required only for the NA condition. A representation of the trials is shown in Figure 4.1. The main differences with the previous experiment were that subjects saw all types of trials in the same block, and the image presented in NN trials had two elements. The task consisted of 400 trials that were preceded by a 40 trial practice. In each trial, participants saw a fixation cross for 1 s and all stimuli except for the images were presented for 300 ms followed by a 300 ms blank screen. The image presented at the end of each trial remained on screen until subjects pressed a key or after 3 s. Following the end of each trial a sound was elicited to encourage participants to blink. The inter-trial interval was randomly varied between 0.8 - 1.5 s.

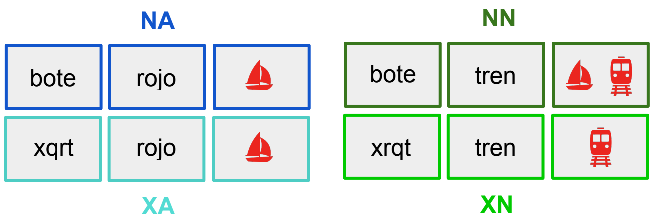


Figure 4.1: Participants were presented with two-word conditions: in NA trials participants were shown a noun followed by an adjective and in NN trials they were presented with two nouns. In the one-word conditions subjects were presented with a non-word followed by an adjective (XA), or a non-word followed by a noun (XN). Finally an image congruent or incongruent to the words was presented.

Stimuli construction

The same pool of words and non-words used in the first experiment was employed to create the stimuli. A python script was coded to combine words and non-words for each subject in order to create 400 trials. For the XN, XA and the NA conditions, a combination of the 11 elements of each pool was made, resulting in 121 pairings with the desired structure. Subsequently, 21 items were randomly discarded. For the NN condition, permutations of the nouns were generated and 10 items were discarded, resulting in 100 pairings. Every subject was presented with 100 trials of each condition, and half of the trials of every condition were incongruent to the image. For the XN and XA conditions, a trial was incongruent if the image did not match the noun or adjective respectively (50 trials). For the NA condition a trial was incongruent if the image did not match the noun (25 trials) or did not match the adjective (25 trials). Finally, for the NN condition a trial was incongruent if the image did not match the first noun (25 trials) or the second noun (25 trials).

4.3.3 Behavioral data analysis

Reaction times

Response times were measured from image onset until subjects pressed no/yes key. In trying to keep the analysis as close to the one implemented for the previous experiment as possible, we grouped the NA and XA into a group named composition and we assembled the NN and XN conditions in another group called List of Nouns, generating a factor named *group*. Following the analysis in Chapter 3 we performed a linear mixed-effects analysis of the relation between reaction times and the interaction between group and number of words.

For each participant, reaction times were analyzed by removing all incorrect and missing responses. Trials in which subjects response time was over or under two standard deviations were discarded (4.72 % of all trials). A model with group as a two-level factor (*composition, list of nouns*) and number of words as a two-level factor (*two-word, one-word*) was fitted. Random intercepts were also fitted for each subject (eq. 4.1). The residual plots showed some deviation from normality therefore a logarithmic transformation of the reaction times was applied.

$$log(RT) \sim Group \times NumberWords + (1|Subject) \tag{4.1}$$

The model was tested with an ANOVA-like Chi-squared test and pairwise comparisons were performed using emmeans R package (Lenth, 2018). Bonferroni method was used to correct for multiple comparisons.

Accuracy

In order to test the relationship between accuracy and the interaction between group and number of words, we perfomed a generalized linear model (binomial) with group as a two-level factor (*composition*, *list of nouns*), number of words as a two-level factor (*two-word*, *one-word*) and image congruency as a two-level factor (*congruent - incongruent*). We incorporated subject as random effect to the model (eq. 4.2).

$$Accuracy \sim Group \times NumberWords + Congruency + (1|Subject) \tag{4.2}$$

The generalized linear model for accuracy was tested with a ANOVA-like Chi-squared test. Pairwise comparisons were performed and Bonferroni method was used to correct for multiple comparisons.

4.3.4 EEG recording and preprocessing

Recording and preprocessing steps implemented in the first experiment were followed for this task to obtain ERPs time-locked to first word onset. See EEG recording and preprocessing in Chapter 3 for a description. The number of rejected electrodes and trials was 4.44 ± 2.21 and 55.75 ± 15.28, respectively. A one-way ANOVA was conducted to test for differences in the number of trials across conditions, no difference was found $F(3, 124) = 0.74, p = 0.53$. The SNRLB ranged from 1.74 to 20.31 dB (mean = 10.99, median = 11.52 dB, SD = 4.78 dB, IQR = 5.78 dB). Subjects 3, 14, 15 and 29 failed to meet the SNRLB criterion (3 dB) and were excluded from the analysis.

4.3.5 Time-frequency analysis

In order to avoid edge artifacts, data was epoched 1 s before and 2 s after first word onset, resulting in a 3 s epoch. As epochs were longer and time intervals in which subjects were allowed to blink were included, we performed an independent component analysis (ICA) to remove eye blink artifacts (Jung et al., 2000). ICA was applied using Fieldtrip, and spatial topography of the components was visually inspected to identify and remove eye movement artifacts. The interval from -0.5 to -0.2 s was chosen to perform a decibel normalization baseline correction (dB) (Cohen, 2014). Following Segaert et al. (2018), we performed a time-frequency analysis using sliding Hanning tapers implemented in Fieldtrip. Time-frequency representations (TFRs) were calculated for each trial using Hanning tapers with a window length of three cycles for each frequency ($\Delta T = 3/f$) for the frequency interval 3:1:30 Hz.

4.3.6 Cluster permutation analysis

Cluster permutation analysis on event-related potentials

A cluster permutation analysis was carried from -0.2 s to 1.2 s (until image presentation). We used two-tailed t-tests to contrast the difference between the NA and the XA condition with the difference between NN and XN conditions. In this way we kept our analysis as similar as possible to the one carried out for the previous experiment. Moreover, same parameters and number of permutations were used.

Cluster permutation analysis on time-frequency decomposition

For each subject the TFR for each condition was averaged and pairs of conditions were substracted just as was done for the ERPs. Subsequently, we applied a permutation cluster analysis to compare the differences in TFR between $NA - XA$ and $NN - XN$. We carried out the analyses within the following frequency bands: theta (4 – 7 Hz), alpha (8 – 12 Hz), low-beta (13–20 Hz) and high-beta (20 – 30 Hz) for the -0.2 to 1.2 s time window. For each analyses we averaged across each frequency band, resulting in four analysis. In each, a two-tailed t-test was done on every sample, t-values were clustered depending on if they exceeded a dependent samples t-test threshold of $p < 0.05$ (two-tailed). Data points' t-values of every cluster were summed in order to obtain a value per cluster. The maximum negative and positive cluster statistic value were kept. This was done for 5000 permutations of the data resulting in a null hypothesis distribution, against which we tested the real data. We considered the critical alpha level here to be 0.025. For all the analysis using this method we set to two the minimum number of electrode neighbors that had to be significant for a given time point in order to be part of a cluster. All cluster permutation analysis were conducted on MATLAB using Fieldtrip toolbox (Oostenveld et al., 2011) and neighbors were defined following Biosemi 64 neighbors template.

4.4 Behavioral results

In this experiment participants completed a single task in which they were presented with four conditions. For the analysis the conditions were grouped into *composition* and *list of nouns*, to perform a similar analysis than the one described in Chapter 3.

Reaction times

For the reaction times, a main effect of group was found ($\chi^2(1) = 2346.53, p < 0.001$) as well as an effect of number of words ($\chi^2(1) = 1972.65, p < 0.001$). Moreover, an interaction between group and number of words yielded to be significant ($\chi^2(1) = 1265.46, p < 0.001$). Pairwise comparisons were tested between number of words for each group. The two-word and one-word conditions in the list of nouns were different ($Z = -55.51, p < 0.001$). Furthermore, there was a significant difference between conditions for the composition group as well ($Z = -6.38, p < 0.001$) (Figure 4.2).

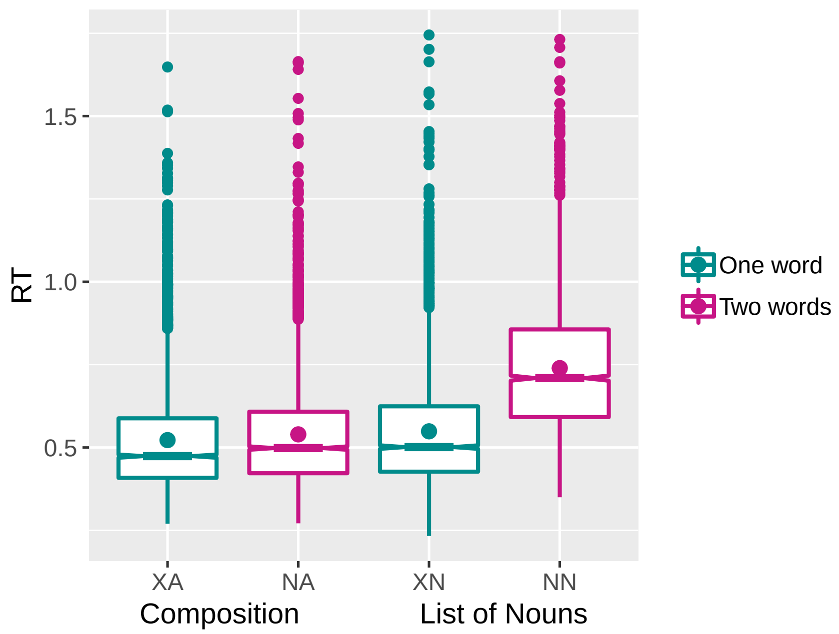


Figure 4.2: Participants' reaction times for the four conditions. For the analysis they were grouped into two groups: composition and list of nouns. Green color indicates one-word conditions, magenta color corresponds to two-word conditions.

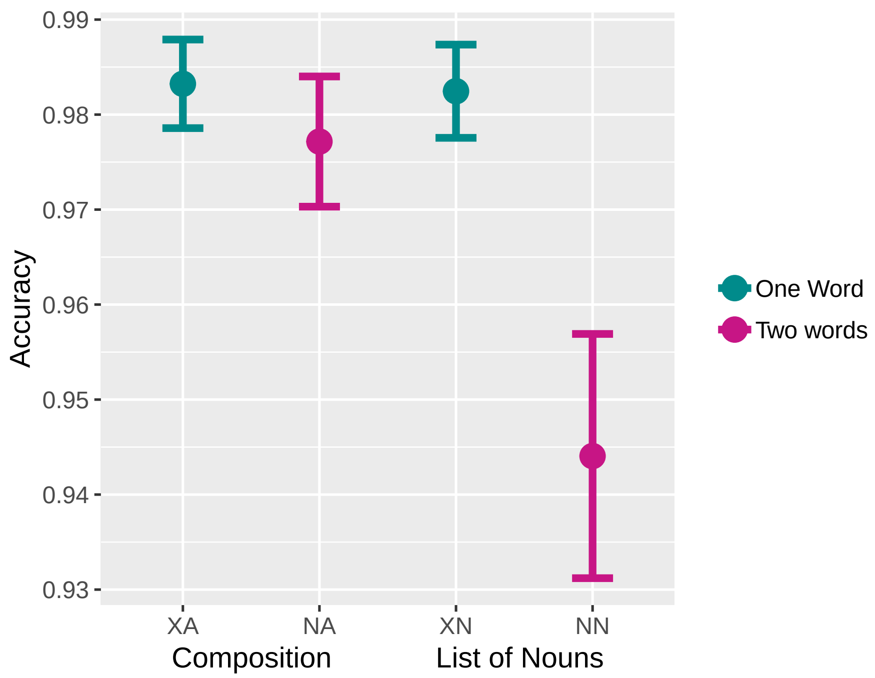


Figure 4.3: Subjects' accuracy for the four conditions. Green represents one-word conditions, and magenta indicates two-word conditions. Bars indicate 95% confidence interval.

Accuracy

In regard to accuracy, (Figure 4.3), there was a main effect of group ($\chi^2(1) = 19.17$, $p < 0.001$) and number of words ($\chi^2(1) = 45.90$, $p < 0.001$), as well as a congruency effect ($\chi^2(1) = 42.04$, $p < 0.001$). The analysis showed a significant interaction between group and number of words ($\chi^2(1) = 15.53$, $p < 0.001$).

Pairwise comparisons revealed no difference between conditions for the composition group ($Z = 1.86$, $p = 0.12$) and a significant difference between conditions in the list of nouns ($Z = 8.25$, $p < 0.001$).

