Auditory and Visual ERP Correlates of Gender Agreement Processing in Dutch and Italian

Srđan Popov

Contents

List of Figures

List of Tables

Chapter 1
General Introduction

1.1 Introduction

The present thesis revolves around the processing of (linguistic) gender and gender agreement in reading/listening. It focuses on real-time processing which is why event-related potentials (ERPs) were chosen as the preferred experimental method. The studies presented here address questions that pertain to psycholinguistic issues related to gender agreement processing, as well as methodological aspects related to the use of ERPs in agreement studies. The majority of agreement studies apply the so-called 'violation paradigm' (e.g., Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004), which means presenting participants with sentences containing agreement errors (agreement mismatch). Since each experiment in this thesis contained a violation paradigm, it is more appropriate to talk about gender disagreement rather than gender agreement processing.

Two central issues act as the backbone of this work. Firstly, we study the relationship between gender disagreement and the repair and reanalysis processes as reflected by the late syntactic component (P600; Friederici, 2002; Kaan & Swaab, 2003). In the first two experiments (Chapters 2 and 3), this is achieved by comparing syntactic and semantic gender in Italian. This study is intertwined with the issue of whether semantics can influence syntactic processing and how this is reflected in the P600 component. The last two experiments (Chapters 4 and 5) also investigate the repair mechanism, but by comparing gender and number disagreement in Dutch.

The second issue investigated in the current thesis is methodological. It relates to the role of input modality (visual or auditory) and real-time processing with ERPs. Each experiment in this thesis was conducted as a reading study first (Chapters 2 and 4) and then as a listening study (Chapters 3 and 5), thus allowing us to manipulate stimulus duration, as well as the violation recognition point while using identical stimuli. By so doing, we were able to compare reading and listening in terms of the presence/absence of language-related ERP components, as well as their temporal and topographic characteristics. By combining the two research goals, we obtained a more detailed picture of the processes underlying gender agreement and their relationship to language-related ERP components.

This introduction provides a general framework for the 4 ensuing experimental chapters. It firstly introduces the technique (ERPs) and its application in the sentence processing field. Afterwards, the theoretical framework for four studies is provided. The introduction ends with the predictions stemming from such framework and with a presentation of the structure of the thesis.

1.2 Event-Related Potentials and Sentence Processing

ERPs have been employed for more than 35 years in the study of language. The first major breakthrough was the discovery of the so-called semantic processing component (N400) by Kutas and Hillyard in 1980. The component was elicited by comparing sentence minimal pairs in a violation paradigm. A minimal pair consisted of a semantically plausible sentence (*He spread the warm bread with butter*) and a semantically anomalous sentence (*He spread the warm bread with socks*; examples taken from Kutas and Hillyard, 1980). Anomalous sentences elicited a negative deflection peaking 400 ms after the stimulus onset in the centro-parietal areas.

In addition to the N400, two syntactically-related components have been identified, namely the left anterior negativity (LAN) and the P600. Both components are most often elicited by morphosyntactic violations (Friederici, Pfeifer, & Hanhne, 1993; Hagoort, Brown, & Osterhout, 1999; Molinaro, Barber, & Carreiras, 2011; Osterhout & Mobley, 1995). The LAN is understood as an automatic response to a morphosyntactic violation. It peaks 300-500 ms post-stimulus onset, and is often left-lateralized and anterior (Gunter, Friederici, & Schriefers, 2000; Hahne & Friederici, 1999; Münte, Heinze, & Mangun, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991). Still, a number of studies have shown that the LAN can have a bilateral distribution (Friederici & Frisch, 2000; Hahne & Jescheniak, 2001; Hahne &

Friederici, 2002). The LAN effect is often followed by a centro-parietal positivity peaking 600 ms after stimulus onset. The P600 is usually described as representing repair and reanalysis processes or structural integration (Friederici, Hahne, & Sadde, 2002; Gouvea, Phillips, Kazanina, & Poeppel, 2010; Osterhout & Holcomb, 1992).

A number of authors distinguish between the early left anterior negativity (eLAN) and the left anterior negativity (LAN) (e.g., Friederici, 2002; Hahne & Friederici, 1999; Hahne & Friederici, 2002). Friederici (2002) proposed an electrophysiological model of auditory sentence processing consisting of 3 phases. In the first phase (app. 100-300 ms post-stimulus onset), the eLAN is elicited in the presence of a word category violation. In the following phase, the parser becomes sensitive to lexical semantics (N400) and morphosyntactic processes (LAN). Finally, the last phase is understood as an integration phase (P600). According to the model, this is the only phase in which syntactic and semantic information are allowed to interact. Also, at this stage the morphosyntactic violation detected previously in the LAN phase has to be repaired or reanalyzed so that structural integration can proceed.

1.3 Syntactic (Grammatical) and Semantic (Biological) Gender

Gender can be described as a noun property that is reflected in the behaviour of other words (Hockett, 1958). There are two reasons for characterizing gender in this way. Firstly, it is often impossible to tell the gender of a noun by the noun's morphological composition or semantics. Instead, overt gender cues, often in the form of inflectional morphemes, are found on elements preceding (e.g., determiners) or following the noun (e.g., adjectives). Therefore, the initial definition refers to the role that gender plays in establishing a syntactic relationship between the noun and other sentential elements. Nouns, together with pronouns, are marked for three essential grammatical features: number, gender, and person. Collectively, these features are called *phi-features* (Adger & Harbour, 2008). They are the building blocks of a syntactic operation called *agree* (Pesetsky & Torrego, 2007). Pronominal elements enter the syntactic derivation with valued phi-features (e.g., feminine and singular) (e.g., Bošković, 2011; Chomsky, 2001; Pesetsky & Torrego 2007). Depending on the language, some of these features have to be morphologically realized on an element different than the noun (e.g., determiners, adjectives). This happens through an agreement relationship between the noun (target) and a subordinate element (probe), whereby feature values are transferred from the goal onto the probe (Chomsky, 2000; 2001; Pesetsky & Torrego, 2007).

When we talk about gender and gender agreement, we talk about a purely formal feature and a strictly syntactic relationship (Hagoort and Brown, 1999). Unlike other pronominal features (e.g., number), gender often does not bring any new semantic information into play. This is especially true in case of inanimate nouns (syntactic gender). As an illustration, the number value represents the numerosity of the entity in question, that is, whether there is only one entity (singular) or more than one entity (plural). Therefore, the number value is retrieved from a conceptual representation (Levelt, Roelofs, & Meyer; 1999). In the case of grammatical gender, however, the gender value cannot be inferred from semantic properties of the noun or from contextual information (Corbett, 1991; 2006). Given the absence of any generalizable factor in assigning grammatical gender, gender values for the same noun show a large variability in related languages. For example, the noun *time* is masculine in Italian (*il tempo*), feminine in German (*die Zeit*), and neuter in Serbian (*to vreme*).

However, in a small but frequent group of nouns, gender is a predictable feature, that is, it is based on semantics. These are mainly nouns referring to people, such as family relations (e.g., mother, father...), professions or functions (e.g., monk, nun), and certain animals (e.g., cow, bull) (Dahl, 2000). In these cases, gender mirrors the biological gender of the noun's referent (Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). Such gender is often referred to as 'semantic' or 'biological' gender, as opposed to 'grammatical' or 'syntactic' gender, which applies to all other cases. In the limited set of nouns with 'biological' gender, then (and only in this set) semantics can aid gender decoding (Vigliocco & Franck, 1999).

Psycholinguistic theory has suggested that people process syntactic and semantic gender differently. In order to illustrate the mechanism, we will use Levelt's speech production model (Levelt, 1989; 1992; 1999; Levelt et al., 1999; Levelt & Schriefers, 1987). In short, the model stipulates that, once it has been retrieved, a concept activates the corresponding lemma and lexeme. The lemma contains lexical semantics and lexical syntax, whereas the lexeme is the spoken word form. Crucially, noun gender information is stored as a node or diacritic at the lemma level, together with other lexical syntactic features (e.g., word class, number). Once the lemma is available, its syntactic information is used for grammatical encoding (narrow syntax) during which syntactic operations, such as agreement, are performed. After the lemma has been activated, and the grammatical encoder has prepared the necessary structure, phonological encoding can begin, which will eventually lead to articulation.

According to the model, gender is retrieved as part of the lemma. Its value is set pre-syntactically (before grammatical encoding), and is invariable. However, Vigliocco and Franck (1999) suggested that in addition to accessing gender from the gender node, the parser can also retrieve gender through conceptual information. This is only possible for the group of nouns with semantic gender. Therefore, gender in nouns with semantic gender can be retrieved via two routes: gender node (lemma) or the word's conceptual information.

A few words on the Italian gender system are required at this point. Italian has two genders: masculine and feminine. The noun agrees in gender with determiners, adjectives (both attributive and predicative), and certain verb forms (e.g., the past participle). For example, the definite article agrees with the noun in gender and number, thus, *il* and *i* are used for masculine singular and plural nouns, respectively (*il treno* 'the train', *i treni* 'the trains'). For feminine nouns, the article *la* is used in singular and *le* in plural (*la mela* 'the apple', *le mele* 'the apples'). However, these are only the default forms of the articles. The article in Italian can have different allomorphic realizations depending on the phoneme(s) the adjacent noun starts with. For example, the singular masculine article *il* is realized as *lo* if the following noun begins with *z /ts, dz/* (e.g., *lo zio* 'uncle'). For the experiments in Chapters 2 and 3, only nouns that take default article forms were used.

Interestingly, Italian is to a large extent a gender-transparent language. Almost 2/3 of all nouns end in a gender transparent vowel, that is, *-o/-i* for masculine singular/plural (*il treno*, *i treni*) and *–a/e* for feminine singular/plural nouns (*la mela*, *la mele*). The rest of the nouns end either in *-e* or in a consonant (Cacciari, 2011; D'Achille & Thornton, 2006, as cited in Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani, & Cacciari, 2015). Nouns ending in *-e* are gender-opaque, as their gender cannot be predicted from morphology (e.g., *il pesce* ' the fish', *la carne* 'the meat'). In addition, the gender of inanimate transparent nouns is invariable, in the sense that one gender morpheme cannot be substituted by another which would correspond to an existing and semantically-related word (e.g., *la sedia*$_{F(eminine)}$ 'the chair' > *il sedio*$_{M(asculine)}$ *'?'). In nouns with semantic gender, however, the gender morpheme is by rule variable. As an illustration, the stem *bambin-* signifies a child, but inflecting it with *-o* (*bambino*) marks the child as a boy, whereas *-a* (*bambina*) means the child is a girl.

1.4 Gender and Number in Dutch

Similar to Italian, the Dutch gender system consists of two genders: common and neuter. Common gender nowadays includes nouns that were masculine and feminine in earlier stages of Dutch (Van Berkum, 1996). The noun agrees with determiners and attributive adjectives in gender and number. The definite article for common nouns is *de* (*de tafel* 'the table'), whereas *het* is used with neuter nouns (e.g., *het huis* 'the house'). The article *de* is used also for both common and neuter nouns in the plural (*de tafels* 'the tables', *de huizen* 'the houses'). In addition to articles, attributive adjectives also agree with the noun, but only in an indefinite or bare noun phrase. If the adjective is preceded by a definite article, it is always inflected with *-e* (e.g., *de groen-e boom* 'the green tree'; *het groen-e huis* 'the green house'). If the article is indefinite or absent, the adjective takes *-e* for common nouns (*een grote stad* 'a big city'), and *-Ø* for neuter nouns (*een klein dorp* 'a small village'). If the noun is plural, the adjective is always inflected with *–e*, regardless of gender (*grote steden* 'big cities', *kleine dorpen* 'small villages').

Dutch nouns are usually not morphologically marked for gender (i.e., they are gender-opaque. The only exceptions are derived nouns that contain derivational suffixes associated with certain gender (e.g., diminutive nouns derived by the suffix *-(t)je* and its allomorphs are always neuter; e.g., *de tafel* 'the table' > *het tafeltje* 'the little table'). Number (plural) is always realized as a suffix. The plural suffix is either *-en* (e.g., *boek* 'book' > *boeken* 'books') or *-s* (*tafel* 'table' > *tafels* 'tables').

1.5 Goal 1: Repair and Reanalysis in Gender Disagreement

The first goal of this thesis is to investigate the relationship between structural repair/reanalysis processes and the P600. As already mentioned, the P600 represents the late syntactic stage in which information integration, as well as repair/reanalysis, takes place. Our view on repair/reanalysis processes is along the following line (see also Friederici, 1996; Hagoort, Brown, & Groothusen, 1993; Kaan & Swaab, 2003): If the parser spots a gender incongruent article in Italian (e.g., *la_F $treno_M$* 'train'), it will repair the article (*la_F $treno_M$ > il_M $treno_M$*) in order for the structure (i.e., the NP) to be further integrated. Since the repair/reanalysis mechanism is tightly tied to the P600, we assume that an increase in the repair/reanalysis complexity modulates the P600 effect in terms of its amplitude or distribution (see Hagoort, 2003a). More precisely, if one condition is more complex in terms of repair/reanalysis than the other, the P600 will either have larger amplitude or a broader distribution in the first than in the second condition. We tested this hypothesis in two languages and two different paradigms. Each paradigm was tested in both the visual and auditory modality.

As a first step, we tested article-noun disagreement in Italian in two groups of nouns: nouns with syntactic gender and nouns with semantic gender. Due to the nature of semantic gender (morphological variability: e.g., *il nonno* ‘the grandpa’, *la nonna* ‘the grandma’), the repair/reanalysis process was expected to be more complex in the semantic than in the syntactic gender violation. More precisely, in the case of syntactic gender, a violation such as *la_F $treno_M$* can only be repaired in one way – by repairing the mismatching feminine article *la* into the masculine article *il*. By contrast, a semantic gender violation offers an additional repair option: *il_M $nonna_F$ > la_F $nonna_F$ but also *il_M $nonna_F$ ‘the grandma’ > il_M $nonno_M$ ‘the grandpa’. That is, due to the morphological alternation in semantic gender, the parser may repair or reanalyze the noun instead of the article. Increased repair complexity was predicted to result in a larger P600 effect for semantic gender than syntactic gender. The findings of this study are also discussed in terms of the interaction between syntactic and semantic information during sentence processing.

In a similar vein, we compared gender and number disagreement within the Dutch noun phrase (NP). The gender feature in Dutch is lexical, as in the large majority of cases there is no word-final gender suffix that varies depending on the biological gender of the referent. In this regard, gender disagreement in Dutch resembles syntactic gender disagreement in Italian. Conversely, the number feature in Dutch is always realized as a suffix when it is plural, meaning that it varies between *-Ø* (singular) and *-en* (plural; we only tested nouns taking *-en* as the plural suffix). The number disagreement condition corresponds to the semantic gender condition in the Italian experiments, as it is a morphologically expressed feature. Similar to the Italian study, we stipulated that gender disagreement in Dutch involves a simpler repair/reanalysis mechanism, which should be reflected in the P600 effect. More precisely, there is only one repair/reanalysis option in gender disagreement (e.g., *een_{SG} $kleine_{C(ommon)}$ $dorp_{N(euter)}$ > een_{SG} $klein_N$ $dorp_N$ ‘a small village’), whereas there are two possibilities for number disagreement (*het_{SG} $kleine_{SG}$ $dorpen_{PL}$ > de_{PL} $kleine_{PL}$ $dorpen_{PL}$ & het_{SG} $kleine_{SG}$

dorpen$_{PL}$ > *het*$_{SG}$ *kleine*$_{SG}$ *dorp*$_{SG}$). Results will be discussed in terms of different underlying mechanism in gender and number (dis)agreement processing.

1.6 Goal 2: The Role of the Presentation Modality

Sentence processing, and especially agreement processing, has been studied mostly by the means of reading experiments (see Molinaro et al., 2011). In these studies, the stimuli, which often consist of entire sentences, are presented word-by-word. In this way, the experimenter is able to control the presentation rate, as well as the general time course of the experiment. This enables the experimenter to pinpoint the exact onset time of each word, thus helping disentangle the effects caused by the target word from those caused by the target word′s neighborhood. This is almost impossible in an auditory experiment, because of the rate at which speech is delivered. Unless the speech signal is significantly modified, the natural speech rate does not allow the experimenter to deliver words and pauses at the same rate as in reading. All in all, reading grants a better control over the experiment at the expense of creating an artificial environment (e.g., the average reading rate in a cognitively intact reader is faster than the average rate of 600 ms/word in a reading experiment). Conversely, listening enables a more ecological stimulus presentation, but at the same time, it allows a larger overlap of ERP components elicited by adjacent words, as there is usually no programmed break between words.

The time-course issue is the reason why we decided to run each experiment both as a reading and as a listening study. The greatest strength of ERPs is their excellent temporal resolution. ERPs allow collecting a new data point at the rate of 1 ms or even faster (Luck, 2005). Paradoxically, the exquisite temporal power of ERPs is often not fully exploited in sentence processing studies. This is understandable in the case of reading studies. When a written stimulus is presented, the participant fixates an entire word in approximately 200-300 ms (Rayner, 1998), after which post-lexical language processing (i.e., the kind of syntactic and semantic processing associated with the P600 and N400, respectively) can commence (Sereno, Rayner, & Posner, 1998; Sereno & Rayner, 2003). Since there is little variability in the time it takes to perceive the physical (visual) stimulus in the form of a written word compared to a spoken word (Rayner & Clifton, 2009), the language components are expected to peak at approximately the same time across studies. But, what happens when the stimulus is delivered in chunks (phonemes) and at the rate in which it would be delivered in natural speech? For example, a one-syllable written word is fixated in the first 200 ms from its visual onset and the duration of the same word when spoken is approximately 200 ms. Therefore, ERP responses are expected to be more or less identical in both modalities. However, an average three-syllable word may last 500 to 600 ms when presented in the auditory modality, whereas its visual recognition will still take only 200 to 300 ms. The question asked here is what happens during these 303-400 ms that are 'lost' when the stimuli are presented only visually. In particular, we are interested in finding out how this 'delay' affects language-related components, for example, their onset time.

As regards the timing issue, we are particularly interested in testing whether the language-related components are violation-aligned. Hagoort and Brown (2000) noticed that the onset of the P600 is remarkably stable in both reading and listening, peaking at approximately 600 ms post-stimulus onset. Such finding can be interpreted in two ways. The P600 is a cyclical process that peaks every 600 ms after a certain moment in time, for example, from the onset of each word in the sentence. Speculatively, this would mean that the parser allows for a new round of revision 500-600 ms after detecting each new lexical input. However, it is equally plausible that the 600 ms time mark is the result of averaging stimuli of various length. If the average duration of the stimuli were approximately 300 ms, the violation would be perceived at approximately the same time in both reading and listening. In this case, the onset of the P600 would be violation-aligned, meaning that the parser starts the revision process depending on when it detects the violation.

So far, no study has tested explicitly whether or not the P600 is violation aligned. A study by O'Rourke and Holcomb (2002) demonstrated that the N400 correlates with the word's recognition point, meaning that the component is aligned to the violation/recognition point. A way to test whether syntactically-related components are violation aligned would be to test two conditions in which the violation points differ considerably in time. For example, in our Italian experiment, syntactic gender violations are recognized before semantic gender violations with a difference of approximately 100 ms. In the Dutch experiment, the difference between the conditions is even larger, approximately 200 ms. This allows ample time to investigate these issues.

The P600 is expected in both studies regardless of input modality. It has been demonstrated that input modality does not affect the elicitation of the P600 (e.g., Hagoort & Brown, 2000). Unlike the P600, the LAN has proven to be more volatile and more difficult to elicit. Whereas the P600 is almost universally present in agreement studies, the LAN has been recorded unsystematically. For example, Barber and Carreiras (2005) elicited the LAN followed by the P600 in gender disagreement in Spanish. However, in another very similar study on Spanish, Wicha, Moreno, and Kutas (2004) failed to elicit the LAN. So far, different authors have offered different psycholinguistic and methodological explanations to account for the volatility of the LAN (see Hagoort and Brown, 1999; Molinaro, Barber, Caffarra, & Carreiras, 2014; Osterhout, 1997). However, none of them have managed to satisfactorily capture the entire issue.

The scope of this study does not include solving the LAN puzzle. Rather, our goal is to provide a satisfactory account irrespective of whether or not we manage to elicit the LAN. If we rely on the auditory sentence comprehension model (Friederici, 2002), we expect to obtain the LAN in both listening experiments. Then, based on the logic that the syntactic processing mechanism is similar in listening and reading, and that this mechanism is reflected in the ERP components, we expect to elicit identical components in listening and reading. However, Hagoort and Brown (2000) demonstrated that the latter is not necessarily true. In a series of experiments in Dutch, the LAN was elicited in auditory but not in visual studies. In the event of conflicting results regarding the LAN in reading and listening, outcomes will be addressed in terms of the temporal difference between these two modalities.

1.7 Predictions and Hypotheses

The current study deals with the issue of (dis)agreement processing. Our first goal is to investigate which processes and which components are elicited during (dis)agreement processing. A summary of expected components for each chapter is given in Table 1.1 at the end of this Section.

<u>Hypothesis 1</u>: (Dis)agreement is processed at the syntactic level. Therefore, it elicits the P600 and LAN.

Prediction 1: Each experiment, being centered around gender disagreement, will elicit the P600. We make no firm predictions about the LAN, due to its volatile nature in agreement studies.

The first prediction assumes the presence of the P600. Since this component is usually described as a marker of repair/reanalysis processing, we are interested in finding out more about how an increase in repair/reanalysis complexity affects the P600 component.

Hypothesis 2: The P600 is a marker of structural repair/reanalysis.

Prediction 2: As such, it is modulated by the complexity of the repair/reanalysis mechanism. Conditions that require more complex repair/reanalysis processing (semantic gender, number disagreement vs. syntactic gender, gender disagreement) are expected to elicit larger P600 effects, in terms of either increased amplitude or broader distribution. These effects should not depend on the presentation modality.

Regarding the role of presentation modality, we are interested in investigating how the difference in timing between the auditory and visual presentation affects language-related components. We assume that the recognition point of the violation is constant in reading studies due to the nature of reading (e.g., one fixation every 200 ms), thus allowing establishing pre-determined time windows in which the LAN (300-500 ms) and the P600 (500 ms on) are obtained. By contrast, the violation recognition point in listening depends on word length and on the position of the violation within the word.

Hypothesis 3a: The P600 and the LAN are violation-aligned.

Prediction 3a: The onset of the P600 and the LAN is tied to the moment in time when the parser detects the morphosyntactic violation. Therefore, the onset of the auditory components will vary depending on word length and on the violation recognition point within the word.

Hypothesis 3b: The P600 and the LAN are not violation-aligned.

Prediction 3b: The onset of the LAN and P600 will be the same for all conditions, irrespective of word length and of the position of the violation within the word.

Lastly, in addition to the temporal characteristics of the components, we investigate whether the presence or absence of a component is modality-dependent.

Hypothesis 4: Morphosyntactic violations are processed similarly regardless of presentation modality.

Prediction 4: The same components should be elicited when stimuli are presented visually and auditorily. In the event of discrepancy between reading and listening, particularly as regards the presence/absence of the LAN, this is most likely related to the difference in the time course between reading and listening.

Table 1.1: An overview of expected ERP components per chapter

Italian			
Chapter 1		**Chapter 2**	
Reading		**Listening**	
Syntactic gender	**Semantic gender**	**Syntactic gender**	**Semantic gender**
LAN?, P600	**LAN?, P600**	**LAN?, P600**	**LAN?, P600**
P600 < P600		**P600 < P600**	
larger P600 effect expected for the semantic gender condition			
Dutch			
Chapter 3		**Chapter 4**	
Reading		**Listening**	
Gender	**Number**	**Gender**	**Number**
LAN?, P600	**LAN?, P600**	**LAN?, P600**	**LAN?, P600**
P600 < P600		**P600 < P600**	
larger P600 effect for the number condition			

1.8 Structure

The experimental part of the thesis opens with Chapter 2, a reading study on semantic and syntactic gender processing in Italian. In Chapter 3 the same stimuli as in the previous chapter are used in a listening experiment. Chapter 4 and Chapter 5 present a gender and number (dis)agreement study in Dutch in the reading and listening modality, respectively. The closing Chapter 6 discusses the results from the experimental chapters and relates them to our predictions, as well as to current research in the field.

Chapter 2
Syntactic and Semantic Gender Processing in Reading: An ERP Study on Italian

2.1 Introduction

2.1.1 Gender Cues

Gender processing is a complex process spanning over a number of available linguistic cues while simultaneously integrating them (for production, see Vigliocco & Franck, 1999 and Vigliocco & Hartsuiker, 2002). During sentence comprehension, gender can be decoded in two ways: it is either deduced from an agreeing element (probe) or from the noun itself (goal). The distinction is crucial as only the goal is inherently (lexically) marked for gender. Once the target enters into an agreement relationship, the value of the gender feature gets copied onto the probe (Bošković, 2011; Pesetsky & Torrego, 2007). Such an agreement relationship is usually morphologically marked as a gender suffix on the probe. However, at the noun level, which is the source of gender information, gender can be represented at multiple levels as: syntactic information (lemma), morphological information (lexeme) or even semantic information (Vigliocco and Franck, 1999). The current study investigates how differences in gender encoding affect real-time gender processing by means of event-related potentials (ERPs).

Unlike number and person, grammatical or syntactic gender does not bring any additional semantic or pragmatic information into context. Therefore, it could be described as semantically void. In turn, this means that its value (e.g., feminine or masculine) cannot be deduced on its own. One reliable strategy to determine the gender value is to look at the probe. A probe is an element (e.g., adjective, verb, determiner), which is in an agreement relationship with the target (noun). This relationship is usually morphologically marked on the probe. Thus, the very first reliable gender cue is morphological. This is also the only gender cue available on an element, which is not a noun.

Whereas probes have to be morphologically marked for gender, nouns may or may not be inflected for gender. For example, the overwhelming majority of Italian nouns ending in *-a* and *-o* are feminine and masculine, respectively, thus being gender-transparent. Still, almost 1/3 of all Italian nouns end in *-e* (Cacciari, 2011; D'Achille & Thornton, 2006, as cited in Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani, & Cacciari, 2015), and are thus gender-opaque (e.g., the noun *carne* 'meat' is feminine, while *pesce* 'fish' is masculine). Going one step further with morphological decomposition, it is often possible to interpret gender from a lexical suffix used for word derivation. In a number of languages, there is a one-to-one correspondence between a derivational suffix and gender (e.g., Hickey, 1999). For example, all Italian nouns ending in *–zione* are feminine, despite the final *-e* being an opaque gender marking.

Whereas morphology can be a salient cue, it is not always enough, as shown above for Italian. Regardless of any overt cues, gender information is first and foremost an intrinsic lexical syntactic feature stored as a lexical representation at the lemma level (Levelt, 1989; 1999; Levelt, Roelofs, & Meyer 1999; Vigliocco, Antonini, & Garrett, 1997; but see Miozzo & Caramazza, 1997 for an alternative account). Everything being equal, gender is often overrepresented; it is always retrievable from the lemma making morphological noun marking redundant in case of nouns.

Finally, there is one more gender cue, available only with a limited number of nouns. Nouns denoting people, professions, and some animals have real world referents that have biological sex (e.g., man and woman). This extra-linguistic feature is often preserved in the gender system of a language, meaning that if the referent is female in the real world, it will have feminine gender (Corbett, 1991). This correspondence between biological sex and lexical gender represents a very reliable cue, with rarely any exceptions. Such gender is usually called 'biological' or 'semantic gender'.

The purpose of this study is twofold. The first goal is to investigate the role of semantic information in nouns with semantic gender. Since this issue is at the interface of morphosyntactic and semantic processing, ERPs seem to be the best method to tackle it. So far, research on semantic gender processing and ERPs has produced conflicting results. We will try to resolve the issue of whether semantic gender retrieval relies on semantic information by measuring the effect on the source of gender information – the noun.

2.1.2 ERP Gender Agreement Studies

Like a vast majority of ERP studies in the field of agreement processing (e.g., Hagoort, Brown, & Groothusen, 1993; Kutas and Hillyard, 1980; Friederici, Steinhauer, & Frisch, 1997; Osterhout & Nicol, 1999), the current one also employs a widely used violation paradigm, whereby instances of grammatical and ungrammatical usage of gender are compared. Such a comparison yields well-documented ERP components, namely N400, LAN, and P600, which are believed to reflect underlying language-related processes. The second goal of the study relates to the last processing stage (P600) usually associated with repair and reanalysis processes (e.g., Friederici, 1995; 2002; Friederici & Jacobsen, 1999; Friederici & Meyer, 2004; Kaan, Harris, Gibson, & Holcomb, 2000; Molinaro, Vespignani, Zamparelli, & Job, 2011). We will investigate in depth whether nouns marked for semantic gender are repaired in a different way than nouns marked for syntactic gender.

In order to process gender, the parser has to apply a number of operations. To reiterate, the source of the gender information is the noun. Gender is part of the lexical syntactic information (lemma) of that noun. Put broadly, there are two possible scenarios regarding gender information: 1) a probe (e.g., articles, demonstratives, and adjectives) precedes a noun (e.g., la_F $sedia_F$ ‘the chair’); 2) a noun is followed by a probe (e.g., verbs and adjectives; *La* $ragazza_F$ *è* $bella_F$. ‘The girl is beautiful.’). In the former scenario, the violation is recognized on the noun, and in the latter, on the probe. When grammatical information is first accessed from a probe (scenario 1), it has to be decoded from the probe’s inflectional morphology. After that, the gender information is carried onto the noun. Once the noun becomes available, the parser accesses the noun’s lemma and/or morphological information and evaluates it against the previously retrieved gender information from the probe. In the second scenario, the processes are the same, just in reverse. Once the two types of gender information are available, they are checked against each other. In case of a mismatch (violated sentences), most studies report a biphasic pattern of LAN and P600 (e.g., Barber & Carreiras, 2005; Barber, Salillas, & Carreiras, 2004; Gunter, Friederici, & Schriefers, 2000; Molinaro, Vespignani, & Job, 2008). This is the expected outcome, taking into account that both LAN and P600 are usually labeled as syntactically-related ERP responses. Since gender processing consists of either morphological or lexical syntax decoding, it is safe to label gender processing as a syntactic process (Hagoort & Brown, 1999).

The presence of the LAN is explained as a marker of gender violation (e.g., Friederici, 2002; Molinaro, Barber, Caffarra, & Carreiras, 2014). That is, the LAN arises as an automatic response to an agreement mismatch, typically 300-500 ms post-stimulus onset. It is followed by a centro-parietal positive shift (P600), from 500 ms on. The P600 is assumed to represent integration difficulty in the form of repair and reanalysis (Friederici, 1995; 2002; Friederici & Jacobsen, 1999; Friederici & Meyer, 2004; Kaan et al., 2000, Molinaro, Vespignani et al., 2011). This is the stage at which the parser tries to repair the syntactic incongruity and integrate it with the rest of the discourse at the structural level.

Still, there is a noticeable disagreement regarding the LAN in the agreement processing literature. Unlike the P600, which is almost unanimously reported in the literature, the LAN seems to be somewhat more volatile (for a review, see Molinaro, Barber, & Carreiras, 2011). For example, in an almost identical paradigm measuring determiner-noun gender agreement in Spanish, Barber and Carreiras (2005) reported the LAN followed by the P600, whereas Wicha, Moreno, and Kutas (2004) reported only the P600. In addition, studies on gender agreement in Italian (e.g., Caffarra et al., 2015; Molinaro et al, 2008) almost always confirm the presence of the LAN, whereas the effect seems to be absent in Dutch (Hagoort & Brown, 1999). The underlying cause may be language specific, that is, it may depend on the way a language encodes the gender feature (Friederici & Weissenborn, 2007). The large majority of Italian nouns end in a gender marking suffix, whereas Dutch nouns are gender opaque, unless they contain a gender specific derivational suffix. Hagoort and Brown (1999) suggested that the LAN is sensitive only to phonologically overt inflectional morphology, which explains the difference between Dutch and Italian. However, such an explanation cannot account for both the absence of the LAN in Spanish as reported by Wicha et al. (2004) and its presence as reported by Barber and Carreiras (2005).

2.1.3 Syntactic and Semantic Gender Processing

Due to some variability in ERP findings on gender processing (e.g., presence or absence of the LAN), researchers started exploring how different gender cues affect gender processing. The most obvious example is the contrast between gender-transparent and gender-opaque nouns. As the former are always marked with a gender suffix, lemma gender retrieval is practically redundant. In the case of nouns with opaque gender, the only way to access the gender feature is through the lemma. Caffarra et al. (2015) compared the processing of gender transparent and gender opaque nouns in Italian. As mentioned, nouns ending in *-o* and *-a* are masculine and feminine, respectively, with very few exceptions. This means that simple morphological decomposition is enough for successful gender retrieval. However, there are few Italian nouns whose gender does not correspond to the suffix (e.g., *la mano*$_F$ 'hand'). For such gender opaque nouns, the only way to retrieve gender is from the lemma. The study showed no difference in processing transparent and opaque gender in a gender mismatch condition, with each condition eliciting identical LAN and P600 effects. Still, when only grammatical instances were compared, a difference in the waveform between transparent and opaque gender was observed. It is possible that the LAN and P600 are attuned to certain processes, such as morphosyntactic violation detection and repair, without reflecting the fine-grained difference between morphological and lemma gender mismatch.

The LAN and P600 may lack power to discern the exact underlying language processes, but still they point to general syntactic processing. The question asked in this study is what happens when a semantic component interacts with syntactic processes. Measuring effects in semantic gender processing should provide an answer. As mentioned, nouns with semantic gender have real-world referents whose biological sex corresponds to their language gender. Therefore, it should be theoretically possible to draw on the semantic information in order to infer the noun's gender. In a behavioral production study, Vigliocco and Franck (1999) and Vigliocco and Zilli (1999) demonstrated that nouns with semantic gender cause both non-brain-damaged participants and people with aphasia to produce fewer errors. Therefore, the authors proposed a dual-route model for nouns with semantic gender, according to which the gender feature can be retrieved both from the lemma and semantic information. For all other nouns, the only route is through the lemma. It is important to highlight that this model is based on production. Regardless, a similar idea of an extra semantic route has been applied in a number of ERP comprehension studies on semantic gender (Barber et al., 2004; Deutsch &

Bentin, 2001; Hammer, Jansma, Lamers, & Münte, 2005; Osterhout, Bersick, & McLaughlin, 1997). The expectations in these studies were that the N400 reflects the activation of the semantic information in semantic gender access. Typically, the N400 indicates difficulty in the integration in a wider semantic and discourse context, and is usually related to semantic processing in general (Kutas & Federmeier, 2011).

The results of the previous studies on semantic gender are not clear-cut. Deutsch and Bentin (2001) studied subject-verb agreement in Hebrew, in which the subject was either animate or inanimate. They reported the N400 only for the animate stimuli, whereas violation of inanimate nouns elicited the P600. The presence of the N400 was interpreted as a hallmark of semantic processing as it was present only with animate nouns. In another subject-predicate agreement processing study, Barber et al. (2004) manipulated the gender type of the subject in Spanish. Based on the results from Deutsch and Bentin's study, Barber et al. expected that gender disagreement should cause the LAN and P600 for nouns with syntactic gender, and the N400 for nouns with semantic gender. Interestingly, both conditions elicited identical effects: LAN and P600. According to the authors, gender processing in general is a syntactic process. The N400 is obtained for outright semantic violations: something that semantic gender violation is not. The different results, as compared to Deutsch and Bentin's study, were explained by a typological distance between Hebrew and Spanish.

Similarly, two studies on pronoun reference processing reported contradictory results regarding the role of semantic gender. Schmitt, Bernadette, Lamers, and Münte (2002) manipulated the pronoun's antecedent in German. In their first condition, the antecedent was an animate referent marked for semantic gender (e.g., der Bub_{M} 'the boy'), while the second condition contained diminutive versions of the nouns. In German, diminutives are derived through suffixation, during which all nouns are assigned neuter gender regardless of the noun's semantic information (e.g., das_{N} $\text{Bübchen}_{\text{N}}$ 'little boy'). In the violated non-diminutive condition, both the N400 and P600 were reported, whereas only the P600 was attested in the violated diminutive condition. The authors concluded that establishing pronoun reference is a syntactic process eliciting the P600 in both conditions. In addition, if the antecedent noun is marked for semantic gender, semantic processes also get activated, which is reflected in the presence of the N400. These findings were challenged by a different pronoun study in German in which no effect of semantic gender was detected (Hammer et al., 2005). Both antecedents with grammatical and semantic gender elicited the P600. There was no N400 associated with the semantic gender condition.

