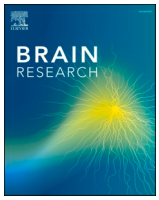
Brain Research 1887 (2026) 150390

Contents lists available at ScienceDirect

Brain Research

journal homepage: www.elsevier.com/locate/brainres



Research paper

Linear distance modulates P600 amplitude for gender agreement: An ERP study in Norwegian

Sarah Cameron [a,*], Natalia Kartushina [a], Björn Lundquist [b], Sendy Caffarra [c,d]

[a] Department of Linguistics and Scandinavian Studies, University of Oslo, Norway
[b] Department of Language and Culture, UiT The Arctic University of Norway, Tromsø, Norway
[c] Department of Biomedical, Metabolic and Neural Sciences, University of Modena and Reggio Emilia, Modena, Italy
[d] Division of Developmental-Behavioral Pediatrics, Stanford School of Medicine, Stanford, CA, United States

ARTICLE INFO

Keywords:
Sentence processing
Gender processing
Morphosyntactic processing
EEG
ERP
Norwegian

ABSTRACT

Former ERP studies have shown that linear distance (number of intervening words between two agreeing elements in a sentence) modulates P600 amplitude for gender agreement violations. However, these studies have primarily focused on across-phrase dependencies, making it difficult to tease apart the effects of linear distance from those of structural distance (number of intervening phrases between two agreeing elements in a sentence). Norwegian, an understudied language in gender processing literature, permits manipulation of linear distance while holding structural distance constant. We tested 36 native Norwegian speakers with EEG, comparing three violation types: noun-suffix (within-word), determiner-noun (within-phrase), and noun-predicative adjective (across-phrase). Our aims were to: 1) assess processing of within-word gender violations, which, to our knowledge has not previously been done; 2) examine the effect of linear distance on within-phrase gender agreement by comparing processing of noun-suffix and determiner-noun dependencies; and 3) establish an ERP baseline for Norwegian gender processing. All three violation types elicited P600 responses, indicating repair/reanalysis processes. Noun–predicative adjective violations additionally produced a left anterior negativity, and determiner–noun violations showed a late frontal negativity. The noun-suffix P600 demonstrates that within-word gender violations recruit similar repair mechanisms as between-word violations. Moreover, the noun-suffix P600 was larger than the determiner–noun P600, likely due to either weaker constituent links resulting from longer distance, or longer distances reducing repairs due to a cost–benefit tradeoff. The findings from this study show that 1) within-word violations engage the same reanalysis processes as between-word violations, 2) within-phrase gender agreement processing is sensitive to linear distance, and 3) Norwegian exhibits similar ERP correlates of gender processing as those found for other languages.

1. Introduction

1.1. Background

Reading a sentence with a violation, such as *the bus drive fast* is believed to trigger cognitive processes of reanalysis and repair of the agreement error between the constituents "bus" and "drive", reflected in neurophysiological responses that can be detected using electroencephalography (EEG; see e.g., Friederici, 2011). It is further believed that the cognitive effort required to process sentences increases with the length and complexity of the sentence, and that higher processing effort elicits larger EEG responses (Barber and Carreiras, 2005; Friederici, 2011). Contrary to this, however, many studies manipulating the distance between the agreeing constituents have found that longer distances, which should add to the cognitive load, produced smaller neurophysiological responses. This finding has been linked to two theories. The first postulates that an increase in distance between two agreeing constituents reduces the link between them, and thus makes the error less salient (e.g., Hagoort, 2013). According to the second theory, reanalysis is governed by a cost-benefit tradeoff: if the cognitive cost of reanalyzing a sentence outweighs the benefit (for example, when the sentence is already interpretable without revision), then the parser will often forgo full reanalysis. As a result, sentences that require substantial effort to repair may not undergo complete reanalysis and

* Corresponding author at: Department of Linguistics and Scandinavian Studies, Postboks 1102, Blindern, 0317 Oslo, Norway.
E-mail addresses: sarah.cameron@iln.uio.no (S. Cameron), natalia.kartushina@iln.uio.no (N. Kartushina), bjorn.lundquist@uit.no (B. Lundquist), caffarra@stanford.edu (S. Caffarra).

https://doi.org/10.1016/j.brainres.2026.150390
Received 22 December 2025; Received in revised form 27 March 2026; Accepted 18 May 2026
Available online 20 May 2026
0006-8993/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

correction (Christianson et al., 2001; Ferreira et al., 2002). This idea of a cost-benefit calculation for sentence parsing aligns with the minimal revision principle of language processing, which advances that the sentence parser operates as a "least-effort device" (Fodor, 1998; Fodor and Inoue, 1998; Fodor and Inoue, 2000).