Although participants' reaction times were faster in both groups for the one-word condition than for the two-word condition, there was a bigger difference between XN and NN conditions. Moreover, there was no difference in accuracy between XA and NA conditions, but a significant difference between XN and NN conditions. These results are in line with the behavioral results shown in Chapter 3, where composition probably posed a processing advantage. Nevertheless, we cannot be certain that the same process is accountable for the results in this experiment as NN condition involved answering to a two-element picture. Therefore, checking both items would probably be more demanding for subjects, resulting in delayed responses and a higher error number.

4.5 Electrophysiological results

Event-related potentials time-locked to first word onset (Figure 4.4, Figure 4.5) and time-frequency representations of power were obtained (Figure 4.6) and cluster permutation analyses were employed to test for differences between conditions.

Cluster permutation analysis on event-related potentials

A cluster permutation analyses was carried out to evaluate an interaction between number of words (two-word - one-word) and group (composition - list of nouns). The cluster permutation analysis yielded 9 positive and 8 negative clusters (Figure 4.7b); however, none of them reached significance ($p > 0.64$) (Figure 4.7a)

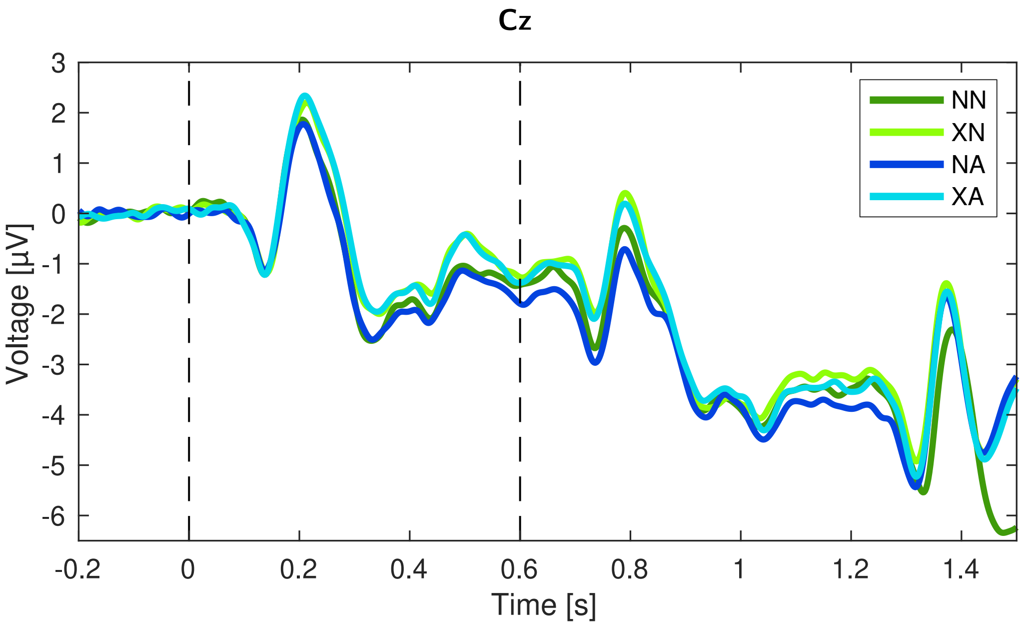


Figure 4.4: ERP grand average for each condition superimposed for Cz electrode. NN: noun - noun condition, XN: non-word - noun condition, NA: noun - adjective condition, XA: non-word - adjective condition. Dotted lines indicate first and second word onset.

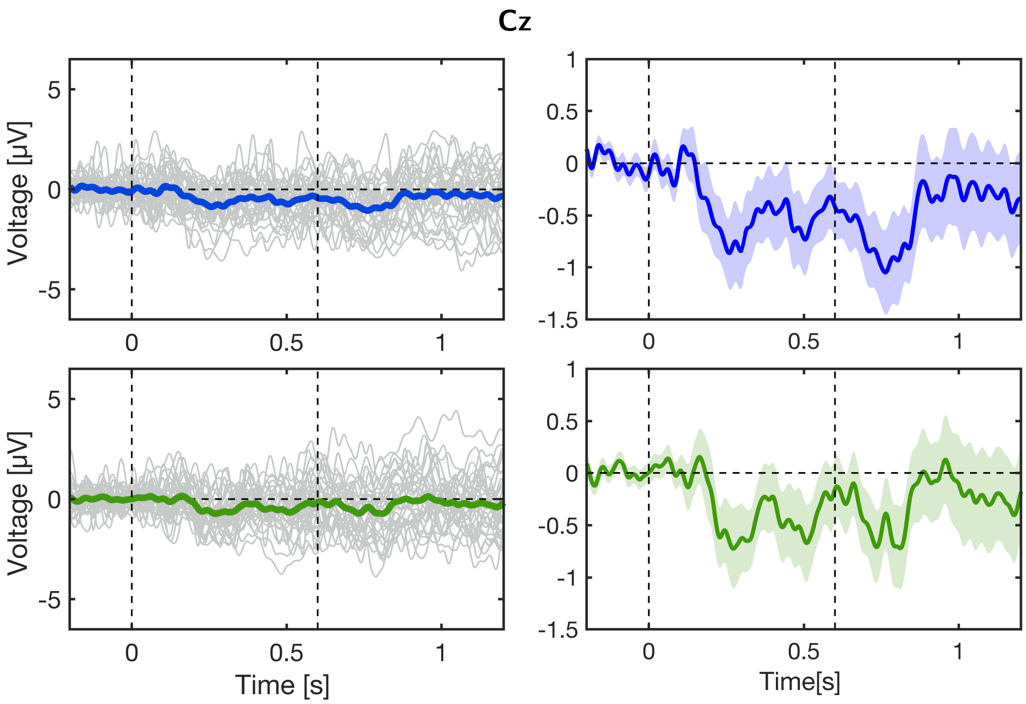


Figure 4.5: Left: Each gray stroke is the average difference between two-word condition and one-word condition for each participant for electrode Cz. Top graph corresponds to NA-XA conditions and bottom graph corresponds to NN-XN. The thick color stroke indicates the mean difference. **Right**: Mean differences between conditions with confidence interval. Vertical dotted lines indicate first and second word onset.

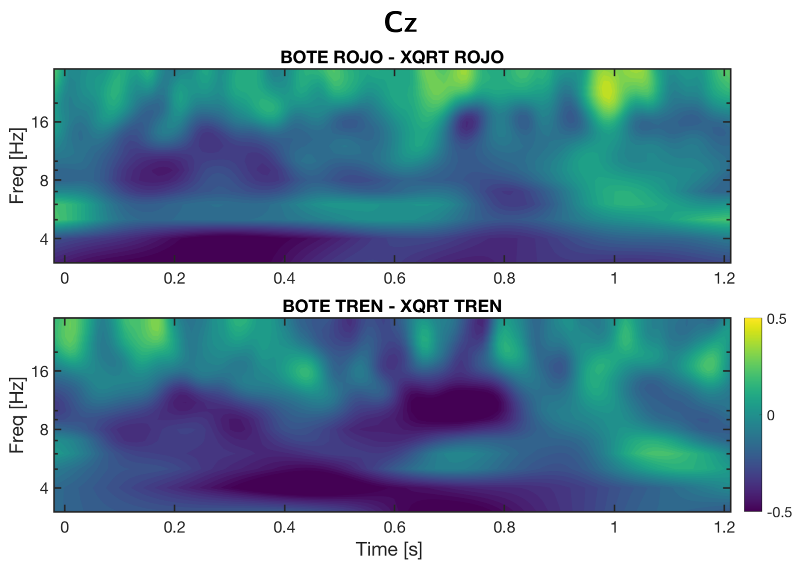


Figure 4.6: Time-frequency spectrum for Cz. **Top**: difference between NA and XA power grand average. **Bottom**: difference between NN and XN power spectrum. The colorbar indicates dB change from baseline

Cluster permutation analysis on time-frequency decomposition

Cluster permutation analyses was carried out on time-frequency power representations obtained by Hanning tapers. For the theta frequency band (4 - 7 Hz) two negative cluster were obtained ($p > 0.18$). The analysis across alpha-band (8 - 12 Hz) yielded five positive clusters ($p > 0.12$) and two negative

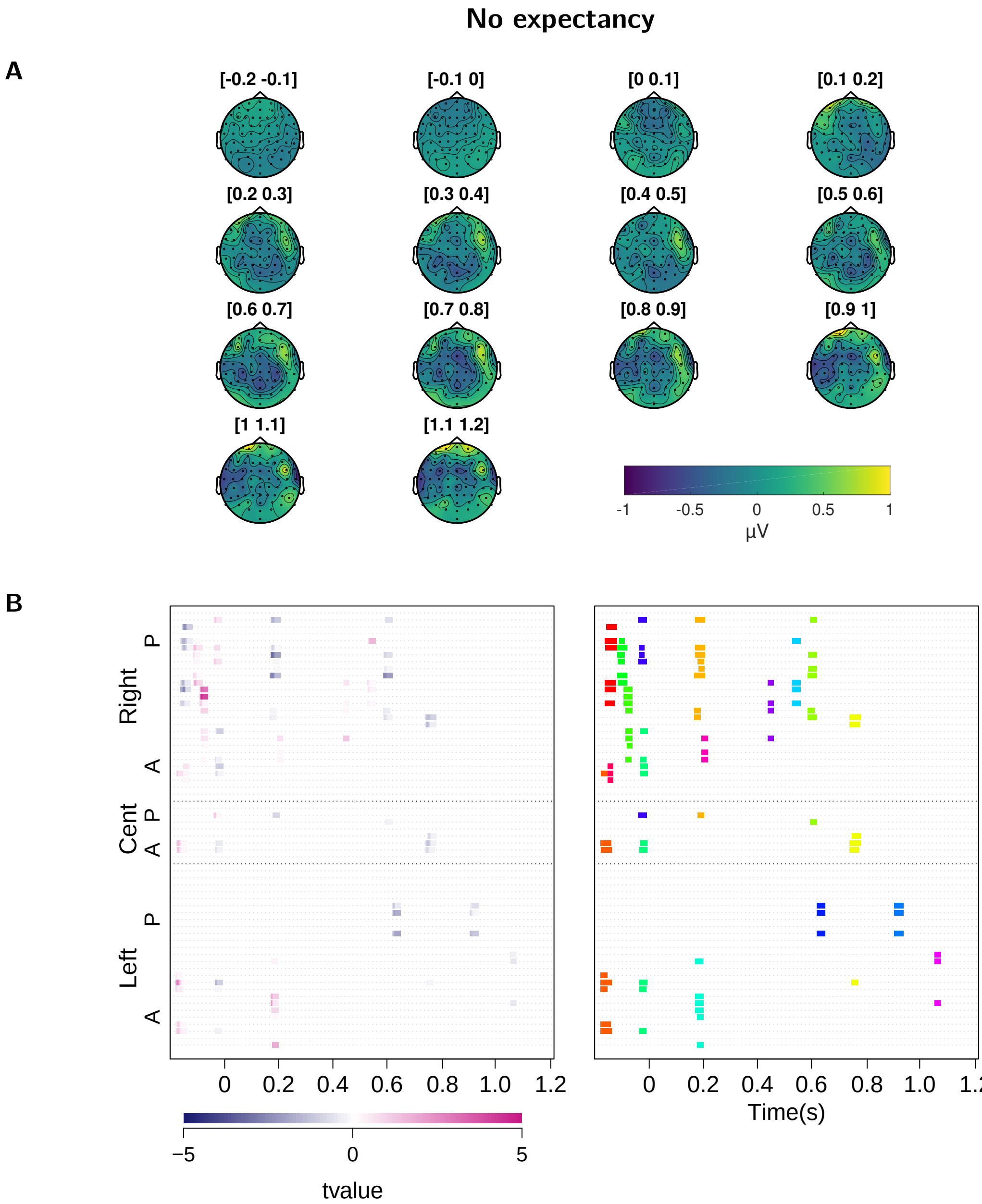


Figure 4.7: **A**. Topography of the difference of the two-word condition minus the one-word condition between the two groups **B**. **Left**: t-values of the clusters obtained for the permutation cluster analysis of the interaction between NA-XA and NN-XN, A refers to anterior and P to posterior electrodes. **Right**: The color code represents data points that participate in a given cluster. There were no significant clusters.

clusters ($p > 0.57$). For the beta frequency range (13 - 20 Hz) three positive clusters ($p > 0.26$) and two negative clusters ($p > 0.62$) were obtained. Finally, the analysis carried for the high beta-band (20 - 30 Hz) resulted in six positive clusters ($p > 0.42$) and six negative clusters ($p > 0.075$). No significant clusters resulted from the analyses conducted over either frequency band.