2.1.4 Expectations and Predictions

It is obvious from the previous studies that the issue of the role of semantic information in gender processing has not been resolved yet. For each study that reports the N400 for semantic gender there is a comparable study that failed to find the same effect. However, all these studies have one thing in common: they measure the violation at the probe (verb, adjective, or pronoun). The source of gender information is always the noun. Once this feature is transmitted onto a probe, its origin is most often irrelevant for any further computation. Therefore, if the aim is to find out whether semantics is activated and engaged during gender processing, the effect should be measured at the noun. That is the novelty of the current study: we created a mismatch between an article followed by a noun, both embedded in a sentence context. This way, we are able to measure gender processing right at the source.

Caffarra et al. (2015) showed that ERPs and the accompanying violation paradigm might not be able to discern whether gender is retrieved from the lemma, morphology, or a combination of the two. Based on their finding, we may assume that the LAN is sensitive to morphosyntactic violation detection, without taking into account how that violation came to be (through morphology, the lemma or semantics). If this assumption is correct, we expect the LAN to be present in semantic gender. Everything being equal, gender disagreement causes a violation that should be legible only at the syntactic level. An open question is whether such a violation can still be legible to the semantic system in case of semantic gender. If this is the case, we should get the N400 instead of the LAN.

The N400 is not the only component that can indicate semantic processes. Almost all agreement studies report the presence of the P600. This component is most often tied to syntactic processing and represents repair and reanalysis processes. However, Gunter et al. (2000) claim that semantic processes can modulate late syntactic processing. This is compatible with Friederici's (2002) model on auditory sentence comprehension in which syntactic and semantic processes are held separate until 500 ms, after which they can interact.

Based on the possibility that syntactic and semantic information interact during late processing stages, we predict that the semantic gender violations will elicit a larger P600 effect, be it in terms of amplitude or distribution (Otten & Rugg, 2004). We expect that violations of nouns with semantic gender cause an increase in the integration load primarily because they are more complex to repair. Let us assume that repair processes function in the following way. The parser encounters a gender mismatch il_M $sedia_F$ 'the chair'. In the P600 time window, possibly in its late stage (Hagoort & Brown, 2000), the parser tries to repair the incongruity, thus turning the incongruous masculine article *il* into the correct feminine form *la*. The parser can only manipulate the article, since the -*a* suffix on the noun is invariable, which is true for all nouns with syntactic gender. The parser cannot repair the noun as *il *sedio*. Yet, this process is a viable option for nouns with semantic gender. If the parser encounters a noun such as *il bambina* 'the girl', it has two options: 1) to repair the article, *il* > *la bambina*; 2) to repair the noun, *il bambina* 'the girl' > *il bambino* 'the boy'. This is possible thanks to the masculine/feminine dichotomy that exists with nouns with semantic gender and their real-world referents. In other words, the semantic information in semantic gender licenses an extra repair process, which should be reflected in an increase of the P600 effect.

2.1.5 Goals

The first goal of the study is to assess the influence of semantic information in processing semantic gender. Behavioral production studies (Vigliocco & Franck, 1999; Vigliocco & Zilli, 1999) indicate that semantic gender can be retrieved through two routes: lexical syntax (lemma) and semantics. We have tried to find evidence for the dual-route gender access in comprehension by using ERPs. The first and most salient indication of semantic processing would be the presence of the N400. This component represents difficulty in semantic integration (Kutas & Federmeier, 2011) and is most often labeled as a marker of semantic processing. Its characteristics are a centro-parietal negative deflection, with a slight right hemisphere bias, peaking at approximately 400 ms. However, if a violation is morphosyntactic in its nature, a different negative component is expected, that is, the LAN (e.g., Friederici, 2002; Molinaro et al., 2014). Unlike the N400, the LAN's distribution is usually anterior and left-lateralized. It is present in the time window of 300 to 500 ms, thus coinciding with the timing of the N400, yet with a more frontal and left-lateralized distribution.

2.2 Method

2.2.1 Participants

For this experiment, 25 native speakers of Italian (17 male, mean age 23.4, range 19-33) were recruited. They were all right handed, which was confirmed with the Edinburgh Handedness Questionnaire (Oldfield, 1971). No participant reported any history of neurological or psychiatric disease. They all had normal or corrected-to-normal vision.

2.2.2 Materials

The experimental materials consisted of 160 sentences, half of which contained nouns marked for semantic gender (semantic condition), and the other half nouns marked with syntactic gender (syntactic condition). The frequency of the two noun types was matched (semantic gender average LogF: 2.08, syntactic gender average LogF: 2.32; $t(38) = 1.33$, $p = 0.19$). Each noun was used four times as: 1) singular grammatical; 2) singular ungrammatical; 3) plural grammatical; 4) plural ungrammatical. Finally, gender of the nouns was balanced, with half of all nouns being masculine, and half feminine.

As a consequence of a balanced design (each noun used as singular and plural, grammatical and ungrammatical), the participant would have been exposed to the same noun 4 times. In order to minimize repetition, the experimental items were divided between 2 lists. This way, every participant saw each noun only twice, once as singular and once as plural. Also, the noun was used once in a grammatical and once in an ungrammatical context. The list design decreased the experiment duration from 40 to 20 minutes.

Each subject listened to 80 experimental and 80 filler sentences. Half of the experimental sentences contained nouns marked with semantic gender and the other half nouns marked with syntactic gender. Each sentence started with a temporal adverb (e.g., often, yesterday, rarely), followed by a definite article, which was followed by a noun. The target noun phrase was always in the subject position. Depending on the gender and number of the noun, the article took one of the 4 forms: *il* masculine singular; *I* masculine plural; *la* feminine singular; *le* feminine plural. Italian articles can have different morphological realizations depending on the phonological context, but we opted for nouns taking the default article forms (see Chapter 1). For the rest of the sentence, the noun was followed by the verb, and then either by an object or an adverbial. Half of the experimental sentences had a morphosyntactic violation in the form of gender disagreement between the article and the noun. For example, a masculine singular noun, such as *nonno* 'grandfather', was preceded by the feminine singular article *la*, resulting in an ungrammatical sentence. Example sentences for each condition are given in (1) and (2).

(1) Semantic gender

Spesso	il	nonno	fumava	la	pipa.
often	the$_{SG.M}$	grandfather$_{SG.M}$	smoked	the	pipe

'The grandfather often smoked his pipe.'

*Spesso	la	nonno	fumava	la	pipa.
often	the$_{SG.F}$	grandfather$_{SG.M}$	smoked	the	pipe

'The grandfather often smoked his pipe.'

(2) Syntactic gender

Generalmente	il	treno	parte	in ritardo.
usually	the$_{SG.M}$	train$_{SG.M}$	leaves	in late

'The train usually leaves late.'

*Generalmente	la	treno	parte	in ritardo.
usually	the$_{SG.F}$	train$_{SG.M}$	leaves	in late

'The train usually leaves late.'

There were 80 fillers per list (3). Each filler item was presented once as grammatical and once as ungrammatical. The ungrammatical sentence lacked subject-verb agreement, having the predicate in infinitive. Thus, each participant heard 80 grammatical and 80 ungrammatical sentences in total.

(3) Le scale lassú sono troppo ripide per me.
the stairs up there are too steep for me.
'The stairs up there are too steep for me.'

*Le scale lassú **essere** troppo ripide per me.
the stairs up there **be** too steep for me.

2.2.3 Procedure

The participants were seated at a comfortable distance from the screen (app. 80 cm) while the EEG was being recorded. Their task was to read the presented sentences carefully and answer the grammaticality judgment question after every sentence. The presentation software used was E-Prime 2.0 (Psychology Software Tools, Inc.). Before the experiment started, the experimenter explained in detail what the task was and provided a few examples. The instructions were presented on the screen, followed by 5 practice sentences. The participants were repeatedly asked if they needed additional clarification before the actual experiment started.

Sentences were presented word-by-word, white letters on a black background. The font used was Arial, and the size was 24 pt. Each stimulus started with a fixation cross (500 ms), followed by blank screen for 300 ms. Each word was presented for 300 ms and was followed by a 300 ms break (black screen). The last word always appeared with a full stop. After the last word, the screen remained blank for 1 second, after which a question mark appeared. The question mark was a cue for the participant to press the appropriate ("p" or "q") button depending on whether the sentence was grammatical or not. The order of the buttons was counterbalanced throughout the experiment. There were 4 blocks, each containing 40 sentences, with a break after each block. Recording took 20 minutes on average.

2.2.4 EEG Recording and Data Processing

The continuous electroencephalogram was recorded from 64 Ag/AgCl scalp electrodes (WaveGuard) using the ASA-Lab system (ANT Neuro Inc, Enschede, The Netherlands). Additional bipolar electrodes were used to record horizontal (HEOG; at the outer canthus of each eye) and vertical (VEOG; above and below the left eye) eye movements. Impedances were kept below 10 kΩ. Data were acquired at 512 Hz sampling rate with the common average reference.

The offline processing was carried out in Brain Vision Analyzer 2.0.4 software (Brain Products, GmbH, Munich, Germany). The data were down-sampled to 256 Hz before being re-referenced to the average of the mastoids. Offline filtering was performed using a band-pass filter (0.1-40 Hz), followed by automatic PCA-based eye-blink correction. The data were segmented into epochs starting 200 ms before the onset of the critical word (the noun) and lasting until 1500 ms post-word onset. Only trials that were judged correctly were segmented. The (automatic) artifact rejection (+/- 100 µV threshold) was performed only on a section of each epoch (-200-1000 ms) that was used in the statistical analysis. The data were corrected relative to the 200 ms prestimulus baseline.

Finally, data were averaged per subject and per condition. No participants were excluded due to excessive noise or artefacts, as they were all above the threshold of 60% of averaged trials in all conditions.

2.2.5 Analysis

Averaged values (in µV) were extracted per participant, per condition, and per region of interest. The scalp electrodes were divided into 9 regions of interest, each containing either 5 or 6 electrodes (Fig. 2.1): left anterior (F7, F5, F3, FC3, FC5), midline anterior (F1, Fz, F2, FC1, FCz, FC2), right anterior (F4, F6, F8, FC4, FC6), left central (TP7, C5, C3, CP5, CP3), midline central (C1, Cz, C2, CP1, CPz, CP2), right central (C4, C6, CP4, CP6, TP8), left posterior (P7, P5, P3, PO7, PO5, O1), midline posterior (P1, Pz, P2, PO3, POz, PO4), and right posterior (P4, P6, P8, PO6, PO8, O2). The analysis was carried out in three independent time windows, from 300 to 450 ms (roughly corresponding to the LAN and N400), followed by the 500-700 ms window (early P600), and ending with the 700-900 ms window (late P600).

For the statistical analysis, repeated measure ANOVAs were used with the following within subject factors: condition (2 levels: syntactic and semantic gender), grammaticality (2 levels: grammatical and ungrammatical), laterality (2 levels: left and right hemisphere), and anteriority (3 levels: anterior, central, and posterior). The significance level was set to p < .05. For each time window, 2 global repeated measures ANOVAs were performed; first for the lateral regions (all factors included), and then for the midline regions (factor laterality excluded). Follow-up ANOVAs were applied to those interactions that turned out at least marginally significant (p < .1), and that included the factor grammaticality. In case the assumption of sphericity was violated, the Geisser and Greenhouse (1959) correction was applied. Lastly, only correctly judged trials were included in the analysis.

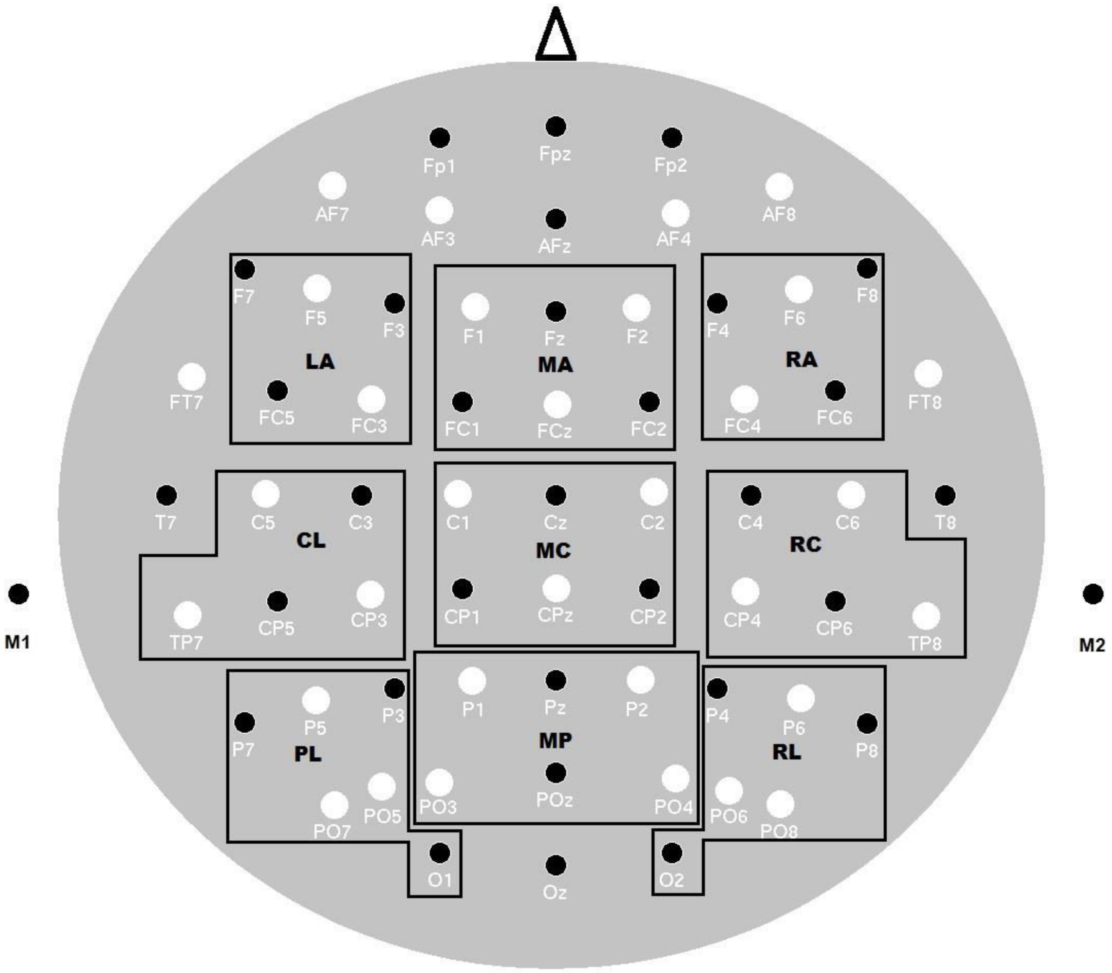

Fig. 2.1 Electrode positions and the 9 regions of interest used in the analysis: left anterior (LA), left central (LC), left posterior (LP), midline anterior (MA), midline central (MC), midline posterior (MC), right anterior (RA), right central (RC), right posterior (RP).

2.3 Results

2.3.1 Accuracy Data

The overall accuracy rate in the grammaticality judgment task was 97.1%. Out of 25 participants, 6 achieved 100% accuracy rate, whereas the number of errors in the rest varied from 1 to 6. No further analysis has been performed due to the very low number of errors and their equal distribution across the conditions.

2.3.2 Behavioural Results

The behavioral data did not reveal any difference in reaction times between syntactic and semantic gender ($F (1, 24) = .022$, $p = .884$), or in the interaction of condition (syntactic and semantic gender) and grammaticality (grammatical and ungrammatical) ($F (1, 24) = .516$, $p = .48$). There was a main effect of grammaticality ($F = 10.039$, $p = .004$), meaning that ungrammatical sentences provoked faster responses.

2.3.3 ERP Results

Visual inspection indicted a presence of a left-lateralized anterior negativity and posterior-central positivity (semantic gender: Figs. 2.2; syntactic gender: Fig. 2.3). The negativity had the same distribution in both conditions encompassing the left anterior region, but also spreading into the left central region. Its emergence and duration were in accordance with most previous reading studies on agreement, meaning that it started approximately 300 ms post-stimulus onset, and lasted until 450 ms. This negativity was immediately followed by a positive effect lasting from 500 ms until almost 1200 ms in both conditions. Visually, it seemed that the positive effect had larger amplitude in semantic gender, as well as a somewhat wider distribution. There seemed to be no difference in amplitude, time-window or distribution of the negative effect between the two conditions.

In the first time window (**300-450 ms**), the overall lateral ANOVA showed an interaction between grammaticality and hemisphere ($F (1, 24) = 14.742$, $p = .001$), as well as a three-way interaction between congruency, hemisphere, and anteriority ($F (2, 48) = 4.291$, $p = .032$). A follow-up ANOVA localized the effect in the anterior and central regions, in which grammaticality interacted with laterality (anterior: $F (1, 24) = 17.851$, $p < .001$; central: $F (1, 24) = 12.184$, $p = .002$). The negative effect was left-lateralized, corresponding to the left anterior and left central regions. The midline analysis did not yield any significant result with the factor grammaticality.

The main effect of grammaticality was significant in the lateral analysis of the **500-700 ms** time window ($F (1, 24) = 20.591$, $p < .001$), with ungrammatical sentences eliciting a more positive waveform. The effect also interacted with anteriority ($F (2, 48) = 11.948$, $p = .001$), as well as anteriority, laterality, and condition ($F (2, 48) = 5.344$, $p = .012$). Follow-up testing revealed that the positivity was most pronounced in the central ($F (1, 24) = 20.532$, $p < .001$) and posterior regions ($F (1, 24) = 31.036$, $p < .001$). Grammaticality interacted with laterality and condition in the posterior regions ($F (1, 24) = 5.821$, $p = .024$), while being marginally significant in the central regions ($F (1, 24) = 4.216$, $p = .051$). In addition, the interaction between condition and grammaticality was marginal in the posterior regions ($F (1, 24) = 3.83$, $p = .062$). Due to a three-way repeated interaction between grammaticality, gender, and laterality, a follow-up ANOVA was performed for each hemisphere independently, with those levels of anteriority in which grammaticality interacted with condition (central and posterior). In the left hemisphere, there was a main effect of grammaticality ($F (1, 24) = 26.527$, $p < .001$), as well as a two-way interaction with anteriority ($F (1, 24) = 5.885$, $p = .023$), denoting that the positivity was significantly stronger in the posterior regions. In the right hemisphere, a main effect of grammaticality was observed ($F (1, 24) = 25.712$, $p < .001$), as well as an interaction with condition ($F (1, 24) = 5.276$, $p = .031$), and anteriority ($F (1, 24) = 8.151$, $p = .009$). The lateral analysis (of the central and posterior regions) showed that in both hemispheres ungrammatical sentences caused a significantly larger positivity in the posterior regions compared to the central regions. In the right hemisphere, the amplitude of the positivity was significantly larger in semantic gender than in syntactic gender.

The midline analysis showed an effect of grammaticality ($F (1, 24) = 31.693$, $p < .001$), which also interacted with anteriority ($F (2, 46) = 13.728$, $p < .001$). Unlike in the lateral analysis, ungrammatical sentences yielded a more positive waveform in all three regions across the anterior-posterior axis (anterior: $F (1, 24) = 9.547$, $p = .005$; central $F (1, 24) = 31.068$, $p < .001$; posterior: $F (1, 24) = 43.84$, $p < .001$).

In the last time window (**700-900 ms**), ungrammatical sentences continued to elicit a more positive waveform, which was confirmed by a main effect of grammaticality ($F(1, 24) = 16.24$, $p < .001$). In addition, grammaticality interacted with the topographic factors of anteriority ($F(2, 48) = 103.878$, $p < .001$) and laterality ($F(1, 24) = 7.919$, $p = .01$), as well as with both of them ($F(2, 48) = 18.173$, $p < .001$). Finally, a marginally significant three-way interaction emerged between condition, grammaticality, and anteriority ($F(2, 48) = 18.173$, $p < .001$). Based on the multiple interactions, we decided to first follow-up test on each level of anteriority separately. In the anterior regions, grammaticality interacted with laterality ($F(1, 24) = 24.117$, $p < .001$), due to ungrammatical sentences eliciting a more negative waveform in the left anterior region ($F(1, 24) = 7.069$, $p = .014$). In the central regions, a main effect of grammaticality (ungrammatical sentences eliciting a more positive waveform) was obtained (grammaticality: $F(1, 24) = 19.214$, $p < .001$), as well as an interaction between grammaticality and laterality ($F(1, 24) = 6.065$, $p = .021$). Further testing show that the positive effect was present in both central regions, but was significantly larger in the right ($F(1, 24) = 25.908$, $p < .001$) than in the left hemisphere ($F(1, 24) = 7.069$, $p = .004$). Also, grammaticality entered a marginally significant interaction with condition in the right hemisphere ($F(1, 24) = 3.125$, $p = .09$), meaning that the effect was marginally stronger in the semantic than in the syntactic gender condition. Finally, a main effect of grammaticality was observed in the posterior regions ($F(1, 24) = 66.852$, $p < .001$), in which ungrammatical sentences elicited a more positive waveform. Most importantly, grammaticality interacted with condition ($F(1, 24) = 4.437$, $p = .046$), indicating that a significantly larger effect was elicited in the semantic gender than in the syntactic gender condition.

In line with the lateral results, the midline analysis yielded an overall effect of grammaticality ($F(1, 24) = 22.193$, $p < .001$), as well as an interaction between grammaticality and anteriority ($F(2, 48) = 92.167$, $p < .001$). Further testing proved that the positivity was present in the central ($F(1, 24) = 26.855$, $p < .001$) and posterior midline region ($F(1, 24) = 67.055$, $p < .001$).

2.3.4 Summary of ERP Results

Ungrammatical sentences in both conditions elicited aleft-lateralized anterior-central negativity in the 300 - 450 ms time window. The effect corresponds to the left anterior negativity. The LAN was followed by a broadly distributed positivity in the subsequent time window (500-700 ms), that is, the P600. The positive effect was present in the posterior and central regions laterally, as well as in all midline regions. It was stronger in the right hemisphere. Most importantly, the statistical analysis confirmed that the effect was larger for semantic gender than syntactic gender in the right hemisphere. The positive effect persisted into the last time window (700-900 ms). Again, the effect was significantly larger for semantic than syntactic gender, this time in the posterior regions.

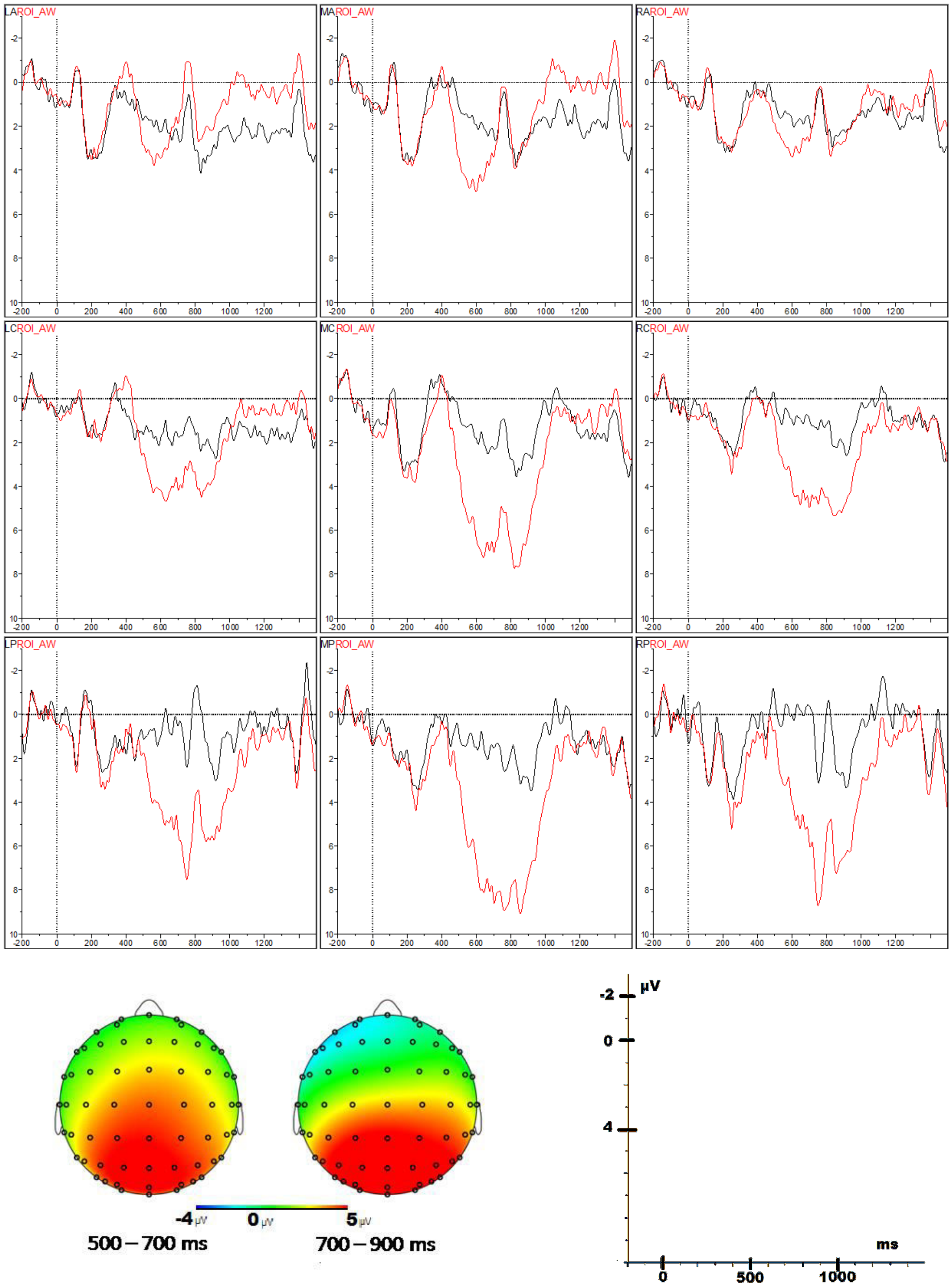

Fig. 2.2 Grand average ERPs for the semantic gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences (gender mismatch). The topographic maps represent a difference between ungrammatical and grammatical sentences in the two P600 windows.

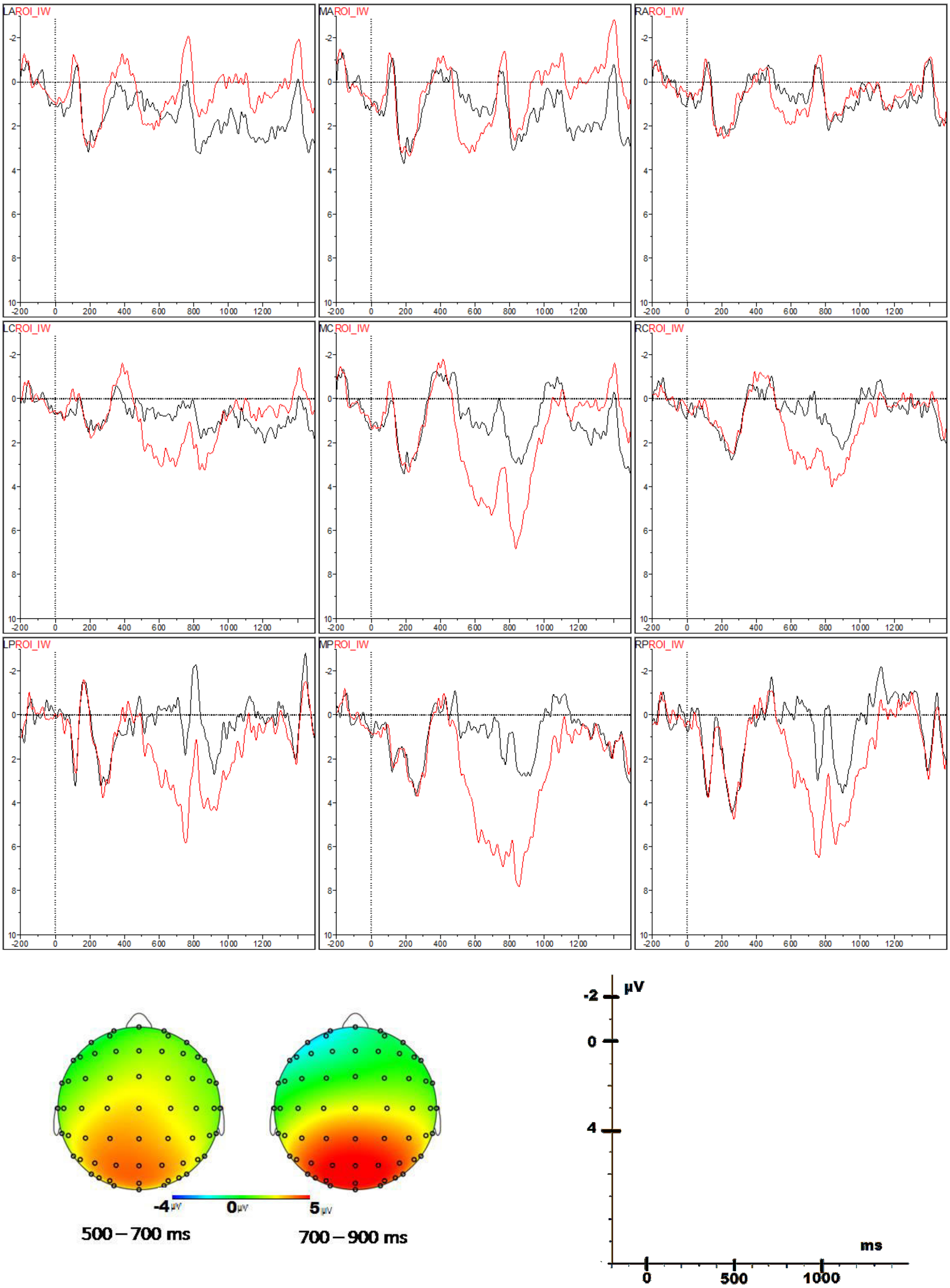

Fig. 2.3 Grand average ERPs for the syntactic gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences (gender mismatch). The topographic maps represent a difference between ungrammatical and grammatical sentences in the two P600 windows.

2.4 Discussion

2.4.1 LAN

The experimental manipulation in this study elicited the LAN in the LAN/N400 time window. A left-lateralized negative effect peaked at approximately 350 ms. The effect was confined to the left lateral and left central region. Its spatial, as well as its temporal distribution, is in line with previous studies reporting the LAN (Friederici, 2002; Molinaro et al., 2014). More importantly, the effect was identical in both syntactic and semantic gender. We offer one explanation for the absence of the N400, one for the presence of the LAN, and one that includes both.

The most obvious explanation for the absence of the N400 is that there is no semantic activation in semantic gender. The parser retrieves the gender information in the same way for both syntactic and semantic gender, possibly through the morphology, the lemma, or both, while completely bypassing semantics. Therefore, the LAN was caused by a morphosyntactic violation in both syntactic and semantic gender.

It is still possible for the LAN to be present despite the parser's reliance on the noun's semantic information. Caffarra et al. (2015) found no difference between transparent (gender retrieved from morphology and/or lemma) and opaque nouns (gender retrieved from lemma) in the violated condition. However, a comparison between grammatical instances of transparent and opaque nouns elicited a difference in the waveform. It is, thus, possible that the underlying cognitive processes represented by the LAN are only sensitive to the end product, which is a morphosyntactic violation even in the case of semantic gender mismatch. Even if the parser went through the semantic route, along with the morphological and lemma route, this would be irrelevant for the processes underpinning the LAN.

Finally, the parser may have retrieved gender through the semantic route, but the resulting violation is not legible at the semantic level. That is, a gender mismatch does not contradict any semantic properties of the noun or the discourse. More precisely, the system encounters difficulty integrating information at the syntactic level only, giving rise to the LAN. Since a gender mismatch does not contradict any lexical semantics, there is no increased demand in meaning integration in general; hence the absence of the N400. Essentially, the system does not recognize a semantic gender violation as an outright semantic violation, which was also suggested by Barber et al. (2004).

2.4.2 P600

The left anterior negativity was followed by a positive deflection with mainly posterior-central distribution, which corresponds to the P600 (e.g., Friederici, 1995; 2002; Friederici & Jacobsen, 1999; Friederici & Meyer, 2004; Kaan et al., 2000, Molinaro, Vespignani et al., 2011). The distribution of the effect was the same for both conditions; it was the strongest in the posterior regions, but it included all central, as well as parts of anterior regions. However, the effect was statistically stronger in the semantic gender condition, particularly in the posterior regions. More precisely, the P600 amplitude was larger for the semantic gender violation. According to Otten and Rugg (2004), if an effect's distribution is the same for two conditions, the engaged cognitive processes are likely to be identical. If, however, the amplitude is larger in one condition, this signals a larger processing load in that condition. This assumption aligns neatly with our hypothesis that semantic gender violations are more difficult to integrate due to more complex reanalysis processes.

The purpose of repair and reanalysis is to enable the computation to continue unhindered, despite an illegal entry in the form of a syntactic violation (Molinaro et al., 2008). In case of a simple article-noun gender mismatch (e.g., il_M $sedia_F$ 'the chair'), the parser will try to repair the article (il_M $sedia_F$ > la_F $sedia_F$). If a disagreeing noun is marked for semantic gender, as well as a gender transparent suffix (*-o* for masculine, *-a* for feminine), the parser has an extra repair option at its disposal. In addition to repairing the article (il_M $bambina_F$ > la_F $bambina_F$ 'the girl'), the parser can also repair the noun (il_M $bambina_F$ 'the girl' > il_M $bambino_M$ 'the boy'). By repairing the noun, the parser actually tackles two different linguistics levels: it alters the noun's gender feature stored at the lemma level (syntax), but it also alters the referent's biological sex (semantics). This explanation is further supported by theories claiming that syntactic and semantic information are allowed to interact in the P600 stage (Friederici, 2002; Gunter et al., 2000).

The current study demonstrated that processing nouns with semantic gender encompasses an interaction between syntactic and semantic processes, as reflected in the increased amplitude of the P600. The P600 is preceded by the LAN in both conditions indicating that gender processing always has a syntactic component regardless of the source of gender. There is one remaining issue, though. The difference found in this study reflects a difference in the P600 and repair and reanalysis processes. However, these processes were caused by the inclusion gender violations in our paradigm. On the basis of the current result, we cannot claim with absolute certainty that semantic gender processing in non-violated environment takes a syntactic (lemma), semantic or both routes. Rather, we can say that the system can fall back on semantic information, provided it is available, in order to repair a syntactic computation. These results are in line with the assumption that syntactic-semantic interface is likely to occur during the late processing stage (P600).

Chapter 3
Syntactic and Semantic Gender Processing in Listening: An ERP Study on Italian

3.1 Introduction

The study in Chapter 2 investigated the processing mechanisms behind syntactic and semantic gender. The idea was that grammatical gender is retrievable from the lemma level, with the gender suffix being invariable. The term ‘invariable’ here means that the word-final vowel cannot alternate between *-o* (masculine gender) and *-a* (feminine gender), creating nouns with different gender forms and possibly different semantics. As an illustration, the stem *tren-* cannot yield two lexemes: *treno* ‘train’ and **trena* ‘?’. The implication is that in a violated context, such as la_{FEM} $tren\boldsymbol{o}_{MASC}$, there is only one possible reanalysis/repair process that would reanalyze the feminine article (*la*) as the correct masculine article (*il*). However, nouns with semantic gender have to be morphologically marked for their gender by a suffix. The root *bambin-* denotes a child but it lacks gender specification. If it is a boy, the suffix will be *-o*, and if it is a girl, the suffix will be *-a*. The suffix alternation opens the door to a dual reanalysis/repair process. Just like in grammatical gender, once the parser spots the mismatch between the article and the noun suffix (**il bambina*), it will try to repair the article (*il* > *la*). However, the parser has an alternative repair option at its disposal, that is, repairing the gender suffix on the noun (**il bambina* > *il bambino*). The more complex repair/reanalysis processes in semantic gender led to a bigger processing load, realized through increased amplitude of the P600 in the previous experiment. In addition to the P600, both syntactic and semantic gender violations caused the LAN, which is taken to represent a violation detection process.

In the current experiment, we used the same materials as in Chapter 2, but in the auditory presentation modality. The goal was to investigate the relationship between syntactically-related ERP components and the presentation modality. We aimed to capitalize on the fact that the violation position is different for syntactic gender (lexical level) and semantic gender (inflectional morphology). This way, we were able to further explore temporal characteristics of syntactically-related ERP components.

3.1.1 Time Course

Accessing gender feature on the noun seems to be available through different routes. Gender can always be retrieved from lexical knowledge (***tren***-*o*), decomposed from morphology (*tren*-***o***), or sometimes deduced on the semantic basis (animate referents with biological gender). If we put it in the real-time processing perspective, the time course of accessing different types of gender information will depend on the presentation modality. When reading, the visual information is perceived during gaze fixations lasting on average 200-250 ms (Rayner & Clifton, 2009). One fixation period is enough to cover 7-8 letter positions, which is usually one word at a time. During a single fixation, the reader is able to take in a whole word, which includes inflectional morphology. Therefore, in visual gender recognition, all gender cues become available at the same time. While interpreting the grammatical information off a suffix, the reader is able to simultaneously accesses semantic and grammatical properties of that word.