Several studies have examined how the distance between the agreeing constituents influences grammatical gender processing, but methodological and design heterogeneity have so far prevented a clear and coherent account of distance effects. The present study addresses gaps in previous research and examines the role of distance on processing of gender violations in Norwegian using EEG. We investigate this by comparing processing of gender violations occurring within the same word to violations across words – increasing the linear distance between the constituents, but within the same phrase to control for effects of structural distance (i.e., the number of phrases between two constituents). As Norwegian is an understudied language within the field of ERP research on agreement processing, an additional aim of this study is to provide a comprehensive account of gender processing in Norwegian which can be used as baseline for future studies in atypical populations and second-language learners of Norwegian.

1.1.1. *Brain responses to violations in agreement*

To study gender agreement processing, EEG studies typically employ violation paradigms, where an agrammatical sentence containing an agreement violation (e.g. *en gult bil* 'a.$_{MASC}$ yellow.$_{NEUT}$ car.$_{MASC}$') is compared to its grammatical counterpart (e.g.,. *en gul bil* 'a.$_{MASC}$ yellow.$_{MASC}$ car.$_{MASC}$'). Any EEG difference emerging from this comparison is linked to the processing of the specific morphosyntactic feature that has been manipulated (or to the processing of its violation). The dependency structures that are typically used in studies on gender processing are combinations of noun and adjective, as in the above example, and determiner-noun (e.g., *den gule bilen 'the.$_{MASC}$ yellow car.$_{MASC}$';* Molinaro et al., 2011). Two brain responses are commonly associated with agreement violations: the P600 and the left anterior negativity (LAN).

The P600 effect (Osterhout and Mobley, 1995) is characterized by a more positive deflection for ungrammatical compared to grammatical sentences in centro-posterior electrode sites starting around 500–600 ms post stimulus and lasting several hundred milliseconds (Friederici, 2002; Swaab et al., 2011). According to the three-phase model of sentence processing (Friederici, 2011, 2012), this response reflects difficulties with integrating syntactic information during the interpretation phase of sentence processing, and the resulting reanalysis. The P600 effect has been elicited with a variety of syntactic and morphosyntactic violations, such as phrase structure anomalies, gender agreement violations, verb tense violations, case marking violations, and number agreement violations (see Swaab et al., 2011 for a more extensive overview). However, the diversity of manipulations which elicit this response, including structure violations in music and mathematical rules, indicates that the P600 does not exclusively reflect linguistic processing, but rather is an index of more general structure building processes or conflict monitoring (Swaab et al., 2011; van de Meerendonk et al., 2010).

LAN effects correspond to a greater negative amplitude for ungrammatical sentences compared to grammatical controls and typically occur in left frontal electrode sites around 350–500 ms after stimulus onset, following the initial syntactic structure building. The LAN is connected to rule-based processes such as assignment of agreement dependencies and thematic roles based on morphosyntactic cues, and is generally elicited by morphosyntactic violations (Friederici, 2011; Molinaro et al., 2011; Royle and Steinhauer, 2023). In addition to morphosyntactic processing, LAN responses have been linked to working memory load and the cost of holding something in working memory while working out dependencies (Fiebach et al., 2002; Kluender and Kutas, 1993; Münte et al., 1998; Vos et al., 2001).

While the P600 has been consistently found with gender agreement violations, the occurrence of the LAN effect is less reliable (Caffarra et al., 2019; Tanner and Van Hell, 2014). It has been found with gender agreement violations in many studies (e.g., Alemán Bañón and Rothman, 2016; Bambini et al., 2021; Barber et al., 2004; Barber and Carreiras, 2005; Caffarra et al., 2015; Caffarra and Barber, 2015; De Resende et al., 2019; Gunter et al., 2000; Martín-Loeches et al., 2006; Molinaro et al., 2008), yet failed to appear in several others (Alemán Bañón et al., 2012; Foucart and Frenck-Mestre, 2011, 2012; Hagoort, 2003; Nevins et al., 2007; Wicha et al., 2004).