This experimental design allowed us to evaluate two-word composition in a free-expectancy context. Although behavioral results indicate a facilitation in the processing of two words that had to be unified into a single concept, this was not accompanied by an electrophysiological correlate. We were not able to identify a composition-related activity in the time nor frequency domain when the expectancy of participants was controlled for. This negative results further supports the hypothesis that anticipatory effects underlie or at least contribute to the activity obtained using Bemis & Pylkkänen's experimental design.

Chapter 5

General discussion

In this work we were interested in studying the fundamentals of language processing by finding the neural correlates of basic composition with EEG. Given the complexity of the experimental tasks usually implemented in language research, the minimal paradigm proposed by Bemis & Pylkkänen came as an interesting addition to the field. By reducing stimuli to pairs of words, variables such as memory demands and high-level sentential processes would not intervene and hinder results' interpretations. Moreover, the basic syntactic-semantic computation theoretically elicited at every step in sentence processing could be addressed. Employing *noun - adjectives*, *non-word - noun* and *noun - noun* pairs, these authors report an increased response in the LATL at around 200 ms after nouns in combinatorial contexts are presented, over conditions in which composition was not possible. In this and subsequent publications, they argue that this activity corresponds to a MEG marker of basic language composition. To our knowledge the only EEG study using this paradigm consists of a recent replication conducted by Neufeld and collaborators (Neufeld et al., 2016). Although the author's interpretation of the results, given the statistical analysis is questionable, the preprocessing procedure used gave us insight into a potential confounding in the experimental design. Specifically, they describe a difference between conditions before second word presentation, when a composition process could not have been elicited. This is not discussed in the original experiment, as MEG responses to the first word are not shown by the authors. Neufeld and collaborators interpret this precombinatorial activity as a syntactic building process, arguing that the adjective-noun syntactic structure is built before second word onset to allocate the expected noun.

In the present study we set out to use EEG to find the composition-related activity described in the MEG literature by applying a robust methodological analysis and a more objective data driven approach than the one carried out in Neufeld's replication. To this aim, we adapted the basic composition paradigm to Spanish by inverting pairs of stimuli and introducing a second control task. Reaction times and accuracy were measured and a cluster permutation analysis was conducted over EEG data considering all time points from first word onset to the moment subjects were asked to elicit a response.

5.1 An alternative explanation to composition

The adaptation of the paradigm allowed us to successfully reproduce the behavioral results previously reported. The binding of two items into a single concept yielded a processing advantage, as shown by faster reaction times and a lower error rate for the composition two-word condition compared to conditions in which composition was not possible. These results were accompanied by an electrophysiological response in line with previous work (Bemis and Pylkkänen, 2011; Neufeld et al., 2016). A cluster permutation analysis comparing brain responses to the composition task and the list of nouns task showed an interaction between task and number of words 260 - 550 ms after second word onset. This effect was driven by a difference between the composition task's two-word and one-word conditions. A similar result in a later time window (410 - 600 ms) was found when comparing the composition and the list of adjectives tasks. Although consisting of different word classes (nouns and adjectives), the clusters obtained in both comparisons show a similar temporal and topographical distribution. Furthermore, no difference was found between the list tasks, suggesting that both tasks were equally appropriate to control for word number. Moreover, these results are supported by a threshold-free cluster analysis in which the number of neighbors and the initial threshold for clustering is not arbitrarily defined. This analysis yielded significant clusters comprised of mostly central electrodes at 280 - 550 ms for the comparison between composition and list of nouns, and at 290 - 590 ms for the interaction between composition and list of adjectives.

The cluster permutation analysis points to a similar observation as the one suggested by Neufeld et al. (2016). It is noticeable that the same electrodes that take part in the interaction clusters obtained from both comparisons, take part in smaller clusters before composition could be elicited. Although these clusters do not reach significance, we suggest that they are indicatory of a process initiated before second word onset. It is possible that the activity measured in the two-word condition in the composition task could in fact be a slow wave that starts after first word onset but reaches actual significance after the presentation of the second stimuli. We argue that the block design in which conditions are presented could be responsible for the introduction of expectancy in the composition task. Specifically, an anticipatory activity could be induced as subject's were required to process the first word in relation to the second word only for the composition two-word condition.

Given the experimental design in which a first stimulus announces a second stimulus and a relation between them has to be established, CNV is a good candidate to be responsible for the obtained results. It is important to address that contrary to experiments in which a typical CNV is observed, Bemis & Pylkkänen paradigm consists of three stimulus and therefore two expectancy processes would be elicited: one between first and second word, and another between second word onset and image presentation. For all tasks and conditions, after presentation of the second word, an anticipatory activity was probably elicited as subjects had to maintain the verbal material available in memory and prepare to give a motor response. We argue that this process would be equal for all conditions, whereas the only condition that would elicit a prior anticipatory process following first word onset, would be the two-word condition in the composition task. We suggest that this initial slow anticipatory potential, would superpose with the image related expectancy process. The addition of these two anticipatory waves would render the initial process detectable by our cluster permutation analysis.

Multiple processes have been postulated to underlie the widely studied CNV. Relevant to the present work is the fact that this component's amplitude is affected by attentional demands and task progression. It has been consistently shown that CNV increases as the contingency between two stimuli is learned, whereas the effect on later trials seems to be task dependent. One theory proposes that CNV's amplitude reflects participant's engagement with the task, initially increasing in all tasks as the pairing is learned, and later only tasks which actively engage subject's attention would continue eliciting this component. Additionally, It has been postulated that CNV directly reflects participant's attention and/or motivation to the task at hand. In this line, we have shown an effect of task progression on the composition-related activity, such that amplitude increased on a trial-to-trial basis. This electrophysiological response occurred without an effect on behavior, as reaction times equally decreased on a trial-to-trial bases for the one-word and two-word conditions. This negative potential time dependency could have different explanations. Following CNV's literature, one possibility is that the contingency between words was still being learned by participants. Although each participant was shown 100 trials in which a combinatorial process was required, and this amount of trials should be enough to grasp the pairing, they were randomly presented with trials in which no composition was needed. It is possible that this form of presentation delayed the learning process. An attentional interpretation could also be appropriate, such that more attention was exerted as task unfolded. It is possible that to maintain the same level of performance between the one-word and the two-word condition, more attentional resources had to be engaged during the composition condition as task progressed and participant's fatigue increased, resulting in a higher amplitude potential.

In order to determine if our results are compatible with a CNV interpretation, the above discussed possibilities should be further explored. For instance, it would be relevant to implement the same task with a bigger number of trials. If a learning effect is accountable for our result, the increase in amplitude of this negative potential as task unfolds should reach a plateau once the contingency is established. Furthermore, another interesting possibility would be to increase the ISI between first word, second word and image presentation. This would allow the negative potentials to develop in time, enabling a comparison with the two component response described for the CNV. In our experiment, the short time that separates the three stimuli, probably produces a superposition of anticipatory waves, preventing a proper characterization of these potentials' components. Moreover, an effect of first word properties would further support our suggestion. For example, designing a task in which the first stimulus would be informative of the probability of seeing a subsequent composable word would help clarify what mechanisms underlie this activity.

Although an effect of task progression was only shown for the interaction cluster between composition and list of adjectives and not for the other list task, this could be a consequence of the number of electrodes averaged in each analysis. Because not all electrodes that comprise each interaction cluster show a significant difference across the entire cluster time window, it is possible that averaging all electrodes without considering their temporal contribution to the cluster introduces noise. Furthermore, this variability would probably increase with the number of electrodes. Therefore, our result could be due to the fact that the interaction cluster between the composition and list of nouns task is comprised of 26 electrodes, whereas the one obtained for the interaction betweeen composition and list of adjectives consists of 16 electrodes. In order to further explore this result, and support our suggestion, future analyses selecting fewer but representative electrodes of the interaction clusters should be conducted.

Additional support for an anticipatory interpretation of the results comes from our original experiment described in Chapter 4. In this task, participant's expectancy would be equally affecting both NA and NN conditions, as on any given trial a noun gave no indication of whether the subsequent word would be a composable or a not composable item. Behavioral data indicates that subjects were unifying both elements when a noun was followed by an adjective. However, a cluster permutation analysis on EEG data yielded no differences between conditions, suggesting that the process detected by adapting Bemis & Pylkkänen's paradigm does not reflect basic composition. Furthermore, linguistic theories and empirical evidence support language combinatory processes to be to some degree automatic (Frazier and Clifton, 1989; MacDonald et al., 1994). In Appendix A we show that when composition is discouraged, no composition-related activity is found. This further supports our claim that the activity measured with EEG using Bemis & Pylkkänen minimal paradigm does not reflect a compositional process.

5.2 Potential influence of image expectation

Language processing experiments have shown that a negative potential is elicited when a delay is introduced before sentence-final words. Evidence on the effect of contextual constrains on this negative potential's amplitude is conflicting (Besson et al., 1997; Kaan and Carlisle, 2014; León-Cabrera et al., 2017). Nevertheless, the influence of expected stimuli on CNV's amplitude opens another interpretation of our results. We have proposed that the second expectancy wave potentially elicited after second word onset in response to image anticipation is common to all conditions. Nevertheless, given the influence of contextual constraint we cannot discard that our results are due to the fact that the two-word composition condition biases participants to a specific concept and therefore to a singular image. In all other conditions, subjects know that multiple images are congruent with the verbal stimuli. For example, during a '*non-word yellow*' condition (with independence of task), participants would be expecting any image of an object colored yellow to answer affirmatively. The same would occur for the two-word conditions in the lists task: a congruent '*car train*' trial could be followed by an image of a car or a train of any color, and a '*gray blue*' trial would be congruent if any gray or blue object was shown. Therefore, the verbal stimuli in the two-word composition condition induces a more constrained context compared to any other condition, and therefore could account for the difference observed. Nevertheless, this would imply that '*non-word noun*' and '*non-word adjective'* trials should elicit a more negative potential than '*adjective adjective'* and '*noun noun'* trials, respectively, as the one-word conditions further reduce the number of expected images. However, our data shows no evidence of this, and therefore we consider this alternative explanation to be unlikely.

5.3 Time-frequency negative result

We conducted a time-frequency analysis following Segaert et al. (2018). In this article the authors find an increase in alpha-band frequency when participants are exposed to pairs of auditory pseudowords that allow syntactic binding. We expected to reproduce their result, as our NA condition would elicit both syntactic and semantic computations. Moreover, theta and beta oscillations have been previously related to language processing and could be informative of a composition process. However, our analysis yielded no significant results for the analyzed frequency bands. We are not clear on the reasons why we were not able to reproduce Segaert and collaborators' results with our task. One possibility that should be further explored is whether it is due to differences in stimuli timing and modality presentation.

5.4 Sample characteristics

The participants involved in our experiments were predominantly women. This is not ideal as in order to avoid differences introduced by gender this variable should be balanced. Nevertheless, the addition of gender as a factor to the behavioral data analysis yielded no effect, rendering less likely a contribution of gender to the electrophysiological response obtained. Moreover, the experiment conducted by Bemis and Pylkkänen (2011) employed a similar sample (25 subjects, 17 female), hence our adaptation and implementation of the paradigm was very similar.

There have been reports of language differences between man and women, mainly in tasks measuring verbal fluency (Weiss et al., 2003), word use evaluated in tasks that place little constraints on language use (Newman et al., 2008) and in the processing of prosodic cues (Schirmer et al., 2002). Nevertheless the difference in verbal fluency is not supported by studies with bigger and more properly controlled samples (Tombaugh et al., 1999) and there is no consensus on gender based neurobiological differences in language regions (see Wallentin (2009) for a review).

It is probable that basic processes such as unifying words evaluated during a highly constrained task such as the one presented here would be consistent across genders. Nevertheless, if a brain activity specific to noun modification is consistently found, gender differences could be tested and further explored.

5.5 Implications for basic composition studies

The experimental paradigm addressed in this work has consistently showed a difference between two-word and one-word conditions only for the composition task. Because this effect has been attributed to a basic composition process, incorporating expectancy as a variable has important consequences on the interpretation of the results. Nonetheless, it is necessary to distinguish the implications of our study to Neufeld's EEG replication and to the MEG literature results.

We have shown evidence that questions Neufelds's reported composition-related activity using this minimal paradigm. We argue that the experimental design and analysis conducted do not allow to isolate composition from a more general anticipatory process. Furthermore, we consider that their identification of a composition activity preceded by a syntactic building process, originates as a result of testing statistical differences on two visually-targeted time windows, and is not due to two actual separate processes taking place. Importantly, no interaction was found between task and number of words, there by making all post-hoc analysis inappropriate. Moreover, our results suggest that this activity shows a time-dependent behavior consistent with the reported in tasks eliciting CNVs. Additionally, by applying the same robust analysis that was able to capture the so-called composition activity in the initial experiment, we found no effect of composition in a task where subject's expectancy was controlled for. Our results indicate that EEG may not be sensitive to detect the neural activity that underlies composition in a two word design.