Spoken language, however, has a different rate of delivering information compared to reading. The typical reading speed is 250-350 words per minute, almost two times faster than the typical speech word production rate of 120-200 words per minute (Rayner & Clifton, 2009). This sets the average auditory word duration in the range of 300 ms to 500 ms. Unlike reading in which a whole word is instantaneously fixated, a spoken word is delivered in smaller chunks (phonemes) in a bottom-up information flow. The system strives to integrate all the incoming phonemes into an existing word that is available in the lexicon.

The nature of the underlying comprehension processes allows the listener to access the target lexeme before it is uttered or heard in its entirety. According to the Cohort Model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), spoken-word comprehension entails two simultaneous processes: acoustic signal decoding and (phonemic) combinatorial processes. It is a bottom-up process starting with the decoding of the initial phoneme followed by the activation of all lexemes in the mental lexicon that start with that phoneme. This first and largest group of potential lexemes is called a word-initial cohort. As speech progresses, the acoustic signal brings in more information (phonemes) that the system tries to integrate. The initial cohort narrows down with the perception of each subsequent phoneme (by excluding now incorrect phoneme combinations) up to the point when there is only one candidate left. This point in time is called 'recognition point' (RP) and it is, quite often, before the end of the word.

Context is a factor that can speed up final lexeme choice. The more constraining the context is, the faster the word will be recognized. As an illustration, a one or two-syllable content word will be correctly selected after 200 ms if there is supportive context, whereas it might take 300 ms in isolation (Grosjean, 1980).

Due to time-course differences in word comprehension, with reading being a faster modality, auditory gender processing should differ from visual gender processing. Assuming that gender morphology in Italian is always word-final, gender cues are delivered at different, albeit subsequent points in time during speech. As soon as the listener has recognized the word, he is able to access its lemma, thereby accessing semantic and grammatical information (for production, see Levelt, Roelofs, & Meyer, 1999). In this case, the gender feature stored at the lemma becomes available before the gender information from the suffix. As the speech progresses up to the final phoneme, the gender feature can be again accessed from inflectional morphology. In this case, it is redundant information and can only reinforce the previously accessed gender feature. The implication is that in a violation paradigm, a mismatch in gender can already be detected at the lemma level, without hearing the gender morpheme. For example, after having heard only *tren-*, the listener can already deduce that the target noun is *treno* and that it is masculine. Hence, in case of **la treno*, the gender violation is noticed before the masculine suffix has been perceived.[1]

There is one exceptional case when gender processing has to completely rely on morphological information. Nouns with semantic gender (gender based on semantic information) can almost always be grouped in gender pairs. These are minimal pairs in terms

[1] There is a small group of nouns in Italian that are lexically different at the level of the gender morpheme (e.g., *il palo* 'post', *la pala* shovel). Such instances are scarce, and they were not included in our stimuli.

of semantics sharing all characteristics but differing at the level of biological gender (e.g., man - a male member of the human species, woman – a female member of the human species). As already mentioned, the distinction in biological gender is often preserved in the gender system of a language. In Italian, a number of nouns marked for semantic gender also share the noun stem, with masculine nouns ending in *-o* and feminine nouns in *-a* (e.g., *il bambin**o** – la bambin**a*** 'the boy - the girl'). In this case, the disambiguating gender information is stored at the inflectional morphology level, and not at the lemma. Therefore, during auditory processing, the listener can assign semantic gender at the lemma level only if there is supporting context (e.g., gender marked article preceding the noun: il_M $bambin$-o_M). However, in order to detect a gender mismatch, the listener has to hear the complete word.

3.1.2 Time Course and ERP Components

The name of two out of the three most relevant components in sentence processing is related to their onset time. The N400 and the P600 are said to typically peak at approximately 400 ms and 600 ms after the stimulus onset, respectively (e.g., Kutas & Federmeier, 2011; Molinaro, Barber, & Carreiras, 2011; Osterhout & Holcomb, 1992). Such uniform time-wise characterization makes sense mainly in reading, in which word length is usually not a significant factor since reading modality allows the reader to fixate almost any word within 200 ms. Also, there should be no major difference whether the target information (e.g., morphology) is at the beginning or at the end of the word because the entire word is processed as a single unit. Taking this fact into account, it is plausible that the language components will peak at approximately the same time in the majority of reading studies. For example, the P600 is almost uniformly reported in the time window of 500-700 ms post-stimulus onset (e.g., Barber, Salillas, & Carreiras, 2004; Dragoy, Stowe, Bos, & Bastiaanse, 2012; Hagoort & Brown, 1999; Molinaro, Vespignani, & Job, 2008).

The idea behind the current study is that there is an intrinsic relationship between the component onset-time and the moment the crucial information is received (recognition point for lexical information or inflectional morphology for grammatical information). This is only relevant for auditory studies in which there is a bottom-up information flow. During the information flow, different pieces of information are delivered at different points in times. ERP language components are supposed to be tied to the moment in time in which the relevant information is delivered (for example, agreement morphology for the LAN). Most auditory studies focusing on the relationship between component onset-time and the recognition point/violation position investigated semantic processes and N400. The onset of the auditory N400 is reported to be earlier than in reading (Hagoort, 2008; Kutas & Federmeier, 2011). This is contrary to our expectations since the recognition point (RP) in spoken words is usually later than the 200 ms needed for a single word fixation. There are two possible reasons for such a finding.

Firstly, according to the Cohort Model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), the RP of a word comes usually before the end of that word. O'Rourke and Holcomb (2002) tested two groups of words in isolation: words with an early RP and words with a late RP. The onset of the N400 was later for the condition with a late RP relative to the word onset. However, when measured relative to the RP, the N400 started at the same time for both groups (app. 50 ms after the RP). The study provided further support for the Cohort Model and the notion of recognition point, at least when it comes to contextless single word-comprehension. The N400 is intrinsically connected to the RP. Therefore, it follows naturally that words with an earlier RP (shorter words) will cause an earlier N400.

Secondly, the presence of context can modulate the onset time of the auditory N400. Not all electrophysiological studies agree that the N400 has to follow the RP. Van den Brink, Brown, and Hagoort (2001) and Van Petten, Coulson, Rubin, Plante, and Parks (1999) found that in sentence context, the N400 appears before the RP. The difference between the studies that report N400 before and those that reported it after the RP can be explained by the presence/absence of context. If there is high to medium constraining context, the N400 can appear before the RP since context can facilitate lexical selection.

It is worth noting that the auditory N400 has been suggested to consist of two phases: early N200 (from 150-200 ms on) and later N400 (peaking at 400 ms; e.g., Connolly & Phillips, 1994; Hagoort, 2008; Van den Brink et al., 2001). The early stage can only appear when context is provided. According to Hagoort (2008), the N200 represents an interface between lexical form and context. Lexical selection is supposed to be a bottom-up process: it starts from the initial phoneme when the initial-word cohort is created. The lexical choice decreases in scope with each incoming phoneme. However, the system is able to simultaneously interact with contextual information while trying to choose the most appropriate lexeme (Van den Brink et al., 2006). If the context is highly or medium constraining, it aids lexical selection.

But if the target word is semantically incongruous in a medium or highly constraining context, the system will strive and fail to find the best fit among the available lexemes. More specifically, since context is provided, the system limits the scope of the available lexemes based both on the phonological information from the incoming speech signal, as well as on the contextual information. The target word may be a phonologically good fit, but it is semantically incongruous with the context, and is therefore eliminated from the pool of available lexemes. This increases the effort of lexical selection, which is reflected in the N200. Finally, when there is enough acoustic information for the definite lexeme choice, the system integrates the lexeme into the context. If the lexeme is congruous with the context, the amplitude of the N400 is smaller than when the word is incongruous.

The role of the RP in auditory studies on semantics (without context) is clear. The N400 is aligned with the recognition point, that is, regardless of the word length, it consistently appears 50-100 ms after the RP (O'Rourke & Holcomb, 2002). The situation is less clear with the syntactically-related components (LAN, P600). In general, there has not been much work on the differences of auditory and visual P600 and LAN, or their exact onset time.

Hagoort and Brown (2000) investigated the relationship between the presentation modality and language components in a set of morphosyntactic experiments. The starting point was that most reading studies had a very slow presentation rate (one word every 500 to 600 ms) compared to natural connected speech (app. 4 words per 1 s). Therefore, their study consisted of two experiments; in the first visual experiment the words in a sentence were presented at a higher rate than usual (every 250 ms), whereas the second experiment was auditory with the same stimuli. The experimental paradigm consisted of three different morphosyntactic manipulations (agreement, subcategorization, and phrase violation). Regarding the presence of the components, the P600 was present in both modalities, whereas the LAN was attested only in listening in 2 out of 3 conditions (agreement and subcategorization). The onset time of the P600 was very similar for both modalities, starting at approximately 500-600 ms. It seems that generally the onset time of P600 is very similar regardless of the presentation modality (e.g., Balconi & Pazzoli, 2010; Friederici, Hahne, & Mecklinger, 1996; Hagoort & Brown, 2000; Meulman, Stowe, Sprenger, Bresser, & Schmid, 2014). The auditory LAN corresponded to the usual LAN time window of 350 to 500 ms.

The onset time of P600 appears to be remarkably stable across different modalities and presentation rates. Hagoort and Brown (2000) concluded that the postlexical sentence processing mechanisms have to be the same regardless of the modality, which results in the same onset time of P600. They also remarked that if the RP is taken into account, which usually comes later than the 200 ms needed for a word fixation, the auditory P600 actually starts earlier than the visual P600. However, no additional explanation was offered on this topic. The large majority of the studies that do not report any onset-time difference between auditory and visual P600 indicate that the P600 may be a cyclical process independent of word duration or recognition point. In other words, the P600 may be tied to the target word onset, always peaking at around 600 ms because that is the constant moment in time when the parser tries to incorporate all the previously gathered information.

There is at least one study that contradicts this view. Osterhout and Holcomb (1993) reported the auditory P600 as early as 300 ms post-stimulus onset in a sentence comprehension experiment. This auditory study was preceded by a reading study in 1992, conducted with the same material, in which P600 was recorded from 500 ms on. A possible explanation for the early auditory P600 could lie in the target word: the infinitival marker *to*. The duration of this word is very short (100 ms), but the morphosyntactic violation becomes apparent at this word. Thus, there is some indication that P600 may be aligned with the (morphosyntactic) RP, instead of being a fixed effect that always peaks at 600 ms.

In addition to the component onset time, the topographic distribution of an ERP component is another important factor that can point to different underlying processes between different presentation modalities. In the abovementioned studies, the topographic characteristics of the P600, N400, and LAN are very similar when reading and listening. One of the major differences is that the LAN has a somewhat broader distribution in listening and may not be exclusively left-lateralized (Faustmann, Murdoch, Finnigan, & Copland, 2005; Friederici et al., 1996). Also, the LAN in listening tends to present itself as longer lasting than in reading (Osterhout & Holcomb, 1993). Due to its persistence into later time windows, the auditory LAN may modulate the distribution of the P600. More precisely, the P600 may prefer the right-hemisphere distribution due to the persistence of LAN that cancels it out in the anterior-central regions in the left hemisphere (Friederici et al., 1996; Osterhout & Holcomb, 1992).

Our goal is to look in depth into the relationship between component characteristics (particularly timing and distribution) and presentation modality. Hagoort and Brown (2000) suggested that the variation in time distribution of ERP components comes from the unnatural presentation rate of written words. Our hypothesis will rather focus on the listening modality, since listening allows for much more variability. Word duration plays a significant role because the RP correlates with word length (the shorter the word, the earlier the RP). Once the parser has reached the RP, all lexical information becomes available (lexical semantics, lexical syntax: gender, aspect). Another time-related factor is the violation position. Inflectional morphology is mostly word-final in the Romance, Germanic, and Slavic languages, meaning that it almost always follows the RP. This way, there is a temporal hierarchy of lexical and grammatical information flow: lexical semantics and lexical syntax are almost always available before grammatical information that is coded as inflectional morphology (e.g., agreement morphology, case).

This study will focus on the effect of violation position within a word. In the stimuli used in this experiment, the violation is either at the lemma level (syntactic gender) or at the level of inflectional gender morphology (semantic gender). In the latter case, the listener must hear the final vowel (gender morphology) of each target word. In the case of syntactic gender, the gender feature is available from the lemma. The actual time difference between the RP, which coincides with both lexeme and lemma recognition, and inflectional morphology depends on the word length, but the average difference is in the range of 100 and 200 ms for our stimuli.[2]

3.1.3 Expectations and Predictions

The current study strives to replicate the findings from Experiment 1, that is, to show the presence of all the components. Most studies that tested the same materials in a visual and an auditory paradigm reported the same components in both modalities (Balconi & Pazzoli, 2010; Friederici et al., 1996; Meulman et al., 2014; but see Hagoort & Brown, 2000). Accordingly, the P600 is expected in both conditions, with semantic gender showing a more pronounced P600 effect. The LAN is expected in both conditions with no major differences in amplitude or distribution. What we expect to differ is the onset time of the components. We

[2] We do not provide exact values as there is some variation inside the conditions regarding word length. The important point is that the RP in syntactic gender is always earlier than in semantic gender on the length-matched words. This is enough information since the ERPs are tied to the stimulus-onset, and if any difference is caused by the violation position, it will be immediately obvious.

start from the assumption that the P600 and LAN are violation aligned. The implications of this claim are straightforward – the sooner the parser registers a violation, the sooner the response (LAN, P600) is expected to arise. The violation should be detected earlier in syntactic gender since it is detectable from the lemma, which coincides with the RP. In semantic gender, the violation is detectable from the inflectional morpheme, which is its RP.

In sum, both the P600 and LAN are expected to start earlier in the syntactic gender condition. The average duration of the critical words is 550 ms in both conditions. As the RP comes before the end of the word, and it takes an additional 50-100 ms for a component to appear (for N400, see O'Rourke & Holcomb, 2002; for (e)LAN, see Hastings & Kotz, 2008; Steinhauer & Drury, 2012), the LAN in syntactic gender is expected in the time window of 400-600 ms. However, the violation is detected later in semantic gender at approximately 500-600 ms. With an additional 100 ms, this would mean that the LAN should be detectable 600 ms post-stimulus onset at the earliest. This is a very late time window in which we already expect P600. It is still unclear how the two components may interact. It is possible that the LAN cancels out the P600 in the left hemisphere, thus leaving a right lateralized P600. It is, however, also plausible that a strong P600, such as the one in the previous experiment in the semantic gender condition, may overshadow the presence of LAN.

3.2 Method

3.2.1 Participants

For this experiment, 27 native speakers of Italian (11 male, mean age 24.2, age range 19-33) were recruited. Due to excessive noise and artifacts, data of 3 participants were excluded from the analysis, leaving a total of 24 participants. All participants were right-handed, as assessed with the Italian version of the Edinburg Handedness Inventory (Oldfield, 1971). None of the participants reported any neurological impairment or psychiatric disorder. They all had normal to corrected-to-normal vision, as well as normal hearing. The participants signed a written consent prior to the experiment, and they received € 20 for the participation.

3.2.2 Materials

The same materials were used as in Chapter 2. We describe them here too for the reader's convenience.

The experimental materials consisted of 160 sentences, half of which contained a noun marked for semantic gender (semantic condition), and the other half a noun marked with syntactic gender (syntactic condition). The frequency of the two noun types was matched (semantic gender average LogF: 2.08, syntactic gender average LogF: 2.32; $t(38) = 1.33$, $p = 0.19$). Each noun was used four times as: 1) singular grammatical; 2) singular ungrammatical; 3) plural grammatical; 4) plural ungrammatical. Finally, gender of the nouns was balanced, with half of all nouns being masculine, and half feminine.

As a consequence of a balanced design (each noun used as singular and plural, grammatical and ungrammatical), the participant would have been exposed to the same noun 4 times. In order to minimize repetition, the experimental items were divided between 2 lists. This way, every participant saw each noun twice only, once as singular and once as plural. Also, the noun was used once in a grammatical and once in an ungrammatical context. The list design decreased the experiment duration from 40 to 20 minutes.

Each subject listened to 80 experimental and 80 filler sentences. Half of the experimental sentences contained nouns marked with semantic gender and the other half nouns marked with syntactic gender. Each sentence started with a temporal adverb (e.g., often, yesterday, rarely), followed by a definite article, which was followed by a noun. The target noun phrase was always in the subject position. Depending on the gender and number of the noun, the article took one of the 4 forms: *il* masculine singular; *i* masculine plural; *la* feminine singular; *le* feminine plural. Italian articles can have different morphological realizations depending on the phonological context, but we opted for nouns taking the default article forms (see Chapter 1). For the rest of the sentence, the noun was followed by the verb and then either by an object or an adverbial. Half of the experimental sentences had a morphosyntactic violation in the form of gender disagreement between the article and the noun. For example, a masculine singular noun, such as *nonno* 'grandfather', was preceded by the feminine singular article *la*, resulting in an ungrammatical sentence. Example sentences for each condition are given in (1) and (2).

(1) Semantic gender

Spesso il nonno fumava la pipa.
often the$_{SG.M}$ grandfather$_{SG.M}$ smoked the pipe
'The grandfather often smoked his pipe.'

*Spesso la nonno fumava la pipa.
often the$_{SG.F}$ grandfather$_{SG.M}$ smoked the pipe
'The grandfather often smoked his pipe.'

(2) Syntactic gender

Generalmente il treno parte in ritardo.
usually the$_{SG.M}$ train$_{SG.M}$ leaves in late
'The train usually leaves late.'

*Generalmente la treno parte in ritardo.
usually the$_{SG.M}$ train$_{SG.M}$ leaves in late
'The train usually leaves late.'

There were 80 fillers per list (3). Each filler item was presented once as grammatical and once as ungrammatical. The ungrammatical sentence lacked subject-verb agreement, having the predicate in infinitive. Thus, each participant heard 80 grammatical and 80 ungrammatical sentences in total.

(3) Le scale lassú sono troppo ripide per me.
the stairs up there are too steep for me.
'The stairs up there are too steep for me.'

*Le scale lassú **essere** troppo ripide per me.
the stairs up there **be** too steep for me.

All sentences were read aloud by a male native speaker of Italian. The speaker was a professional musician who was instructed to produce both grammatical and ungrammatical sentences with natural intonation. However, it is possible that there are some subtle variations in pitch in ungrammatical sentences, especially around the ungrammatical region - the article and the noun. No cross-splicing was applied to ungrammatical sentences (cutting out the article from a correct sentence and combining it with the target noun) because we believed the resulting phonological environment would not be ecological: an article together with the following noun constitutes one phonological phrase since articles are not stressed. Furthermore, there is often co-articulation, especially between the singular masculine article (*il*) that ends in a liquid and consonant-starting nouns (all nouns in our experiment). Splicing in such an environment causes phonologically audible disruptions that may outweigh the possible gain (see Steinberg, Truckenbrodt, & Jacobsen, 2012). All audio files were normalized to 70 dB, and noise reduction was applied.

3.2.3 Procedure

During the experiment, participants sat in front of the screen (70-80 cm distance) while an electroencephalograph was recorded. The task was to listen to sentences played through headphones and to answer a grammaticality judgment question after every sentence. The experiment was presented with E-Prime 2.0 (Psychology Software Tools, Inc) software. The experiment opened with instructions written on the screen. The experimenter repeated the instructions in more detail. After the practice session, he made sure that the participant had understood the task. Each stimulus was introduced with a fixation cross on the screen for 400 ms after which a sentence started. The experimental sentences were 3 seconds long on average. After the sentence finished, the screen remained blank for 1 second, and then a question mark appeared signaling the grammaticality judgment question. The participant was supposed to press the appropriate button (‘p’ or ‘q’, counterbalanced across participants) depending on whether the sentence was correct or not. There was no time limit for pressing the button. The experiment was divided into 4 blocks, each containing 40 sentences and lasting 5 min. The experiment took 25 min on average.

3.2.4 EEG Recording and Data Processing

The continuous electroencephalogram was recorded from 64 Ag/AgCl scalp electrodes (WaveGuard) using the ASA-Lab system (ANT Neuro Inc, Enschede, The Netherlands). Additional bipolar electrodes were used to record horizontal (HEOG; at the outer canthus of each eye) and vertical (VEOG; above and below the left eye) eye movements. Impedances were kept below 10 kΩ. Data were acquired at 512 Hz sampling rate with the common average reference.

The offline processing was carried out in Brain Vision Analyzer 2.0.4 software (Brain Products, GmbH, Munich, Germany). The data were down-sampled to 256 Hz before being re-referenced to the average of the mastoids. Offline filtering was performed using a band-pass filter (0.1-40 Hz), followed by automatic PCA-based eye-blink correction. The data were segmented into epochs starting 200 ms before the onset of the critical word (the noun) and lasting until 1500 ms post-word onset. Only trials that were judged correctly were segmented. The (automatic) artifact rejection (+/- 100 µV threshold) was performed only on a section of each epoch (0-1000 ms) that was used in the statistical analysis.

The baseline was corrected starting from 0 ms (the onset of the noun) until 100 ms. There were a number of factors affecting the choice of the baseline. Firstly, all data were plotted without correcting the baseline. Visual data inspection revealed that the ungrammatical conditions were slightly more positive in the article regions (app. 150-200 ms before the noun). The effect may be attributed to subtle phonological variations in the auditory signal. Secondly, the more standard pre-stimulus baseline was applied (-200-0 ms), and it introduced very early effects (starting from 50 ms on) that were not present in the non-baseline-corrected signal. In order to avoid this ‘spill-over’ effect, it was necessary to choose a not so commonly used baseline. Based on the visual inspection, as well as some earlier studies (e.g., Friederici et al., 1996), the 0-100 ms post-stimulus onset interval was chosen. The choice seems acceptable, as there are no effects this early that could substantially affect the signal.

Finally, data were averaged per subject and per condition. In case a participant had fewer than 60% of averaged trials in one or more conditions, their data were excluded from the analysis. This led to the exclusion of two participants. In addition, one participant was excluded from the analysis due to technical problems. Therefore, the data of 24 participants were used for the analysis.

3.2.5 Analysis

Averaged values (in μV) were extracted per participant, per condition, and per region of interest. The scalp electrodes were divided into 9 regions of interest, each containing either 5 or 6 electrodes (Fig. 3.1): left anterior (F7, F5, F3, FC3, FC5), midline anterior (F1, Fz, F2, FC1, FCz, FC2), right anterior (F4, F6, F8, FC4, FC6), left central (TP7, C5, C3, CP5, CP3), midline central (C1, Cz, C2, CP1, CPz, CP2), right central (C4, C6, CP4, CP6, TP8), left posterior (P7, P5, P3, PO7, PO5, O1), midline posterior (P1, Pz, P2, PO3, POz, PO4), and right posterior (P4, P6, P8, PO6, PO8, O2). The analysis was carried out in three independent time windows, from 400 to 600 ms (roughly corresponding to LAN and N400), followed by the 600-750 ms window (early P600), and ending with the 750-1000 ms window (late P600). The time windows are somewhat later than those traditionally used. This is a listening study in which the average word duration is 550 ms. A listener needs to have heard approximately 2/3 of a word in order to recognize it. This means that 350 ms or more are necessary for a speaker to identify a violation (provided it is obvious at the lexical level). Thus, it is only reasonable to expect effects from 400 ms or later, which means that all subsequent effects (such as the P600) will appear later.

For the statistical analysis, repeated measures ANOVAs were used with the following within subject factors: condition (2 levels: syntactic and semantic gender), grammaticality (2 levels: grammatical and ungrammatical), laterality (2 levels: left and right hemisphere), anteriority (3 levels: anterior, central, and posterior). In case a follow-up comparison consisted of comparing two means only, a one-tailed *t* test was performed. The significance level was set to $p < .05$. For each time window, 2 global repeated measures ANOVAs were performed; first for the lateral regions (all factors included), and then for the midline regions (factor laterality excluded). Follow-up ANOVAs were applied to those interactions that turned out at least marginally significant ($p < .1$), and that included the factor grammaticality. In case the assumption of sphericity was violated, the Geisser and Greenhouse (1959) correction was applied. Lastly, only correctly judged trials were included in the analysis.

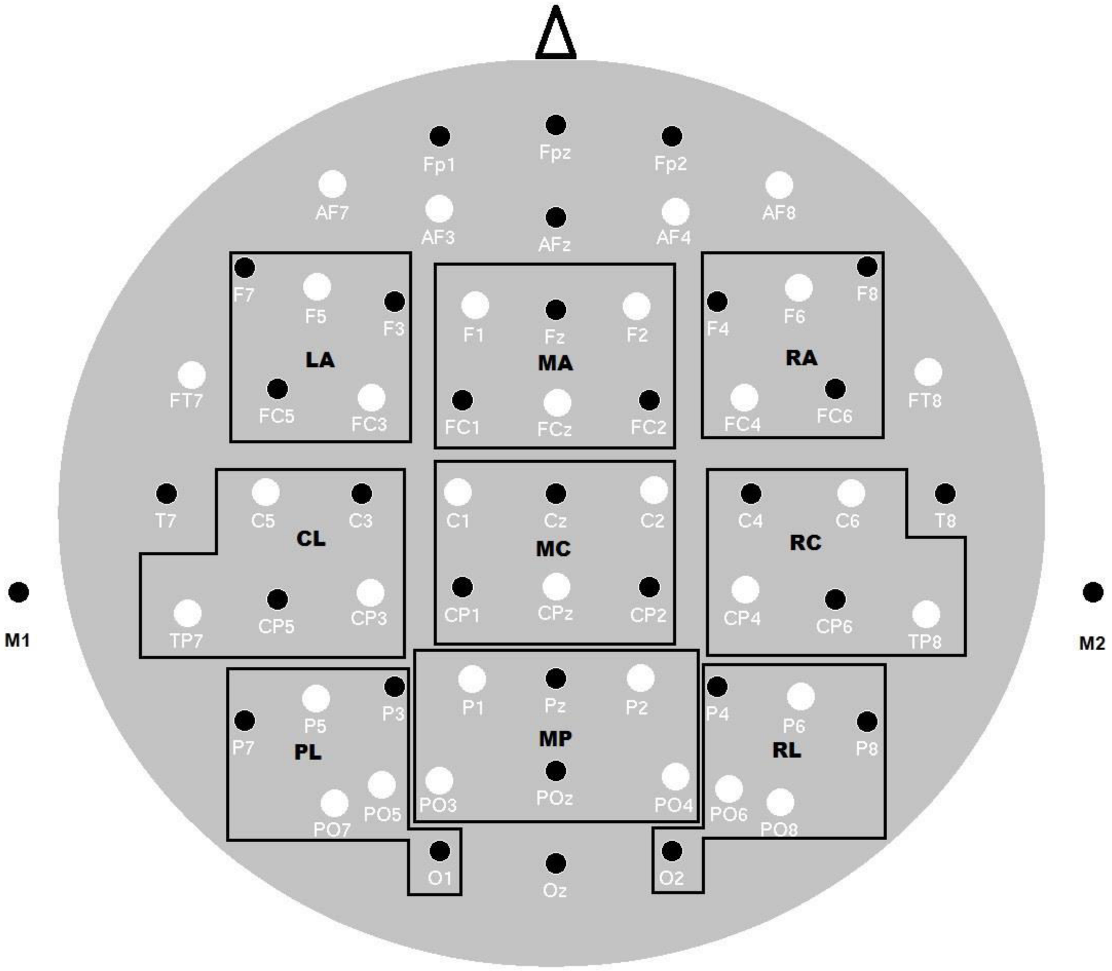

Fig. 3.1 Electrode positions and the 9 regions of interest used in the analysis: left anterior (LA), left central (LC), left posterior (LP), midline anterior (MA), midline central (MC), midline posterior (MC), right anterior (RA), right central (RC), right posterior (RP).

3.3 Results

3.3.1 Accuracy data

All participants performed almost at ceiling in the grammaticality judgment task, with an average accuracy rate of 98.6% (78.9 out of 80 items). Half of participants (14/27) did not have a single incorrect response, 12 participants were in the range of 1 to 3 errors, and one participant made 4 errors. Since the error rates are minimal and evenly distributed across conditions, no further analysis has been performed.

3.3.2 Behavioural results

There was no difference in reaction times between the responses to syntactic and semantic gender (log transformed, $F(1, 23) = 1.725$, $p = .202$), or in the interaction between gender type and grammaticality ($F(1, 23) = .885$, $p = .357$). The only significant result was found for the factor grammaticality on its own ($F(1, 23) = 26.203$, $p < .001$), meaning that participants responded faster to ungrammatical items.

3.3.3 ERP Results

Visual analysis revealed that gender violation caused a positive shift from approximately 700 ms on, somewhat later than reported in most agreement reading studies. The positivity lasted longer in the semantic gender condition, until 1500 ms post-stimulus onset, almost 200 ms longer than the positive deflection for syntactic gender. By glancing at the data, the positivity in semantic gender seemed to be of larger amplitude in most regions, as well as reaching further into the anterior regions from its source in the posterior regions. Whereas a positive effect was evident in both conditions, a wide-spread negativity from approximately 400 ms until the emergence of the positive effect was attested for syntactic gender only. Its distribution started wide-spread across the scalp, with its absolute voltage maximum in the central regions. Its topography became more anterior as the positivity started to emerge (semantic gender: Fig. 3.2, syntactic gender: Fig. 3.3).

In the first time window **(400-600 ms)**, the lateral omnibus ANOVA revealed a main effect of grammaticality ($F (1, 23) = 4.511$, $p = .045$), as well as a marginal interaction between condition, grammaticality, and anteriority ($F (2, 46) = 3.368$, $p = .072$). A follow-up ANOVA in each of the three levels of anteriority was performed, revealing a marginal effect of grammaticality ($F (1, 23) = 2.992$, $p = .097$), as well as a marginal interaction between condition and grammaticality ($F (1, 23) = 3.46$, $p = .076$) in the anterior regions. The same pattern was observed in the central regions, with grammaticality reaching significance ($F (1, 23) = 5.475$, $p = .028$), while marginally interacting with condition ($F (1, 23) = 3.188$, $p = .087$). No significant results that included the factor grammaticality were obtained in the posterior regions. All the effects in the anterior and central regions were driven by syntactic gender only. Firstly, no violation effect was expected in this time window for semantic gender as violations become obvious at approximately 550 ms from stimulus onset. Secondly, absolute mean voltage values showed that there was almost no effect of grammaticality in the semantic gender condition, whereas ungrammatical sentences caused a substantially more negative response in the syntactic gender condition.

The midline analysis showed an overall effect of grammaticality ($F (1, 23) = 5.047$, $p = .035$) that did not interact with anteriority or condition. Even though the interaction between grammaticality and condition did not turn out to be significant ($F (1, 23) = 2.035$, $p = .167$), we proceeded with two separate ANOVAs for syntactic and semantic gender. Again, it is impossible for any effect to arise in this time window (400-600 ms) in semantic gender, and the absolute voltage difference between grammatical and ungrammatical sentences was negligible. As expected, the semantic gender ANOVA did not show an effect of grammaticality ($F (1, 23) = .12$, $p = .732$), or an interaction between grammaticality and anteriority ($F (2, 46) = 1.492$, $p = .236$). The ANOVA for syntactic gender was in line with the results from the lateral analysis; grammaticality proved to be a significant factor ($F (1, 23) = 5.351$, $p = .03$), and approaching significance when interacting with anteriority ($F (2, 46) = 3.203$, $p = .084$). This trend did not prove meaningful in further follow-up testing, as the difference between grammatical and ungrammatical sentences was significant in each region.

The overall lateral ANOVA in the subsequent time-window **(600-750 ms)** revealed an interaction between condition, grammaticality, and anteriority ($F (2, 46) = 12.021$, $p = .002$). Follow-up testing revealed a trend towards an interaction between condition and grammaticality in the anterior regions ($F(1, 23) = 3.1$, $p = .092$), with ungrammatical sentences in the syntactic gender condition being more negative than in the semantic gender condition. The same pattern was observed in the midline ANOVA; an interaction between condition, grammaticality, and anteriority turned out significant ($F(2, 46) = 11.979$, $p = .001$). Upon further testing, an interaction between condition and grammaticality in the anterior regions approached significance ($F(1, 23) = 3.914$, $p = .06$). This trend was caused by ungrammatical sentences having a more negative deflection in the syntactic gender than in the semantic gender condition.

In the last time window **(750-1000 ms)**, ungrammatical sentences in both conditions elicited larger positivity than grammatical sentences ($F(1, 23) = 10.995$, $p = .003$). The factor grammaticality also interacted with anteriority ($F(2, 46) = 26.672$, $p < .001$), as well as with condition and anteriority ($F(2, 46) = 12.574$, $p = .047$). Based on the latter interaction, follow-up ANOVAs were used to determine the distribution of positivity in the anterior-posterior topographic space separately for syntactic and semantic condition. Ungrammatical sentences in both conditions were more positive in the posterior regions (semantic gender: $F(1, 23) = 13.172$, $p = .001$; syntactic gender: $F(1, 23) = 15.506$, $p = .001$). Moreover, the distribution of positivity in syntactic gender showed a trend towards favouring the right hemisphere (grammaticality × hemisphere: $F(1, 23) = 3.179$, $p = .088$). The positivity reached the central ($F(1, 23) = 11.253$, $p = .003$) and anterior ($F(1, 23) = 4.279$, $p = .05$) regions in semantic gender, while being only marginally significant in the central regions in syntactic gender ($F(1, 23) = 3.194$, $p = .087$), and completely absent from the anterior regions ($F(1, 23) = .763$, $p = .391$).

The midline analysis reflected the difference between grammatical and ungrammatical sentences in the lateral analysis, with ungrammatical sentences causing a more positive shift ($F(1, 23) = 17.722$, $p < .001$). Grammaticality also interacted with anteriority ($F(2, 46) = 14.22$, $p < .001$), yielding significant results in the posterior ($F(1, 23) = 36.296$, $p < .001$) and central regions ($F(1, 23) = 17.6$, $p < .001$), and was marginally significant in the anterior regions ($F(1, 23) = 3.159$, $p = .089$). Finally, we tested whether the effect of grammaticality was significant in the anterior region by splitting the ANOVA into one for semantic and one for syntactic gender. We hypothesized that semantic gender has a broader distribution, like in the lateral analysis, and that there is a difference in the distribution of the positivity between the two conditions. This is also supported by a marginally significant interaction between condition, grammaticality, and anteriority (F (2, 46) = 3.089, p = .08). Thus, the anterior ANOVAs revealed a significant effect of grammaticality for semantic gender (t (23) = -2.37, p = .013), and no effect for syntactic gender (t (23) = .039, p = .422).

3.3.4 Summary of ERP Results

The statistical analysis confirmed the presence of two components, a fronto-central negativity and the P600. The negativity was recorded in the syntactic gender condition only. Functionally, the negativity corresponded to the LAN whereas its distribution resembled the N400. The effect was significant 400 to 750 ms after the stimulus onset.

The P600 was confirmed for both conditions from 750 ms on. In line with the visual inspection, the effect had a broader distribution in the semantic gender condition than in the syntactic gender condition. In the syntactic gender condition, the effect was significant in all the posterior regions, as well as the midline central regions. In the semantic gender condition, in addition to the posterior regions, the effect spread over all central and anterior regions.