It is still unclear what exactly elicits a LAN response, but it has been linked to aspects such as markedness effects (difference in default status between features, see e.g. Alemán Bañón and Rothman, 2016) and frequency effects (De Resende et al., 2019). According to Friederici (2011), the occurrence of a LAN response is contingent on the amount of morphosyntactic markings in a language and the importance of using the morphosyntactic cues to interpret the sentence. This can be related to a prediction account in which the LAN is seen as a response to failed syntactic predictions (Molinaro et al., 2011). For example, seeing a neuter determiner would trigger the expectations of a neuter noun, but when this expectation is not met (e.g., a masculine noun appears instead), a LAN response is elicited. Logically, the strength of the prediction should be modulated by the amount of morphosyntactic marking, as well as the importance of these markings for establishing syntactic relations between the constituents in the sentence. The overtness of the morphosyntactic markings has also been hypothesized to impact LAN occurrence, however, support for this is not evident, as several studies using nouns with overt cues have failed to detect any frontal negativities (e.g., Alemán Bañón et al., 2012), while others have found a LAN response to gender agreement violations regardless of the transparency of the cues (Caffarra et al., 2015; Caffarra and Barber, 2015; De Resende et al., 2019; Zeller et al., 2022).

Given the large variability in morphosyntactic markings across languages, difference in their realisation, strength of markedness and frequencies, the conditions under which a LAN is elicited have thus been difficult to establish with any certainty. To complicate things further, the inconsistency persists even in studies on the same language, using similar violation types (e.g., compare predicative adjective violations in Alemán Bañón et al., 2012, which did not elicit a LAN, to those in Barber and Carreiras, 2005, which did). In addition to cross-linguistic variation, there is considerable variability across individuals that might contribute to the inconsistency in LAN appearances (Beatty-Martínez et al., 2021; Tanner and Van Hell, 2014). It has even been suggested that the LAN is not a separate response, but rather an incidental effect caused by grand averaging together individual biphasic N400-P600 responses varying in their negative-positive dominance (Osterhout et al., 2004; Tanner and Van Hell, 2014; but see Caffarra et al., 2019 and Fromont et al., 2020, for arguments against this view.). Finally, the variability in the LAN response has also been connected to methodological choices such as stimuli presentation (auditory versus written) and choice of reference electrode (Molinaro et al., 2011). Despite the uncertainties surrounding the emergence of LAN, a biphasic LAN-P600 response, reflecting the detection of morphosyntactic violations and the subsequent reanalysis and repair of these, appears to be the most common response pattern for gender agreement violations in sentence processing (Molinaro et al., 2011; Royle and Steinhauer, 2023).

In addition to LAN and P600, a second early negativity known as the N400 is occasionally found in studies on morphosyntax, although the N400 is most commonly elicited by semantic anomalies and is believed to index morpho-lexical and lexico-semantic processes (see e.g., Royle and Steinhauer, 2023). Like the LAN, the N400 occurs around 350–500 ms, but with a broader distribution across central and centro-parietal sites (Friederici, 2011; Kutas and Federmeier, 2011; Molinaro et al., 2011). The temporal overlap between the LAN and the N400 suggest that they might index processing mechanisms happening in the same time window, but with at least partially different neural sources, and for separate linguistic properties – that is, morphosyntax and lexical processing, respectively (Barber and Carreiras, 2005; Molinaro et al., 2011).

1.1.2. Effects of structural and linear distance on agreement processing

Most research on the effects of distance (between the constituents) on sentence processing reports that larger distances lead to a reduced P600 amplitude. However, it remains unclear what drives the effect, as linear and structural distances have been largely confounded in most previous research. Linear distance refers to the number of words between the agreeing constituents (Hammer et al., 2008; O'Rourke and Van Petten, 2011) – for example *The **friends have** dinner together* versus *The **friends** who have known each other since they were kids **have** dinner together*. Structural distance refers to the number of phrases between the agreeing constituents – compare *la casa muy vieja* 'the house very old (lit.)', where the agreeing constituents form part one single noun phrase, and *la casa es muy vieja* 'the house is very old', where the agreeing constituents are spread across a noun phrase and a verb phrase (Alemán Bañón et al., 2012). Disentangling the contributions of linear and structural distance therefore remains an important goal for understanding the mechanisms underlying agreement processing.