In regard to Pylkkänen and collaborators' studies, different versions of this paradigm have been used to explore basic composition. In numerous experiments their design allows for a process of contingency to be elicited as first word is informative of the subsequent word, therefore announcing to participants when a trial requires them to relate the two words. This confounding is shown when stimuli are presented visually (Bemis and Pylkkänen, 2011, 2013a,c), in auditory modality (Bemis and Pylkkänen, 2013a) and in a task evaluating LATL's response to verbs that require argument saturation for both English and Arabic (Westerlund et al., 2015). Moreover, activity elicited at first word onset is not analyzed or plotted, and a potential confounding with expectancy processes is never addressed. In order to claim LATL's increased activity as a marker of basic composition in this two-word paradigm, anticipatory processes should be discarded. However, it is possible to some extent that by implementing such a specific statistical analysis over selected ROIs they manage to isolate a composition response. This is supported by studies of this group where they show evidence for this combinatorial effect using experimental designs that lack this obvious methodological confounding. Studies in which expectancy would equally affect two-word constructions show that compared to more general nouns, nouns higher in specificity do not elicit this activity (Westerlund and Pylkkänen, 2014) and similar results on modifier type and specificity were shown (Zhang and Pylkkänen, 2015; Ziegler and Pylkkänen, 2016). Furthermore, these authors show an increase in LATL's activity in a production task only when subjects had to describe pictures corresponding to adjective-noun constructions and not pictures representing two concepts (Prato and Pylkkänen, 2014; Pylkkänen et al., 2014), and a corresponding result was found for sign language (Blanco-Elorrieta et al., 2018). Nevertheless, the statistic used in most of these experiments is based on the initial study findings, therefore it would be of interest to revise the original result.

Conclusions

This work aimed to find the electrophysiological correlates of elementary composition with EEG by adapting to Spanish a linguistically minimal paradigm recently introduced in the MEG literature by Bemis & Pylkkänen. These authors have shown that words presented in combinatorial contexts elicit an increased activity in the LATL 184-255 ms after word presentation. In this study we were able to successfully reproduce the composition-related activity on EEG data by applying a robust statistical analysis with a data-driven approach. Furthermore, the preprocessing and analyses carried out allowed us to explore the time course of this activity, which was indicative of a potential confound in the experimental design. We propose that this paradigm allows for an expectancy processes to develop only for words in combinatorial contexts, and that this could account for the results reported in the literature. In this line, we show a task progression effect on the obtained activity, which follows the behavior described for the anticipatory potential CNV. Furthermore, we designed a task in which expectancy contributions were controlled for, and show that this composition-related activity is not elicited when participants are unaware of whether an incoming trial requires them to compose or not. Together, this results provide evidence of a confounding effect of expectancy introduced by the block design of the experimental paradigm proposed by Bemis & Pylkkänen.

Showing evidence of basic composition in a free-expectancy context is fundamental to determine LATL's actual involvement in this process. We suggest that the simple experiment presented in this work to study basic composition while controlling for anticipatory effects, could prove useful to determine this; an adaptation of our paradigm to English could be implemented using MEG by simply inverting the pairs of stimuli, and constructing pairs of adjectives to control for number of words.

Furthermore, we consider that this work shows the importance of carrying robust and objective analyses when working with EEG data. A previous EEG study, replicating Bemis & Pylkkänen's experiment, in which statistics were applied on specific time windows, failed to properly replicate their result. Moreover, the analyses conducted caused authors to attribute their results to two independent operations instead of reflecting a unique and continuous process.

Finally, we were not able to obtain an unequivocally composition-related activity with a two-word paradigm; therefore, future studies will be necessary to determine if EEG is sensitive to this basic composition process.

Bibliography

Bares, Martin and Rektor, Ivan (2001). "Basal ganglia involvement in sensory and cognitive processing. A depth electrode CNV study in human subjects". *Clinical Neurophysiology*, **112(11)**, pp. 2022–2030. ISSN 13882457. doi: 10.1016/S1388-2457(01)00671-X.

Baron, Sean G. and Osherson, Daniel (2011). "Evidence for conceptual combination in the left anterior temporal lobe". *NeuroImage*, **55(4)**, pp. 1847–1852. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.01.066.

Baron, Sean G.; Thompson-Schill, Sharon L.; Weber, Matthew and Osherson, Daniel (2010). "An early stage of conceptual combination: Superimposition of constituent concepts in left anterolateral temporal lobe". *Cognitive Neuroscience*, **1(1)**, pp. 44–51. ISSN 17588928. doi: 10.1080/17588920903548751.

Bastiaansen, Marcel; Magyari, Lilla and Hagoort, Peter (2010). "Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension". *Journal of Cognitive Neuroscience*, **22(7)**, pp. 1333–1347. ISSN 0898929X. doi: 10.1162/jocn.2009.21283.

Bates, Douglas; Mächler, Martin; Bolker, Ben and Walker, Steve (2015). "Fitting Linear Mixed-Effects Models Using {lme4}". *Journal of Statistical Software*, **67(1)**, pp. 1–48. doi: 10.18637/jss.v067.i01.

Bemis, D. K. and Pylkkänen, Liina (2011). "Simple Composition: A Magnetoencephalography Investigation into the Comprehension of Minimal Linguistic Phrases". *Journal of Neuroscience*, **31(8)**, pp. 2801–2814. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.5003-10.2011.

Bemis, D. K. and Pylkkänen, Liina (2013a). "Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading". *Cerebral Cortex*, **23(8)**, pp. 1859–1873. ISSN 10473211. doi: 10.1093/cercor/bhs170.

Bemis, Douglas K. and Pylkkänen, Liina (2013b). "Combination across domains: An MEG investigation into the relationship between mathematical, pictorial, and linguistic processing". *Frontiers in Psychology*, **3(JAN)**, pp. 1–20. ISSN 16641078. doi: 10.3389/fpsyg.2012.00583.

Bemis, Douglas K. and Pylkkänen, Liina (2013c). "Flexible Composition: MEG Evidence for the Deployment of Basic Combinatorial Linguistic Mechanisms in Response to Task Demands". *PLoS ONE*, **8(9)**. ISSN 19326203. doi: 10.1371/journal.pone.0073949.

Besson, Mireille; Faita, Frederique; Czternasty, Claire and Kutas, Marta (1997). "What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences". *Biological Psychology*, **46(1)**, pp. 3–23. ISSN 03010511. doi: 10.1016/S0301-0511(96)05215-5.

Birbaumer, N; Elbert, T; Canavan, A G and Rockstroh, B (1990). "Slow potentials of the cerebral cortex and behavior." *Physiological Reviews*, **70(1)**, pp. 1–41. ISSN 0031-9333. doi: 10.1152/physrev.1990.70.1.1.

Blair, R. Clifford and Karniski, Walt (1993). "An alternative method for significance testing of waveform difference potentials". *Psychophysiology*, **30(5)**, pp. 518–524. ISSN 0048-5772. doi: 10.1111/j.1469-8986.1993.tb02075.x.

Blanco-Elorrieta, Esti; Kastner, Itamar; Emmorey, Karen and Pylkkänen, Liina (2018). "Shared neural correlates for building phrases in signed and spoken language". *Scientific Reports*, **8(1)**, pp. 1–10. ISSN 20452322. doi: 10.1038/s41598-018-23915-0.

Bosque, I and Demonte, V (1999). *Gramática Descriptiva de la Lengua Española*. Number v. 1 in Nebrija y Bello. Librería Tirant lo Blanch. ISBN 9788423979172.

Brennan, Jonathan and Pylkkänen, Liina (2012). "The time-course and spatial distribution of brain activity associated with sentence processing". *NeuroImage*, **60(2)**, pp. 1139–1148. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.01.030.

Brunia, C.H.M. and Vingerhoets, A.J.J.M. (1981). "Opposite hemisphere differences in movement related potentials preceding foot and finger flexions". *Biological Psychology*, **13(C)**, pp. 261–269. ISSN 03010511. doi: 10.1016/0301-0511(81)90041-7.

Brunia, Cornelis H.M.; Hackley, Steven A.; van Boxtel, Geert J.M.; Kotani, Yasunori and Ohgami, Yoshimi (2011). "Waiting to perceive: Reward or punishment?" *Clinical Neurophysiology*, **122(5)**, pp. 858–868. ISSN 13882457. doi: 10.1016/j.clinph.2010.12.039.

Burgess, Adrian P. (2012). "Towards a Unified Understanding of Event-Related Changes in the EEG: The Firefly Model of Synchronization through Cross-Frequency Phase Modulation". *PLoS ONE*, **7(9)**. ISSN 19326203. doi: 10.1371/journal.pone.0045630.

Buzsáki, György; Anastassiou, Costas A. and Koch, Christof (2012). "The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes". *Nature Reviews Neuroscience*, **13(6)**, pp. 407–420. ISSN 1471003X. doi: 10.1038/nrn3241.

Buzsáki, György C N (2006). *Rhythms of the brain*. ISBN 9780199863716. doi: 10.1093/acprof:oso/9780195301069.001.0001.

Callaway, E (1975). *Brain electrical potentials and individual psychological differences*. Grune & Stratton. ISBN 9780808908821.

Caplan, David; Chen, Evan and Waters, Gloria (2008). "Task-dependent and task-independent neurovascular responses to syntactic processing". *Cortex*, **44(3)**, pp. 257–275. ISSN 00109452. doi: 10.1016/j.cortex.2006.06.005.

Cohen, Jerome (1969). "Very slow brain potentials relating to expectancy: The CNV." In: *Average evoked potentials: Methods, results, and evaluations.*, pp. 143–198. US National Aeronautics and Space Administration, Washington. doi: 10.1037/13016-004.

Cohen, Jerome and Walter, W. Grey (1966). "the Interaction of Responses in the Brain To Semantic Stimuli". *Psychophysiology*, **2(3)**, pp. 187–196. ISSN 14698986. doi: 10.1111/j.1469-8986.1966.tb02642.x.

Cohen, X Mike (2014). *Analyzing Neural Time Series Data*. 1. ISBN 9780874216561. doi: 10.1007/s13398-014-0173-7.2.

Critchley, Hugo D; Mathias, Christopher J and Dolan, Raymond J (2001). "Neural Activity in the Human Brain Relating to Uncertainty and Arousal during Anticipation". *Neuron*, **29(2)**, pp. 537–545. ISSN 08966273. doi: 10.1016/S0896-6273(01)00225-2.

Critchley, Hugo D.; Mathias, Christopher J.; Josephs, Oliver; O'Doherty, John; Zanini, Sergio; Dewar, Bonnie-Kate; Cipolotti, Lisa; Shallice, Tim and Dolan, Raymond J. (2003). "Human cingulate cortex and autonomic control: converging neuroimaging and clinical evidence". *Brain*, **126(10)**, pp. 2139–2152. ISSN 1460-2156. doi: 10.1093/brain/awg216.

Damen, E. J.P. and Brunia, C. H.M. (1987). "Changes in Heart Rate and Slow Brain Potentials Related to Motor Preparation and Stimulus Anticipation in a Time Estimation Task". *Psychophysiology*, **24(6)**, pp. 700–713. ISSN 14698986. doi: 10.1111/j.1469-8986.1987.tb00353.x.

Donald, Merlin W. (1980). "Memory, Learning and Event-Related Potentials". 1966, pp. 615–627. doi: 10.1016/S0079-6123(08)61681-7.

Doya, Kenji (2008). "Modulators of decision making". *Nature Neuroscience*, **11**, p. 410.

Duchon, Andrew; Perea, Manuel; Sebastián-Gallés, Nuria; Martí, Antonia and Carreiras, Manuel (2013). "EsPal: One-stop shopping for Spanish word properties". *Behavior Research Methods*, **45(4)**, pp. 1246–1258. ISSN 1554351X. doi: 10.3758/s13428-013-0326-1.

Dunnett, Charles W (1955). "Journal of the American Statistical Association,". *Journal of the American Statistical Association,*, **50(272)**, pp. 1096–1121. ISSN 00221899. doi: 10.2307/2286841.

Elbert, Thomas (1993). "Slow Cortical Potentials Reflect the Regulation of Cortical Excitability". *Slow Potential Changes in the Human Brain*, pp. 235–251. doi: 10.1007/978-1-4899-1597-9_15.

Frazier, Lyn and Clifton, Charles (1989). *Successive Cyclicity in the Grammar and the Parser*. volume 4. ISBN 0169-0965. doi: 10.1080/01690968908406359.

Friederici, A. D. (2011). "The Brain Basis of Language Processing: From Structure to Function". *Physiological Reviews*, **91(4)**, pp. 1357–1392. ISSN 0031-9333. doi: 10.1152/physrev.00006.2011.