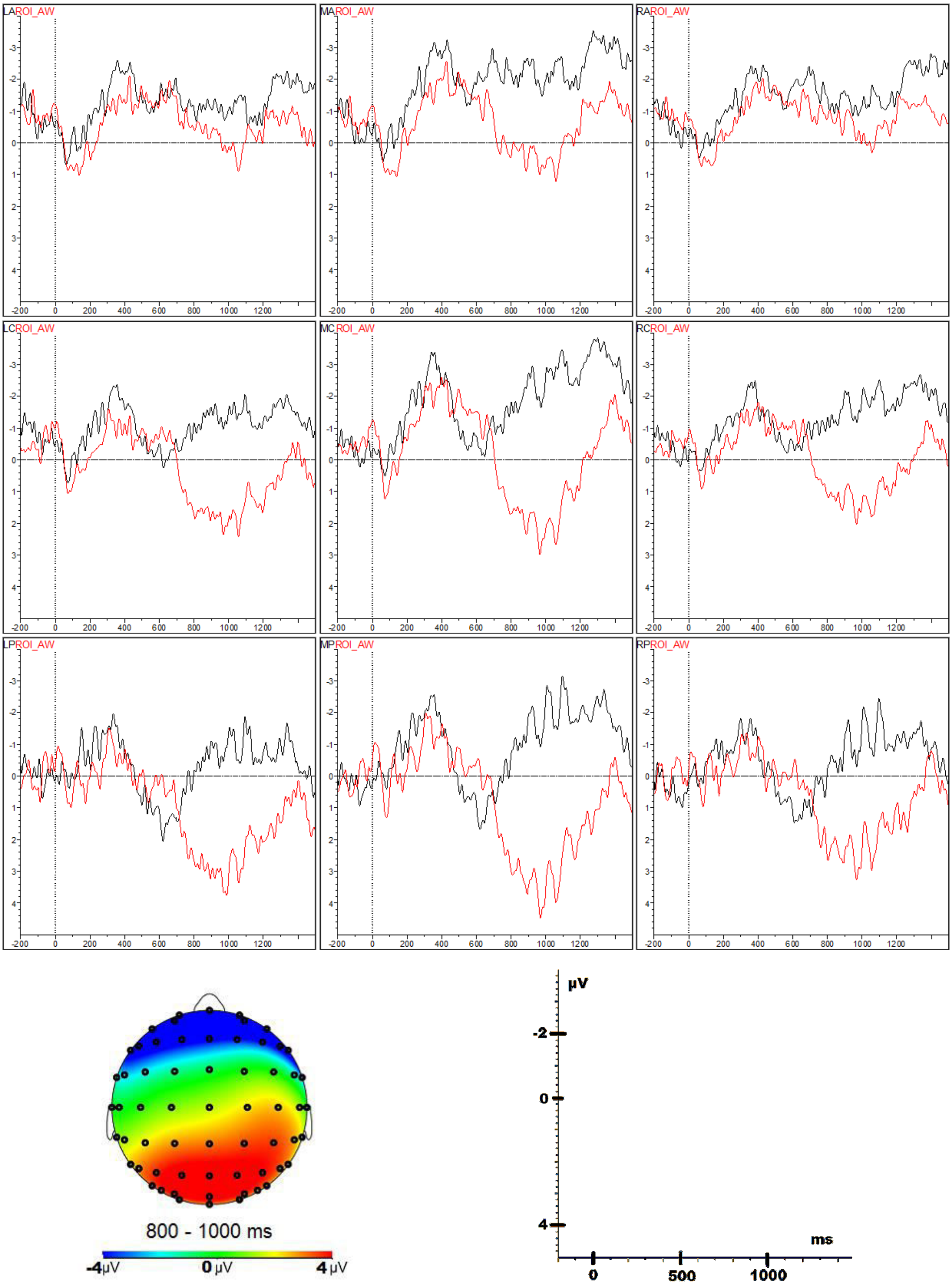

Fig. 3.2 Grand average ERPs for the semantic gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences (gender mismatch). The topographic map represent a difference between ungrammatical and grammatical sentences in the later part of the P600 windows.

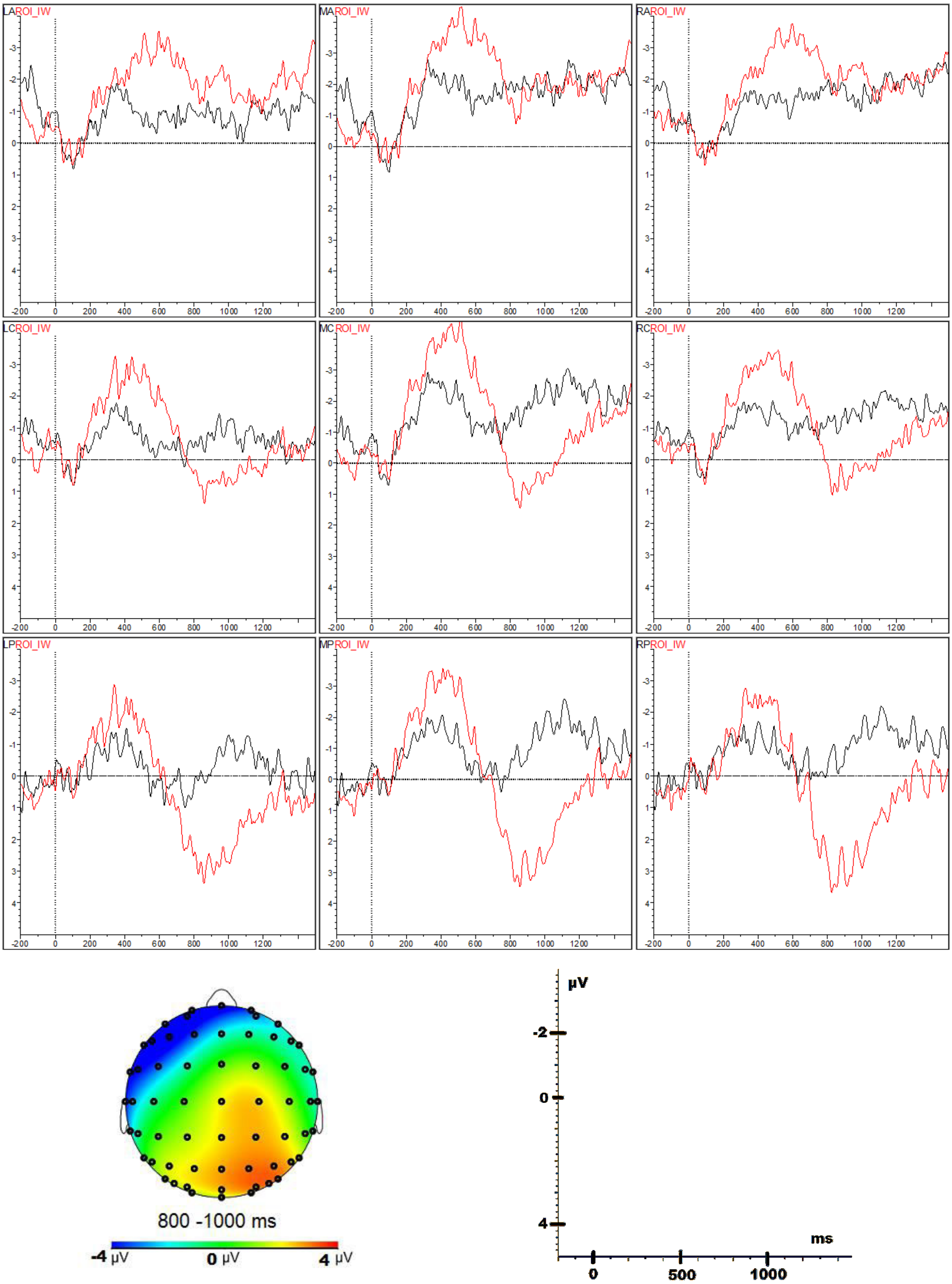

Fig. 3.3 Grand average ERPs for the syntactic gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences (gender mismatch). The topographic map represent a difference between ungrammatical and grammatical sentences in the later part of the P600 windows.

3.4 Discussion

3.4.1 Negativity

A wide-spread negativity was recorded in syntactic gender only. The effect was also long lasting, from 400 ms until the emergence of P600 at approximately 700 ms. Its distribution did not resemble the prototypical LAN distribution, which is described as left-lateralized and centered around the anterior and sometimes the anterior-central regions (Friederici et al., 1996). Rather, the distribution of the negative component was widespread, starting in the anterior regions and peaking in the central regions. It did not reach posterior regions. Due to its distribution, the component resembles an auditory N400 as described by Kutas and Federmeier (2011). Such an N400 is more anterior, lacks right-lateralization, and has a longer duration.

However, regardless of the seemingly N400 like distribution, this negative component is the result of a morphosyntactic violation detection. It is not related to lexical semantics and in no way can it indicate failed discourse integration, as syntactic gender processing is a purely syntactic process (Hagoort & Brown, 1999). Its function is undoubtedly that of the LAN, meaning registering gender mismatch (Molinaro et al., 2011). A possible explanation that would bridge the chasm between the role of N400 in semantic processing and morphosyntactic violations may lie in the nature of the gender feature in Italian. Assuming that gender is stored at the lemma level, its retrieval does encompass lemma access. According to the speech production model by Levelt, et al., 1999, lexical semantics is also stored at the lemma level. Despite being fundamentally different, lexical semantics and lexical syntactic features (e.g., gender, countability, aspect) are both related to lexical/lemma access. Therefore, it is possible that accessing any lexical feature causes a negativity that has a more central distribution.

There are two pieces of evidence that support this hypothesis. Hagoort and Brown (1999) posit that the LAN can only be detected if the violation includes morphophonologically salient cues. That is, in order for the LAN to appear, the violation must be caused by mismatching inflectional morphology. If, as in this case, the feature is lexical or retrieved lexically, the LAN is not expected. Instead, due to lexical access, the effect is a centralized negativity. Another supporting finding comes from a study by Molinaro et al. (2008) on phonotactic and gender violation in Italian (article-noun violations). The morphological realization of an Italian article depends on the initial phoneme of the following word. For example, the masculine article *il* has to be realized as *lo* when followed by a word beginning with *z*. Violating this rule causes a clearly left-lateralized negativity. This is in line with Hagoort and Brown's claim about phonological realization. However, in the gender condition, the expected LAN resembled an N400 with a more central distribution. Molinaro et al. (2008) argued that it was due to a prevalent number of animate nouns in the gender condition, which invoked semantic processing. Yet, upon inspecting the materials used in the study, it became obvious that all the target nouns had the oblique gender marker *-e*. It is impossible to deduce gender from morphology when a noun ends in *–e*, as it can be both masculine and feminine. The only resort is to retrieve gender from lexical information. And this is exactly when the negativity starts showing a more central and bilateral distribution.

Our first explanation pertaining to the atypical LAN distribution is that lexical access and retrieval, regardless of the lexical content, evokes centralized bilateral negativity. However, this explanation is challenged by several (syntactic) studies also reporting N400 like negativity without what we call 'lexical access'. Meulman et al. (2014) reported N400 with both native speakers and language learners in subject-verb mismatch condition. The negativity is explained as mirroring morphosyntactic processing in the way the LAN would. Faustmann et al. (2005) conducted a study in phrase structure violation in line with earlier eLAN studies. They found a very broadly distributed negativity followed by the P600. The authors speculated that either the component was distinct from the N400, or the participants really perceived the violation as a semantic incongruity.

It is also possible that the broad LAN/N400 distribution was caused by the experiment modality. ERP components can have a different distribution when stimuli are presented auditorily or visually. The N400 has a more anterior distribution without the right hemisphere preference in auditory materials (Kutas & Federmeier, 2011; Kutas & Van Petten, 1994). Friederici et al. (1996) noticed that the same material caused a more lateralized negativity/LAN when presented visually as opposed to a broader negativity in the auditory modality. A less clear lateralization, coupled with the anteriorly shifted N400 in the auditory modality, may completely bleach the topographic boundary between LAN and N400. Since distribution is the only available criterion for distinguishing LAN and N400 when looking at a topographic map, we can only tell the two apart based on the stimuli and stipulated underlying language processes.

Unfortunately, the results of this experiment cannot disentangle the LAN/N400 issue in syntactically-related studies. Still, one conclusion is airtight: the LAN/N400 in the current experiment is caused by gender incongruity. Whether its broad distribution is caused by the lexical nature of gender or auditory presentation modality remains to be seen.

The negativity was present only in the syntactic gender condition. There was no statistically detectable negativity in any time window in the semantic gender condition. The expectation was that semantic gender would also cause a negative effect, albeit in a later time window. The lack of negativity is even more unexpected given that negativity was found in both conditions in Chapter 2. Moreover, if the LAN is really tied to overt inflectional morphology (Hagoort & Brown, 1999), its presence was expected in semantic gender rather than syntactic gender.

One possible explanation is that the LAN could not be realized in this case due to its late emergence. The average stimulus duration is 550 ms, plus another 100 ms until the effect starts to emerge (Hastings & Kotz, 2008; O'Rourke & Holcomb, 2002; Steinhauer & Drury, 2012). This is already more than 600 ms from the word onset, which is the same the same moment in time when the P600 is expected to start emerging. We speculate that the P600, which is more broadly distributed in semantic than syntactic gender, masks or cancels out the expected negativity. The distribution of the P600 in this condition reaches central and anterior regions, which is where the negativity is expected to arise. Some authors put forth the idea that the P600 is often right-lateralized due to the preceding LAN (Friederici et al., 1996; Osterhout & Holcomb, 1992). Specifically, the negativity in the left hemisphere may cancel out the positivity in the right hemisphere. We suggest that the reverse can also be possible, given that the P600 in our case had such a broad distribution and high amplitude. Moreover, unlike in reading studies, the onset time of the LAN and P600 was expected to be the same in the present experiment, which could also play a role.

3.4.2 P600

The centro-parietal positivity (P600) was statistically significant in the time window from 750 ms until 1000 ms. The onset of the component is a little earlier, at approximately 670 ms, which is obvious from the graphs (Fig. 3.2 and Fig. 3.3), but not yet statistically significant. This is somewhat later than reported in a large majority of studies, and contrary to the claim that P600 reaches its maximum at 600 ms post-stimulus onset (Barber et al., 2004; Dragoy et al., 2012; Hagoort & Brown, 1999; Molinaro et al., 2008). The supposed delay is most likely caused by the word duration and auditory nature of the experiment. As already stated, the average stimulus duration is 550 ms, which is also the time needed for the violation to be recognized (in semantic gender). As it may take some time for a component to arise as a response to a certain process (Hastings & Kotz, 2008; O'Rourke & Holcomb, 2002; Steinhauer & Drury, 2012), the delayed timing of the P600 is not unexpected.

The distribution of the effect is in line with previous research (Hagoort & Brown, 2000). The positivity starts from the occipito-parietal regions and spreads towards the anterior regions. It reaches the maximum in the parieto-central regions in the midline. In the semantic gender condition, the effect is also statistically significant in both central and anterior regions. In the syntactic gender condition, the effect is only marginally significant in the central regions and it is completely absent from the anterior regions. The more broadly spread distribution in semantic gender is understood as representing more intricate reanalysis/repair processes due to the morphological nature of semantic gender (see Chapter 2).

An interesting trend surfaced with syntactic gender only; the P600 was (marginally) stronger in the right hemisphere. The right hemisphere bias of P600 is well documented, and could be interpreted as a result of the interaction between LAN and P600 (Friederici et al., 1996; Osterhout & Holcomb, 1992). The preceding LAN is supposed to cancel out or weaken the effect of the P600 in the left hemisphere. Therefore, the trend can be taken as an indication that the wide-spread negativity in the previous time windows is indeed LAN that was not clearly differentiated at the time.

Our initial hypothesis predicted that there would be a difference in the P600 in the two conditions, either in terms of amplitude or distribution differences. The hypothesis was supported, as the P600 had a much broader distribution in semantic gender. The second hypothesis addressed the timing of the P600. We expected that the onset of the syntactic P600 would be earlier due to earlier violation detection. This was not confirmed, as the P600 started at the same time in both conditions.

We offer three possible scenarios. The first one is the simplest; we were unable to control for word duration in our stimuli. The very limited number of nouns with semantic gender did not allow us to have a homogenous set (e.g., only disyllabic words). We tried to match the nouns in semantic and syntactic gender for overall duration (530 ms for syntactic and 540 ms for semantic gender). However, there is still some inter-stimuli variation that may have caused the P600 to start at the same time.

A psycholinguistic explanation would be that the parser needs to integrate both the lexical and morphological gender information. Since listening is an incremental process, the recognition point usually starts before the gender suffix. At the recognition point, all lexical information becomes available, gender included. The parser accesses this information but withholds complete integration until the end of the unit. This may be required in case there are lexical morphological tools (derivational suffixes or compounding) that can be applied to that lexeme and that can change its gender, which was not the case in the present experiment.

Finally, the P600 may be a cyclical process that is not violation-aligned. In such a scenario, it is possible that the P600 happens at certain intervals that are not necessarily bound to the violation detection but some higher level of integration (e.g., phrasal integration). This issue will be further addressed in Chapter 5.

Chapter 4

Gender and Number Agreement Processing in Reading: An ERP Study on Dutch

4.1 Introduction

4.1.1 Theoretical Background

Both gender and number are grammatical properties of nouns. In linguistic theory, they are often grouped together with person and labelled ‘phi-features’ (Adger & Harbour, 2008). All phi-features are always further specified with an appropriate value for their category, such as singular and plural for number. Quite often, this feature value needs to be marked on an element different than the noun. For example, the article in Dutch has to have identical feature values as the noun it modifies. Therefore, if the noun is plural, the article has to assume the plural form. However, the numerosity of the noun is obvious from the noun itself and feature reduplication onto the article is redundant in this case as it does not bring any new information (Corbett, 2003). The process of establishing a relationship between two elements by transferring feature values from one to another is called *agreement* (e.g., Pesetsky & Torrego, 2007). The focus of this study is to investigate how number and gender features are decoded and processed in article/adjective-noun disagreement

Agreement is a syntactic process whereby a relationship is established between elements at the syntactic level. But before agreement happens, the phi-features on the noun have to be valued. We will limit the discussion to number and gender only, as person is outside the scope of this paper. The way number and gender features receive and instantiate their value is where these features diverge most. In Dutch, gender is pervasively a lexical feature (e.g., Van Berkum, 1996). In terms of production theories, the gender value is stored as a separate node at the lemma level (e.g., Levelt, Roelofs, & Meyer, 1999). Therefore, the gender value is part of the noun's lexical syntax, and it is invariable. In some languages (e.g., Spanish and Italian), the gender feature is also overtly marked on the noun as a gender morpheme. Overt gender marking in Dutch is possible only when the noun contains a derivational suffix that is always associated with a specific gender value (e.g., nouns with the diminutive suffix *-(t)je* are always neuter). Number, however, is valued differently. Firstly, the speaker needs to assess the numerosity of the chosen concept: that is, whether there is only one or more than one entities. Once the value is determined, it is realized as a number morpheme on the noun (e.g., Roelofs, 1997). In case of Dutch, only plural nouns are morphologically marked (*-en*, *-s*).

Thus, we have established that number and gender differ in at least one crucial respect, namely the way the feature value is determined. Gender is an inherent, lexically stored feature. It is invariable and cannot be valued based on any semantic properties.[3] Number, however, is determined based on semantics (conceptual information), which means it can alternate between singular and plural. We will argue that due to feature variability (most nouns can be both singular and plural), number exerts a higher processing demand. Since we would like to quantify the difference in the processing load in real-time, we opted for the event-related potentials (ERPs) as the method of choice.

4.1.2 Previous ERP Research on Agreement

Previous studies on ERPs and sentence processing have mostly reported the presence of three language-related components: N400, left anterior negativity (LAN), and P600 (Coulson, King, & Kutas; 1998; Friederici, 1995; Hagoort, Brown, & Groothusen, 1993; Kutas & Hillyard, 1980; Osterhout & Holcomb, 1992). These components are usually elicited through violation paradigms consisting of two sets of identical sentences, differing minimally at a single point. The first set is meaningful grammatical sentences (baseline) against which identical sentences containing a grammatical or semantic violation are compared. In case of a semantic violation, the expected response is the N400. This component is construed as a marker of semantic and discourse integration difficulties (Friederici, 2002; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980).

[3] We gloss over the issue of semantic (biological) gender in this Chapter. More information on the distinction between grammatical and semantic gender is given in Chapter 2.

The LAN and P600 are usually described as markers of syntactic processing (Friederici, 2002), which makes them expected components in agreement studies. Indeed, a large number of studies on agreement reported a biphasic response to agreement mismatch in the form of the LAN followed by the P600 (Barber & Carreiras, 2005; Barber, Salillas, & Carreiras, 2004; Gunter, Friederici, & Schriefers, 2000; Molinaro, Vespignani, & Job, 2008). As the name says, the left anterior negativity is a negative deflection peaking between 300 and 500 ms. It is usually left lateralized with anterior distribution (but see Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004). As to its function, it is characterized as being sensitive to morphosyntactic errors (e.g., Friederici, 2002; Molinaro et al., 2011). The LAN is usually followed by a positive deflection peaking at 600 ms (P600). Some authors make a distinction between the early and late P600 (e.g., Hagoort & Brown, 2000). The early P600 lasts from 500 ms to 700 ms post-stimulus onset and has a broad distribution, whereas the subsequent late P600 is the strongest in the parietal regions. In addition to different topography, the two stages are stipulated to be somewhat functionally different. The early stage reflects integration difficulty, which is followed by reanalysis and repair in the late stage.

Based on the previous study, the most reliable agreement processing marker seems to be the P600. It is almost unanimously reported, which is not the case with the LAN (e.g., Bañón, Fiorentino, & Gabriele, 2012; Nevins, Dillon, Malhotra, & Phillips, 2007). As an illustration, several studies on Italian and Spanish gender and/or number agreement reported the LAN followed by the P600 (e.g., Caffarra & Barber, 2015; Caffarra, Siyanova-Chanturia, Pesciarelli, Vespignani, & Cacciari, 2015; Douwens, Vergara, Barber, & Carreiras, 2009; Molinaro et al., 2008; O'Rourke & Van Petten, 2011). However, studies on determiner-noun agreement in Dutch failed to report either LAN or the N400 (Loerts, Stowe, & Schmid 2013; Meulman, Stowe, Sprenger, Bresser, & Schmid, 2014). Hagoort and Brown (1999) proposed that the LAN can only be elicited by a morphologically overt violation, at least in case of gender agreement. This explains the absence of the effect in gender violations in Dutch in which gender is a lexical feature. It still does not explain the lack of effect in number violations. Similarly, two studies on Spanish determiner-noun gender agreement found conflicting results regarding the LAN. Barber and Carreiras (2005) reported the LAN followed by the P600, whereas Wicha, Moreno, and Kutas (2004) only found the P600. The volatility of the LAN in terms of its seemingly random distribution across studies has not been explained yet even though several accounts have been offered.

In addition to the functional explanation by Hagoort and Brown (1999), Molinaro, Barber, Caffarra, and Carreiras (2014) indicated that methodological and technical factors could play a role regarding the LAN, such as the choice of the reference electrode. Osterhout (1997), Tanner (2015), and Tanner and Van Hell (2014) proposed that the presence of the LAN might be due to individual variations among participants (see also Pakulak & Neville, 2010), as well as to the averaging nature in obtaining ERP components.

4.1.3 Gender and Number Agreement in Spanish

Both number and gender are nominal phi-features. As such, they are often believed to affect processing mechanisms in the same way, to the extent that some studies lumped sentences with number and gender violations together as one syntactic condition (e.g., Hagoort, 2003b; Martín-Loeches, Nigbur, Casado, Hohfeld, & Sommer, 2006). Indeed, agreement studies investigating gender (e.g., Gunter et al., 2000; Molinaro et al., 2008) and number (e.g., Münte, Szentkuti, Wieringa, Matzke, & Johannes, 1997) separately often come to the same results and conclusions: both gender and number violations elicit the P600, which is sometimes preceded by the LAN. The processing mechanism is, hence, understood to be identical: the morphosyntactic violation is identified by the LAN already 300 ms post-stimulus onset, after which the violation is repaired, as indicated by the P600. However, in this way it is impossible to compare the effect size (e.g., amplitude size of the P600) between number and gender, which is as important as the presence/absence of a component or its distribution.

Barber and Carreiras (2005) tested Spanish determiner-noun and noun-adjective number and gender agreement in a single study. Gender in Spanish is a lexical feature, being part of the noun's lemma. However, unlike in Dutch, nouns in Spanish are overwhelmingly gender transparent. Most nouns end in either *-o* or *-a* indicating that they are masculine or feminine, respectively. If the noun is used in plural, a suffix *-s* is added onto the gender suffix. The study showed that both gender and number disagreement elicit the LAN and the P600. Crucially, number and gender disagreement differed in the late P600 stage, in which the effect was larger for gender. The authors suggested that repair processes in gender are costlier due to the lexical nature of gender (Faussart, Jakubowicz, & Costes, 1999; Ritter, 1991, 1993). However, Bañón, Fiorentino, and Gabriele (2012) looked into processing number and gender agreement in Spanish, and failed to find any difference. Their rationale was that the parser processes both features in a similar fashion regardless of their inherent differences (Nevins et al., 2007).

An important aspect of gender in Spanish is that it is almost always transparent, with the word-final vowel indicating gender. This fact is of great importance for reading studies and processing accounts based on reading. In an ERP reading paradigm on sentence processing, words are presented one-by-one on the screen with an average duration of 300-350 ms. This is enough just for one fixation (Rayner & Clifton, 2009), that is, the eyes will fixate the word as a whole and the visual system will perceive it as one unit. An integral part of that unit is the gender morpheme, as well as the number inflection. In other words, as soon as the system perceives the word, it has the word's gender and number information at its disposal. Therefore gender, just like number, can be available from the suffix in addition to retrieving its value from the lemma (see Caffarra & Barber, 2015; Caffarra et al., 2015). Of course, this is only true in case of gender transparent languages.

4.1.4 Predictions and Expectations

Regardless of the exact mechanism, overt morphology can affect how the gender feature is accessed. In order to eliminate this confound, we decided to test number and gender disagreement in Dutch. Barring several exceptions (e.g., diminutives), nouns in Dutch are mainly gender-opaque, whereas plural nouns are always morphologically marked. As mentioned earlier, the gender feature is invariable, meaning that a noun, such as *het boek*$_{N(euter)}$ 'the book' cannot become **de boek*$_{C(ommon)}$ under any circumstances. Number, however, is derived from higher order semantics and our knowledge about the numerosity of the object in question. If the object is singular, it is not morphologically marked (*het boek*), and if there is more than one object, its form becomes *de boeken*. Based on behavioural data, Lukatela, Kostić, Todorović, Carello, & Turvey (1987) proposed that the parser behaves in a binary way regarding syntactic violations. More precisely, the parser is only sensitive to the presence or absence of a violation, without any more detailed decomposition of the violation source. This is applicable to agreement mismatches, in which the parser detects the violation disregarding whether it is number, gender or case violation. In line with Lukatela et al.'s proposal are ERP results by Nevins et al. (2007), who found identical effects for both gender and number disagreement.

Based on the idea of parser's binarity, as well as on previous results, we expect that number and gender disagreement elicit the same syntactically-related components, that is, the LAN and P600. Functionally, the first effect should be the LAN as a marker of morphosyntactic incongruence. However, due to the lack of the LAN in previous studies on gender and number processing in Dutch (Loerts et al., 2013; Meulman et al., 2014), it is equally possible that the effect will be absent. Therefore, we are focusing on the P600 and its different stages. If the parser is sensitive to the way gender and number are encoded, that is, with if it responds differently to lexical versus morphological features, we expect to see an effect in the very early stage of the P600. Also, if gender and number are structurally integrated in a different way, this should be reflected in the early stage of the P600 (Hagoort & Brown, 2000; Kaan, Harris, Gibson, & Holcomb, 2000). The largest effect, however, is expected in the late stage of the P600 in which repair and reanalysis processes take place (Friederici, 2002; Hagoort & Brown, 2000). We predict that number disagreement is more complex to repair as it offers more reanalysis options. For example, in gender disagreement the neuter noun *boek* is preceded by an adjective marked for common gender **een* $grote_C$ $boek_N$ 'a big book'. The parser repairs the incongruity by correcting the gender inflection on the adjective (**een* $grote_C$ *boek* > een $groot_N$ $boek_N$). In number disagreement, in addition to repairing the singular article into the plural article (equivalent to repairing the gender inflection on the adjective) (*het_{SG} $boeken_{PL}$ > de_{PL} $boeken_{PL}$), the parser can also repair the inflectional morpheme on the noun (*het_{SG} $boeken_{PL}$ > het_{SG} $boek_{SG}$). In other words, the parser can apply only one operation on gender disagreement, but two operations on number disagreement. As a consequence of the increased processing load due to multiple repair processes in number, we expect that the P600 effect will be larger for number than gender disagreement.

4.2 Method

4.2.1 Participants

Thirty participants were tested for this experiment (8 male; mean age 22.1, age range 19-33). Out of the total number, 4 participants were excluded prior to analysis due to low scores on the grammaticality judgment task (accuracy below 80 %). Furthermore, 2 participants were excluded due to an excessive number of artifacts, leaving 24 participants whose results are reported in this study. All participants were right-handed, which was asserted with a Dutch version of the Edinburgh Handedness Inventory (Oldfield, 1971). The participants declared no reading impairment, or any history of psychiatric or neurological illness. They all had normal or corrected-to-normal vision. All the participants signed a consent form prior to the experiment, and received a € 20 voucher for their participation.

4.2.2 Acceptability Ratings for the Materials

An acceptability judgment survey was carried out in order to verify the grammaticality/ungrammaticality of experimental stimuli and fillers. Both grammatical and ungrammatical sentences were included in the survey. The question asked was: ‘Is this sentence grammatically correct?’ The offered answers were ‘yes’ and ‘no’. The instructions included an example. There were a total of 480 sentences divided over 4 lists. The survey was set up on an online platform SurveyMonkey (www.surveymonkey.com). In total, 32 native speakers of Dutch took the survey (9 male; mean age 33.18, age range 21-62). Each participant filled only one survey. In order to include a stimulus into the experimental set, it had to have an approval rate of at least 80%, that is, 80% of the participants had to correctly judge the sentence as grammatical or ungrammatical. Out of 480 sentences, 16 sentences did not reach the inclusion threshold. These sentences were modified and assessed by another 3 speakers of Dutch who unanimously approved the modified sentences.

4.2.3 Materials

The materials used in the experiment consisted of 320 experimental sentences and 160 fillers. The experimental sentences were created on a basis of 40 unique nouns, half of which were monosyllabic and the other half trisyllabic. The nouns were controlled for lemma frequency (Celex), noun-verb homophony, phonological alternations, and animacy. The average mean (log lemma frequency per million) of monosyllabic nouns was 1.36 (SD = .33), and of trisyllabic nouns 1.3 (SD = .35). Frequency of usage of monosyllabic and trisyllabic nouns was comparable ($t(38) = .53$, $p = .6$). In addition, all nouns had to use the suffix *-en* exclusively for the plural form, which in turn could not be homophonous to an infinitival verb form (e.g., *boek* ‘book’ > *boeken* ‘books’, but also ‘to book’). Nouns could not display any phonological alternations between the singular and the plural form (e.g, voicing *huis* ‘house’ > *huizen* ‘house’ or irregular plurals *stad* ‘city’ > *steden* ‘cities’), and the vowel length had to be maintained (e.g., no nouns with short-long vowel alternation, such as *pad* 'path' > *paden* 'paths'). Finally, all nouns were inanimate.

Each noun yielded 4 discrete sentences, with each sentence used once as grammatical and once as ungrammatical. If the full repertoire of nouns had been used, a participant would have been exposed to the same noun 8 times. In order to reduce this, as well as to make the experiment shorter, stimuli were divided over 2 lists. Each participant was exposed to only one list. Each target noun appeared 4 times in a list, always in a different sentence. Items were counterbalanced between lists in such a way that if a grammatical sentence was in the first list, its ungrammatical counterpart was in the second. Consequently, each participant read 160 experimental and 80 filler sentences.

All experimental sentences were divided into two conditions: gender (80 sentences per list) and number (80 sentences per list). In Dutch, the common gender article *de* is homophonous with the plural article *de*, which is used for both common and neuter nouns in plural. This could lead to a possible ambiguity in the gender violation condition, such as **de_C mes_N* 'the knife', in which the violation becomes apparent only at the end of the noun. Since *de* is always used as the definite plural article, the parser may expect to encounter *de messen* 'the knives'. Once it becomes obvious that there is no plural suffix *-en*, the parser registers a violation. Without any additional context, this violation is ambiguous between number and gender violation. More precisely, since *mes* is a neuter (*het*) noun, **de mes* may be recognized as a gender violation provided that *de* is reanalyzed as a singular common article. However, if the parser classifies *de* as a plural article, the violation is perceived as a number violation. We also do not rule out a possibility that the parser may perceive **de mes* as a double violation. In order to overcome this obstacle, gender violations were created between the indefinite article *een* (used only in singular for both genders) and a gender marked adjective followed by a noun; for example: **een* $mooie_C$ $dorp_N$ 'a beautiful village'. The indefinite article indicates to the parser that it should expect a singular noun. The following adjective (*mooie*) is marked with *-e*, which according to the inflectional rules indicates the noun has to be of common (*de*) gender. By combining the two pieces of information, the parser is ready to encounter a singular *de* noun. However, in the violated sentences the following noun is always a singular *het* noun. Therefore, a clear gender violation is created between the indefinite article-adjective complex and the target noun.

There was an additional reason for including only nouns of the *het* type. As already mentioned, Dutch gender is a lexical feature, that is, it has to be retrieved from the lemma. The only exception is nouns derived with a gender specific suffix, such as the diminutive suffix. Almost all Dutch nouns can have a diminutive form (derived through the suffix *–tje* and its allomorphs), the usage of which is also relatively high (Shetter, 1959). All diminutives are of *het* type (e.g., *de tafel* 'the table', *het tafeltje* 'the little table'). Consequently, in a violated condition, such as *een *rood*$_N$ *tomaat*$_C$ 'a red tomato' the parser recognizes the violation only at the end of the word. Again, the indefinite article means the noun should be singular, whereas the adjective form indicates the noun is of the *het* type. However, since almost any Dutch noun can be used as a diminutive, it is possible that the parser's strategy is to expect a diminutive noun. In order to avoid this possible strategy, we decided not to use *de* nouns as experimental items.

All gender sentences were created in two structural 'molds'. In the first mold (1), the sentence started with an expletive subject (e.g., *er* 'it/there') or a general place adverbial (e.g., *hier* 'here'). The subject was followed by a verb, either lexical or auxiliary, after which the indefinite article *een* 'a/one' was presented. The second sentence type (2) started with a personal pronoun (e.g., *hij/zij* 'he/she') followed by a verb, which was followed by the indefinite article. In both sentence types, the indefinite article was followed by an adjective. In non-violated sentences, the adjective was always inflected with the suffix -e. In violated sentences, the adjective had a zero marking used with *het* nouns. The target noun was placed after the adjective. Also, the target noun was never at the end of a sentence; it was always followed by a prepositional phrase, adverbial, or a lexical verb.

(1)	grammatical:	Er	lag	een	mooi	dorp	vlakbij	de	grote	stad.
		there	lay	a	beautiful$_N$	village$_N$	near	the	big	city
		'A beautiful village was close to the big city.'								
	ungrammatical:	*Er	lag	een	mooi**e**	dorp	vlakbij	de	grote	stad.
		there	lied	a	beautiful$_C$	village$_N$	near	the	big	city.

(2) grammatical: Zij verliet een prachtig dorp met pijn in haar hart.
she left a beautiful$_N$ village$_N$ with pain in her heart
'She left a beautiful village with an aching heart.'

ungrammatical: *Zij verliet een prachtig**e** dorp met pijn in haar hart.
she left a beautiful$_C$ village$_N$ with pain in her heart

All sentences in the number condition had the same structure (3, 4). They began with the plural article *de* in grammatical sentences and the singular neuter article *het* in violated sentences. The article was followed by an inflected adjective, which is the correct form for both singular *het* and plural *de* nouns after which the target noun was presented. Like in the gender condition, the noun was always followed by a prepositional phrase, adverbial phrase, or a lexical verb.

(3) grammatical: De gezellige dorpen trekken veel toeristen in de zomer.
the$_{PL}$ nice$_{PL}$ villages$_{PL}$ attract many tourist in the summer
'The nice villages attract many tourists in the summer.'

ungrammatical: *Het gezellige dorpen trekken veel toeristen in de zomer.
the$_{SG}$ nice$_{SG}$ villages$_{PL}$ attract many tourist in the summer

(4) grammatical: De noordelijke dorpen hebben last van aardbevingingen.
the$_{PL}$ northern$_{PL}$ villages$_{PL}$ have trouble from earthquakes
'The northern villages are troubled by earthquakes.'

ungrammatical: *Het noordelijke dorpen hebben last van aardbevingen.
the$_{SG}$ northern$_{SG}$ villages$_{PL}$ have trouble from earthquakes

We explained the reasons for having only *het* nouns as experimental items. A consequence of this choice is that it can facilitate a learning strategy in participants. In other words, when a participant encounters a sentence starting with *het*, he may learn after a few items that such sentences are always ungrammatical. Similarly, any instance of a zero-marked adjective indicates an ungrammatical sentence. To prevent this possibility, 160 filler item sentences with a pattern reversed to that of the experimental stimuli were included. Half of the filler items contained *het* nouns (5) and the other half *de* nouns (6). The *de* noun group was used to counterbalance the gender condition, whereas the *het* noun-group counterbalanced the number condition.[4] Consequently, it was impossible to judge a sentence as grammatical or ungrammatical based only on the article or the adjective. The participant had to pay attention to everything preceding the noun, as well as the noun itself, in order to correctly judge the sentence.