The idea that agreement processing is more demanding when the dependencies are farther apart is grounded in the dependency locality theory (Gibson, 1998, 2000). This theory posits that the distance between two constituents directly affects the cost of processing, due to both the added storage costs of keeping the properties of the first dependency constituent in memory while anticipating the second, as well as a higher cost of integration – that is, fitting the current input into the existing structure and matching across dependencies and other syntactic relations. Both electrophysiological research and fMRI studies have found support for this theory: longer distance between the dependencies does indeed make processing more demanding, leading to an increase in brain activity (e.g., Barber and Carreiras, 2005; for an overview of the fMRI literature, see Friederici, 2011).

However, several other studies manipulating the distance between the agreeing constituents have reported smaller ERP responses to increased distance and complexity (e.g., Alemán Bañón et al., 2012; Foucart and Frenck-Mestre, 2012; Hammer et al., 2008; Kolk et al., 2003; O'Rourke and Van Petten, 2011). Two accounts have been suggested to explain these findings. Firstly, a decreased sensitivity account proposed in Hagoort's unification model (Hagoort, 2013). It posits that a longer distance between the dependencies leads to a decrease in sensitivity to violations, possibly as a result of a gradual decay of the links between the elements of a sentence as it unfolds. This would imply that longer distance between the dependencies would decrease the salience of the violation. Secondly, an effort-repair trade-off account (Christianson et al., 2001; Ferreira et al., 2002; Fodor and Inoue, 1998; Fodor and Inoue, 2000) posits that the increase in processing effort makes us less inclined to perform repairs. This account implies that problems repaired by the parser are repaired only to the extent that it is necessary for conflict resolution, and if the effort required to reanalyze and repair the ungrammatical structure outweighs the advantages of making that repair for sentence comprehension, the repair may be omitted.

Relatively few studies have investigated the effects of distance on gender processing. A finding which seems to be consistent across studies, though, is that the LAN response appears not to be affected by distance, neither linear or structural,as LAN responses, where present, have not varied systematically with distance in either studies reporting a LAN (e.g. Barber and Carreiras, 2005; O'Rourke and Van Petten, 2011) or those that did not (Alemán Bañón et al., 2012). However, results for P600 responses are less clear. Investigating linear distance effects, Hammer and colleagues (2008) tested across-phrase noun-pronoun gender violations and found a P600 for dependencies with a short linear distance (separated by 3 words), but no P600 for long-distance dependencies (separated by 7 words). On the other hand, Kaan (2002) found no P600 difference for subject-verb number agreement violations of shorter (2 words) or longer (5 words) linear distance – possibly because the establishment of subject-verb agreement is more critical for sentence interpretation. Further, O'Rourke and Van Petten (2011) compared Spanish dependencies across several distances: determiner + noun (within-phrase), noun + adjective (within-phrase), noun + predicative adjective (across-phrase), and noun + relative clause + predicative adjective. Crucially, they observed a reduced P600 for within-phrase adjective violations relative to determiner violations, despite comparable structural and linear distances, suggesting that differences in syntactic category (e.g., determiner vs. adjective as the agreeing element) can modulate repair-related processing independently of distance. A syntactic category difference has also been reported for French, though with the opposite pattern: Royle et al. (2013) found larger P600 amplitudes for adjective violations than for determiner violations, despite no difference in linear or structural distance between the two conditions. The reversal relative to Spanish may reflect cross-linguistic differences in the transparency and regularity of adjectival gender marking, which could affect the morphological salience of each violation type. Interestingly, O'Rourke and Van Petten did not find a difference between the within-phrase adjective condition and the across-phrase adjective condition, possibly because they did not employ a grammaticality judgement task, the use of which has been found to increase sensitivity to grammatical errors and thereby increasing the P600 response (e.g., Gascon et al., 2011; Gunter et al., 1997; Osterhout and Mobley, 1995). Finally, for the long-distance across-phrase condition, they did not find a P600 at all, in line with the idea that longer distance decreases sensitivity to error detection.

However, in O'Rourke and Van Petten's materials, the structures with greater linear distance also spread across phrases, meaning that structural distance – a factor likewise known to modulate P600 magnitude – was confounded with linear distance. Alemán Bañón and colleagues (2012) found a decrease in P600 amplitude for across-phrase adjective agreement as compared to within-phrase adjective agreement, when the linear distance between the constituents was short (1 word). This could indicate that structural distance on its own appears to impact processing, and that even the shortest possible manipulation (dependencies in adjacent phrases) appears to bring forth this effect (but note that the authors also suggest that the difference could stem from different levels of predictability between conditions). However, the effect of structural distance did not interact with grammaticality, meaning that distance appears to affect P600 amplitude regardless of whether the agreement is ungrammatical or not. This means that the differences that have been found in response amplitude to linear distance could have interacted with an initially higher cognitive load due to long structural distance. Thus, we cannot know based on the existing literature if linear distance alone has a similar impact on processing of gender agreement.