Friederici, A. D. and Hahne, A. (2002). "Differential task effects on semantic and syntactic processes as revealed by ERPs". *Cognitive Brain Research*, **13(3)**, pp. 339–356.

Friederici, Angela D; Meyer, Martin and Cramon, D Yves Von (2000). "Auditory Language Comprehension : An Event-Related fMRI Study on the Processing of Syntactic and Lexical Information". *Brain and Language*, **300**, pp. 289–300. doi: 10.1006/brln.2000.2313.

Gaillard, A.W.K. (1976). “Effects of warning-signal modality on the contingent negative variation (CNV)”. *Biological Psychology*, **4(2)**, pp. 139–153. ISSN 03010511. doi: 10.1016/0301-0511(76)90013-2.

Goodman, Steven N. (1998). “Multiple comparisons, explained”. *American Journal of Epidemiology*, **147(9)**, pp. 807–812. ISSN 14766256. doi: 10.1093/oxfordjournals.aje.a009531.

Hagoort, Peter; Hald, Lea and Bastiaansen, Marcel (2004). “Integration of Word Meaning and World Knowledge in Language Comprehension”, **304(April)**, pp. 438–441. doi: 10.1126/science.1095455.

Hagoort, Peter and Indefrey, Peter (2014). “The Neurobiology of Language Beyond Single Words.” *Annual review of neuroscience*, **37(May)**, pp. 347–362. ISSN 1545-4126. doi: 10.1146/annurev-neuro-071013-013847.

Hald, Lea A.; Bastiaansen, Marcel C M and Hagoort, Peter (2006). “EEG theta and gamma responses to semantic violations in online sentence processing”. *Brain and Language*, **96(1)**, pp. 90–105. ISSN 0093934X. doi: 10.1016/j.bandl.2005.06.007.

Hämäläinen, Matti; Hari, Riitta; Ilmoniemi, Risto J.; Knuutila, Jukka and Lounasmaa, Olli V. (1993). “Magnetoencephalography theory, instrumentation, and applications to noninvasive studies of the working human brain”. *Reviews of Modern Physics*, **65(2)**, pp. 413–497. ISSN 00346861. doi: 10.1103/RevModPhys.65.413.

Hansen, Peter C.; Kringelbach, Morten L. and Salmelin, Ritta (2010). *MEG: An Introduction to Methods*. Oxford University Press, New York. ISBN 9780199863990. doi: 10.1093/acprof:oso/9780195307238.003.0002.

Hillyard, Steven A. (1969). “Relationships between the contingent negative variation (CNV) and reaction time”. *Physiology and Behavior*, **4(3)**, pp. 351–357. ISSN 00319384. doi: 10.1016/0031-9384(69)90188-7.

Holcomb, Phillip J (1993). “Semantic priming and stimulus degradation : Implications for the role of the N400 in language processing”. *Psychophysiology*, **30**, pp. 47–61. doi: 10.1111/j.1469-8986.1993.tb03204.x.

Humphries, Colin; Binder, Jeffrey R and Medler, David A (2006). “Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension”. *Journal of Cognitive Neuroscience*, **18(14)**, pp. 665–679. doi: 10.1162/jocn.2006.18.4.665.

Humphries, Colin; Love, Tracy; Swinney, David and Hickok, Gregory (2005). “Response of Anterior Temporal Cortex to Syntactic and Prosodic Manipulations During Sentence Processing”. *Human Brain Mapping*, **138**, pp. 128–138. doi: 10.1002/hbm.20148.

Ikeda, Akio; Shibasaki, Hiroshi; Kaji, Ryuji; Terada, Kiyohito; Nagamine, Takashi; Honda, Manabu and Kimura, Jun (1997). “Dissociation between contingent negative variation (CNV) and Bereitschaftspotential (BP) in patients with parkinsonism”. *Electroencephalography and Clinical Neurophysiology*, **102(2)**, pp. 142–151. ISSN 00134694. doi: 10.1016/S0921-884X(96)95067-5.

Ioannides, A. A.; Fenwick, P. B.C.; Lumsden, J.; Liu, M. J.; Bamidis, P. D.; Squires, K. C.; Lawson, D. and Fenton, G. W. (1994). "Activation sequence of discrete brain areas during cognitive processes: results from magnetic field tomography". *Electroencephalography and Clinical Neurophysiology*, **91(5)**, pp. 399–402. ISSN 00134694. doi: 10.1016/0013-4694(94)90125-2.

Irwin, Donald A.; Knott, John R.; McAdam, Dale W. and Robert, Charles S. (1966). "Motivational determinants of the "contingent negative variation"". *Electroencephalography and Clinical Neurophysiology*, **21(6)**, pp. 538–543. ISSN 00134694. doi: 10.1016/0013-4694(66)90172-6.

J. Valone, Thomas (2006). "Are animals capable of Bayesian updating? An empirical review". *Oikos*, **112(2)**, pp. 252–259. ISSN 00301299. doi: 10.1111/j.0030-1299.2006.13465.x.

Jackendoff, Ray (2000). "The representational structures of the language faculty and their interactions". In: *The Neurocognition of Language*, pp. 37–79. Oxford University Press. doi: 10.1093/acprof:oso/9780198507932.003.0003.

Jacobson, Gary P and Gans, Donald P (1981). "The Contingent Negative Variation as an Indicator of Speech Discrimination Difficulty". *Journal of Speech Language and Hearing Research*, **24(3)**, p. 345. ISSN 1092-4388. doi: 10.1044/jshr.2403.345.

Jahanshahi, M and Hallet, M (2003). *The Bereitschaftspotential: Movement-Related Cortical Potentials*. Kluwer Plenum.

Jarvilehto, T. and Fruhstorfer, H. (1970). "Differentiation between slow cortical potentials associated with motor and mental acts in man". *Experimental Brain Research*, **11(3)**, pp. 309–317. ISSN 0014-4819. doi: 10.1007/BF01474389.

Jasper, Herbert Henri (1958). *The Tenn Twenty Electrode System of The International Federation*. ISBN 0013-4694. doi: 10.1016/0013-4694(58)90053-1.

Jung, T-p; Makeig, S; Lee, T-w; McKeown, M J and Iragui, V (2000). "Removing Electroencephalographic Artifacts By Blind Source Separation". *Psychophysiology*, **37**, pp. 163–178.

Junghöfer, Markus; Elbert, Thomas; Tucker, Don M. and Rockstroh, Brigitte (2000). "Statistical control of artifacts in dense array EEG/MEG studies". *Psychophysiology*, **37(4)**, pp. 523–532. ISSN 00485772. doi: 10.1017/S0048577200980624.

Kaan, Edith and Carlisle, Evan (2014). "ERP indices of stimulus prediction in letter sequences". *Brain Sciences*, **4(4)**, pp. 509–531. ISSN 20763425. doi: 10.3390/brainsci4040509.

Keenan, E L (1979). *On Surface Form and Logical Form*. Linguistic agency. Series. A. Linguistic Agency, University of Trier.

Keuleers, Emmanuel and Brysbaert, Marc (2010). "Wuggy: A multilingual pseudoword generator". *Behavior Research Methods*, **42(3)**, pp. 627–633. ISSN 1554351X. doi: 10.3758/BRM.42.3.627.

Kim, Albert and Osterhout, Lee (2005). "The independence of combinatory semantic processing: Evidence from event-related potentials". *Journal of Memory and Language*, **52(2)**, pp. 205–225. ISSN 0749596X. doi: 10.1016/j.jml.2004.10.002.

Klorman, Rafael and Ryan, Richard M. (1980). "Heart Rate, Contingent Negative Variation, and Evoked Potentials during Anticipation of Affective Stimulation". doi: 10.1111/j.1469-8986.1980.tb02290.x.

Kornhuber, Hans H. and Deecke, Lüder (1965). "Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale". *Pflügers Archiv für die Gesamte Physiologie des Menschen und der Tiere*, **284(1)**, pp. 1–17. ISSN 00316768. doi: 10.1007/BF00412364.

Krause, Christina M.; Lang, Heikki A.; Laine, Matti; Helle, Seppo I.; Kuusisto, Mika J. and Pörn, Bodil (1994). "Event-Related Desynchronization evoked by auditory stimuli". *Brain Topography*, **7(2)**, pp. 107–112. ISSN 08960267. doi: 10.1007/BF01186768.

Kujala, Jan; Pammer, Kristen; Cornelissen, Piers; Roebroeck, Alard; Formisano, Elia and Salmelin, Riitta (2007). "Phase coupling in a cerebro-cerebellar network at 8-13 Hz during reading". *Cerebral Cortex*, **17(6)**, pp. 1476–1485. ISSN 10473211. doi: 10.1093/cercor/bhl059.

Kuperberg, Gina R. (2007). "Neural mechanisms of language comprehension: Challenges to syntax". *Brain Research*, **1146(1)**, pp. 23–49. ISSN 00068993. doi: 10.1016/j.brainres.2006.12.063.

Kutas, M (1988). "Event-related brain potential studies of language." *Advances in Psychophysiology*, **3**, pp. 129–187.

Kutas, M and Hillyard, S A (1980). "Reading senseless sentences: brain potentials reflect semantic incongruity". *Science*, **207(4427)**, pp. 203 LP ——— 205.

Kutas, M and Van Petten, C. K. (1994). "Psycholinguistics electrified: Event-related brain potential investigations". In: M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, pp. 83–143. Academic Press, San Diego, CA, US.

Kutas, Marta and Federmeier, Kara D. (2011). "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)". *Annual Review of Psychology*, **62(1)**, pp. 621–647. ISSN 0066-4308. doi: 10.1146/annurev.psych.093008.131123.

Lau, Ellen F.; Phillips, Colin and Poeppel, David (2008). "A cortical network for semantics: (de)constructing the N400". *Nature Reviews Neuroscience*, **9(12)**, pp. 920–933. ISSN 1471-003X. doi: 10.1038/nrn2532.

Lee, Kyoung Min; Chang, Kee Hyun and Roh, Jae Kyu (1999). "Subregions within the supplementary motor area activated at different stages of movement preparation and execution". *NeuroImage*, **9(1)**, pp. 117–123. ISSN 10538119. doi: 10.1006/nimg.1998.0393.

Lenth, Russell (2018). "emmeans: Estimated Marginal Means, aka Least-Squares Means".

León-Cabrera, Patricia; Rodríguez-Fornells, Antoni and Morís, Joaquín (2017). "Electrophysiological correlates of semantic anticipation during speech comprehension". *Neuropsychologia*, **99(February)**, pp. 326–334. ISSN 18733514. doi: 10.1016/j.neuropsychologia.2017.02.026.

Loveless, N. E. (1975). "The effect of warning interval on signal detection and event-related slow potentials of the brain". *Perception & Psychophysics*, **17(6)**, pp. 565–570. ISSN 00315117. doi: 10.3758/BF03203970.

Loveless, N. E. and Sanford, A. J. (1974). "Slow potential correlates of preparatory set". *Biological Psychology*, **1(4)**, pp. 303–314. ISSN 03010511. doi: 10.1016/0301-0511(74)90005-2.

Low, Morton D.; Coats, Alfred C.; Rettig, Gayle M. and McSherry, Joseph W. (1967). "Anxiety, attentiveness-alertness: A phenomenological study of the CNV". *Neuropsychologia*, **5(4)**, pp. 379–384. ISSN 00283932. doi: 10.1016/0028-3932(67)90009-7.

Luck, Steven J. (Steven John) (2005). *An introduction to the event-related potential technique*. ISBN 0262122774. doi: 10.1118/1.4736938.

Luks, Tracy L.; Simpson, Gregory V.; Feiwell, Robert J. and Miller, William L. (2002). "Evidence for anterior cingulate cortex involvement in monitoring preparatory attentional set". *NeuroImage*, **17(2)**, pp. 792–802. ISSN 10538119. doi: 10.1016/S1053-8119(02)91210-3.

Macar, F.; Lejeune, H.; Bonnet, M.; Ferrara, A.; Pouthas, V.; Vidal, F. and Maquet, P. (2002). "Activation of the supplementary motor area and of attentional networks during temporal processing". *Experimental Brain Research*, **142(4)**, pp. 475–485. ISSN 00144819. doi: 10.1007/s00221-001-0953-0.

Macar, Françoise and Vidal, Franck (2002). "Time processing reflected by EEG surface Laplacians". *Experimental Brain Research*, **145(3)**, pp. 403–406. ISSN 00144819. doi: 10.1007/s00221-002-1103-z.

Macar, Françoise and Vidal, Franck (2003). "The CNV peak: An index of decision making and temporal memory". *Psychophysiology*, **40(6)**, pp. 950–954. ISSN 00485772. doi: 10.1111/1469-8986.00113.