(5)	grammatical:	Er	ligt	een	rotte	tomaat	in de	koelkast.
		there	lie	a	rotten$_{C}$	tomato$_{C}$	in the	fridge
		‘There is a rotten tomato in the fridge.’						
	ungrammatical:	*Er	ligt	een	rot	tomaat	in de	koelkast.
		there	lie	a	rotten$_{N}$	tomato$_{C}$	in the	fridge

(6)	grammatical:	Het	oude	paspoort	is niet	meer geldig.
		the$_{N.SG}$	old$_{N.SG}$	passport$_{N.SG}$	is not	more valid
		‘The old passport is not valid anymore.’				
	ungrammatical:	*De	oude	paspoort	is niet	meer geldig.
		the$_{C/PL}$	old$_{C/PL}$	passport$_{N.SG}$	is not	more valid

4.2.4 Procedure

Participants were seated in front of the screen at a distance of 70 to 80 cm. The experiment was presented in E-Prime (Psychology Software Tools, Inc). The passive task was to read the sentences presented word-by-word on the screen. The active task was to reply to a randomly assigned grammaticality judgment question. On average, a question appeared once for every five sentences. The purpose of the active task was to keep participants focused. The experiment opened with written instructions that were repeated by the experimenter. There was a brief practice session (4 sentences) in order to ensure that participants had understood the instructions and were able to follow the stimulus presentation. After the last practice item,

[4] The violation in *het* nouns was ambiguous between number and gender. However, this was not important since the only purpose of the fillers was to prevent a learning strategy disregarding the exact nature of the violation.

participants had a chance to ask for clarifications or more detailed instructions. Once they were ready, they could proceed to the experimental part by pressing any keyboard button. Each trial opened with a fixation cross (500 ms) and a break (200 ms), after which the first word was presented (400 ms). The stimulus onset asynchrony (the time between the onsets of two subsequent words) was 600 ms. The last word in the sentence was presented with a full stop. Sentences were shown on a black background with white letters. The letter font was Arial and letter size was 24 pt. After the last word was presented, the screen remained blank for 500 ms. In case there was a grammaticality judgment question, a question mark appeared after the 500 ms break. The question remained on the screen for 3 s, during which participants were supposed to press either ‘p’ or ‘q’ (counterbalanced across participants), depending on whether the previous sentence was grammatical or not.

The experiment lasted approximately 30 min. Stimuli were divided into 4 blocks, each containing 40 experimental and 20 filler items. There were 12 grammaticality judgment questions per block, 8 for experimental and 4 for filler items. The presentation order within a block was random, as determined by the software. The participants were advised to take a short break after each block.

4.2.5 EEG Data Acquisition and Processing

Continuous EEG data were recorded using the ASA-Lab system (ANT Neuro Inc, Enschede, The Netherlands) from 64 Ag/AgCl scalp electrodes fitted in an elastic cap (WaveGuard). Electrodes were positioned according to the extended 10-20 system. Eye movements were recorded using one bipolar channel for horizontal movements (HEOG; the electrodes were placed at the outer canthus of the eyes) and one for vertical movements (VEOG; placed above and below the left eye). Electrode impedances did not exceed 10 kΩ, and were kept at 5 kΩ or below in the large majority of cases. Data were sampled at 512 Hz with the common average reference.

Data were pre-processed with Brain Vision Analyzer 2.04 (Brain Products, GmbH, Munich, Germany). The first step was to down-sample the data to 256 Hz in order to speed up the analysis; this was followed by re-referencing the offline data to the average of the left and right mastoid. Afterwards, a band-pass filter was applied (0.1 – 40 Hz) after which an automatic ocular PCA-based correction was performed. The continuous data were segmented into 1700 ms long epochs, starting 200 ms before the trigger marker (target noun onset). The automatic artifact rejection (+/- 100 μV threshold, minimal activity 0.1 μV) was performed in the interval of -100 ms to 1000 ms for each epoch. Electrodes with a high artifact contamination rate (>20%) were interpolated (1 electrode in 5 participants). Finally, the baseline correction was applied starting -100 ms until 0 ms after which data were averaged per subject and per condition. If a participant had fewer than 70 % averaged trials in one or more conditions, his data were excluded from the analysis. This resulted in excluding the data of 2 participants.

4.2.6 Analysis

For the analysis, we used averaged participant values (in μV) per condition, level of grammaticality, and regions of interest (ROI). Regions of interest (Fig. 4.1) were created by averaging the values of 5 to 6 adjacent electrodes (50 in total), which resulted in 9 ROIs: left anterior (F7, F5, F3, FC3, FC5), midline anterior (F1, Fz, F2, FC1, FCz, FC2), right anterior (F4, F6, F8, FC4, FC6), left central (TP7, C5, C3, CP5, CP3), midline central (C1, Cz, C2, CP1, CPz, CP2), right central (C4, C6, CP4, CP6, TP8), left posterior (P7, P5, P3, PO7, PO5, O1), midline posterior (P1, Pz, P2, PO3, POz, PO4), and right posterior (P4, P6, P8, PO6, PO8, O2). For the statistical analysis, 4 time windows were created based on the literature and visual data inspection: 1) 300 – 450 ms corresponding to the LAN/N400; 2) 450 – 600 ms onset of the P600; 3) 600 – 800 ms early P600; 4) 800 - 1000 late P600.

The following within subject factors were included in a repeated measure ANOVA: 1) condition (2 levels: gender and number); 2) grammaticality (2 levels: grammatical and ungrammatical); 3) hemisphere (2 levels: left and right); 4) anteriority (3 levels: anterior, central, and posterior). The global analysis for each time window was performed by two separate ANOVAs. The first ANOVA analyzed only lateral the regions and it included all 4 factors. A second omnibus ANOVA was run on the midline regions only, excluding the factor hemisphere. In case the assumption of sphericity was violated, the Geisser and Greenhouse (1959) correction was applied. The significance level was set to $p < .05$. Follow-up ANOVAs were performed only for interactions that were at least marginally significant ($p < .1$) and that included factor grammaticality. Finally, incorrectly judged trials were not included in the analysis.

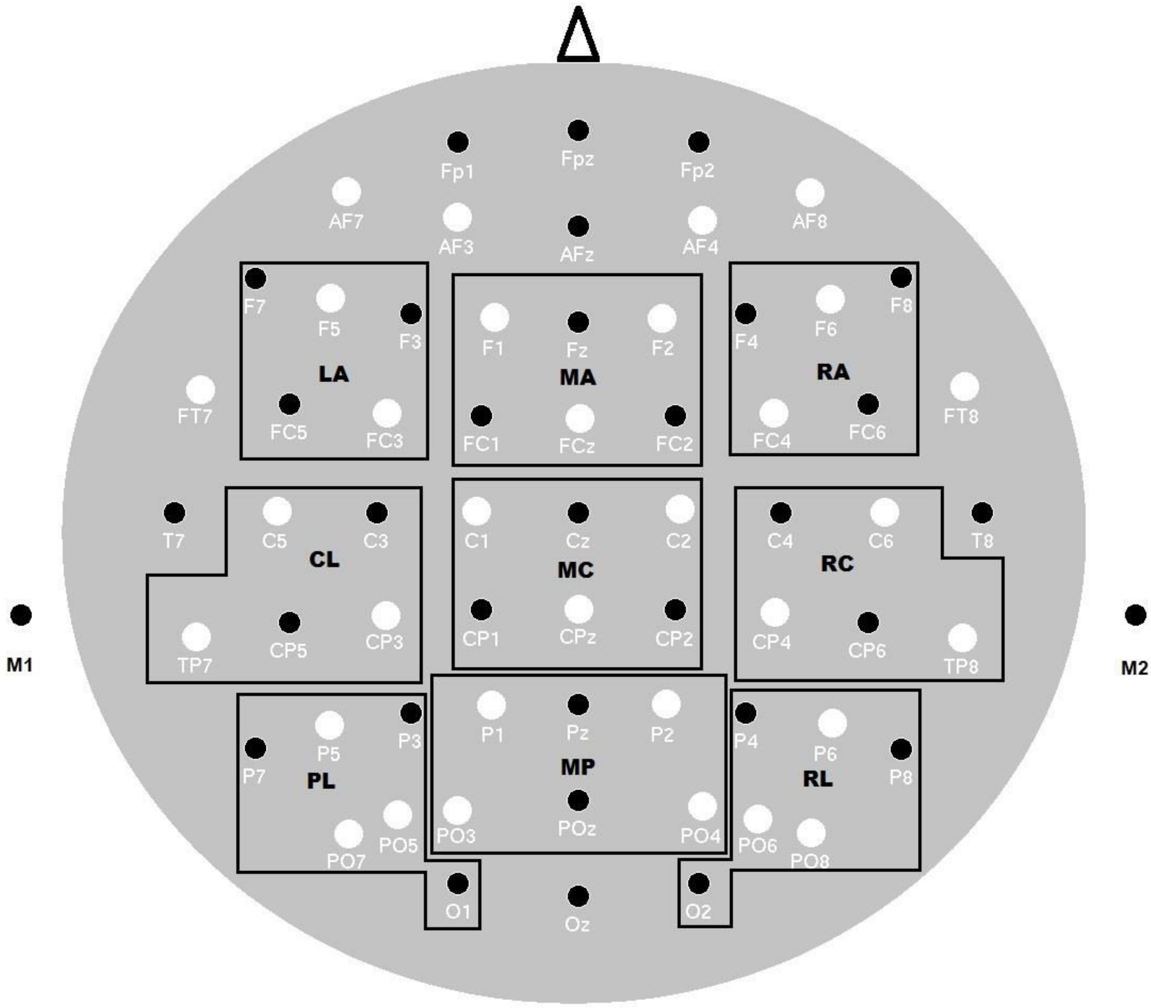

Fig. 4.1 Electrode positions and the 9 regions of interest used in the analysis: left anterior (LA), left central (LC), left posterior (LP), midline anterior (MA), midline central (MC), midline posterior (MC), right anterior (RA), right central (RC), right posterior (RP).

4.3 Results

4.3.1 Accuracy Results

The cut-off for including a participant in the data analysis was set at 80 %, meaning that each participant had to correctly respond to 51 out of 60 questions. Of the 30 participants, 4 were excluded due to a low score on the grammaticality judgment task. The remaining 26 participants had an accuracy rate of 94% (average number of errors: 3.6, SD 2.6). Four participants performed at ceiling. Since the only purpose of the grammaticality judgment question was to ensure the participants' alertness throughout the entire experiment, these data were not further analyzed.

4.3.2 ERP Results

A visual inspection of the waveforms indicated a centro-parietal positive effect from approximately 500 ms post-stimulus onset. The effect was caused by ungrammatical sentences in both conditions. The positivity seemed to display somewhat larger amplitude, as well as an earlier onset, in the number condition (Fig. 4.2) compared to the gender condition (Fig. 4.3). Contrary to expectations, the positivity was not preceded by a left-lateralized negative effect in either condition.

The first time window (**300-450 ms**) did not yield any significant effects or interactions. This holds true for both the lateral and midline analysis. Since the LAN was expected in this time window, we performed a hypothesis-driven ANOVA in the left anterior region only. Still, the factor grammaticality did not reach significance ($F(1, 23) = .278$, $p = .603$).

In the following time window (**450-600 ms**), the omnibus ANOVA on the lateral regions revealed a main effect of grammaticality ($F(1, 23) = 4.313$, $p = .049$), with ungrammatical sentences showing a more positive waveform. Since visual inspection, as well as voltage values, did not indicate any effect for gender, we performed two additional ANOVAs on gender and number separately. The main effect of grammaticality was confirmed for number ($F(1, 23) = 5.503$, $p = .028$), but was absent for gender ($F(1, 23) = .064$, $p = .803$).

The midline results mirrored the lateral results. The global midline ANOVA produced a main effect of grammaticality ($F(1, 23) = 8.448$, $p = .008$). However, visual inspection and voltage values failed to show an effect of grammaticality in both conditions despite the lack of an interaction between condition and grammaticality ($F(1, 23) = 1.083$, $p = .309$). The assumption that the positive effect was present in the number condition only was confirmed by separate tests for gender (grammaticality: $F(1, 23) = 2.218$, $p = .15$) and number (grammaticality: $F(1, 23) = 7.555$, $p = .011$).

The positive effect continued into the **600-800 ms** time window, with ungrammatical sentences in both conditions yielding more positive waveforms (grammaticality: $F(1, 23) = 10.45$, $p = .004$). In addition, the effect of grammaticality interacted with anteriority ($F(2, 46) = 14.22$, $p = .001$), and it also entered into a marginal three-way interaction with anteriority and hemisphere ($F(2, 46) = 3.283$, $p = .066$). Further follow-up tests along the anterior-posterior axis confirmed the main effect of grammaticality in the central ($F(1, 23) = 10.481$, $p = .004$) and posterior regions ($F(1, 23) = 20.312$, $p < .001$). In the anterior regions, grammaticality interacted marginally with condition and hemisphere ($F(1, 23) = 3.637$, $p = .069$). However, the follow-up tests did not produce any significant results. Lastly, an interaction between grammaticality and hemisphere turned out to be marginally significant in the central regions ($F(1, 23) = 4.247$, $p = .051$), indicating a larger effect in the right hemisphere.

The midline analysis revealed a main effect of grammaticality ($F(1, 23) = 13.519$, $p = .001$), which also interacted with anteriority ($F(2, 46) = 22.123$, $p < .001$). Just like in the lateral analysis, the positive deflection was statistically confirmed in the central ($F(1, 23) = 13.694$, $p = .001$) and posterior regions ($F(1, 23) = 25.582$, $p < .001$).

Finally, the positive deflection caused by ungrammatical sentences in both conditions persisted in the latest time window of **800-1000 ms**, but only in an interaction with anteriority ($F(2, 46) = 43.49$, $p < .001$) and hemisphere ($F(1, 23) = 16.602$, $p < .001$) or both ($F(2, 46) = 5.581$, $p = .016$). The main effect of grammaticality was only marginally significant ($F(1, 23) = 3.572$, $p = .071$). In addition to being more lateralized and more narrowly distributed in this time window, the positive effect had a tendency towards being stronger in the number condition (condition × grammaticality: $F(1, 23) = 3.674$, $p = .068$). Two sets of follow-up ANOVAs were thus performed based on the significant and close-to-significant interactions in the omnibus ANOVA: for each hemisphere and for each level of anteriority.

A follow-up test focusing on the left hemisphere revealed an interaction between anteriority and grammaticality ($F(2, 46) = 38.68$, $p < .001$). This interaction was also obtained in the right hemisphere ($F(2, 46) = 35.28$, $p < .001$), in which a main effect of grammaticality was also recorded ($F(1, 23) = 7.243$, $p = .013$). In addition, the positive deflection in the number condition was stronger in amplitude than in the gender condition in the right hemisphere, as demonstrated through an interaction between condition and grammaticality ($F(1, 23) = 5.171$, $p = .033$). A second set of follow-up tests for each level of anteriority confirmed a positive deflection in the central (grammaticality: $F(1, 23) = 5.582$, $p = .027$) and posterior regions ($F(1, 23) = 21.654$, $p < .001$). The effect of grammaticality interacted with hemisphere in the anterior ($F(1, 23) = 16.302$, $p = .001$) and central regions ($F(1, 23) = 22.543$, $p < .001$). However, this interaction did not indicate a lateralization of the positive effect in the anterior regions. Rather, it was caused by a left-lateralized negativity in the anterior regions (grammaticality: $F(1, 23) = 6.864$, $p = .015$). In contrast, the positive effect reached the central regions with larger amplitude in the right hemisphere. Finally, ungrammatical sentences in the number condition elicited a significantly larger effect than the gender condition in the posterior regions as confirmed by an interaction between condition and grammaticality ($F(1, 23) = 4.83$, $p = .038$). This interaction was marginally significant in the central regions ($F(1, 23) = 4.024$, $p = .057$).

The midline analysis revealed a main effect of grammaticality ($F(1, 23) = 5.087$, $p = .034$), as well as in an interaction between grammaticality and anteriority ($F(2, 46) = 48.22$, $p < .001$). Most importantly, a significant interaction surfaced between condition and grammaticality ($F(1, 23) = 4.47$, $p = .046$) with ungrammatical sentences in the number condition causing a significantly larger positive effect than ungrammatical sentences in the gender condition. A follow-up analysis for each level of anteriority was carried out in order to localize the effect. The anterior analysis did not yield any significant results, which proved that the positive effect did not reach the anterior region. In contrast, a main effect of grammaticality was present in both the central ($F(1, 23) = 6.539$, $p = .018$) and posterior region ($F(1, 23) = 24.781$, $p < .001$). Finally, condition and grammaticality interacted significantly in the central region ($F(1, 23) = 4.382$, $p = .048$), and almost reached significance in the posterior region ($F(1, 23) = 4.07$, $p = .055$).

4.3.3 Summary of ERP Results

The statistical analysis confirmed a significant centro-parietal positivity elicited by ungrammatical sentences (P600). The P600 was not preceded by either a left lateralized negativity (LAN) or a central negativity (P600). The onset of the P600 was earlier in the number condition (from 450 ms on) than in the gender condition (from 600 ms on). Its distribution included all posterior and central regions, favouring the right over the left hemisphere. The positive effect was larger for the number condition in the last time window, particularly in lateral posterior regions, as well as in the midline central region. The analysis also showed the presence of a left-lateralized negativity in the last time window (800-1000 ms). Due to its late onset, the negative effect may have been elicited by the word following the target word.

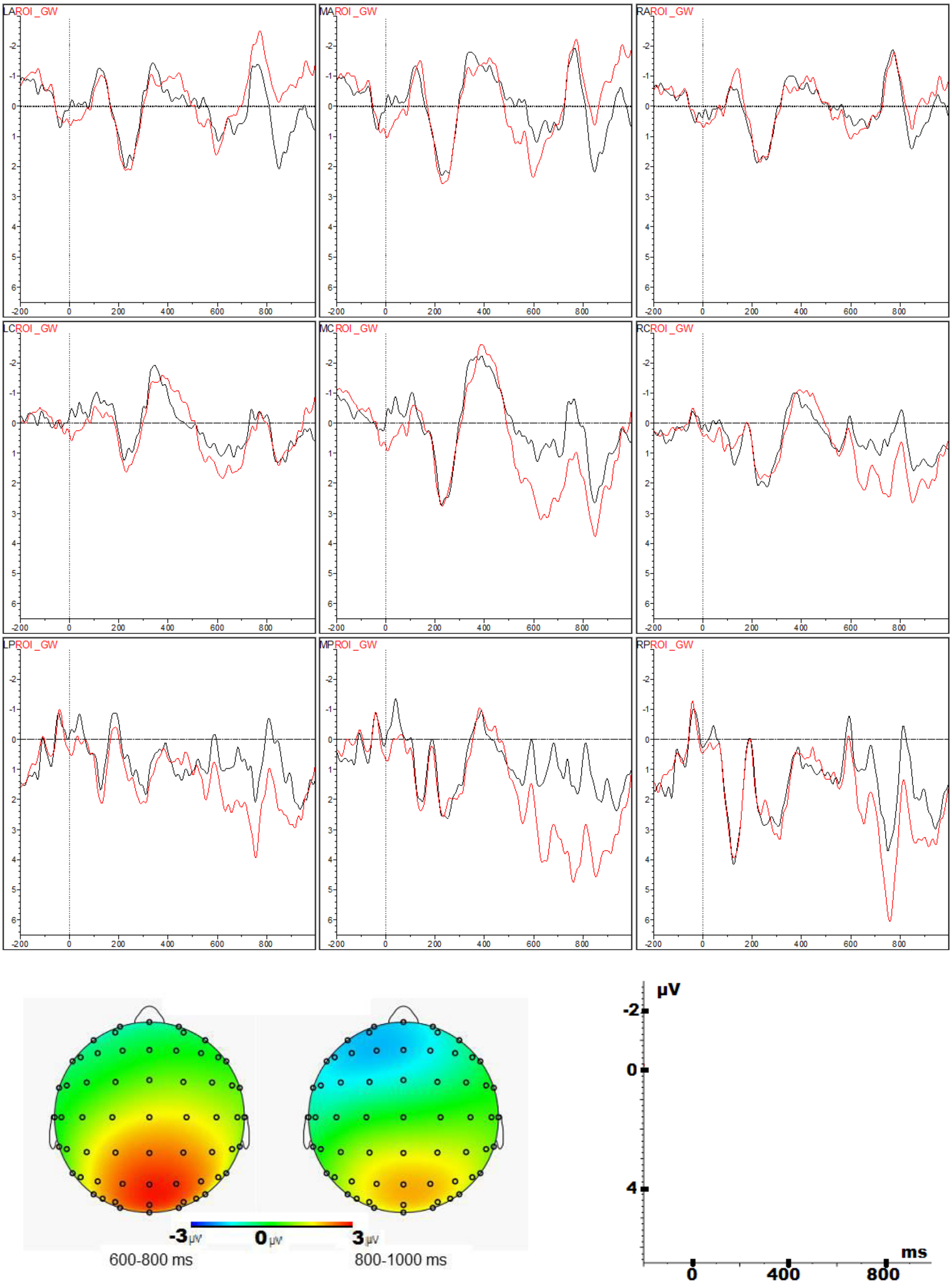

Fig. 4.2 Grand average ERPs for the gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences. The topographic maps represent a difference between ungrammatical and grammatical sentences in the two P600 windows.

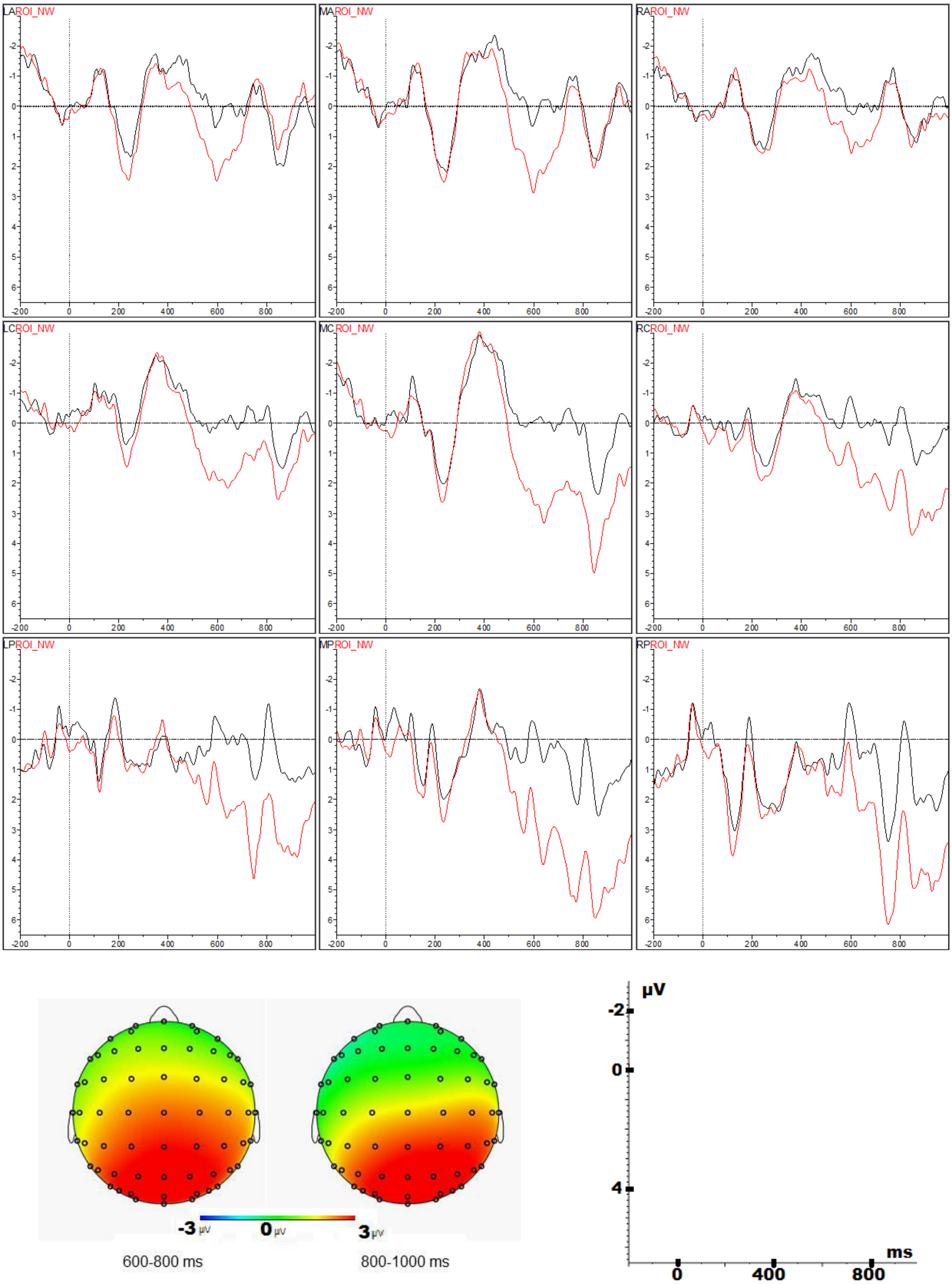

Fig. 4.3 Grand average ERPs for the number condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences. The topographic maps represent a difference between ungrammatical and grammatical sentences in the two P600 windows.

4.4 Discussion

We tested processing of gender and number disagreement between the article, the adjective and the noun in Dutch. Results are in line with previous research in that a robust posterior positivity (P600) was elicited by ungrammatical sentences (Barber et al., 2004; Gunter et al., 2000; Molinaro et al., 2008). The P600 is interpreted as a stage in which repair and reanalysis take place (Friederici, 2002; Hagoort & Brown, 2000). It is often preceded by the LAN, which arises in response to morphosyntactic violations (Friederici, 2002; Molinaro et al., 2011; Molinaro et al., 2014). However, even though the current experiment contained a morphosyntactic violation, it failed to elicit the LAN.

4.4.1 Lack of Biphasic Response

A number of authors talk about a biphasic response to agreement violations, that is, the LAN followed by the P600 (e.g., Barber & Carreiras, 2005; Molinaro et al., 2011). In the first stage (LAN), the parser automatically identifies the morphosyntactic violation, after which it tries to integrate and repair it (P600) during the late syntactic stage (Friederici, 2002). The presence of the P600 has been reported in almost all agreement studies, which is not the case with the LAN. As an illustration, Barber and Carreiras (2005) reported the LAN followed by the P600 in number and gender disagreement in Spanish, both for the article-noun and noun-adjective pairs. Conversely, Wicha et al. (2004) tested gender disagreement between article-noun in Spanish, which produced only the P600. Similarly, Bañón et al. (2012) tested adjective-noun and noun-adjective disagreement in Spanish, neither of which elicited the LAN.

There are at least three accounts that attempted to explain the inconsistency in obtaining the LAN. Hagoort and Brown (1999) suggested that the LAN is sensitive only to phonologically overt morphosyntactic violations. In terms of the present study, this means that the LAN should have been recorded in the number condition only. Since gender is a lexical feature in Dutch, there was no inconsistency between the article/adjective and the noun's inflectional morphology for gender as there was none. However, the number violation was a typical example of a determiner disagreeing with the noun because of the noun's number morphology (e.g., *het boek* 'the book', **het boek-en* 'the books'). Therefore, our findings refute the proposed account.

Molinaro et al. (2014) proposed that methodological factors may influence whether or not the LAN can be elicited or not. For example, the reference choice may play a role, since studies using the left mastoid as the reference reported the LAN less frequently than studies using the average of the mastoids. This account is not supported by the current experiment since we used the average of the mastoids. As for other methodological factors, we used a fairly common number of stimuli per condition (80 per condition), as well as the often-used stimulus asynchrony (600 ms). Thus, failure to elicit the LAN cannot be accounted for by either the Hagoort and Brown (1999) or Molinaro et al. (2014) account. A third account, proposed by Osterhout (1997), focuses on individual differences and ERP response. According to this hypothesis, some people react to violation either in the form of a positive (P600) or negative deflection (N400). Once the individual data are averaged, they result in a P600, which is sometimes preceded by the LAN. The LAN is seen as a residual distribution of the negative deflection that some participants exhibit and that was cancelled out in other regions by the P600 (see also Tanner & Van Hell, 2014; Tanner, 2015). Even though the data may be compatible with the account, any claim in its favour or against it would require an in-depth analysis of individual data, which is outside the scope of this study.

Regardless of the absence of the LAN, current results are in line with a large body of studies that failed to elicit the biphasic pattern, and only yielded the P600 (e.g., Bañón et al., 2012; Nevins et al., 2007; Wicha et al., 2004). The theoretical issues regarding the LAN are beyond the scope of this chapter and call for further research in terms of both psycholinguistic theory and methodology.

4.4.2 P600 in Gender and Number Disagreement

Due to all the issues associated with the LAN, the main hypotheses of the current study relied on the P600. We predicted that both kinds of violation would elicit the P600, with possible differences in distribution or amplitude. Differences were, indeed, recorded in two out of three time windows in which the P600 was inspected. Firstly, the effect started earlier in the number condition (from 450 ms on) than in the gender condition (from 600 ms on) to become indistinguishable in the two conditions in the 600-800 ms time window. Secondly, the amplitude of the P600 became significantly larger in the number condition than in the gender condition in the 800-1000 ms time-window. The difference was particularly pronounced in the lateral posterior regions, as well as in the midline central region.

We predicted that the parser is sensitive to two differences between gender and number disagreement. The first difference is related to the way each feature is encoded. In Dutch, gender is mainly a lexical feature, whereas number is an inflectional feature. We propose that the onset of the P600 may be modulated by the way the targeted feature is encoded. Barber and Carreiras (2005) suggested that accessing gender is costlier than accessing number due to the lexical (gender) – morphological (number) opposition. Namely, retrieving a lexical feature is a more arduous process than decoding a feature from a suffix. If this is indeed the case, it is possible that the parser is more delayed in the P600 stage, as it needs to go back to the lemma to retrieve gender whereas a simple morphological decomposition suffices for decoding number.

Interestingly, the P600 in the ensuing time window was identical in both conditions. This time window corresponds roughly to the early stage of the P600, which is stipulated to represent integration processes (Hagoort & Brown, 1999). Since no difference was detected, we assume that both gender and number are integrated in the same manner. In spite of a delay in the gender condition, which was probably due to the lexical nature of gender, the parser leans towards a binary behaviour (Lukatela et al., 1987). That is, it seems that the parser is only sensitive to the presence/absence of an agreement violation. There is no evidence that the parser takes into account the fact that gender is a formal lexical feature, whereas number is based on semantics.

4.4.3 P600 as a Marker of Repair/Reanalysis

Our final prediction regarded repair and reanalysis processes that are believed to take place in the late P600 stage (Hagoort & Brown, 1999). Namely, due to the inflectional nature of number, number disagreement should be a more complex process to repair as it offers two repair options compared to a single repair option in gender. As an illustration, article/adjective-noun gender disagreement can trigger only one repair process in which the adjective has to be repaired (**een grote*$_C$ *boek* > *een groot*$_N$ *boek*$_N$ ‘a large book’). This is equivalent to the first repair option in gender in which the preceding singular article is repaired into the plural article (*het*$_{SG}$ *boeken*$_{PL}$ > *de*$_{PL}$ *boeken*$_{PL}$ 'the books'). However, the parser has an additional repair option at its disposal, and that is repairing (deleting) the plural suffix on the noun (*het*$_{SG}$ *boeken*$_{PL}$ > *het*$_{SG}$ *boek*$_{PL}$ 'the book'). The double repair option in number may be more demanding than a single repair in gender. The increase in processing demand was expected to be reflected as a difference in the late stage of the P600, either as higher amplitude or a broader distribution in the number condition. The former scenario turned out to be true in the current study. These results are not in line with what Barber and Carreiras (2005) reported for number and gender disagreement in Spanish. They found that the P600 in its late stage was larger for gender, which they attributed to the lexical nature of gender which makes it more difficult to process. However, such results were not replicated by Bañón et al. (2012) in a similar study on Spanish in which no difference between gender and number disagreement was reported. The only explanation we can offer is that gender is most often realized as a transparent morpheme in Spanish (*-o* for masculine and *–a* for feminine) on top of which a plural number suffix can be added. It is, thus, plausible that reanalysis and repair processes are affected in a different way in Spanish and Dutch due to the morphological nature of gender.

The results of this study are in line with the majority of agreement studies. The experimental manipulation (article/adjective-noun gender or number disagreement) elicited the P600. The LAN was not observed in the current study, which is in line both with a large number of studies (e.g., Hagoort & Brown, 1999; Wicha et al., 2004), but simultaneously at odds with equally substantial body of research (e.g., Barber & Carreiras, 2005; Molinaro et al., 2011). Clearly, the issue is still controversial and requires further research. The lack of the LAN seems to be part of a larger controversy revolving around this component that is well worth further research. The effects observed in the current study were reflected in the onset time and amplitude difference of the P600 component. Namely, the P600 had an earlier onset in the number condition than in the gender condition. This, we believe, is a consequence of the morphological nature of number as opposed to the lexical nature of gender. Finally, number disagreement elicited a P600 with significantly stronger amplitude in its late stage compared to gender disagreement. We interpret this result as a marker of increased processing load due to more complex repair and reanalysis processes in number.

Chapter 5
Gender and Number Agreement Processing in Listening: An ERP Study on Dutch

Introduction 5.1

The previous chapter focused on processing gender and number agreement within the noun phrase in Dutch. We hypothesized that number disagreement is more complex to repair than gender disagreement due an extra repair step in number processing. The results seem to point in this direction.

Both the number and gender condition elicited the P600 component. Its onset was earlier in number than in gender, probably reflecting the difference in the inflectional nature of number as opposed to the lexical nature of gender. Crucially, the amplitude of the P600 differed between the conditions in the last time window. In the number condition, the amplitude of the P600 was significantly larger than in the gender condition, particularly in the posterior regions. We interpret this difference as a marker of increased processing demand due to a more complex repair mechanism in number than in gender.

In this chapter, we replicated the experiment from Chapter 4 as a listening study. The first goal was to investigate gender and number processing in Dutch in the auditory modality. However, the main focus was to investigate the relationship between ERP components and the presentation modality. Thus, the bulk of this study was dedicated to examining the characteristics (e.g., onset time) of the disagreement-elicited ERP components, and how these characteristics vary depending on the presentation modality.

5.1.1 Presentation Modality in ERP Sentence Processing Studies

Event-related potentials (ERPs) are a widely employed technique for studying sentence processing (Hagoort, Brown, & Osterhout, 1999; Kaan, 2007; Kutas & Federmeier, 2007; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004). In this subfield of psycholinguistics, ERPs are mainly used in comprehension experiments, both in a visual (reading) and an auditory (listening) modality. The current experiment tested gender and number disagreement in the Dutch noun phrase in the auditory modality. In addition to investigating which language-related ERP components are elicited by disagreement, we were also interested in determining the exact onset time of the components. For example, we investigate whether the onset of a component is tied to the word onset or a syntactic recognition (uniqueness) point. This is only possible in a listening paradigm since listening is an incremental process in which phonemes are perceived one-by-one (Raynar & Clifton, 2009). This means that the listener is first able to access the word's semantic information, as well as lexical syntax, followed by whichever grammatical information is expressed through an inflectional morpheme. Conversely, written words are fixated and processed as a whole, a consequence of which is that all different pieces of information conveyed by the written word can be accessed simultaneously (Molinaro, Barber, & Carreiras, 2011).

There are only few studies that directly compared the two presentation modalities in an ERP experiment using the same stimuli (e.g., Balconi & Pazzoli, 2010; Friederici, Hahne, & Mecklinger, 1996; Hagoort & Brown, 2000; Meulman, Stowe, Sprenger, Bresser, & Schmid, 2014). This is surprising knowing that some of these studies have reported different effects for reading and listening (e.g., Hagoort & Brown, 2000). Therefore, the second goal of the current study is to examine the degree of consistency and overlap in the experimental outcome in relation to the presentation modality. In Chapter 4 we investigated determiner/article-noun gender and number disagreement in written Dutch. We used part of the stimuli from that experiment for the current study.

5.1.2 Reading vs. Listening

Both reading and listening have been employed in sentence comprehension paradigms. Each modality has its own advantages and disadvantages, yet reading seems to be somewhat preferred over listening. One of the reasons is that the natural flow of speech cannot be controlled, at least not to a large extent, meaning that words come in succession without a controlled pause. Therefore, it may sometimes seem unclear if a late effect was elicited by the target word or the following word. Contrary to auditory experiments, the presentation rate in reading can be controlled, which makes a neat design possible in which words are presented one-by-one always at the same pace. However, this results in a non-ecological stimulus presentation. In addition to being presented one-by-one, words are usually presented at a slow pace, for example, one word every 600 ms. Such a presentation rate is very slow for an average reader (Hagoort & Brown, 2000).

One aspect of listening versus reading seems to be overlooked often. Namely, during reading, our eyes are able to fixate 8-9 letter positions simultaneously in just 200 ms; this means that an average word including word-final inflectional morphology is processed in one fixation. In listening, however, the speech signal is delivered phoneme-by phoneme, meaning that the word stem (e.g., *interpret*) becomes available before the inflectional morpheme (*interpret*-s). Consequently, a trisyllabic word can be visually processed in only 200 ms, whereas it takes approximately 500 to 600 ms when the same word is presented auditorily.

According to the Cohort Model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980), the listener does not need to hear a word in its entirety in order to identify it. This model postulates that the system activates an initial cohort when it perceives the first phoneme. This cohort consists of words beginning with that phoneme. With each subsequent phoneme, the system discards all phonologically irrelevant combinations. Once there are no more possible combinations (phonological neighbours), the system selects the correct word. This often happens before all the phonemes have been heard, particularly in longer words (Radeau, Morais, Mousty, & Bertelson, 2000). As an illustration, the listener only needs to hear */spəg-/* in order to deduce that the target word is *spaghetti*, as there are no other English word beginning with the same sequence (Brysbaert & Dijkstra, 2006). The point at which there are no more phonological neighbours, and there is only one word that can be selected is called the Recognition Point (RP). Once the system has reached the RP, it can access the word's lemma. The lemma contains the word's meaning (semantics) but also lexical syntactic information, such as word class (Levelt, Roelofs, & Meyer, 1999).

A number of studies have looked into the application of the Cohort Model during real time auditory processing and ERPs (e.g., O'Rourke & Holcomb, 2002; Van den Brink, Brown, &

Hagoort, 2006). A component has been recognized, which arises in response to the contextual influence on cohort building and lexical selection. That is, if all the phonological neighbours in a cohort contradict the context semantically, the N200 arises before the N400, which marks the incongruence of the final lexical choice (Connolly & Phillips, 1994; Hagoort, 2008). However, the N200 can only be elicited auditorily, as there is no cohort building in a visual paradigm. In addition, O'Rourke and Holcomb (2002) demonstrated that the onset of the N400 correlates with the word's RP rather than always peaking at 400 ms.

Since gender and number in Dutch are coded lexically and morphologically, respectively, they are conveyed at different moments in time. Therefore, we can say that their syntactic recognition points do not align, with the syntactic RP of number always following that of gender. In more concrete terms, an average duration of the 3-syllable words used in the current experiment is 600 ms, whereas it is 700 ms for the same words in plural. Bearing in mind that the gender RP is often recognized before the end of the word (Radeau et al., 2000), the difference in time between accessing gender and number can be 200 ms or even more; this means that there is a large period of time during which significant temporal information is lost if stimuli are presented visually. The current experiment will exploit the large difference in time between accessing gender and number in an attempt to find out more about processing mechanisms, and to achieve a more precise timing of the language-related ERP components.

5.1.3 Agreement in Auditory ERP Studies

In line with the majority of reading studies on agreement, all auditory studies reported a centro-parietal positive deflection, peaking approximately 600 ms after the onset of the stimulus (P600; Balconi & Pazzoli, 2010; Demestre, Meltzer, Garcia-Albea, & Vigil, 1999; Friederici et al., 1996; Hagoort & Brown, 2000; Loerts, Stowe, & Schmid, 2013; Meulman et al., 2014). This component is believed to represent a repair and reanalysis stage and is often considered to be a marker of sentence processing (Friederici, 2002). As such, it is almost always elicited in both reading and auditory studies. For example, Loerts et al. (2013) investigated determiner-noun gender agreement in Dutch. The auditorily presented stimuli elicited a positive deflection, peaking at approximately 600 ms. Similarly, in Meulman et al. (2014) determiner-noun gender disagreement elicited the P600 in both native speakers and language learners of Dutch.

Another component often associated with agreement processing is the left-lateralized negativity (LAN). It precedes the P600, and is usually reported in the time-window of 300 to 500 ms after the onset of the stimulus. It arises in response to a morphosyntactc violation (Friederici, 2002; Molinaro, Barber, & Carreiras, 2011). Hagoort and Brown (1999) suggested that the morphosyntactic error has to be in the form of an erroneous inflectional morpheme, thus eliminating syntactic errors related to lexical syntax (e.g., gender). However, the presence of the LAN in auditory studies is as inconsistent as in reading studies. In other words, unlike the P600, the LAN is not a reliable marker of a morphosyntactic violation, such as disagreement. In addition to its volatile nature, the LAN in auditory studies tends not to be clearly differentiated in topographic terms. That is, the negativity preceding the P600 is often either not lateralized or not confined to anterior regions only. For example, Demstre et al. (1999) reported a broad negativity in the anterior and central regions in PRO-adjective gender disagreement. Since the negativity was elicited by a morphosyntactic violation and it preceded the P600, it was functionally interpreted as the LAN. In a study on Dutch, Hagoort and Brown (2000) tested three conditions, each with a different morphosyntactic violation (agreement, subcategorization, and phrase structure violation). Only agreement and subcategorization violations elicited anterior negativities. Interestingly, the anterior negativity in the agreement condition was left lateralized, whereas it was undifferentiated in the subcategorization condition.

Several studies investigated whether reading and listening elicit identical components with identical or similar characteristics using the same materials for both modalities. In one such study, Friederici et al. (1996) tested sentences containing syntactic category violations or syntactic ambiguity. When presented auditorily, both conditions elicited the P600, with the LAN present only in the syntactic category violation. Identical results were obtained for the visual modality. However, an N400 component was also identified for the syntactic ambiguity condition. The presence of the N400 was related to the absence of phonetic-prosodic cues in reading.

In the study by Hagoort and Brown (2000), three types of morphosyntactic violations were investigated in Dutch in both modalities. The most remarkable difference was the absence of the LAN in the visual modality, whereas two out of three conditions elicited the LAN in the auditory modality. Unfortunately, the authors did not offer any further explanation on this inconsistency. Unlike the LAN, the P600 was obtained in both conditions. Interestingly, the

P600 had the same onset time in both modalities, measured from the onset of the critical word, even though the recognition point for the syntactic violation was later in listening than in reading.

We have already established that the LAN is an unpredictable component across studies (Demestre et al., 1999; Loerts et al., 2013), but it also seems to be unpredictable depending on the presentation modality (Hagoort and Brown, 2000). Conversely, the P600 is fairly stable and uniform across both reading and listening studies. Surprisingly, it is so consistent that most studies report its onset time between 500 and 600 ms. Hagoort and Brown (2000) already noticed that in reading the P600 started at the approximately the same time as in listening, even though the violation is recognized later in listening than in reading. This raises an interesting question, which is whether the P600 is violation aligned or not; we assume that language-related ERP components are either violation-aligned or they occur in regular time windows (e.g., LAN: 300-500 ms, P600: 500 ms onwards). The time-windows are attested for reading since most studies use similar presentation rates which allow for one fixation and consequently one word or phrase. However, the violation recognition point varies greatly in listening studies. An average monosyllabic word may last 200-250 ms whereas a trisyllabic word may last 600 ms. Provided that the violation is word-final, the violation recognition points are almost 400 ms apart in the given examples. In the case of the stimuli in the current experiment, the violation recognition point is 200-300 ms earlier in the gender condition than in the number condition. This substantial time difference enables us to further study the issue of the component onset time.

5.1.4 Predictions

In line with the goals of the study, we will make two sets of predictions, one regarding the difference in processing number and gender and one regarding the component characteristics. Firstly, we expect that number disagreement is more complex to reanalyze than gender disagreement. This difference should be reflected in the P600, which should have either a broader distribution or larger amplitude in the number condition compared to the gender condition.

Secondly, we expect that the same components be elicited in both modalities. However, Hagoort and Brown (2000) demonstrated that the same stimuli can elicit the LAN in listening only. Therefore, we cannot exclude the possibility that the LAN will be elicited in the listening experiment, even though it was absent in the reading modality. If present, the LAN is equally likely to be restricted to the left anterior region only or to have a broader distribution. Also, the LAN should be present in both conditions as they both contain a syntactic violation. Still, there is a large difference between the conditions in terms of the violation recognition point. In the gender condition, the violation can already be detected approximately 400-500 ms post-stimulus onset, whereas the detection in the number condition is delayed to approximately 700 ms post-stimulus onset. If the violation is stimulus-aligned, its onset should be later in the number condition than the gender condition.

Our strongest prediction is that both conditions will elicit the P600. However, in addition to expecting a larger P600 in number than in gender, we hypothesize that the onset of the P600 may differ between the conditions, starting later in the gender than the number condition. Such an outcome would confirm that the P600 is violation aligned.

5.2 Method

5.2.1 Participants

A total number of 30 participants were tested for this experiment (8 male; mean age 22.1, age range 19-33). Out of the total number, 4 participants were excluded prior to analysis due to low scores on the grammaticality judgment task (accuracy below 80 %), leaving 26 participants whose results are reported in this study. All participants were right-handed, which was asserted with a Dutch version of the Edinburg Handedness Inventory (Oldfield, 1971). The participants declared no reading impairment, or any history of psychiatric or neurological illness. They all had normal to corrected-to-normal vision. All the participants signed a consent form prior to the experiment, and received a € 20 voucher for their participation.

5.2.2 Acceptability Ratings for the Materials

An acceptability judgment survey was carried out in order to verify the grammaticality/ungrammaticality of experimental stimuli and fillers. Both grammatical and ungrammatical sentences were included in the survey. The question asked was: ‘Is this sentence grammatically correct?'. The offered answers were 'yes' and 'no'. The instructions included an example. There were a total of 480 sentences divided over 4 lists. The survey was set up on an online platform SurveyMonkey (www.surveymonkey.com). In total, 32 native speakers of Dutch took the survey (9 male; mean age 33.18, age range 21-62). Each participant filled in only one survey. In order to include a stimulus into the experimental set, it had to have an approval rate of at least 80%, that is, 80% of the participants had to correctly judge the sentence as grammatical or ungrammatical. Out of 480 sentences, 16 sentences did not reach the inclusion threshold. These sentences were modified and assessed by another 3 speakers of Dutch who unanimously approved the modified sentences.

5.2.3 Materials

The materials were the same as in Chapter 4. However, due to a difference in duration between monosyllabic and trisyllabic words, only trisyllabic words were used as experimental items.

There were 80 experimental sentences and 240 fillers. The experimental sentences were created from a basis of 20 unique nouns. Nouns were controlled for plural morphology, noun-verb homophony, phonological alternations, and animacy. All nouns had to use the suffix *-en* exclusively for the plural form, which in turn could not be homophonous to an infinitival verbal form (e.g., *boek* 'book' > *boeken* 'books', but also 'to book'). Nouns could not display any phonological alternations between the singular and the plural form (e.g, voicing *huis* 'house' > *huizen* 'houses' or irregular plurals *stad* 'city' > *steden* 'cities'), and the vowel length had to be maintained (e.g., no nouns with short-long vowel alternation, such as *pad* 'path' > *paden* 'paths'). Finally, all nouns were inanimate.

Each noun yielded 4 unique sentences, with each sentence being used once as grammatical and once as ungrammatical. This means that a participant would have been exposed to the same noun 8 times. In order to reduce this, as well as to make the experiment shorter, stimuli were divided over 2 lists. Each participant was exposed to only one list. This way, a target noun appeared 4 times in a list, always in a different sentence. Items were counterbalanced in such a way that if a grammatical sentence was in the first list, its ungrammatical counterpart was in the second list. Consequently, each participant read 80 experimental and 160 filler sentences.

The experimental sentences were divided over two conditions: gender (40 sentences per list) and number (40 sentences per list). In Dutch, the common gender article *de* is homophonous with the plural article *de*, which is used for both common and neuter nouns in plural. This could lead to a possible ambiguity in the gender violation condition, such as *de_C mes_N 'the knife', in which the violation becomes apparent only at the end of the noun. Since *de* is always used as the definite plural article, the parser may expect to encounter *de messen* 'the knives'. Once it becomes obvious that there is no plural suffix *-en*, the parser registers a violation. Without any additional context, this violation is ambiguous between number and gender violation. More precisely, since *mes* is a neuter (*het*) noun, **de mes* may be recognized as a gender violation provided that *de* is reanalyzed as a singular common article. However, if the parser classifies *de* as a plural article, the violation is perceived as a number violation. We also do not rule out a possibility that the parser may perceive **de mes* as a double violation. In order to overcome this obstacle, gender violations were created between the indefinite article *een* (used only in singular for both genders) and a gender marked adjective followed by a noun; for example: *een $mooie_C$ $dorp_N$ 'a beautiful village'. The indefinite article indicates to the parser that it should expect a singular noun. The following adjective (*mooie*) is marked with *-e*, which according to the inflectional rules indicates that the noun has to be of common (*de*) gender. By combining the two pieces of information, the parser is ready to encounter a singular *de* noun. However, in the violated sentences, the following noun is always a singular *het* noun. Therefore, an unambiguous gender violation is created between the indefinite article-adjective complex and the target noun.

There was an additional reason as to why all the target nouns were of the *het* type. As already mentioned, Dutch gender is a lexical feature, that is, it has to be retrieved from the lemma. The only exceptions are nouns derived with a gender specific suffix, such as the diminutive suffix. Almost all Dutch nouns can have a diminutive form (derived through the suffix *–tje* and its allomorphs), the usage of which is also relatively high (Shetter, 1959). All diminutives are of *het* type (e.g., *de tafel* 'the table', het *tafeltje* 'the little table'). Consequently, in a violated condition, such as *een* *$rood_N$ $tomaat_C$ 'a red tomato' the parser recognizes the violation only at the end of the word. Again, the indefinite article means the noun should be singular, whereas the adjective form indicates the noun is of the *het* type. However, since almost any Dutch noun can be used as a diminutive, it is possible that the parser's strategy is to expect a diminutive noun. In order to avoid this possible strategy, we decided not to use *de* nouns as experimental items.

All gender sentences were created in two structural 'molds'. In the first mould (1), the sentence started with an expletive subject (e.g., *er* 'it/there') or a general place adverbial (e.g., *hier* 'here'). The subject was followed by a verb, either lexical or auxiliary, after which the indefinite article *een* was presented. The second sentence type (2) started with a personal pronoun (e.g., *hij/zij* 'he/she'), followed by a verb, which was followed by the indefinite article. In both sentence types, the indefinite article was followed by an adjective. In non-violated sentences, the adjective was always inflected with the suffix -e. In violated sentences, the adjective had a zero marking used with *het* nouns. The target noun was placed after the adjective. Also, the target noun was never at the end of a sentence; it was always followed by a prepositional phrase, an adverb, or a lexical verb.

(1)	grammatical:	Dat	was	een	mooi	compliment	over	haar	werk.
		that	was	a	beautiful$_N$	compliment$_N$	on	her	work.
		'That was a beautiful compliment for her work.'							
	ungrammatical:	*Dat was		een	mooie	compliment	over	haar	werk.
		that was		a	beautiful$_C$	compliment$_N$	on	her	work.

(2)	grammatical:	Hij	gaf	een	groot compliment	aan	zijn	leerling.
		he	gave	a	big$_N$ compliment$_N$	to	his	students
		'He gave a big compliment to his students.'						
	ungrammatical:	*Hij	gaf	een	grote compliment	aan	zijn	leerling.
		he	gave	a	big$_C$ compliment$_N$	to	his	students

All sentences in the number condition had the same structure (3, 4). They began with the plural article *de* in grammatical sentences, and the singular neuter article *het* in violated sentences. The article was followed by an inflected adjective, which is the correct form for both singular *het* and plural *de* nouns, after which the target noun was presented. Like in the gender condition, the noun was always followed by a prepositional phrase, adverbial phrase, or a lexical verb.

(3)	grammatical:	De	onverwachte	complimenten	zijn	vaak	het leukst.

the$_{PL}$ unexpected$_{PL}$ compliments$_{PL}$ are often the best
'The unexpected compliments are often the best.'

ungrammatical: *Het onverwachte complimenten zijn vaak het leukst.
the$_{SG}$ unexpected$_{SG}$ compliments$_{PL}$ are often the best

(4) grammatical: De welgemeende complimenten werken motiverend.
the$_{PL}$ well-meant$_{PL}$ compliments$_{PL}$ work motivating
'The well-meant compliments are motivating.'

ungrammatical: *Het welgemeende complimenten werken motiverend.
the$_{SG}$ well-meant$_{SG}$ compliments$_{PL}$ work motivating

We explained the reasons for having only *het* nouns as experimental items. A consequence of this choice is that it can facilitate a learning strategy in participants. In other words, when a participant encounters a sentence starting with *het*, he may learn after a few items that such sentences are always ungrammatical. Similarly, any instance of a zero-marked adjective indicates an ungrammatical sentence. A solution was to include a number of fillers that have a reverse pattern to the experimental stimuli. For this reason, we included 160 filler item sentences. Half of the filler items contained *het* nouns (5) and the other half *de* nouns (6). The *de* noun-group was used to counterbalance the gender condition, whereas the *het* noun group counterbalanced the number condition.[5] Consequently, it was impossible to judge sentence as grammatical or ungrammatical based only on the article or the adjective. The participant had to pay attention to everything preceding the noun, as well as the noun itself, in order to correctly judge the sentence.

(5) grammatical: Er ligt een rotte tomaat in de koelkast.
there lie a rotten$_{C}$ tomato$_{C}$ in the fridge
'There is a rotten tomato in the fridge.'

ungrammatical: *Er ligt een rot tomaat in de koelkast.

[5] The violation in *het* nouns was ambiguous between number and gender. However, this was not important since the only purpose of the fillers was to prevent a learning strategy disregarding the exact nature of the violation.

there lie a rotten$_{N}$ tomato$_{C}$ in the fridge

(6) grammatical: Het oude paspoort is niet meer geldig.
the$_{N.SG}$ old$_{N.SG}$ passport$_{N.SG}$ is not more valid
'The old passport is not valid anymore.'

ungrammatical: *De oude paspoort is niet meer geldig.
the$_{C/PL}$ old$_{C/PL}$ passport$_{N.SG}$ is not more valid

In addition, there were 80 filler items whose structure was identical to the experimental items (gender and number violation), but whose target noun was monosyllabic (7).

(7) grammatical: Er lag een mooi dorp vlakbij de grote stad.
there lay a beautiful$_{N}$ village$_{N}$ near the big city.
'A beautiful village was close to the big city.'

ungrammatical: *Er lag een mooi**e** dorp vlakbij de grote stad.
there lay a beautiful$_{C}$ village$_{N}$ near the big city.

5.2.4 Procedure

Participants were seated in front of the screen at a distance of 70 to 80 cm. The experiment was presented in E-Prime (Psychology Software Tools, Inc). The passive task was to listen to the sentences presented through headphones. The active task was to reply to a randomly assigned grammaticality judgment question. On average, a question appeared once for every five sentences. The purpose of the active task was to keep participants focused. The experiment opened with written instructions that were repeated by the experimenter. There was a brief practice session (4 sentences) in order to ensure the participants had understood the instructions and were able to follow the stimulus presentation. After the last practice item, the participants had a chance to ask for clarifications or more detailed instructions. Once they were ready, they could proceed to the experimental part by pressing any keyboard button. Each trial opened with a fixation cross (500 ms) after which a sentence was played. The screen remained blank until the end of the sentence. In case there was a grammaticality judgment question, a question mark appeared on the screen after the sentence finished. The question mark remained on the screen for 3 s, during which participants were supposed to press either ‘p’ or ‘q’ (counterbalanced among participants), depending on whether the previous sentence was grammatical or not.

The experiment lasted approximately 25 minutes. The stimuli were divided into 4 blocks, each block containing 20 experimental and 40 filler items. There were 12 grammaticality judgment questions per block, 4 for experimental and 8 for filler items. The presentation order within a block was random, as determined by the software. The participants were advised to take a short break after each block.

5.2.5 EEG Data Acquisition and Processing

Continuous EEG data were recorded using the ASA-Lab system (ANT Neuro Inc, Enschede, The Netherlands) from 64 Ag/AgCl scalp electrodes fitted in an elastic cap (WaveGuard). The electrodes were positioned according to the extended 10-20 system. Eye movements were recorded using one bipolar channel for horizontal movements (HEOG; the electrodes were placed at the outer canthus of the eyes) and one for vertical movements (VEOG; placed above and below the left eye). Electrode impedances did not exceed 10 kΩ, and were kept at 5 kΩ or below in the large majority of cases. The data were sampled at 512 Hz with the common average reference.

The data were pre-processed with Brain Vision Analyzer 2.04 (Brain Products, GmbH, Munich, Germany). The first step was to down-sample the data to 256 Hz in order to speed up the analysis. This was followed by re-referencing the offline data to the average of the left and right mastoid. Afterwards, a band-pass filter was applied (0.1 – 40 Hz), after which an automatic ocular PCA-based correction was performed. The continuous data were segmented into 1700 ms long epochs, starting 200 ms before the trigger marker (target noun onset). The automatic artifact rejection (+/- 100 µV threshold, minimal activity 0.1 µV) was performed in the interval of -200 ms to 1500 ms for each epoch. Finally, the baseline correction was applied starting -200 ms until 0 ms, after which the data were averaged per subject and per condition. If a participant had fewer than 70 % of averaged trials in one or more conditions, his data would have been excluded from the analysis. No data were excluded based on this criterion.

5.2.6 Analysis

For the analysis, we used averaged participant values (in μV) per condition, level of grammaticality, and regions of interest (ROI). Regions of interest (Fig. 5.1) were created by averaging the values of 5 to 6 adjacent electrodes (50 in total), which resulted in 9 ROIs: left anterior (F7, F5, F3, FC3, FC5), midline anterior (F1, Fz, F2, FC1, FCz, FC2), right anterior (F4, F6, F8, FC4, FC6), left central (TP7, C5, C3, CP5, CP3), midline central (C1, Cz, C2, CP1, CPz, CP2), right central (C4, C6, CP4, CP6, TP8), left posterior (P7, P5, P3, PO7, PO5, O1), midline posterior (P1, Pz, P2, PO3, POz, PO4), and right posterior (P4, P6, P8, PO6, PO8, O2).

In order to determine the most suitable time windows for the analysis of the P600 in each condition, a series of one-tailed *t* tests was performed in the period of 500 ms to 1000 ms post-stimulus onset, every 100 ms. The *t* test was carried out on the midline posterior region, which is the most representative region for the detection of the P600.The tests compared the mean voltage values of grammatical and ungrammatical sentences in both conditions. More detail on establishing time windows is given in the Result section.

Based on the *t* tests, different time windows were created for the two conditions.For the gender condition, the P600 was analyzed in the following time windows: 600-800 ms (onset), 800-1000 ms (early), 1000-1200 ms (late). For the number condition, the time windows were slightly later: 900-1100 ms (onset), 1100-1300 ms (early), 1300-1500 ms (late). Lastly, a time window was chosen in each condition for the analysis of the LAN, in case it preceded the P600. The LAN time window in the gender condition was 400-600 ms, and in the number condition it was 700-900 ms post-stimulus onset.

The statistical analysis using repeated measures ANOVA was carried out separately for each condition. The following within subject factors were included in a repeated measures ANOVA: 1) grammaticality (2 levels: grammatical and ungrammatical); 2) hemisphere (2 levels: left and right); 3) anteriority (3 levels: anterior, central, and posterior)._The global analysis for each time window and each condition was performed by two separate ANOVAs.

The first ANOVA analyzed only lateral the regions and it included all 4 factors. A second omnibus ANOVA was run on the midline regions only, excluding the factor hemisphere. In case the assumption of sphericity was violated, the Geisser and Greenhouse (1959) correction was applied. The significance level was set to $p < .05$. Follow-up ANOVAs were performed only for interactions that were at least marginally significant ($p < .1$) and that included factor grammaticality. In case a follow up analysis consisted of testing two levels of a single factor, a one-tailed *t* test was performed. Finally, incorrectly judged trials were not included in the analysis.

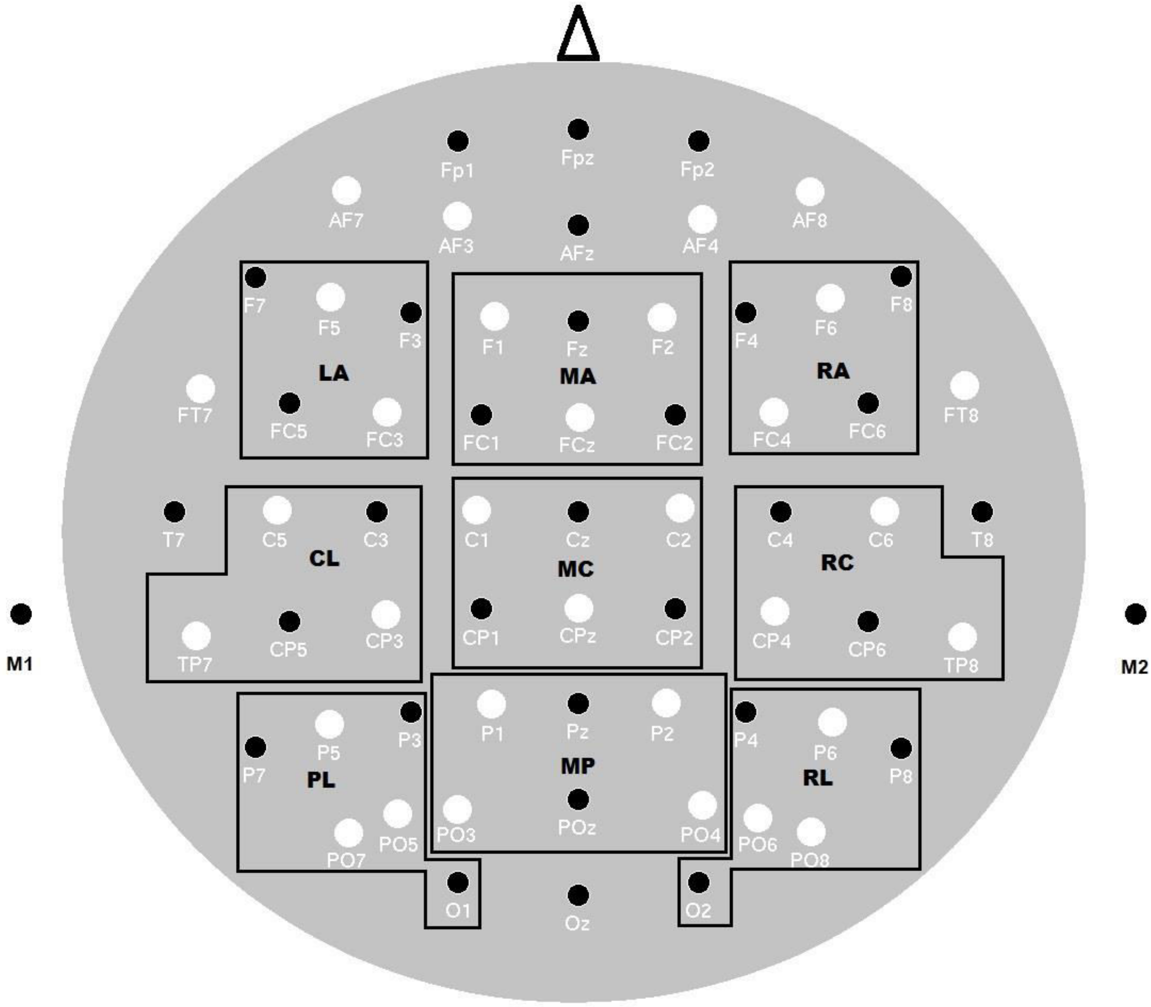

Fig. 5.1 Electrode positions and the 9 regions of interest used in the analysis: left anterior (LA), left central (LC), left posterior (LP), midline anterior (MA), midline central (MC), midline posterior (MC), right anterior (RA), right central (RC), right posterior (RP).

5.3 Results

5.3.1 Accuracy Results

From the 30 participants, 5 were excluded from the analysis due to a low score on the grammaticality judgment task. The cut-off was set at 80 %, meaning that each participant had to correctly respond to 51 out of 60 questions. The remaining 25 participants had an accuracy rate of 95% (average number of errors: 3, SD 2.4). One participant performed at ceiling. The only purpose of the grammaticality judgment question was to ensure the participants' alertness throughout the entire experiment. Therefore, the data were not further analyzed.

5.3.2 ERP Results

Visual inspection suggested that violated sentences elicited a positive shift in both gender (Fig. 5.2) and number condition (Fig. 5.3). The effect differed in terms of its onset time in the two conditions. The positivity in the gender condition emerged at approximately 600 ms after the stimulus-onset, some 200 to 300 ms earlier than in the number condition. The effect showed a slightly broader distribution in the number condition. In addition, a left-lateralized anterior negativity was present in the gender condition only. It had a relatively early onset time (approximately 400 ms post-stimulus onset), and persisted throughout all time windows.

5.3.2.1 Determining Time Windows

Before the component analysis was performed, it was necessary to establish the onset of the P600 component in order to create appropriate time windows. This was done by carrying out 5 one-tailed *t* tests on the mean voltage values of ungrammatical and grammatical sentences in both conditions. The test was performed 5 times (every 100 ms) in the period of 500 ms to 1000 ms after the onset of the stimulus on the midline posterior region. The analysis showed that the onset of the positive effect was earlier in the gender condition (600 ms on) than in the number condition (900 ms on). The results are summarized in Table 5.1.

Table 5.1: A summary of the *t* tests performed on subsequent time windows (from 500 ms to 1000 ms after the stimulus onset) used for the detection of the onset of the P600 in the gender and number condition

Time Windows				
500 – 600 ms	**600 – 700 ms**	**700 – 800 ms**	**800 – 900 ms**	**900 – 1000 ms**
		gender		
X	*	*	*	*
		number		
X	X	X	X	*

X no significant effect (p > .05)

* significant effect (p < .05)

5.3.1.2 ERP Results: Gender

The first time window (**450-600 ms**) was chosen to investigate whether gender disagreement elicits the LAN preceding the P600. A lateral ANOVA, however, did not reveal any significant interactions. A hypothesis driven *t* test was subsequently performed on the left anterior region without producing a significant result.

The following time window (**600-800 ms**) was designated for the detection of the onset of the P600 in the gender condition. A global lateral ANOVA yielded a significant interaction between grammaticality and anteriority (F (2, 48) = 7.790, p = .008). Follow-up tests revealed a marginally significant effect in the posterior regions in which grammatical sentences were more positive (F (1, 24) = 4.224, p = .051). The effect did not reach the central regions. A hypothesis driven *t* test was performed on the left anterior region, in which ungrammatical sentences caused a significant negative effect (t (24) = 1.77, p = .044) In addition, a midline analysis revealed a significant interaction between grammaticality and anteriority (F (2, 48) = 7.293, p = .009). Further testing produced a significant effect in the posterior region (t (24) = -2.17, p = .02), in which ungrammatical sentences were more positive.

The presence of a positive effect in the gender condition persisted in the subsequent time window of **800-1000 ms** post-stimulus onset. A lateral ANOVA yielded a significant interaction between grammaticality and anteriority (F (2, 48) = 15.861, p < .001). Additional tests confirmed that ungrammatical sentences caused a negative effect in the anterior regions

(F (1, 24) = 6.632, p = .017), but also a positive effect in the posterior regions (F (1, 24) = 7.613, p = .011). In addition, the factor grammaticality entered a close-to significant interaction with the factor hemisphere in the central regions (F (1, 24) = 2.975, p = .097), for which no significant result was obtained in the follow-up. A global midline ANOVA uncovered an interaction between grammaticality and anteriority (F (2, 48) = 23.906, p < .001). The interaction was caused by a significant difference between ungrammatical (more positive) and grammatical sentences in the posterior region (t (24) = -2.88, p = .003), but also by a marginal significance in the anterior region, in which ungrammatical sentences were more negative (t (24) = 1.68, p = .053).

In the last time window in the gender condition analysis(**1000-1200 ms**), a lateral analysis revealed an interaction between grammaticality and anteriority (F (2, 48) = 11.187, p = .002), which, like in the previous time windows, was caused by ungrammatical sentences being more positive in the posterior regions (F (1, 24) = 4.283, p = .049), and more negative in the anterior regions (F(1, 24) = 4.474, p = .045), specifically in the left hemisphere (t (24) = 2.72, p = .006). Similar to the previous time window, a midline analysis yielded an interaction between grammaticality and anteriority (F (2, 48) = 20.913, p < .001). In the posterior region, ungrammatical sentences elicited a significant positive effect (t (24) = 2.27, p = .016), whereas in the anterior region a marginally significant effect of negative polarity was detected (t (24) = 1.36, p = .09).

5.3.1.3 ERP Results: Number

The first time window (**500-700 ms**) was designated to check whether number disagreement elicited left-lateralized negativity which preceded the P600. Therefore, since a lateral analysis did not produce any significant result, a planned *t* test was performed on the left lateral region. Still, the test failed to detect any effect.

The **900-1100 ms** time window was intended for testing for the onset stage of the P600 in the number condition. A lateral omnibus ANOVA produced an interaction between grammaticality and anteriority (F (2, 48) = 6.718, p = .012). The follow-up tests revealed a main effect of grammaticality in the posterior region only, in which ungrammatical sentences were more positive (F (1, 24) = 6.331, p = .019). Since the positive effect was expected to have either a broader distribution or larger amplitude in the number condition than in the gender condition, additional planned *t* tests were performed on the central regions since they are immediately adjacent to the posterior regions. The tests showed that the positive effect was significant in the right central region (t (24) = -1.82, p = .041) and marginally significant in the left central region (t (24) = -1.44, p = .081). The midline analysis produced comparable results, with a significant interaction between grammaticality and anteriority (F (2, 48) = 6.83, p = .012). Once again, ungrammatical sentences were more positive in the posterior region (t (24) = 2.54, p = .009), but also marginally significant in the central region (t (24) = 1.67, p = .054).

In the following time window (**1100-1300 ms**), a lateral ANOVA showed an interaction between grammaticality and anteriority (F (2, 48) = 16.453, p <.001), which was driven by ungrammatical sentences being more positive than grammatical sentences in the lateral posterior regions (F (1, 24) = 7.953, p = .009). In the central regions, there was a marginally significant effect in the right hemisphere (t (24) = -1.37, p = .092). A midline analysis revealed an interaction between grammaticality and anteriority (F (2, 48) = 19.695, p < .001) which was caused by a positive effect (ungrammatical sentences more positive) in the posterior region (t (24) = -2.81, p = .005).

In the last time window **(1300-1500 ms)** a lateral analysis confirmed an interaction between anteriority and grammaticality (F (2, 48) = 6.664, p = .01), caused by a positive effect in the posterior regions (F (1, 24) = 6.205, p = .02). Similar to the previous time window, the positive effect was only marginally significant in the right central region (t (24) = -1.45, p = .08). Finally, a midline analysis also yielded an interaction between grammaticality and anteriority (F (2, 46) = 10.556, p = .002), driven by the positive effect in the posterior region (t (24) = -2.33, p = .014).

5.3.3 Summary of ERP Results

The statistical analysis confirmed the presence of two language-related components, namely the LAN and the P600. The LAN was present in the gender condition only. It was detected at 600 ms and it persisted throughout all subsequent time windows with mainly left-lateralized distribution.

The P600 was confirmed for both conditions. In line with the visual inspection, the effect had an earlier onset in the gender condition (from 600 ms on). In its onset, the effect was confined to the posterior midline region with a tendency to spread to the other posterior regions, which occurred in the following time window. The effect remained in the posterior regions until the end of measuring.

In the number condition, the positive effect was first detected in the time window from 900 ms, which is 300 ms later than in the gender condition. In its onset stage (900-1100 ms), the effect was robustly present in all posterior regions and it also included the right central region with a tendency towards the midline central region. In the following two time windows, the effect remained in the posterior regions, with a tendency towards the right central region.