Macar, Françoise; Vidal, Franck and Casini, Laurence (1999). "The supplementary motor area in motor and sensory timing: Evidence from slow brain potential changes". *Experimental Brain Research*, **125(3)**, pp. 271–280. ISSN 00144819. doi: 10.1007/s002210050683.

MacDonald, Maryellen C.; Pearlmutter, Neal J. and Seidenberg, Mark S. (1994). "The lexical nature of syntactic ambiguity resolution." *Psychological Review*, **101(4)**, pp. 676–703. ISSN 1939-1471. doi: 10.1037/0033-295X.101.4.676.

Maidhof, Clemens and Koelsch, Stefan (2011). "Effects of selective attention on syntax processing in music and language". *Journal of Cognitive Neuroscience*, **23(9)**, pp. 2252–2267. ISSN 0898929X. doi: 10.1162/jocn.2010.21542.

Maris, Eric and Oostenveld, Robert (2007). "Nonparametric statistical testing of EEG- and MEG-data". *Journal of Neuroscience Methods*, **164(1)**, pp. 177–190. ISSN 01650270. doi: 10.1016/j.jneumeth.2007.03.024.

Martin, Laura E.; Potts, Geoffrey F.; Burton, Philip C. and Montague, P. Read (2009). "Electrophysiological and hemodynamic responses to reward prediction violation". *NeuroReport*, **20(13)**, pp. 1140–1143. ISSN 09594965. doi: 10.1097/WNR.0b013e32832f0dca.

Mazoyer, B M; Tzourio, N; Frak, V; Syrota, A; Murayama, N and Salamon, G (1993). "The Cortical Representation of Speech". *Journal of Cognitive Neuroscience*, **5(4)**, pp. 467–479. doi: 10.1162/jocn.1993.5.4.467.

McAdam, Dale W. (1966). "Slow potential changes recorded from human brain during learning of a temporal interval". *Psychonomic Science*, **6(9)**, pp. 435–436. ISSN 00333131. doi: 10.3758/BF03330974.

McCallum, W. C.; Papakostopoulos, D.; Gombi, R.; Winter, A. L.; Cooper, R. and Griffith, H. B. (1973). "Event Related Slow Potential Changes in Human Brain Stem". *Nature*, **242(5398)**, pp. 465–467. ISSN 0028-0836. doi: 10.1038/242465a0.

McCallum, W. Cheyne (1993). "Human Slow Potential Research: A Review". In: *Slow Potential Changes in the Human Brain*, volume 39, pp. 1–12. Springer US, Boston, MA. ISBN 9781489915993. doi: 10.1007/978-1-4899-1597-9_1.

Mensen, Armand and Khatami, Ramin (2013). "Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics". *NeuroImage*, **67**, pp. 111–118. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.10.027.

Mento, Giovanni; Tarantino, Vincenza; Sarlo, Michela and Bisiacchi, Patrizia Silvia (2013). "Automatic Temporal Expectancy: A High-Density Event-Related Potential Study". *PLoS ONE*, **8(5)**. ISSN 19326203. doi: 10.1371/journal.pone.0062896.

Murphy, Gregory L (2002). *The Big Book of Concepts*. MIT Press, Boston, Mass.. ISBN 9780262134095. doi: 10.1016/S0001-6918(03)00022-2.

Nagai, Y.; Critchley, H. D.; Featherstone, E.; Fenwick, P. B C; Trimble, M. R. and Dolan, R. J. (2004). "Brain activity relating to the contingent negative variation: An fMRI investigation". *NeuroImage*, **21(4)**, pp. 1232–1241. ISSN 10538119. doi: 10.1016/j.neuroimage.2003.10.036.

Neufeld, Chris; Kramer, Stephanie E.; Lapinskaya, Natalia; Heffner, Christopher C.; Malko, Anton and Lau, Ellen F. (2016). "The electrophysiology of basic phrase building". *PLoS ONE*, **11(10)**, pp. 1–22. ISSN 19326203. doi: 10.1371/journal.pone.0158446.

Newman, Matthew L.; Groom, Carla J.; Handelman, Lori D. and Pennebaker, James W. (2008). "Gender differences in language use: An analysis of 14,000 text samples". *Discourse Processes*, **45(3)**, pp. 211–236. ISSN 0163853X. doi: 10.1080/01638530802073712.

Okada, Yoshio; Lähteenmäki, Airi and Xu, Chibing (1999). "Comparison of MEG and EEG on the basis of somatic evoked responses elicited by stimulation of the snout in the juvenile swine". *Clinical Neurophysiology*, **110(2)**, pp. 214–229. ISSN 13882457. doi: 10.1016/S0013-4694(98)00111-4.

Oostenveld, Robert; Fries, Pascal; Maris, Eric and Schoffelen, Jan Mathijs (2011). "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data". *Computational Intelligence and Neuroscience*, **2011**. ISSN 16875265. doi: 10.1155/2011/156869.

Osterhout, Lee and Holcomb, Phillip J. (1992). "Event-related brain potentials elicited by syntactic anomaly". *Journal of Memory and Language*, **31(6)**, pp. 785–806. ISSN 0749596X. doi: 10.1016/0749-596X(92)90039-Z.

Pallier, Christophe; Devauchelle, Anne-Dominique and Dehaene, Stanislas (2011). "Cortical representation of the constituent structure of sentences". *Proceedings of the National Academy of Sciences*, **108**, pp. 2522–2527. doi: 10.1073/pnas.1018711108.

Parks, Nathan A.; Gannon, Matthew A.; Long, Stephanie M. and Young, Madeleine E. (2016). "Bootstrap Signal-to-Noise Confidence Intervals: An Objective Method for Subject Exclusion and Quality Control in ERP Studies". *Frontiers in Human Neuroscience*, **10(February)**, pp. 1–15. ISSN 1662-5161. doi: 10.3389/fnhum.2016.00050.

Partee, Barbara. H. (1984). "Compositionality". In: F Veltman and F Landmand (Eds.), *Varieties of formal semantics*, Foris publications, Dordrecht.

Partee, Barbara H (2007). "Compositionality and coercion in semantics: The dynamics of adjective meaning", pp. 145–161.

Peirce, Jonathan W. (2007). "PsychoPy-Psychophysics software in Python". *Journal of Neuroscience Methods*, **162(1-2)**, pp. 8–13. ISSN 01650270. doi: 10.1016/j.jneumeth.2006.11.017.

Pernet, C. R.; Latinus, M.; Nichols, T. E. and Rousselet, G. a. (2014). "Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study". *Journal of Neuroscience Methods*, **250**, pp. 85–93. ISSN 01650270. doi: 10.1016/j.jneumeth.2014.08.003.

Peters, Jon F.; Billinger, Thomas W. and Knott, John R. (1977). "Event Related Potentials of Brain (CNV and P300) In A Paired Associate Learning Paradigm". *Psychophysiology*, **14(6)**, pp. 579–585. ISSN 0048-5772. doi: 10.1111/j.1469-8986.1977.tb01203.x.

Pfeuty, Micha; Ragot, Richard and Pouthas, Viviane (2005). "Relationship between CNV and timing of an upcoming event". *Neuroscience Letters*, **382(1-2)**, pp. 106–111. ISSN 01650009. doi: 10.1007/s10584-005-6886-z.

Picton, Terence W.; Bentin, S.; Berg, P.; Donchin, E.; Hillyard, S. A.; Johnson, R.; Miller, G. A.; Ritter, W.; Ruchkin, D. S.; Rugg, M. D. and Taylor, M. J. (2000). "Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria". *Psychophysiology*, **37(2)**, pp. 127–152. ISSN 00485772. doi: 10.1017/S0048577200000305.

Poon, Leonard W.; Thompson, Larry W.; Williams, Redford B. and Marsh, Gail R. (1974). "Changes of Antero-Posterior Distribution of CNV and Late Positive Component as a Function of Information

Processing Demands". *Psychophysiology*, **11(6)**, pp. 660–673. ISSN 0048-5772. doi: 10.1111/j.1469-8986.1974.tb01135.x.

Poortman, Eva B. and Pylkkänen, Liina (2016). "Adjective conjunction as a window into the LATL's contribution to conceptual combination". *Brain and Language*. ISSN 0093934X. doi: 10.1016/j.bandl.2016.07.006.

Potter, Mary C. and Faulconer, Barbara A. (1979). "Understanding noun phrases". *Journal of Verbal Learning and Verbal Behavior*, **18(5)**, pp. 509–521. ISSN 00225371. doi: 10.1016/S0022-5371(79)90274-3.

Prato, Paul Del and Pylkkänen, Liina (2014). "MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production". *Frontiers in Psychology*, **5(June)**, pp. 1–11. doi: 10.3389/fpsyg.2014.00524.

Proulx, G. B. and Picton, T. W. (1980). "The CNV during Cognitive Learning and Extinction". *Progress in Brain Research*, **54(C)**, pp. 309–313. ISSN 18757855. doi: 10.1016/S0079-6123(08)61640-4.

Pulvermüller, Friedemann; Shtyrov, Yury; Hasting, Anna S. and Carlyon, Robert P. (2008). "Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing". *Brain and Language*, **104(3)**, pp. 244–253. ISSN 0093934X. doi: 10.1016/j.bandl.2007.05.002.

Pylkkänen, Liina (2008). "Mismatching meanings in brain and behavior". *Linguistics and Language Compass*. ISSN 1749818X. doi: 10.1111/j.1749-818X.2008.00073.x.

Pylkkänen, Liina (2016). "Composition of Complex Meaning: Interdisciplinary Perspectives on the Left Anterior Temporal Lobe". *Neurobiology of Language*, pp. 621–631. doi: 10.1016/B978-0-12-407794-2.00050-X.

Pylkkänen, Liina; Bemis, Douglas K and Blanco, Estibaliz (2014). "Building phrases in language production : An MEG study of simple composition". *Cognition*, **133**, pp. 371–384. doi: 10.1016/j.cognition.2014.07.001.

Pylkkänen, Liina; Brennan, Jonathan and Bemis, Douglas K (2011). "Language and Cognitive Processes Grounding the cognitive neuroscience of semantics in linguistic theory", **(November 2013)**, pp. 37–41. doi: 10.1080/01690965.2010.527490.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rebert, Charles S. (1972). "Cortical and subcortical slow potentials in the monkey's brain during a preparatory interval". *Electroencephalography and Clinical Neurophysiology*, **33(4)**, pp. 389–402. ISSN 00134694. doi: 10.1016/0013-4694(72)90119-8.

Rebert, Charles S; McAdam, Dale W and Knott, John R (1967). "Slow potential change in human brain related to level of motivation.". doi: 10.1037/h0024146.

Ridderinkhof, K. Richard; Van Den Wildenberg, Wery P.M.; Segalowitz, Sidney J. and Carter, Cameron S. (2004). “Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning”. *Brain and Cognition*, **56(2 SPEC. ISS.)**, pp. 129–140. ISSN 02782626. doi: 10.1016/j.bandc.2004.09.016.

Ritter, Walter; Rotkin, Laurence and Vaughan, Herbert G. (1980). “The Modality Specificity of the Slow Negative Wave”. *Psychophysiology*, **17(3)**, pp. 222–227. ISSN 0048-5772. doi: 10.1111/j.1469-8986.1980.tb00138.x.

Rohrbaugh, John W. and Gaillard, Anthony W.K. (1983). “13 Sensory and Motor Aspects of the Contingent Negative Variation”. pp. 269–310. doi: 10.1016/S0166-4115(08)62044-0.

Rohrbaugh, John W.; Syndulko, Karl and Lindsley, Donald B. (1976). “Brain wave components of the contingent negative variation in humans”. *Science*, **191(4231)**, pp. 1055–1057. ISSN 00368075. doi: 10.1126/science.1251217.

RStudio Team (2016). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Ruchkin, D.S; Sutton, S; Mahaffey, D and Glaser, J (1986). “Terminal CNV in the absence of motor response”. *Electroencephalography and Clinical Neurophysiology*, **63(5)**, pp. 445–463. ISSN 00134694. doi: 10.1016/0013-4694(86)90127-6.

Schirmer, Annett; Kotz, Sonja A. and Friederici, Angela D. (2002). “Sex differentiates the role of emotional prosody during word processing”. *Cognitive Brain Research*, **14(2)**, pp. 228–233. ISSN 09266410. doi: 10.1016/S0926-6410(02)00108-8.

Segaert, Katrien; Mazaheri, Ali and Hagoort, Peter (2018). “Binding language: Structuring sentences through precisely timed oscillatory mechanisms”. *European Journal of Neuroscience*, **1**, pp. 1–12. ISSN 14609568. doi: 10.1111/ejn.13816.