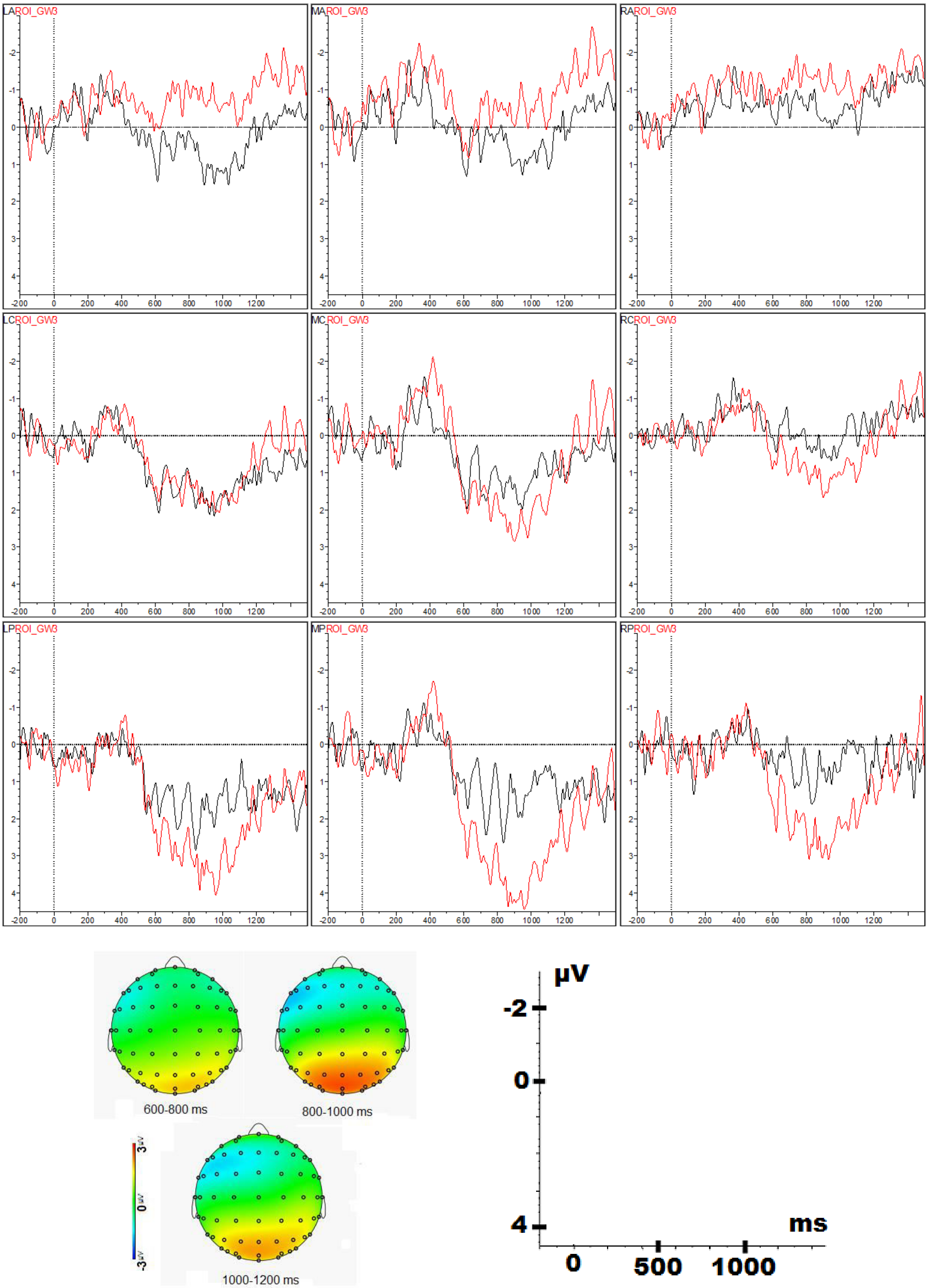

Fig. 5.2 Grand average ERPs for the gender condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences. The topographic maps represent a difference between ungrammatical and grammatical sentences in the three P600 windows.

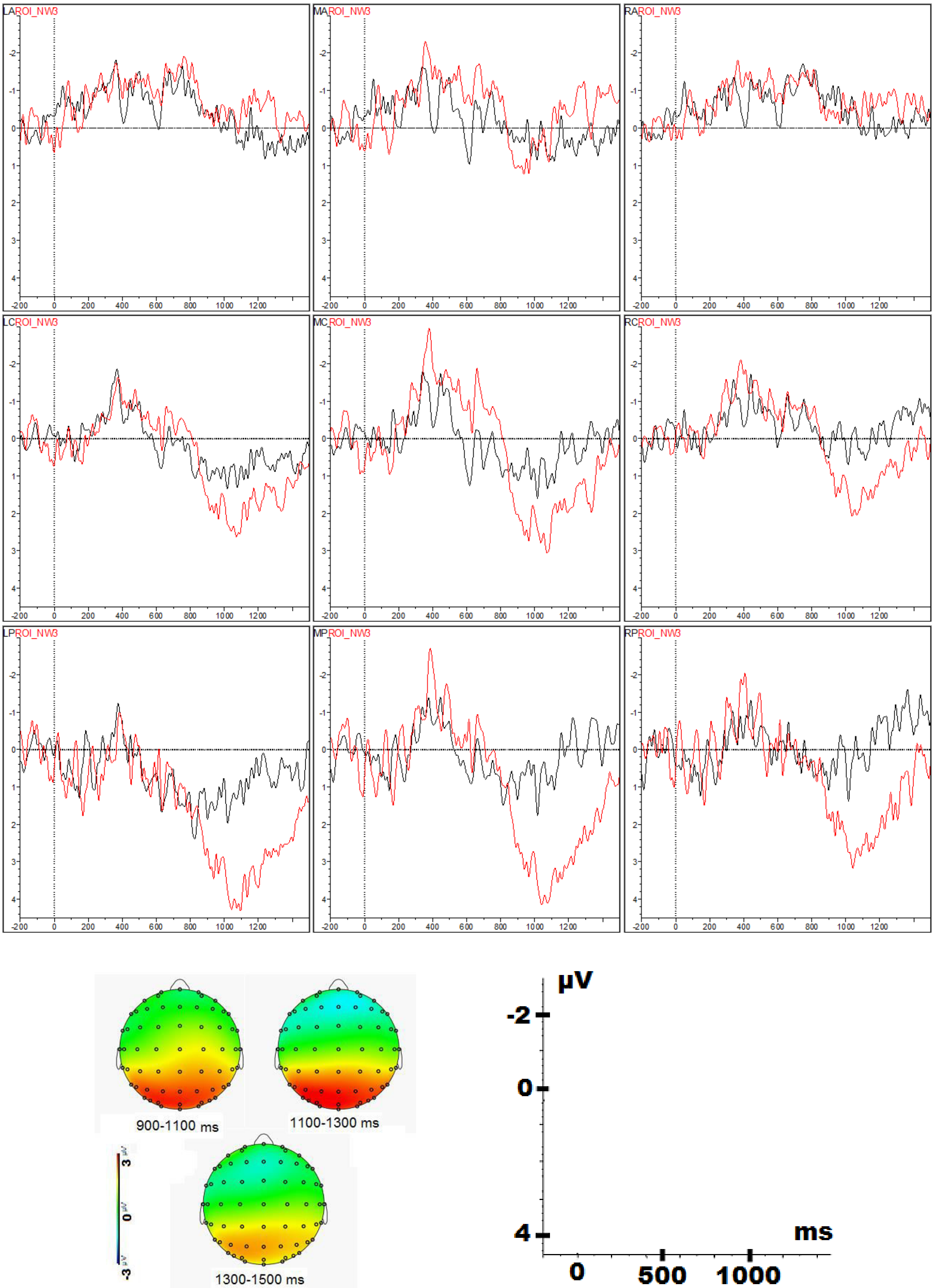

Fig. 5.3 Grand average ERPs for the number condition across all 9 ROIs: black line represents correct sentences and red line represents violated sentences. The topographic maps represent a difference between ungrammatical and grammatical sentences in the three P600 windows

5.4 Discussion

We tested auditory processing of gender and number agreement in the Dutch noun phrase. As expected, results confirmed the presence of the P600 in both conditions. The effect differed between conditions in terms of both distribution and onset time. In addition, a left-lateralized anterior negativity (LAN) was detected in the gender condition only. Therefore, the comparison between auditory gender and number processing resulted in three effects: 1) the LAN was present in the gender condition only; 2) the P600 had an earlier onset in the gender condition compared to the number condition; 3) the P600 was more broadly distributed in the number condition than in the gender condition. We will offer an interpretation of the effects in this order.

5.4.1 LAN

The left anterior negativity is understood as an automatic response to a morphosyntactic violation (Friederici, 2002; Molinaro et al., 2011). However, it is not a very reliable marker since it is not always elicited in the presence of morpho-syntactic errors (Osterhout et al., 2004). Hagoort and Brown (1999) proposed that the LAN is only sensitive to violations due to erroneous inflectional morphology. This claim is contrary to our results, since the LAN was elicited for a morphologically non-overt violation (gender) and not for an overt one (number). Rather than looking for a psycholinguistic explanation related to an intrinsic difference between number and gender, we suggest that the presence/absence of the LAN is a timing issue. More precisely, a gender violation is recognized at the word's recognition point, which usually precedes the last phoneme (on average 400-500 ms from the onset of the word). A number violation, however, is recognized only word-finally from the number suffix (700 ms). Once the gender violation is recognized, and the parser can focus on the violation. But, if the parser recognizes the violation word-finally (number), it is likely that a following word is already being uttered. In this case, the parser needs to perform multiple operations, such as wrapping up the previous noun phrase, and starting to integrate the incoming word. In addition, simultaneously with the wrap-up and novel-word integration a syntactic repair process also takes place.

Taking into account a multitude of processes occurring almost simultaneously, we suggest that the LAN as an automatic process could have been suppressed in the number condition. Previous studies comparing number and gender agreement reported the same components in both conditions (Barber and Carreiras, 2005). However, these were reading studies, which is exactly why we believe the presence/absence of the LAN is connected with the presentation modality. The most obvious difference between visual and auditory presentation is the time course of information, with auditory processes delivering information at a slower rate. Therefore, the LAN is most likely induced by a combination of the temporal flow of grammatical information and the position violation in gender and number.

Two additional issues regarding the LAN are also relevant for this study. The LAN is a controversial component in that it cannot be consistently elicited (Osterhout et al., 2004). The current experiment is at odds with two auditory studies on article-noun agreement in Dutch that failed to obtain the LAN (Loerts et al., 2013; Meulman et al., 2014). The only explanation we can offer pertains to the design. The gender violation in both previous studies on Dutch was created between a mismatching article and the noun. Since the article *de* is used for both common gender and all plural nouns, all violations between the article *de* and a *het* noun are not perceived as violation until the end of the word when the absence of the plural morpheme is detected. Even then, the violation is ambiguous between a number and a gender violation. Similarly, a violation between the *het* article and a *de* noun is probably not recognized as a violation until the end of the noun since almost all Dutch nouns can be marked with a diminutive suffix. The stimuli used in the current eliminated these two confounds (ambiguity and delayed violation detection). In line with our explanation for the absence of the LAN in the number condition, it is possible that word-final violations are less likely to elicit the LAN due to several simultaneous processes that occur at that moment in time.

The last point regarding the LAN is its distribution. Unlike most listening studies that managed to elicit the LAN, the LAN in the gender condition was left lateralized and anterior, fitting its textbook definition. Most auditory studies that managed to elicit the LAN reported a scalp-broad non-lateralized negativity (Demestre et al, 1999; Meulman et al., 2014). However, Hagoort and Brown (1999) and Friederici et al. (1996) elicited a clear left-lateral negativity. All of these studies, including ours, differed at both methodological and theoretical levels, thus preventing a definite conclusion. Thus, it remains to be seen whether the distribution of the LAN in auditory studies reflects partly distinguishable computational processes or methodological differences (e.g., number of stimuli, presentation rate, stimulus length).

5.4.2 P600 and Violation-Alignment

The second finding speaks to the issue of whether or not the language-related ERP components are violation-aligned or not. Since only the P600 was elicited in both conditions, we will focus on this component. The target words in the current study were all trisyllabic. The rationale behind this decision was to make the stimuli as long as possible in order to augment the difference in time between the violation in gender and number. As already said, the gender violation correlates with the word's recognition point. On average, the recognition point of a 600 ms trisyllabic word is in the range of 400-500 ms. The number violation was created between a singular article and a plural trisyllabic word. Since the plural suffix was -*en*, the trisyllabic words contained four syllables in the violated condition (700 ms). Therefore, the average difference between detecting a number and a gender violation was approximately 200 ms. This difference of 200 ms was expected to be reflected in the onset of the P600, provided that the P600 is violation aligned. This expectation turned out to be true, as the onset of the P600 in the gender condition occurred 200 ms earlier than in the number condition. In addition, the onset of both components was somewhat later than usual, approximately 800 ms for the gender condition and 1000 ms for the number condition. The late onset is caused by a late violation detection, which was in the range of 400-500 ms post-stimulus onset for gender, and 700 ms for number. This additional evidence supports the view that the P600 is violation-aligned.

5.4.3 P600 as a Marker of Repair/Reanalysis

Lastly, the P600 had a broader distribution in the number condition than in the gender condition. The P600 started off as a broad posterior positivity in the number condition, extending to the right central region with a tendency towards the midline central region. The P600 remained as a robust posterior effect with a strong tendency towards the central regions throughout the subsequent time windows. In contrast, the P600 in the gender condition was weak in its onset, mainly confined to the midline posterior region. It became more robust in the following time windows, spreading across all the posterior regions. However, it never reached any central region.

We anticipated that in the number condition the P600 would have either a broader distribution or larger amplitude, due to differences in processing load. Namely, the P600 is often characterized as a marker of repair and reanalysis processes (Friederici et al, 1996). That is, it is taken to reflect the parser's effort to repair and integrate syntactically anomalous content. Due to the inflectional nature of number, the parser has two repair options (to repair the article and/or the noun suffix) as opposed to a single repair option in the case of gender (to repair the article). Therefore, we propose that the P600 had a broader distribution in the number condition than in the gender condition because repairing number requires a more elaborate mechanism.

Chapter 6
General Discussion & Conclusion

6.1 Overview

The discussion addresses the hypotheses and predictions postulated in the Chapter 1 in the same order. In order to make it easier for the reader to follow the discussion, results from all experimental chapters are summarized in Table 6.1.

Table 6.1: Summary of results

Italian					
Chapter 1				**Chapter 2**	
Reading				**Listening**	
Syntactic gender		**Semantic gender**		**Syntactic gender**	**Semantic gender**
LAN, P600		**LAN, P600**		**(L)AN, P600**	**P600**
LAN				**LAN**	
300-450 ms; left anterior and left central regions				**400-750 ms; anterior and central regions**	**x**
P600				**P600**	
500-700 ms	**700-900 ms**	**500-700 ms**	**700-900 ms**	**750-1000 ms**	**750-1000 ms**
*	*	***	***	*	*
posterior and central regions				**posterior regions**	**posterior and some central regions**

Dutch											
Chapter 3						**Chapter 4**					
Reading						**Listening**					
Gender			**Number**			**Gender**			**Number**		
P600			**P600**			**LAN, P600**			**P600**		
LAN						**LAN**					
x						**600-1200 ms; left anterior region**			**x**		
P600						**P600**					
450 - 600 ms	**600 – 800**	**800 - 1000**	**450-600**	**600-800**	**800-1000**	**600-800**	**800-1000**	**1000-1200**	**900-1100**	**1100-1300**	**1300-1500**
x	*	*	*	*	***	*	*	*	*	*	*
posterior and central regions						**(some) posterior regions**			**posterior and some central regions**		

x indicates the absence of an effect

* indicates the presence of an effect in a given time window

*** indicates the presence of an effect in a given time window; it also indicates that the effect in the given time window was significantly larger in terms of amplitude than the same effect in the same time window in the other condition

6.3 Gender (Dis)agreement

6.1.1 P600

This thesis revolves around the issue of gender (dis)agreement. Processing gender and gender (dis)agreement is a pervasively syntactic phenomenon which is expected to elicit syntactically-related language components. Indeed, in each condition of each experiment, the P600 component was successfully elicited regardless of presentation modality. The P600 is often understood as a marker of repair/reanalysis processes occurring after the detection of a morphosyntactic error (Friederici, 2002; Kaan & Swaab, 2003). In addition to the P600, a number of studies have reported another syntactically-related component, that is, the left anterior negativity (LAN). The LAN precedes the P600, and arises as a response to the detection a morphosyntactic violation detection (Friederici, 2002). In the current thesis, the LAN was elicited in half of the conditions.

6.1.2 LAN: Reading

The P600 can be characterized as a reliable marker of syntactic processing, as it is consistently elicited across agreement studies. By contrast, the LAN has been reported inconsistently in studies on agreement (see Molinaro et al., 2011). So far, there is no precise account for such variability. The current thesis, unfortunately, does not provide solutions to this issue. The LAN was elicited in the reading study on syntactic and semantic gender in Italian (Chapter 2). Both conditions (syntactic gender and semantic gender (dis)agreement) elicited a small negative left-lateralized effect, consistent with the LAN. There was no difference in the characteristics of the effects across conditions. This finding is in line with several studies on agreement in Italian that reported the LAN (Caffarra et al., 2015; Molinaro et al., 2008).

However, gender and number disagreement in reading in Dutch (Chapter 4) did not elicit the LAN. Again, this is in line with the majority of studies on gender in Dutch that failed to obtain the LAN effect (Hagoort & Brown, 1999; Loerts et al., 2013; Meulman et al., 2014). Hagoort and Brown (1999) suggested that the LAN can be elicited only in the presence of a morphologically overt morphosyntactic violation. This can explain why there is no LAN in

Dutch gender disagreement (no gender morpheme) as opposed to Italian gender disagreement (gender morpheme present in our stimuli). Still, this explanation does not account for the lack of the LAN in number disagreement in Dutch, in which the morphosyntactic error is caused by an incongruous number suffix. Therefore, we can only conclude that our results are in line with a number of studies that elicited the LAN affected in Italian (Molinaro et al., 2008, Caffarra et al., 2015), as well as with a number of studies that failed to obtain the LAN effect in Dutch (Hagoort & Brown, 1999; Loerts et al., 2013; Meulman et al., 2014).

6.1.3 LAN: Listening

The auditory results did not parallel the findings from the reading studies. In the experiment on Italian, the LAN-like effect was recorded only in the syntactic gender condition. Similarly, in the Dutch experiment, the LAN was present in the gender disagreement condition. This issue will be addressed again in more detail later in this chapter.

6.3 Repair and Reanalysis

Since it is assumed that gender (dis)agreement is a syntactic phenomenon, we expected it to elicit the P600 as a reliable marker of syntactic processing. We were particularly interested in investigating how manipulating repair/reanalysis complexity affects the P600 effect. We predicted that an increase in difficulty would result in a larger effect, in terms of either amplitude or distribution. The prediction was borne out in the data – an increase in repair/reanalysis complexity resulted in an increase of the P600 effect.

There are two more issues of interest regarding the relationship between the repair/reanalysis processes and the P600. Firstly, an increase in repair/reanalysis complexity yielded increased amplitude of the P600 in the two reading studies (Chapters 2 and 4). In addition, in the listening studies (Chapters 3 and 5), a broader distribution of the P600 was observed in those conditions (semantic gender and number disagreement), which were costlier to repair/reanalyze. According to our prediction, a change in either amplitude or distribution indicates a difference in processing load. It is, thus, obvious from the experiments that the characteristics of the components are affected by presentation modality. As to why reading was associated with a change of amplitude and listening with a change of distribution, only speculations are possible. A tendency was observed when comparing reading to listening experiments: the difference in amplitude was always larger in reading. Furthermore, the effects also seemed to have a broader distribution in reading than in listening. These two issues might be interrelated. Namely, an increase in amplitude may indicate an increase in processing load, whereas a different distribution may indicate different underlying generators (Otten & Rugg, 2005). However, it is highly unlikely that the same P600 generator yields both low- (single repair/reanalysis) and high-demand (double repair/reanalysis) responses in reading whereas different P600 generators are involved in listening. Rather, an increase in processing demand might modulate amplitude in both modalities. Statistically, this emerges differently in different modalities.

Hagoort and Brown (2000) noted that the auditory P600 has a more posterior distribution than the visual P600. This could be a consequence of a smaller effect size in listening than in reading. As already said, in the experiments in this thesis the P600 tended to have smaller amplitude in listening than in reading. The topographic origin of the repair/reanalysis P600 is most likely in the posterior regions (Kaan & Swaab, 2003) from where it spreads over central and towards anterior regions. The effect starts to deteriorate in strength, as seen from its voltage, the farther away it spreads from its posterior source. Since the effect is already stronger in reading, it can reach statistical significance in the central regions in reading but not in listening. Then, if the processing load is affected by increasing repair/reanalysis complexity, the P600 amplitude increases. Since the P600 is strongest in posterior regions, adding more power to one condition will yield a statistically larger effect in the posterior regions in reading. If the effect is already large, the significance may spill over to central regions as well. By contrast, in the listening experiments, an increase in voltage of the P600 in one condition is not enough to make that condition significantly larger than the other condition in the posterior regions, since the effect is weak to start with. However, it may be just enough to tip the scales in the central regions. That is, an increase of the P600 effect in one condition in listening adds just enough power to the already trending effect in the central regions to reach significance. Therefore, we believe that an increase in repair/reanalysis complexity modulates the amplitude of the P600. Statistically, however, this effect may manifest itself as an increase in distribution, as was the case in the listening experiments.

The second issue pertains to the distinction of the early and late phase of the P600. Hagoort and Brown (2000) suggested that the P600 consists of two functionally and morphologically distinct phases. The early phase (500-700 ms) is characterized by a broadly distributed positive effect and is primarily connected with syntactic integration. It is followed by a late stage with a more posterior distribution, in which repair/reanalysis processes take place. We cannot test this in the listening experiments, as the difference in stimulus duration did not allow us to create two distinct time windows. Therefore, the discussion that follows refers to the reading experiments only.

In the experiment on syntactic and semantic gender processing in Italian, the difference in the P600 was statistically significant in both time windows. However, this does not rule out that repair/reanalysis takes place in the late time window. Rather, due to the nature of the semantic gender, the experiment actually combined information integration and repair/reanalysis processes. That is, in the semantic gender condition, the parser integrated conceptual and syntactic information in order to retrieve gender information. This is followed by a complex repair/reanalysis mechanism. Based on the results from Italian, it is impossible to confirm or rule out the early/late P600 distinction. However, the experiment on gender and number disagreement in Dutch supports the aforementioned distinction. Namely, the P600 in the number condition was stronger than the P600 in the gender condition, only in the late time window. Since the main difference between the two conditions lies in the complexity of the repair/reanalysis mechanism, the results of this experiment favour the early/late P600 hypothesis as suggested by Hagoort and Brown (2000).

6.4 Component Onset Time

In Chapter 1, it was stipulated that auditory comprehension studies allow for a better time resolution of sentence processing. Therefore, only the results from the auditory studies (Chapters 3 and 5) are discussed here. The initial assumption was that, like the N400 (O'Rourke & Holcomb, 2002), the syntactically-related components are violation-aligned. That is, rather than occurring in cycles (e.g., every 600 ms), the components arise after a certain time period from the detection of the violation. Data consistent with this claim were obtained as regards the LAN. In both experiments, the LAN was elicited soon after the violation was detected (see also Hastings & Kotz, 2008). As an illustration, the LAN was detectable 600 ms after the stimulus onset in the experiment on Dutch (Chapter 5) rather than in the usual time window of 300-500 ms post-stimulus onset.

The situation is more complex in the case of the P600. The onset time of this component was identical in both conditions (syntactic and semantic gender) in the experiment on Italian (Chapter 3), despite the earlier recognition point in the syntactic gender than in the semantic gender. However, the results were not replicated in the study on Dutch (Chapter 5). In the latter experiment, the onset time of the P600 in the number condition was delayed by 200 ms as compared to the gender condition. This corresponds to the difference between the recognition points in the number and gender condition, which were approximately 200 ms apart.

The results from the two studies seem to contradict each other. However, we believe that there is further evidence in support of the P600 being violation-aligned like the LAN. The simplest explanation is that the results from the first study are skewed due to a variation in stimulus length. Unlike the study on Dutch, in which all target words were tri-syllabic, individual target words in the experiment on Italian varied in length. That is, the average stimulus duration was held constant and comparable across the two conditions (app. 550 ms in both cases), but it was impossible to control for individual word duration due to the very limited number of nouns with semantic gender. This can be corrected by measuring the ERPs from the (violation) recognition point rather than from the onset of the noun. Taking into account the limitations of the study on Italian, data provide some evidence consistent with the P600 being violation-aligned.

6.5 Comparison of Presentation Modalities

The initial hypothesis concerning the relationship of ERP components in reading and listening was as follows. Each syntactically-related ERP component reflects a phase in language processing. In case of (dis)agreement, both the LAN and the P600 are expected, as they represent two discrete processing phases (Friederici, 2002). More precisely, the LAN is understood as an automatic response to the detection of the violation and is followed by a repair/reanalysis phase represented by the P600. We predicted that if both processes are present in one modality, they should also be present in the other. Consequently, these two processes should elicit the same components. Thus, if reading and listening entail the same processes, they should also elicit identical components, in line with Friederici et al. (1993; 1996). However, at least one study (Hagoort & Brown, 2000) that tested the same materials in listening and reading showed that this is not necessarily true. That is, two out of three morphosyntactic conditions elicited the LAN in listening but not in reading.

Earlier on, we showed that the LAN could not be consistently elicited by gender disagreement in reading studies, with inconsistent results not only across different languages (see Molinaro, 2011), but also across studies in the same language (e.g., Barber & Carreiras, 2005; Wicha et al., 2004). In addition, this thesis demonstrated that the LAN is inconsistently elicited by the same materials in different presentation modalities, which is in line with Hagoort and Brown's (2000) study. By summing up the results from all four experiments, we will offer a few speculations on some aspects underlying the LAN issue.

First of all, we do not have an adequate explanation for the presence of the LAN in Italian and its absence in Dutch. Since such results are in line with a number of studies, which usually report the LAN in Italian but not in Dutch (e.g., Meulman at al., 2014; Molinaro et al., 2008), one possibility is that the difference may be language-specific. Perhaps the morphosyntactic violations in Dutch were not as straightforward as the violations in Italian, due to the homophony of the definite articles (*de* is used both for singular noun with common gender and for all plural nouns), which delayed or reduced the effect of violation detection. However, if a gender violation in Dutch was so weak that it did not trigger the LAN in reading, it probably should have had the same effect in the auditory modality.

An even more perplexing finding is the absence of the LAN in the semantic gender condition in Italian and its presence in the gender condition in Dutch in the listening modality. In the former case, the result triggers the question 'Why is LAN not present in both conditions?' and in the latter 'Why is LAN present at all?' It is obvious that there is a relationship between presentation modality and the presence/absence of the LAN, but which intrinsic characteristic of either modality affects the LAN remains unclear. We believe that the simplest explanation is that the LAN is highly susceptible to the difference in timing between reading and listening. In both auditory studies, the LAN was recorded only in those conditions in which the violation was not word-final. More precisely, both the violations that elicited the LAN (syntactic gender disagreement in Italian and gender disagreement in Dutch) are lexical violations, meaning that they usually precede the end of the word. Conversely, the violations in semantic gender disagreement and number disagreement are all word-final, at the level of an inflectional morpheme. The distance between the word-final and non word-final violations ranged between 100 and 200 ms. In all cases, the last word of the noun phrase was the target noun. In non word-final violations, the parser had ample time (100-200 ms) to recognize the violation before reaching the noun phrase boundary. Word-final violations, however, corresponded to the noun phrase boundary. In this case, the parser had to recognize the violation while simultaneously performing phrasal integration and preparing for the repair/reanalysis. Possibly, the additional processing load at the phrasal boundary obfuscated the detection mechanism signalled by the LAN.

The conditions that elicited the LAN share another feature. Namely, both syntactic gender disagreement in Italian and gender disagreement in Dutch elicited a weaker P600 effect, as compared to semantic gender disagreement in Italian and number disagreement in Dutch. There may be a relationship between the size of the P600 effect and the LAN - a larger P600 effect might cancel out the LAN (cf. Osterhout, 1997). Even though the LAN precedes the P600 in time, the time difference between components may have been affected by the auditory nature of the experiment. All in all, a minimized temporal difference between components and a large P600 effect may have cancelled out the LAN effect in the conditions in which it was not elicited.

Both accounts are based on the assumption that the LAN should have been present in all experiments and conditions but that certain factors prevented its elicitation. We offered two possible accounts of the LAN in the auditory modality. However, we cannot explain the behaviour of the LAN in the visual modality, which remains a question for further research.

6.6 Conclusion and Future Research

A number of issues were addressed in the current thesis, some of which were successfully solved, whereas others require further investigation. We presented a study on a syntactic phenomenon, that is (dis)agreement, and its relationship with ERP components traditionally associated with syntactic processing. To provide a more comprehensive account, we used both auditory and visual stimuli. This approach helped, to an extent, to disentangle language-related effects from effects that may depend on extra-linguistic factor, such as the time course of stimulus presentation.

As a first issue, we addressed the components elicited by gender disagreement processing. In line with most studies in the field, the P600 was obtained in all experiments. Not entirely unexpectedly, the LAN was elicited only in 50% of the conditions. This replicates a number of reading studies that reported the LAN in disagreement processing in Italian but not in Dutch. Unfortunately, there is not enough evidence to explain the seemingly inconsistent behaviour of the LAN, which warrants further research.

Next, we managed to document a relationship between the size of the P600 effect and the complexity of the repair/reanalysis mechanism. Furthermore, in line with previous research, we provided some evidence showing that the repair/reanalysis process occurs in the late stage of the P600. However, the distinction between early and late P600 could not be entirely upheld, and it deserves additional attention in the future.

Finally, we demonstrated the role of the presentation modality in studying sentence processing, and particularly the need to take into account the time course of the auditory information flow. We provided evidence supporting the idea that the auditory P600 and the LAN are violation-aligned, which is relevant for language processing theories, as well as for experimental design. In addition, we showed that the presentation modality can influence whether or not the LAN is elicited. Rather than interpreting the presence or absence of the LAN as a marker of the presence or absence of language processes, we claim that the LAN can be influenced by extra-linguistic factors, such as the time course of stimulus presentation or an interaction with the P600.

References

Adger, D., & Harbour, D. (2008). Why Phi. In D. Harbour, D. Adger, & S. Bejar (Eds.), *Phi-theory: phi features across interfaces and modules* (pp. 1-34). Oxford: Oxford University Press.

Balconi, M., & Pozzoli, U. (2005). Comprehending semantic and grammatical violations in Italian. N400 and P600 comparison with visual and auditory stimuli. *Journal of Psycholinguistic Research*, *34*, 71–98.

Bañon, J. A., Fiorentino, R., & Gabriele, A. (2012). The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance. *Brain Research*, *1456*, 49–49.

Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience, 17*, 137-153.

Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *On-line study of sentence comprehension; eye-tracking, ERP and beyond* (pp. 309–328). Brighton, UK: Psychology Press.

Bošković, Ž. (2011). On unvalued uninterpretable features. In *Proceedings of the North East Linguistics Society Annual Meeting (Vol. 39)*, 109-120.

Brysbaert, M., & Dijkstra, T. (2006). Changing views on word recognition in bilinguals. In J. Morais & G. d'Ydewalle (Eds.), *Bilingualism and second language acquisition* (25-37). Brussels, Belgium: The Royal Academies for Science and the Arts of Belgium..

Cacciari, C. (2011). *Psicologia del linguaggio [Psychology of language]*. Bologna, Italy: Il Mulino.

Caffarra, S., & Barber, H. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research*, *1605*, 83-92.

Caffarra, S., Siyanova-Chanturia, A., Pesciarelli, F., Vespignani, F., & Cacciari, C. (2015). Is the noun ending a cue to grammatical gender processing? An ERP study on sentences in Italian. *Psychophysiology*, *52*, 1019-1030.

Chomsky, N. (2000). Minimalist inquiries. In R. Martin, D. Michaels, & J. Uriagereka (Eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik* (pp. 89–156). Cambridge, MA: The MIT Press.

Chomsky, N. (2001). Derivation by phase. In Michael Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: The MIT Press.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256‑266.

Corbett, G. G. (1991). *Gender.* Cambridge: Cambridge University Press.

Corbett, G. G. (2003). Agreement: Canonical instances and the extent of the phenomenon. In *Morphology: Selected papers from the Third Mediterranean Morphology Meeting. September 20-22, 2001* (pp. 109-128).

Corbett, G. G. (2006). Gender, grammatical. *Encyclopedia of language & linguistics*, 749-756.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, *13*, 21-58.

D'Achille, P., & Thornton, A. M. (2006). I nomi femminili in –o [Feminine nouns ending in –o]. In E. Cresti (Ed.), *Prospettive nello studio del lessico italiano, Atti SILFI* (pp. 473–481). Firenze, Italy: FUP.

Dahl, O. (2000). Animacy and the notion of semantic gender. *Trends in Linguistics Studies and Monographs*, *124*, 99-116.

Demestre, J., Meltzer, S., Garcia-Albea, J. E., & Vigil, A. (1999). Identifying the null subject: Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *28*, 293-312.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing genderagreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*, 200-224.

Dowens, M. G., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*, 1870-1887.

Dragoy, O., Stowe, L. A., Bos, L. S., & Bastiaanse, R. (2012). From time to time: Processing time reference violations in Dutch. *Journal of Memory and Language*, *66*, 307-325.

Faussart, C., Jakubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista di Linguistica*, *11*, 75-101.

Faustmann, A., Murdoch, B. E., Finnigan, S. P., & Copland, D. A. (2005). Event-related brain potentials elicited by semantic and syntactic anomalies during auditory sentence processing. *Journal of the American Academy of Audiology*, *16*, 708-725.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language*, *50*, 259–281.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78-84.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language*, *43*, 476-507.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). The temporal structure of syntactic parsing: Early and late effects elicited by syntactic anomalies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1219‑1248.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*, 45-63.

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research*, *28*, 467-484.

Friederici, A. D., & Meyer, M. (2004). The brain knows the difference: Two types of grammatical violations. *Brain Research*, *1000*, 72‑77.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 2*, 183-192.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, *27*, 438-453.

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Research*, *1146*, 50-58.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*, 149-188.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*, 267–283.

Gunter, T., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556-568.

Hagoort, P. (2003a). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, *20*, 18-29.

Hagoort, P. (2003b). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15*, 883–899.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*(1493), 1055-1069.

Hagoort, P., & Brown, C. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*, 715–728.

Hagoort, P., & Brown, C. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*, 1531-1549.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*,439-483.

Hagoort, P., Brown, C., & Osterhout, L. (1999). The neural architecture of syntactic processing. In C. M. Brown & P. Hagoort (Eds.), *Neurocognition of language* (pp. 273-317). Oxford, UK: Oxford University Press.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*, 194-205.

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

Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, *11*, 199-212.

Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2005). Pronominal reference in sentences about persons or things: An electrophysiological approach. *Journal of Cognitive Neuroscience*, *17*, 227–239.

Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*, 1207-1219.

Hickey, R. (1999). The phonology of gender in Modern German. In Rissanen, Matti, & Unterbeck (Eds.), *Gender in grammar and cognition* (pp. 621-663). Berlin: Mouton-de Gruyter.

Hockett, C. F. (1958). *A course in modern linguistics*. New York: Holt/Rinehart and Winston.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass, 1*, 571–591.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159-201.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*, 98-110.

Kutas, M., & Federmeier, K. D. (2007). Event-related brain potential (ERP) studies of sentence processing. In G. Gaskell (Ed.), *Oxford handbook of psycholinguistics* (pp. 385-406). Oxford: Oxford University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics Electrified: Event-related potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83-143). Academic Press.

Levelt, W. (1989). *Speaking: From intention to articulation.* Cambridge, MA: The MIT Press.

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*, 1-22.

Levelt, W. J. (1999). Models of word production. *Trends in Cognitive Sciences*, *3*, 223-232.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-38.

Levelt, W. J., & Schriefers, H. (1987). Stages of lexical access. In *Natural language generation* (pp. 395-404). Springer Netherlands.

Loerts, H., Stowe, L. A., & Schmid, M. S. (2013). Predictability speeds up the re-analysis process: An ERP investigation of gender agreement and cloze probability. *Journal of Neurolinguistics*, *26*, 561-580.

Luck, S. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: The MIT Press.

Lukatela, G., Kostić, A., Todorović, D., Carello, C., & Turvey, M. T. (1987). Type and number of violations and the grammatical congruency effect in lexical decision. *Psychological Research*, *49*, 37-43.

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

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1-71.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research*, *1093*, 178-189.

Meulman, N., Stowe, L. A., Sprenger, S. A., Bresser, M. & Schmid, M. S. (2014). An ERP study on L2 syntax processing: When do learners fail? *Frontiers in Psychology*, *5: 1072*.

Miozzo, M., & Caramazza, A. (1997). Retrieval of lexical–syntactic features in tip-of-the tongue states. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1410-1423.

Molinaro, N., Barber, H., Caffarra, S., & Carreiras, M. (2014). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex*, *66*, 156-159.

Molinaro, N., Barber, H., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*, 908–930.

Molinaro, N., Vespignani, F., & Job, R. (2008). A deeper reanalysis of a superficial feature: An ERP study on agreement violations. *Brain Research*, *1228*, 161-176.

Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011). Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language*, *64*, 211-232.

Münte, T. F., Heinze, H. J., & Mangun, G. R. (1993). Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience*, 335-344.

Münte, T. F., Szentkuti, A., Wieringa, B. M., Matzke, M., & Johannes, S. (1997). Human brain potentials to reading syntactic errors in sentences of different complexity. *Neuroscience Letters*, *235*, 105-108.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151-165.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research*, *1164*, 81-94.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

O'Rourke, P. L., & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research*, *1392*, 62-79.

O'Rourke, T. B., & Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology*, *60*, 121-150.

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494-522.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition, 25*, 273-285.

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785-806.

Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, *8*, 413-437.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739-773.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 271-308). Psychology Press.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, *14*, 283-317.

Otten, L. J., & Rugg, M. D. (2005). Interpreting Event-Related Brain Potentials. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 1-16). Cambridge, MA: The MIT Press.

Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, *22*, 2728-2744.

Pesetsky, D., & Torrego, E. (2007). The syntax of valuation and the interpretability of features. In S. Karimi, V. Samiian, & W. D. Wilkins (Eds.), *Phrasal and clausal architecture: Syntactic derivation and interpretation* (pp. 262-294). John Benjamins Publishing.

Radeau, M., Morais, J., Mousty, P., & Bertelson, P. (2000). The effect of speaking rate on the role of the uniqueness point in spoken word recognition. *Journal of Memory and Language*, *42*, 406-422.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Rayner, K., & Clifton, C. (2009). Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research. *Biological Psychology*, *80*, 4-9.

Ritter, E. (1991). Two functional categories in noun phrases. *Syntax and Semantics*, *25*, 37-62.

Ritter, E. (1993). Where is gender? *Linguistic Inquiry*, *24*, 795–803.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*, 249-284.

Schmitt, B. M., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research*, *14*, 333-346.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Neuroreport*, *9*, 2195-2200.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*, 489-493.

Shetter, W. Z. (1959). The dutch diminutive. *The Journal of English and Germanic Philology*, *58*, 75-90.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2012). The role of stimulus cross-splicing in an event-related potentials study. Misleading formant transitions hinder automatic phonological processing. *The Journal of the Acoustical Society of America*, *131*, 3120-3140.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*, 135-162.

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A Commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex*, *66*, 149-155.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289-301.

Van Berkum, J. (1996). *The psycholinguistics of grammatical gender: Studies in language comprehension and production.* Doctoral dissertation, Max Planck Institute for Psycholinguistics. Nijmegen, Netherlands: Nijmegen University Press.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967-985.

Van den Brink, D., Brown, C., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 364-372.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning Memory & Cognition*, *25*, 394-417.

Vigliocco, G., Antonini, T., & Garrett, M. F. (1997). Grammatical gender is on the tip of Italian tongues. *Psychological Science*, *8*, 314–314.

Vigliocco, G., & Franck, J. (1999). When sex and syntax go hand in hand: Gender agreement in language production. *Journal of Memory and Language*, *40*, 455–455.

Vigliocco, G., & Hartsuiker, R. J. (2005). Maximal input and feedback in production and comprehension. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones*, (pp. 209-228). Mahwah, NJ: Erlbaum.

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*, 501.

Vigliocco, G., & Zilli, T. (1999). Syntactic accuracy in sentence production: The case of gender disagreement in Italian language-impaired and unimpaired speakers. *Journal of Psycholinguistic Research*, *28*, 623-648.

Wicha, N. Y. Y., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*, 1272–1288.

A Appendix to Chapter 1 and Chapter 2

Matrerials used in the reading (Chapter 1) and listening experiment (Chapter 2) on Italian. Both experiments used the same experimental and filler items.

Experimental Items (N = 160)

Oggi le figlie vanno in piscina.
Spesso le ragazze cucinano dolci.
Generalmente le bambine ballano molto.
Spesso le nonne fanno la maglia.
Generalmente le vecchie infastidiscono i vicini.
Ieri i vedove hanno comprato dei fiori.
Ieri i fidanzate hanno spedito gli inviti di nozze.
Adesso i cugine hanno un cane.
Spesso i monache consolano i fedeli.
Domani i cognate portano i pasticcini.
Ieri la vedova ha comprato dei fiori.
Ieri la fidanzata ha spedito gli inviti di nozze.
Adesso la cugina ha un cane.
Spesso la monaca consola I fedeli.
Domani la cognata porta I pasticcini.
Oggi il figlia va in piscina.
Spesso il ragazza cucina dolci.
Generalmente il bambina balla molto.
Spesso il nonna fa la maglia.
Generalmente il vecchia infastidisce I vicini.
Ieri i vedovi hanno pianto tanto.
Finalmente i fidanzati hanno comprato degli anelli.
Adesso i cugini hanno soldi.
Spesso i monaci puliscono il convento.
Domani i cognati accompagnano lo sposo.
Ora le figli stanno aiutando la mamma.
Spesso le ragazzi guardano la tv.
Generalmente le bambini dormono tranquilli.
Spesso le nonni fumavano la pipa.
Generalmente le vecchi russavano molto.
Ora il figlio sta aiutando la mamma.
Spesso il ragazzo guarda la tv.

Generalmente il bambino dorme tranquillo.
Spesso il nonno fumava la pipa.
Generalmente il vecchio russava molto.
Ieri la vedovo ha pianto tanto.
Ieri la fidanzato ha comprato un anello.
Adesso la cugino ha soldi.
Spesso la monaco pulisce il convento.
Domani la cognato accompagna lo sposo.
Adesso le bambole piacciono alla bambina.
Raramente le lettere arrivano in tempo.
Ieri le bottiglie sono cadute dal tavolo.
Spesso le finestre scricchiolano rumorosamente.
Generalmente le valige pesano tanto.
Ora i statue dominano sulla città.
Spesso i mele aiutano la digestione.
Ieri i camice le stavano bene.
Oggi i bombe esploderanno in piazza.
Generalmente i macchine consumano molta benzina.
Ora la statua domina sulla città.
Spesso la mela aiuta la digestione.
Ieri la camicia le stava bene.
Oggi la bomba esploderà in piazza.
Generalmente la macchina consuma molta benzina.
Adesso il bambola piace alla bambina.
Raramente il lettera arriva in tempo.
Ieri il bottiglia è caduta dal tavolo.
Spesso il finestra scricchiola rumorosamente.
Generalmente il valigia pesa tanto.
Generalmente i giochi piacciono ai bambini.
Ieri i dati hanno rivelato la verità.
Adesso i telefoni squillano di continuo.
Spesso i piatti cadono dalla dispensa.
Generalmente i treni partono in ritardo.
Generalmente le cuscini puzzavano di fumo.
Spesso le pacchi arrivano in ritardo.
Ieri le numeri combaciavano con la funzione.
Spesso le palazzi scricchiolavano per il vento.
Ora le documenti sembrano falsi.
Generalmente il cuscino puzzava di fumo.
Spesso il pacco arriva in ritardo.
Ieri il numero combaciava con la funzione.
Spesso il palazzo scricchiolava per il vento.
Ora Il documento sembra falso.
Generalmente la gioco piace ai bambini.
Ieri la dato ha rivelato la verità.
Adesso la telefono squilla di continuo.
Spesso la piatto cade dalla dispensa.
Generalmente la treno parte in ritardo.
Ieri le vedove hanno comprato dei fiori.
Ieri le fidanzate hanno spedito gli inviti di nozze.

Adesso le cugine hanno un cane.
Spesso le monache consolano i fedeli.
Domani le cognate portano i pasticcini.
Oggi i figlie vanno in piscina.
Spesso i ragazze cucinano dolci.
Generalmente i bambine ballano molto.
Spesso i nonne fanno la maglia.
Generalmente i vecchie infastidiscono i vicini.
Oggi la figlia va in piscina.
Spesso la ragazza cucina dolci.
Generalmente la bambina balla molto.
Spesso la nonna fa la maglia.
Generalmente la vecchia infastidisce I vicini.
Ieri il vedova ha comprato dei fiori.
Ieri il fidanzata ha spedito gli inviti di nozze.
Adesso il cugina ha un cane.
Spesso il monaca consola I fedeli.
Domani il cognata porta I pasticcini.
Ora i figli stanno aiutando la mamma.
Spesso i ragazzi guardano la tv.
Generalmente i bambini dormono tranquilli.
Spesso i nonni fumavano la pipa.
Generalmente i vecchi russavano molto.
Ieri le vedovi hanno pianto tanto.
Finalmente le fidanzati hanno comprato degli anelli.
Adesso le cugini hanno soldi.
Spesso le monaci puliscono il convento.
Domani le cognati accompagnano lo sposo.
Ieri il vedovo ha pianto tanto.
Ieri il fidanzato ha comprato un anello.
Adesso il cugino ha soldi.
Spesso il monaco pulisce il convento.
Domani il cognato accompagna lo sposo.
Ora la figlio sta aiutando la mamma.
Spesso la ragazzo guarda la tv.
Generalmente la bambino dorme tranquillo.
Spesso la nonno fumava la pipa.
Generalmente la vecchio russava molto.
Ora le statue dominano sulla città.
Spesso le mele aiutano la digestione.
Ieri le camice le stavano bene.
Oggi le bombe esploderanno in piazza.
Generalmente le macchine consumano molta benzina.
Adesso i bambole piacciono alla bambina.
Raramente i lettere arrivano in tempo.
Ieri i bottiglie sono cadute dal tavolo.
Spesso i finestre scricchiolano rumorosamente.
Generalmente i valige pesano tanto.
Adesso la bambola piace alla bambina.
Raramente la lettera arriva in tempo.

Ieri la bottiglia è caduta dal tavolo.
Spesso la finestra scricchiola rumorosamente.
Generalmente la valigia pesa tanto.
Ora il statua domina sulla città.
Spesso il mela aiuta la digestione.
Ieri il camicia le stava bene.
Oggi il bomba esploderà in piazza.
Generalmente il macchina consuma molta benzina.
Generalmente i cuscini puzzavano di fumo.
Spesso i pacchi arrivano in ritardo.
Ieri i numeri combaciavano con la funzione.
Spesso i palazzi scricchiolavano per il vento.
Ora I documenti sembrano falsi.
Generalmente le giochi piacciono ai bambini.
Ieri le dati hanno rivelato la verità.
Adesso le telefoni squillano di continuo.
Spesso le piatti cadono dalla dispensa.
Generalemente le treni partono in ritardo.
Generalmente il gioco piace ai bambini.
Ieri il dato ha rivelato la verità.
Adesso il telefono squilla di continuo.
Spesso il piatto cade dalla dispensa.
Generalemente il treno parte in ritardo.
Generalmente la cuscino puzzava di fumo.
Spesso la pacco arriva in ritardo.
Ieri la numero combaciava con la funzione.
Spesso la palazzo scricchiolava per il vento.
Ora la documento sembra falso.

Fillers (N = 80)

Il flauto d'argento era suonato da un membro dell'orchestra
Il burrocacao é stato applicato sulle labbra screpolate.
La lettera d’affari é stata scritta dalla segretaria.
La spina non é nella presa e per questo non succede nulla.
Le scale lassú sono troppo ripide per me.
Nessuno ha piú esperienza tranne forse Carlo.
Da questo momento non puoi piú fumare nel mio salotto.
Sembra meglio di com’é perché non sei qui.
C’é una bellissima poesia scritta sulla lavagna.
Le domestiche stanno preparando il letto prima che il presidente arrivi.
La ragazza disperata si é suicidere ieri.
La valigia ha causare a tutti mal di schiena.
Il tempo essere bello in questo villaggio per tutto il mese.
Io preferire essere su un’isola tropicale.
Speriamo che non causare troppo dolore al tuo fidanzato.
Giovanni non ha mai vedere tulipani cosí belli.
La vela principale é stata issare velocemente dai marinai.
I denti bianchi sono stati lavare dal bambino.
Gli italiani non bere mai direttamente dalla bottiglia ma bevono sempre dal bicchiere.

I turisti stanno cavalcare un asino prima di andare via.
Il maglione di lana é stato lavorato a maglia da mia nonna.
Questo olandese non ha gusto chiaramente.
La casa fatiscente é stata finalmente venduta e rinnovata.
Le carte sono state mischiate dai giocatori d'azzardo.
Spero che gli impiegati mi aiutino con questi moduli complicati.
Lei ha trovato il regalo semplice molto bello.
Il grano é stato macinato da forti mugnai.
I mocciosi accendono sempre pericolosi fuochi d'artificio.
Il pane integrale era fresco di forno ogni giorno.
Il criminale violento ha infranto la legge.
I netturbini raccogliere l'immondizia ogni settimana.
Il fornaio di paese essere un vero gentiluomo.
Le storie emozionanti vengono raccontare intorno al faló.
Queste arancie non produrre molto succo.
Tutta la famiglia ha giocare al nuovo gioco da tavola fino a tarda sera.
L'ambiente é molto inquinare dalle fabbriche sporche.
La birra migliore é fare fermentare al sud.
La giovane lepre é stata sparare dal cacciatore.
Questa musica é stata comporre da grandi esperti.
Gli scudieri mangiare un uovo bollito al giorno.
Generalmente la gioco piace ai bambini.
La valigia ha causato a tutti mal di schiena.
Il tempo sará bello in questo villaggio per tutto il mese.
Preferirei essere su un'isola tropicale.
Speriamo che non causi troppo dolore al tuo fidanzato.
Giovanni non ha mai visto tulipani cosí belli.
I denti bianchi sono stati lavati dal bambino.
Gli italiani non bevono mai direttamente dalla bottiglia ma bevono sempre dal bicchiere.
I turisti stanno cavalcando un asino prima di andare via.
Le domestiche stanno preparando il letto prima che il presidente arrivi.
Il flauto d'argento era suonare da un membro dell'orchestra.
Il burrocacao é stato applicare sulle labbra screpolate.
La lettera d'affari é stata scrivere dalla segretaria.
La spina non essere nella presa e per questo non succede nulla.
Le scale lassú essere troppo ripide per me.
Nessuno avere piú esperienza tranne forse Carlo.
Da questo momento voi non potere piú fumare nel mio salotto.
Sembra meglio di com'é perché ti non essere qui.
Le domestiche stanno preparare il letto prima che il presidente arrivi.
C'essere una bellissima poesia scritta sulla lavagna.
I netturbini raccolgono l'immondizia ogni settimana.
Il fornaio di paese é un vero gentiluomo.
Le storie emozionanti vengono raccontate intorno al faló.
Queste arancie non producono molto succo.
Tutta la famiglia ha giocato al nuovo gioco da tavola fino a tarda sera.
L'ambiente é molto inquinato dalle fabbriche sporche.
La birra migliore é fatta fermentare al sud.
La giovane lepre é stata sparata dal cacciatore.
Questa musica é stata composta da grandi esperti.

Gli scudieri mangiavano un uovo bollito al giorno.
Il maglione di lana é stato lavorare a maglia da mia nonna.
Questo olandese non avere gusto chiaramente.
La casa fatiscente é stata finalmente vendere e rinnovata.
Le carte sono state mischiare dai giocatori d'azzardo.
Spero che gli impiegati mi aiutare con questi moduli complicati.
Lei ha trovare il regalo semplice molto bello.
Il pane integrale essere fresco di forno ogni giorno.
Il grano é stato macinare da forti mugnai.
Il criminale violento ha infrangere la legge.
I mocciosi accendere sempre pericolosi fuochi d'artificio.

Practice Items (N = 5)

La canzone famosa era cantata dai muratori.
Il bimbo ha fatto un grosso pupazzo di neve.
Il mais é stato falciare dagli agricoltori.
La maglietta sgualcita é stata stirare dal lavoratore.
L'ufficiale fare una multa alta.

B Appendix to Chapter 3 and Chapter 4

Materials used in the reading (Chapter 3) and listening experiment (Chapter 4) on Dutch. In the reading experiment (Chapter 3), only threesyllabic experimental items were used (21-40), wheraeas all experimental items (1-40) were used in the listening experiment (Chapter 4). All filler items (1-20) were used in both experiments.

Experimental Items (Chapter 3, N = 320; Chapter 4, N = 160)

1. Doek

Er is een wit doek in de bioscoop.
Hij heeft een enorm doek beschilderd.
De oude doeken beschermen de parketvloer tijdens het klussen.
De gedrapeerde doeken dienen ter decoratie.
Er is een witte doek in de bioscoop.
Hij heeft een enorme doek beschilderd.
Het oude doeken beschermen de parketvloer tijdens het klussen.
Het gedrapeerde doeken dienen ter decoratie.

2. Gas

Er is een giftig gas in de atmosfeer terecht gekomen.
Hij heeft een gevaarlijk gas ingeademd.
De schadelijke gassen zijn slecht voor het milieu.
De reukloze gassen zijn moeilijk waar te nemen.
Er is een giftig gas in de atmosfeer terecht gekomen.
Hij heeft een gevarlijke gas ingeademd.
Het schadelijke gassen zijn slecht voor het milieu.
Het reukloze gassen zijn moeilijk waar te nemen.

3. Fort

Hier stond een prachtig fort in de Middeleeuwen.
Zij heeft een indrukwekkend fort laten bouwen.
De grote forten staan verspreid over het landschap.
De oude forten werden vernield tijdens de oorlog.
Hier stond een prachtige fort in de Middeleeuwen.
Zij heeft een indrukwekkende fort laten bouwen.

Het grote forten staan verspreid over het landschap.
Het oude forten werden vernield tijdens de oorlog.

4. Hek

Hij heeft een lang hek laten verven.
Er stond een hoog hek om het landgoed.
De sierlijke hekken werden mooi verguld.
De stevige hekken werden snel gebouwd.
Hij heeft een lange hek laten verven.
Er stond een hoge hek om het landgoed.
Het sierlijke hekken werden mooi verguld.
Het stevige hekken werden snel gebouwd.

5. Front

Dat is een politiek front tegen windmolens.
Zij vormden een onverslaanbaar front tegen de vijand.
De radicale fronten vochten tegen elkaar.
De vijandelijke fronten zijn in gesprek.

Dat is een politieke front tegen windmolens.
Zij vormden een onverslaanbarefront tegen de vijand.
Het radicale fronten vochten tegen elkaar.
Het vijandelijke fronten zijn in gesprek.

6. Mes

Er lag een bloederig mes op de grond.
Zij legde een vlijmscherp mes op tafel.
De botte messen moeten geslepen worden.
De nieuwe messen snijden gemakkelijk.
Er lag een bloederige mes op de grond.
Zij legde een vlijmscherpe mes op tafel.
Het botte messen moeten geslepen worden.
Het nieuwe messen snijden gemakkelijk.

7. Nest

Er lag een kapot nest op de grond.
Zij bouwden een klein nest hoog in de boom.
De mooie nesten zijn vernield.
De lege nesten waren verspreid over het bos.
Er lag een kapotte nest op de grond.
Zij bouwden een kleine nest hoog in de boom.
Het mooie nesten zijn vernield.
Het lege nesten waren verspreid over het bos.

8. Net

Er hing een stevig net onder de trapeze.
Hij spande een klein net over de boomgaard.
De grote netten liggen in de zee.
De beschadigde netten werden gerepareerd.
Er hing een stevige net onder de trapeze.

Hij spande een kleine net over de boomgaard.
Het grote netten liggen in de zee.
Het beschadigde netten werden gerepareerd.

9. Oor

Er is eennieuw oor aan het kopje gelijmd.
Hij heeft een blauw oor door de kou.
De grote oren hielpen de wolf om beter te horen.
De gevoelige oren konden het harde geluid niet verdragen.
Er is een nieuwe oor aan het kopje gelijmd.
Hij heeft een blauwe oor door de kou.
Het grote oren hielpen de wolf om beter te horen.
Het gevoelige oren konden het harde geluid niet verdragen.

10. Park

Dat was een gevaarlijk park voor jonge kinderen.
Zij ontdekten een rustig park midden in de stad.
De mooie parken zijn druk op zomeravonden.
De groene parken moeten behouden blijven.
Dat was een gevaarlijke park voor jonge kinderen.
Zij ontdekten een rustige park midden in de stad.
Het mooie parken zijn druk op zomeravonden.
Het groene parken moeten behouden blijven.

11. Plein

Er was een groot plein voor de kerk.
Zij bezochten een prachtig plein in de oude stad.
De drukke pleinen zijn makkelijke doelwitten.
De kleine pleinen zijn met elkaar verbonden door steegjes.
Er was een grote plein voor de kerk.
Zij bezochten een prachtige plein in de oude stad.
Het drukke pleinen zijn makkelijke doelwitten.
Het kleine pleinen zijn met elkaar verbonden door steegjes.

12. Dorp

Er lag een mooi dorp vlakbij de grote stad.
Zij verliet een prachtig dorp met pijn in haar hart.
De gezellige dorpen trekken veel toeristen in de zomer.
De noordelijke dorpen hebben last van aardbevingen.
Er lag een mooie dorp vlakbij de grote stad.
Zij verliet een prachtige dorp met pijn in haar hart.
Het gezellige dorpen trekken veel toeristen in de zomer.
Het noordelijke dorpen hebben last van aardbevingen.

13. Touw

Er lag een kapot touw in de garage.
Hij bond een stevig touw om de mast.
De lange touwen hingen over de reling.
De dikke touwen waren om de paal geknoopt.
Er lag een kapotte touw in de garage.

Hij bond een stevige touw om de mast.
Het lange touwen hingen over de reling.
Het dikke touwen waren om de paal geknoopt.

14. Vak
Dat is een lastig vak volgens veel leerlingen.
Zij heeft een ander vak gekozen op school.
De moeilijke vakken werden het minst gewaardeerd.
De nieuwe vakken worden goed beoordeeld door studenten.
Dat is een lastige vak volgens veel leerlingen.
Zij heeft een andere vak gekozen op school.
Het moeilijke vakken werden het minst gewaardeerd.
Het nieuwe vakken worden goed beoordeeld door studenten.

15. Vlak
Er zat een vergeeld vlak op de muur.
Hij kleurde een groot vlak in met een stift.
De blauwe vlakken zijn opvallend.
De donkere vlakken moeten opnieuw worden geverfd.
Er zat een vergeelde vlak op de muur.
Hij kleurde een grote vlak in met een stift.
Het blauwe vlakken zijn opvallend.
Het donkere vlakken moeten opnieuw worden geverfd.

16. Wiel
Er zat een nieuw wiel op de fiets.
Zij verving een kapot wiel van de auto.
De kleine wielen reden over de weg.
De piepende wielen werden vervangen.
Er zat een nieuwe wiel op de fiets.
Zij verving een kapotte wiel van de auto.
Het kleine wielen reden over de weg.
Het piepende wielen werden vervangen.

17. Lint
Hier is een lang lint langs de weg gespannen.
Zij strikte een rood lint om haar vlecht.
De gekleurde linten hangen in de bomen.
De rafelige linten werden afgeknipt.
Hier is een lange lint langs de weg gespannen.
Zij strikte een rode lint om haar vlecht.
Het gekleurde linten hangen in de bomen.
Het rafelige linten werden afgeknipt.

18. Meer
Daar is een diep meer vol vissen.
Zij bezocht een rustig meer in de bossen.
De grote meren zijn echte trekpleisters voor toeristen.
De mooie meren worden het meest bezocht.
Daar is een diepe meer vol vissen.

Zij bezocht een rustige meer in de bossen.
Het grote meren zijn echte trekpleisters voor toeristen.
Het mooie meren worden het meest bezocht.

19. Pact
Dat was een historisch pact tussen Oost en West.
Zij sloten een geheim pact om dat bedrijf te saboteren.
De omvangrijke pacten zorgden voor veel problemen.
De belangrijke pacten veranderden de geschiedenis.
Dat was een historische pact tussen Oost en West.
Zij sloten een geheime pact om dat bedrijf te saboteren.
Het omvangrijke pacten zorgden voor veel problemen.
Het belangrijke pacten veranderden de geschiedenis.

20. Ras
Het is een nieuw ras van Nederlandse bodem.
Hij heeft een sterk ras gefokt.
De verschillende rassen reageren anders op prikkels.
De dure rassen zijn niet altijd beter.
Het is een nieuwe ras van Nederlandse bodem.
Hij heeft een sterke ras gefokt.
Het verschillende rassen reageren anders op prikkels.
Het dure rassen zijn niet altijd beter.

21. Apparaat
Er ligt een ongebruikt apparaat in de la.
Hij kocht een duur apparaat voor in de keuken.
De kleine apparaten zijn vaak even effectief als de grote.
De nieuwe apparaten zijn voorzien van technische snufjes.
Er ligt een ongebruikte apparaat in de la.
Hij kocht een dure apparaat voor in de keuken.
Het kleine apparaten zijn vaak even effectief als de grote.
Het nieuwe apparaten zijn voorzien van technische snufjes.

22. Avontuur
Dat was een spannend avontuur voor alle reisgenoten.
Hij beleeft een groot avontuur op safari.
De nieuwe avonturen worden beschreven in het boek.
De gevaarlijke avonturen zijn vaak het leukst.
Dat was een spannende avontuur voor alle reisgenoten.
Hij beleeft een grote avontuur op safari.
Het nieuwe avonturen worden beschreven in het boek.
Het gevaarlijke avonturen zijn vaak het leukst.

23. Compliment
Dat was een mooi compliment over haar werk.
Hij gaf een groot compliment aan zijn leerling.
De onverwachte complimenten zijn vaak het leukst.
De welgemeende complimenten werken motiverend.

Dat was een mooie compliment over haar werk.
Hij gaf een grote compliment aan zijn leerling.
Het onverwachte complimenten zijn vaak het leukst.
Het welgemeende complimenten werken motiverend.

24. Continent
Er ligt een enorm continent onder de evenaar.
Hij ontdekte een nieuw continent in de 15^e eeuw.
De kleine continenten zijn goed vertegenwoordigd.
De indrukwekkende continenten werden door haar omschreven.
Er ligt een enorme continent onder de evenaar.
Hij ontdekte een nieuwe continent in de 15^e eeuw.
Het kleine continenten zijn goed vertegenwoordigd.
Het indrukwekkende continenten werden door haar omschreven.

25. Document
Er is een oud document verdwenen uit het archief.
Zij leest een belangrijk document over klimaatverandering.
De uitgebreide documenten worden besproken.
De handige documenten zijn altijd binnen handbereik.
Er is een oude document verdwenen uit het archief.
Zij leest een belangrijke document over klimaatverandering.
Het uitgebreide documenten worden besproken.
Het handige documenten zijn altijd binnen handbereik.

26. Etiket
Hier is een groot etiket op geplakt.
Hij las een verkeerd etiket op de fles.
De mooie etiketten worden eerder verkocht.
De goedkope etiketten zien er minder goed uit.
Hier is een grote etiket op geplakt.
Hij las een verkeerde etiket op de fles.
Het mooie etiketten worden eerder verkocht.
Het goedkope etiketten zien er minder goed uit.

27. Formulier
Er ligt een belangrijk formulier in de kast.
Zij moest een lang formulier invullen voor school.
De medische formulieren moeten ingevuld worden.
De verplichte formulieren liggen op de stapel.
Er ligt een belangrijke formulier in de kast.
Zij moest een lange formulier invullen voor school.
Het medische formulieren moeten ingevuld worden.
Het verplichte formulieren liggen op de stapel.

28. Huwelijk
Dat was een vervelend huwelijk voor beide partijen.
Zij heeft een uitstekend huwelijk met haar echtgenoot.
De moderne huwelijken zijn anders dan vroeger.
De kerkelijke huwelijken zijn verbintenissen voor het leven.

Dat was een vervelende huwelijk voor beide partijen.
Zij heeft een uitstekende huwelijk met haar echtgenoot.
Het moderne huwelijken zijn anders dan vroeger.
Het kerkelijke huwelijken zijn verbintenissen voor het leven.

29. Incident

Dat was een ernstig incident tijdens de operatie.
Zij maakte een grappig incident mee in de kroeg.
De dagelijkse incidenten worden opgelost in een gesprek.
De recente incidenten worden besproken in de vergadering.
Dat was een ernstige incident tijdens de operatie.
Zij maakte een grappige incident mee in de kroeg.
Het dagelijkse incidenten worden opgelost in een gesprek.
Het recente incidenten worden besproken in de vergadering.

30. Instituut

Er is een officieel instituut voor statistiek.
Zij heeft een nieuw instituut voor vluchtelingenhulp opgericht.
De bekende instituten gaan anders te werk.
De christelijke instituten voor onderwijs werken samen.
Dat is een officiële instituut voor statistiek.
Zij heeft een nieuwe instituut voor vluchtelingenhulp opgericht.
Het bekende instituten gaan anders te werk.
Het christelijke instituten voor onderwijs werken samen.

31. Instrument

Er lag een zeldzaam instrument op het podium.
Hij gebruikte een klassiek instrument voor het concert.
De dure instrumenten zijn heel kwetsbaar.
De exotische instrumenten zijn moeilijk te bespelen.
Er lag een zeldzame instrument op het podium.
Hij gebruikte een klassieke instrument voor het concert.
Het dure instrumenten zijn heel kwetsbaar.
Het exotische instrumenten zijn moeilijk te bespelen.

32. Salaris

Er was een nieuw salaris vastgesteld voor die baan.
Hij verdiende een goed salaris met zijn baan.
De maandelijkse salarissen zijn op tijd uitbetaald.
De verlaagde salarissen zorgden voor veel onvrede.
Er was een nieuwe salaris vastgesteld voor die baan.
Hij verdiende een goede salaris met zijn baan.
Het maandelijkse salarissen zijn op tijd uitbetaald.
Het verlaagde salarissen zorgden voor veel onvrede.

33. Manuscript

Er lag een origineel manuscript van het boek in de kast.
Zij schreef een interessant manuscript over mode.
De afgekeurde manuscripten belandden bij het oud papier.
De complexe manuscripten waren te moeilijk voor de lezer.

Er lag een originele manuscript van het boek in de kast.
Zij schreef een interessante manuscript over mode.
Het afgekeurde manuscripten belandden bij het oud papier.
Het complexe manuscripten waren te moeilijk voor de lezer.

34. Monument
Het is een ontroerend monument geworden.
Hij bezocht een oud monument op de heuvel.
De moderne monumenten zijn zeker de moeite waard.
De beroemde monumenten bestaan al heel lang.
Het is een ontroerende monument geworden.
Hij bezocht een oude monument op de heuvel.
Het moderne monumenten zijn zeker de moeite waard.
Het beroemde monumenten bestaan al heel lang.

35. Pensioen
Het is een welverdiend pensioen voor deze werknemer.
U kunt een aanvullend pensioen opbouwen.
De verlaagde pensioenen zorgden voor veel onrust.
De nieuwe pensioenen zijn ongunstig voor veel werknemers.
Het is een welverdiende pensioen voor deze werknemer.
U kunt een aanvullende pensioen opbouwen.
Het verlaagde pensioenen zorgden voor veel onrust.
Het nieuwe pensioenen zijn ongunstig voor veel werknemers.

36. Ritueel
Er is een lang ritueel in het bos uitgevoerd.
Hij voerde een oud ritueel volgens traditie uit.
De heidense rituelen zijn onderdeel van de Halloweenviering.
De heilige rituelen zijn overbodig geworden.
Er is een lange ritueel in het bos uitgevoerd.
Hij voerde een oude ritueel volgens traditie uit.
Het heidense rituelen zijn onderdeel van de Halloweenviering.
Het heilige rituelen zijn overbodig geworden.

37. Resultaat
Er staat een onverwacht resultaat op de cijferlijst.
Zij kreeg een slecht resultaat terug.
De wekelijkse resultaten variëren altijd.
De teleurstellende resultaten zijn te wijten aan het onderwijs.
Er staat een onverwachte resultaat op de cijferlijst.
Zij kreeg een slechte resultaat terug.
Het wekelijkse resultaten variëren altijd.
Het teleurstellende resultaten zijn te wijten aan het onderwijs.

38. Schilderij
Er hing een goedkoop schilderij aan de muur.
Zij maakte een verrassend schilderij in de cursus.
De besmeurde schilderijen werden gerestaureerd.

De verkleurde schilderijen lagen op zolder.
Er hing een goedkope schilderij aan de muur.
Zij maakte een verrassende schilderij in de cursus.
Het besmeurde schilderijen werden gerestaureerd.
Het verkleurde schilderijen lagen op zolder.

39. Visioen
Dat was een helder visioen over de toekomst.
Hij kreeg een angstaanjagend visioen na het ongeluk.
De opmerkelijke visioenen zijn gebundeld in een boek.
De unieke visioenen worden uitgelegd.
Dat was een heldere visioen over de toekomst.
Hij kreeg een angstaanjagende visioen na het ongeluk.
Het opmerkelijke visioenen zijn gebundeld in een boek.
Het unieke visioenen worden uitgelegd.

40. Attribuut
Er is een nieuw attribuut in het assortiment.
Zij pakte een handig attribuut uit de keukenla.
De benodigde attributenliggen al klaar.
De geschikte attributen zijn hier te verkrijgen.
Er is een nieuwe attribuut in het assortiment.
Zij pakte een handige attribuut uit de keukenla.
Het benodigde attributen liggen al klaar.
Het geschikte attributen zijn hier te verkrijgen.

Fillers (N = 80)

1. Tomaat
Er ligt een rotte tomaat in de koelkast.
Zij heeft een rode tomaat van de plant geplukt.
Er ligt een rot tomaat in de koelkast.
Zij heeft een rood tomaat van de plant geplukt.

2. Crisis
Er ontstond een acute crisis bij het bedrijf.
Hij had een flinke crisis na het trauma.
Er ontstond een acuut crisis bij het bedrijf.
Hij had een flink crisis na het trauma.

3. Citroen
Er hangt een rijpe citroen aan de boom.
Zij kocht een zure citroen in de supermarkt.
Er hangt een rijp citroen aan de boom.
Zij kocht een zuur citroen in de supermarkt.

4. Balans
Er is een stevige balans nodig voor die oefening.
Ze maakte een nauwkeurige balans op voor het bedrijf.
Er is een stevig balans nodig voor die oefening.

Ze maakte een nauwkeurig balans op voor het bedrijf.

5. Viool
Er hing een zeldzame viool in de etalage.
Zij bespeelde een kostbare viool tijdens het concert.
Er hing een zeldzaam viool in de etalage.
Zij bespeelde een kostbaar viool tijdens het concert.

6. Tempel
Er staat een heilige tempel in de oude stad.
Ze bezochten een schitterende tempel op reis.
Er staat een heilig tempel in de oude stad.
Ze bezochten een schitterend tempel op reis.

7. Kliniek
Er is een uitstekende kliniek in die stad.
Hij bezit een beroemde kliniek voor drugsverslaafden.
Er is een uitstekend kliniek in die stad.
Hij bezit een beroemd kliniek voor drugsverslaafden.

8. Kopie
Hier ligt een geschikte kopie van de cijferlijst.
Zij maakte een duidelijke kopie van het testament.
Hier ligt een geschikt kopie van de cijferlijst.
Zij maakte een duidelijk kopie van het testament.

9. Impuls
Dat gaf een enorme impuls aan de ontwikkeling.
Zij wilde een grote impuls geven aan de kwaliteit van het onderwijs.
Dat gaf een enorme impuls aan de ontwikkeling.
Zij wilde een groot impuls geven aan de kwaliteit van het onderwijs.

10. Stoornis
Het is een ernstige stoornis die alleen voorkomt bij vrouwen.
Zij kreeg een milde stoornis in het spreken.
Het is een ernstig stoornis die alleen voorkomt bij vrouwen.
Zij kreeg een mild stoornis in het spreken.

11. Cursus
Er wordt een jaarlijkse cursus gegeven op school.
Zij verzorgt een intensieve cursus voor gepensioneerden.
Er wordt een jaarlijks cursus gegeven op school.
Zij verzorgt een intensief cursus voor gepensioneerden.

12. Status
Er is een bepaalde status die hij wil behalen.
Hij heeft een hoge status binnen de sector.
Er is een bepaald status die hij wil behalen.
Hij heeft een hoog status binnen de sector.

13. Woestijn
Er ligt een koude woestijn in Mongolië.
Zij bezoeken een hete woestijn tijdens de vakantie.
Er ligt een koud woestijn in Mongolië.
Zij bezoeken een heet woestijn tijdens de vakantie.

14. Sigaar
Er zat een kostbare sigaar in het doosje.
Hij stak een dure sigaar aan met de aansteker.
Er zat een kostbaar sigaar in het doosje.
Hij stak een duur sigaar aan met de aansteker.

15. Planeet
Dat is een onbekende planeet in ons zonnestelsel.
Hij tekende een grote planeet in het boek.
Dat is een onbekend planeet in ons zonnestelsel.
Hij tekende een groot planeet in het boek.

16. Fabriek
Er is een grote fabriek te vinden op het platteland.
Zij verbouwden een onveilige fabriek na veel protest.
Er is een groot fabriek te vinden op het platteland.
Zij verbouwden een onveilig fabriek na veel protest.

17. Ketting
Er ligt een unieke ketting bij de juwelier.
Hij gaf een dure ketting aan zijn vriendin.
Er ligt een uniek ketting bij de juwelier.
Hij gaf een duur ketting aan zijn vriendin.

18. Rivier
Daar stroomt een grote rivier door het landschap.
Zij staken een woeste rivier over op een vlot.
Daar stroomt een groot rivier door het landschap.
Zij staken een woest rivier over op een vlot.

19. Structuur
Het is een ingewikkelde structuur om uit te leggen.
Hij probeerde een simpele structuur te beschrijven.
Het is een ingewikkeld structuur om uit te leggen.
Hij probeerde een simpel structuur te beschrijven.

20. Factor
Dat is een onmisbare factor voor het proces.
Hij bezit een belangrijke factor om succesvol te worden.
Dat is een onmisbaar factor voor het proces.
Hij bezit een belangrijk factor om succesvol te worden.

Practice Items (N = 5)

Er is een uitstekende kliniek in die stad.
Hij probeerde een simpele structuur te beschrijven.
Daar stroomt een groot rivier door het landschap.
De zwakke signaal verbeterde langzaam.
Het noodzakelijke vetten staan op de dieetlijst