Shaffer, J P (1995). “Multiple Hypothesis Testing”. *Annual Review of Psychology*, **46(1)**, pp. 561–584. ISSN 0066-4308. doi: 10.1146/annurev.ps.46.020195.003021.

Shtyrov, Yury; Pulvermüller, Friedemann and Na, Risto (2002). “Grammar Processing Outside the Focus of Attention : an MEG Study”. *Journal of Cognitive Neuroscience*, pp. 1195–1206.

Simons, Robert F.; ÖHman, Arne and Lang, Peter J. (1979). “Anticipation and Response Set: Cortical, Cardiac, and Electrodermal Correlates”. *Psychophysiology*, **16(3)**, pp. 222–233. ISSN 14698986. doi: 10.1111/j.1469-8986.1979.tb02982.x.

Sitnikova, Tatiana; Holcomb, Phillip J; Kiyonaga, Kristi A and Kuperberg, Gina R (2008). “Two Neurocognitive Mechanisms of Semantic Integration during the Comprehension of Visual Real-world Events”. *Journal of cognitive neuroscience*, pp. 2037–2057.

Skinner, James E and Yingling, Charles D (1976). "Regulation of slow potential shifts in nucleus reticularis thalami by the mesencephalic reticular formation and the frontal granular cortex". *Electroencephalography and Clinical Neurophysiology*, **40(3)**, pp. 288–296. ISSN 00134694. doi: 10.1016/0013-4694(76)90152-8.

Smith, Stephen M. and Nichols, Thomas E. (2009). "Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference". *NeuroImage*, **44(1)**, pp. 83–98. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.03.061.

Snijders, Tineke M; Vosse, Theo; Kempen, Gerard; Berkum, J A Van; Petersson, Karl Magnus and Hagoort, Peter (2009). "Retrieval and Unification of Syntactic Structure in Sentence Comprehension : an fMRI Study Using Word-Category Ambiguity". *Cerebral Cortex*, **19(July)**, pp. 1493–1503. doi: 10.1093/cercor/bhn187.

Spruston, Nelson (2008). "Pyramidal neurons: dendritic structure and synaptic integration". *Nature Reviews Neuroscience*, **9(3)**, pp. 206–221. ISSN 1471-003X. doi: 10.1038/nrn2286.

Tallon-Baudry, Catherine and Bertrand, Olivier (1999). "Oscillatory gamma activity in humans and its role in object representation". *Trends in Cognitive Sciences*, **3(4)**, pp. 151–162. ISSN 13646613. doi: 10.1016/S1364-6613(99)01299-1.

Tecce, J. J. and Scheff, N. M. (1969). "Attention Reduction and Suppressed Direct-Current Potentials in the Human Brain". *Science*, **164(3877)**, pp. 331–333. ISSN 0036-8075. doi: 10.1126/science.164.3877.331.

Tecce, Joseph J. (1972). "Contingent negative variation (CNV) and psychological processes in man". *Psychological Bulletin*, **77(2)**, pp. 73–108. ISSN 00332909. doi: 10.1037/h0032177.

Tecce, Joseph J.; Savignano-Bowman, June and Meinbresse, Debbie (1976). "Contingent negative variation and the distraction-arousal hypothesis". *Electroencephalography and Clinical Neurophysiology*, **41(3)**, pp. 277–286. ISSN 00134694. doi: 10.1016/0013-4694(76)90120-6.

Tombaugh, Tom N.; Kozak, Jean and Rees, Laura (1999). "Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming". *Archives of Clinical Neuropsychology*, **14(2)**, pp. 167–177. ISSN 08876177. doi: 10.1016/S0887-6177(97)00095-4.

van Rijn, Hedderik; Kononowicz, Tadeusz W.; Meck, Warren H.; Ng, Kwun Kei and Penney, Trevor B. (2011). "Contingent negative variation and its relation to time estimation: a theoretical evaluation". *Frontiers in Integrative Neuroscience*, **5(December)**, pp. 1–5. ISSN 1662-5145. doi: 10.3389/fnint.2011.00091.

Von Helmholtz, Hermann Ludwig Ferdinand (2004). "Some laws concerning the distribution of electric currents in volume conductors with applications to experiments on animal electricity". *Proceedings of the IEEE*, **92(5)**, pp. 868–870. ISSN 00189219. doi: 10.1109/JPROC.2004.826598.

Wallentin, Mikkel (2009). "Putative sex differences in verbal abilities and language cortex: A critical review". *Brain and Language*, **108(3)**, pp. 175–183. ISSN 0093934X. doi: 10.1016/j.bandl.2008.07.001.

Walter, W Grey; Cooper, R; Aldrige, V J; McCallum, W C and Winter, A L (1964). "Contingent Negative Variation : An Electric Sign of Sensori-Motor Association and Expectancy in the Human Brain". *Nature*, **203**, p. 380.

Weerts, Theodore C. and Lang, Peter J. (1973). "The effects of eye fixation and stimulus and response location on the Contingent Negative Variation (CNV)". *Biological Psychology*, **1(1)**, pp. 1–19. ISSN 03010511. doi: 10.1016/0301-0511(73)90010-0.

Weiss, Elisabeth M.; Kemmler, Georg; Deisenhammer, Eberhard A.; Fleischhacker, W. Wolfgang and Delazer, Margarete (2003). "Sex differences in cognitive functions". *Personality and Individual Differences*, **35(4)**, pp. 863–875. ISSN 01918869. doi: 10.1016/S0191-8869(02)00288-X.

Weiss, Sabine; Mueller, Horst M.; Schack, Baerbel; King, Jonathan W.; Kutas, Martha and Rappelsberger, Peter (2005). "Increased neuronal communication accompanying sentence comprehension". *International Journal of Psychophysiology*, **57(2)**, pp. 129–141. ISSN 01678760. doi: 10.1016/j.ijpsycho.2005.03.013.

Westerlund, Masha; Kastner, Itamar; Al, Meera and Pylkkänen, Liina (2015). "Brain & Language The LATL as locus of composition : MEG evidence from English and Arabic". *BRAIN AND LANGUAGE*, **141**, pp. 124–134. ISSN 0093-934X. doi: 10.1016/j.bandl.2014.12.003.

Westerlund, Masha and Pylkkänen, Liina (2014). "The role of the left anterior temporal lobe in semantic composition vs. semantic memory". *Neuropsychologia*, **57(1)**, pp. 59–70. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2014.03.001.

Yingling, Charles D and Skinner, James E (1976). "Selective regulation of thalamic sensory relay nuclei by nucleus reticularis thalami". *Electroencephalography and Clinical Neurophysiology*, **41(5)**, pp. 476–482. ISSN 00134694. doi: 10.1016/0013-4694(76)90059-6.

Zaccarella, Emiliano and Friederici, Angela D. (2015). "Merge in the human brain: A sub-region based functional investigation in the left pars opercularis". *Frontiers in Psychology*, **6(NOV)**, pp. 1–9. ISSN 16641078. doi: 10.3389/fpsyg.2015.01818.

Zhang, Linmin and Pylkkänen, Liina (2015). "NeuroImage The interplay of composition and concept speci fi city in the left anterior temporal lobe : An MEG study". *NeuroImage*, **111**, pp. 228–240. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2015.02.028.

Ziegler, Jayden and Pylkkänen, Liina (2016). "Neuropsychologia Scalar adjectives and the temporal unfolding of semantic composition : An MEG investigation". *Neuropsychologia*, **89**, pp. 161–171. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2016.06.010.

Appendix A

No-composition task

Early compositional processes elicited by grammatical expressions have been shown to be automatic and fairly independent of task and attentional demands. This is indicated by electrophysiological (Friederici and Hahne, 2002; Maidhof and Koelsch, 2011; Pulvermüller et al., 2008), hemodynamical (Caplan et al., 2008) and MEG (Shtyrov et al., 2002) experiments. Furthermore, Bemis and Pylkkänen (2011) minimal paradigm has been implemented in a study in which composition was not required to do the task. In this version of the experiment, the LATL response was maintained (Bemis and Pylkkänen, 2013c).

In this section we design a task using the same words and non-words employed during the experiment described in Chapter 3 to further explore if the activity previously obtained reflects a process of basic composition. Given the automatic nature of language processing, this activity should be elicited during a task in which participants are not required to compose.

A.1 Materials and methods

A.1.1 Participants

All participants that took part in the experiment described in Chapter 3, except for one subject, undertook another task in which composition was discouraged (28 subjects). This no-composition task was done at the end of each experimental session.

A.1.2 Stimuli and procedure

Participants were presented with pairs of stimuli in three conditions. A NA condition (noun - adjective), a AA condition (adjective - adjective) and a XA (non-word - adjective) condition. Words were presented one at a time, and in each trial subjects saw a probe word delimited by question marks. They were instructed to answer if it matched any of the words previously presented (Figure A.1). The task was comprised of 100 trials per condition (300 trials in total) and half of the trials required a 'no' answer. For the NA condition, in 25 trials the probe corresponded to the noun, and in 25 trials it matched the adjective. As for the 'no' trials, 25 were followed by an incorrect noun and 25 by an incorrect adjective. For the AA condition, in 25 trials the probe matched the adjective in the first position, and in 25 trials it matched the adjective in the second position. The remaining trials were followed by an adjective different to both words. The XA condition consisted of 50 trials in which the probe was equal to the adjective, and 50 trials in which a different adjective was shown. The XA condition was incorporated to maintain the experiment as similar as possible to the one shown in Chapter 3.

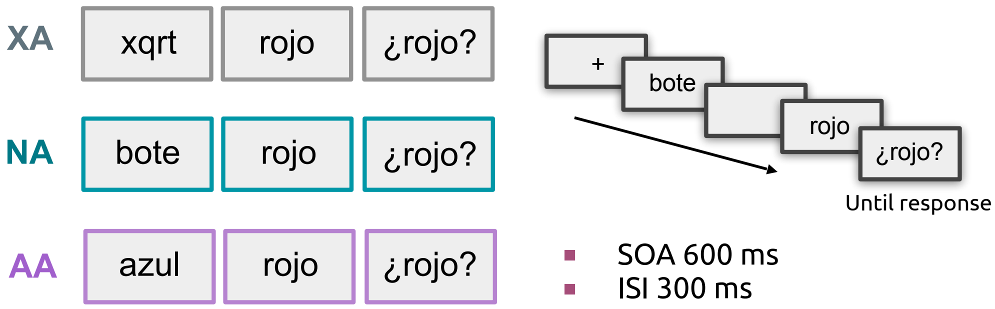


Figure A.1: No-composition task. Participants were presented with three conditions: in XA they saw a non-word followed by an adjective. In the AA condition subjects saw two adjectives and in the NA they saw a noun followed by an adjective. In each trial, after each word pairing subjects had to decide if a probe word had already been presented.

We employed the same word and non-words as described in materials and methods in Chapter 3 to construct the stimuli for each participant. During each trial participants saw a fixation cross for 1 s, and all stimuli except for the probe word were presented for 300 ms followed by a 300 ms blank screen. The probe word at the end of each trial remained on screen until subjects pressed a key or after 3 s. At the end of each trial a sound was elicited to encourage participants to blink at that time. The inter-trial interval was randomly varied between 0.8 - 1.5 s. Stimuli presentation was coded in Psychopy (Peirce, 2007) and displayed on a CRT monitor with a 60 Hz refresh rate.

A.1.3 Data preprocessing and analysis

EEG data was preprocessed and analyzed as described in Chapter 3. The number of rejected electrodes and trials was 3.36 ± 1.68 and 35.71 ± 11.74, respectively. A one-way ANOVA was conducted to test for differences in the number of trials across conditions, no difference was found $F(2, 81) = 0.66, p = 0.52$. For the no-composition task the SNRLB for subjects' ERPs ranged from 4.76 to 22.99 dB (mean = 11.76, median = 10.44 dB, SD = 5.06 dB, IQR = 5.53 dB). All 28 subjects were kept for the cluster permutation analysis.

Reaction times were analyzed by fitting a generalized linear mixed effect model according with equation A.1. The model was tested with a type two ANOVA and pairwise comparisons were obtained by emmeans R package.

$$log(RT) \sim Condition + (1|Subject) \quad \text{(A.1)}$$

A.1.4 Behavioral results

In this experiment participants performed a simple task in which they saw two words and had to report if a probe word had already appeared. As described in the methods section, a linear mixed model was fitted to the logarithm of subjects reaction times, with condition as fixed factor and subject as random effect.

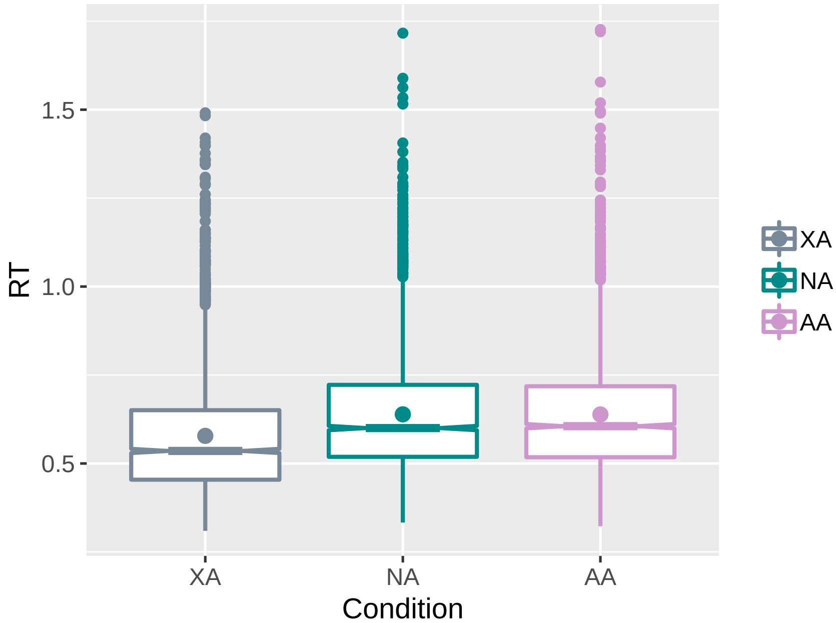


Figure A.2: Participants reaction times. In XA subjects were shown a non-word followed by an adjective. In NA condition participants saw a noun followed by an adjective, and in AA subjects were presented with two adjectives. Reaction times were measured from probe onset to subject's response

An effect of condition was found ($\chi^2(2) = 517.22, p < 0.001$). Multiple comparisons were tested between conditions. There was no difference between AA and NA conditions ($Z = -0.32, p = 1$). However, there were differences between NA and XA ($Z = 19.47, p < 0.001$), and between AA and XA ($Z = 19.82, p < 0.001$).

A.1.5 Electrophysiological results

ERPs aligned to first word onset were obtained (Figure A.3, Figure A.4) A cluster permutation analysis was carried on the epoch from -0.2 to 1.2 s to test for differences between NA and AA conditions. A positive significant cluster (p=0.018) and a negative significant cluster (p=0.003) were obtained (Figure A.5b.

A.1.6 Discussion

The objective of this task was to further asses if the activity obtained in Chapter 3 is in fact a composition-related activity. During this no-composition task, subjects were presented with three conditions comprised of two elements. The conditions NA and AA were contrasted with a cluster permutation analysis. This test revealed two significant clusters limited to latencies before second word onset and with a different topography than the potential composition activity previously obtained. Therefore, is probable that the process elicited in the first experiment does not match the one implicated here. In this task, there is no difference in expectancy across conditions, as second word is always an adjective. Furthermore, second

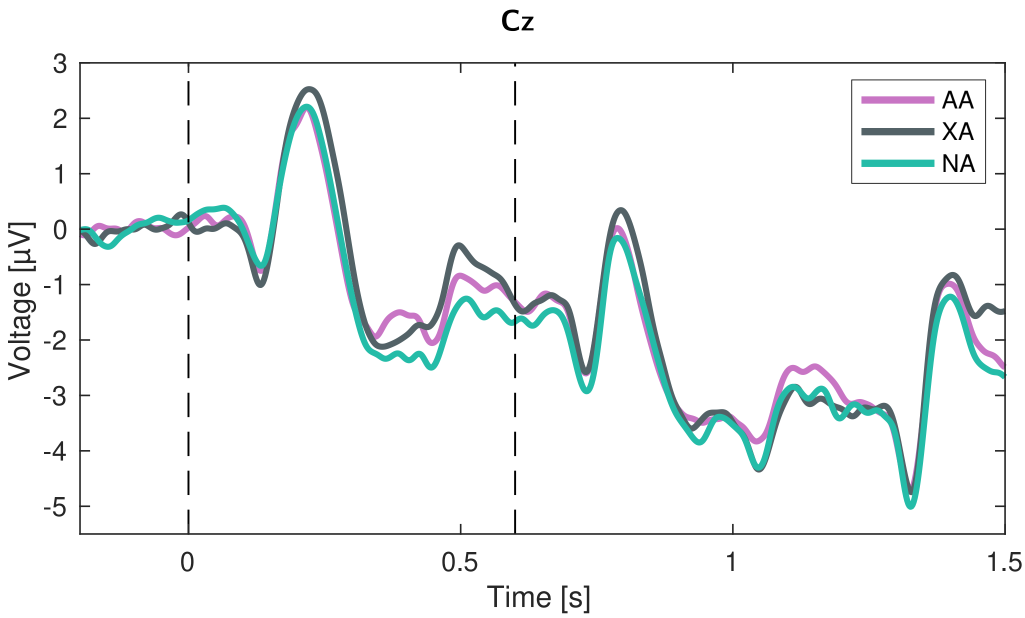


Figure A.3: ERP grand average for each condition superimposed. C1W: composition one-word condition, XA non-word - adjective condition, NA: noun - adjective condition, AA: adjective - adjective condition.

words in conditions NA and AA had to be equally handled in relation to first word to answer correctly. It is possible that the result obtained in the simple composition experiment reflects only an anticipatory process and therefore in this new task, in which expectancy is not a factor, it is not present. However, we can not exclude the alternative that we do not replicate the result previously obtained because this task forces participants to maintain both words as independent elements and therefore is extremely discouraging for a compositional process to be engaged. This is supported by reaction times showing no difference between NA and AA conditions, indicating a similar processing of both conditions. However, behavioral and electrophysiological responses have been showed to be independent: a similar behavioral result is shown in Bemis and Pylkkänen (2013c), where composition is discouraged and nonetheless combinatorial activity is reported in the LATL. Therefore, the fact that the composition-related activity obtained in Chapter 3 is not elicited in this experiment could indicate its incorrect identification in Neufeld et al. (2016) replication.

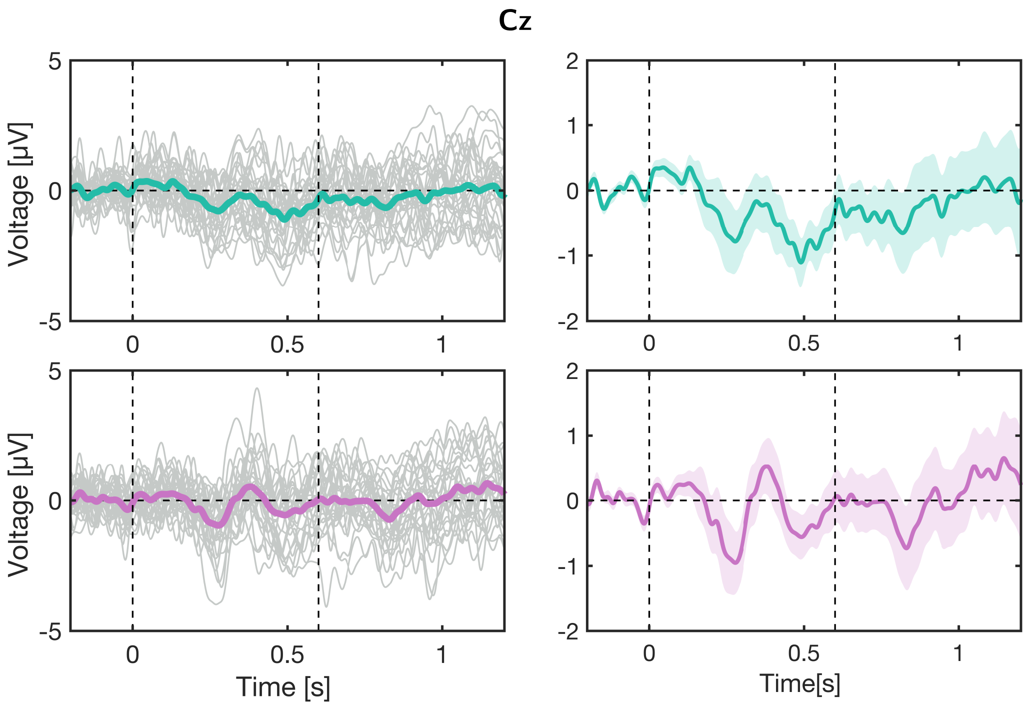


Figure A.4: Top left: Each gray stroke is the average difference between NA and XA for each participant for electrode Cz. **Bottom left**: Each gray stroke is the average difference between AA and XA for each participant for electrode Cz. **Right**: Mean of the differences between conditions with 95% confidence interval.

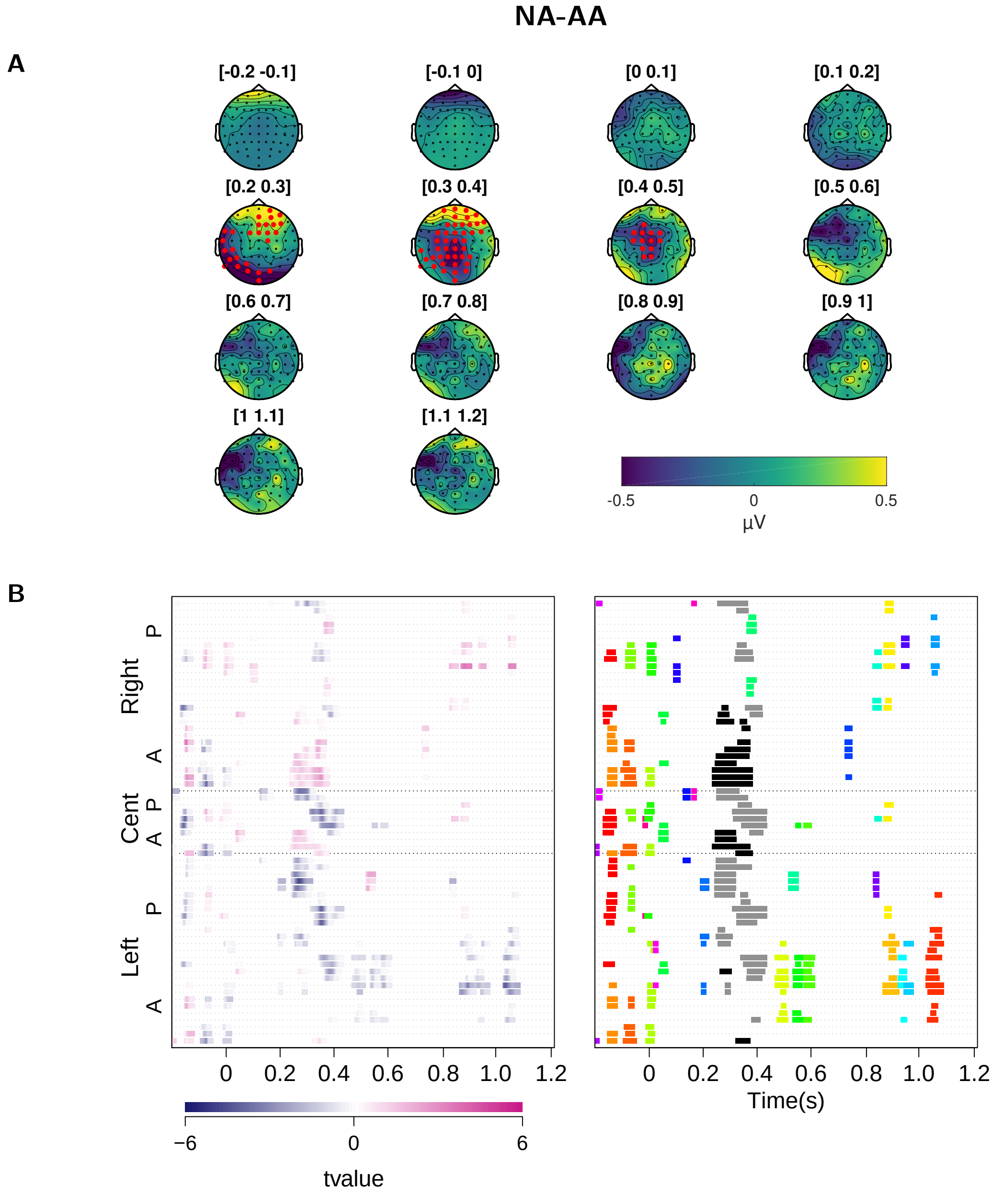


Figure A.5: A. Topography of the difference of the two-word condition minus the one-word condition between the NA and the AA condition. Red points mark electrodes that take part of the significant cluster. **B**. **Left**: t-values of the clusters obtained for the permutation cluster analysis of the interaction between XA and AA conditions. **Right**: The color code represents data points that participate in a given cluster. Grey and black colored clusters indicate the significant clusters.