Another unsolved question is whether we can observe changes to the P600 as a result of even shorter linear distance. As noted above, an effect of structural distance was found even when the dependencies were found in adjacent phrases (and only separated by one word). Given that there is an independent effect of linear distance, can this be observed even when the linear distance difference is minimal? The present study aims to address these questions using within-phrase manipulations of linear distance. Norwegian presents an opportunity to investigate this, as definiteness is marked directly on the noun by a suffix which also conveys information about the gender of the noun, for example, bil**en** 'car-**the.**$_{MASC}$' versus hus**et** 'house-**the.**$_{NEUT}$' (see section 1.2 for a fuller description of the Norwegian gender system). This feature makes it possible to create gender violations within the same word, which can be used to examine the effects of linear distance while controlling for structural distance.

1.1.3. Findings from within-word agreement processing

Given the suggestion above that P600 amplitudes decrease with increasing distance – consistent with an effort-reduction strategy – testing a smaller distance change (from across-word to within-word) would show how sensitive the parser's cost-benefit calculations are to such fine-grained differences in distance. Relatively few studies have investigated agreement processing where the two agreeing constituents

are part of the same word, and a relevant question in this context is whether processing of within-word grammatical violations is similar to processing of ungrammaticalities across words, or whether the adhesive relationship between word and suffix might lead to a stronger reliance on lexical processing – that is, recalling information stored in the mental lexicon rather than using combinatorial rule-based processing.

Several previous studies have shown that within-word violations involving verb formation violations (Penke et al., 1997; Rodriguez-Fornells et al., 2001) or number agreement violations (Ciaccio et al., 2023; Lück et al., 2006; Weyerts et al., 1997), elicit neural responses similar to those elicited by across-word violations, indicating that they undergo the same processing. For example, studies on German nouns (Ciaccio et al., 2023; Lück et al., 2006; Weyerts et al., 1997), German verb particles (Penke et al., 1997) and Catalan verb participles (Rodriguez-Fornells et al., 2001) all found a LAN response for irregular stems with regular endings, and either an N400 (Ciaccio et al., 2023; Weyerts et al., 1997) or no response (Penke et al., 1997; Rodriguez-Fornells et al., 2001) to regular stems with irregular endings. These results are generally taken as evidence in favor of dual-route models of language processing, which suggest that regular forms undergo a combinatorial (rule-based) processing pattern, resulting in the typical ERP signatures of morphosyntactic processing (LAN + P600). Processing of irregular forms, on the other hand, relies on memory-based processes, leading to higher demand in lexical processing which is realized as an N400 (see e.g. Caffarra, 2023 for a more detailed account). P600s were found for both irregular noun-regular ending and regular noun-irregular ending in several studies (Ciaccio et al., 2023; Lück et al., 2006; Rodriguez-Fornells et al., 2001).[1] It is worth noting that irregular verb forms have also been found to elicit N400/LAN combinations, and that individual factors impact processing of regular verbs: Steinhauer and Ullman (2002) reported variability in ERP responses to regular morphological violations, with a LAN-P600 combination observed in some participants and an N400-P600 pattern in others. This underlines the complexity of language processing and suggests that neural responses to morphological violations are not just driven by linguistic factors, but also individual differences with regards to processing strategies. Overall, these findings suggest that within-word grammatical violations are treated similarly to across-word violations, resulting in a P600, often preceded by a LAN – or, alternatively, an N400 when a more lexical processing route is triggered.

To our knowledge, no studies have so far investigated within-word gender agreement processing. However, since most studies comparing gender and number processing across words have found that these two features are processed similarly (e.g., Alemán Bañón et al., 2012; Nevins et al., 2007), it seems reasonable to expect a similar response pattern for regular within-word gender violations, that is, a P600 response, possibly preceded by a LAN. As noted above, this response pattern indicates that regular within-word violations rely on rule-based or combinatorial processing rather than lexical processing. As will be discussed in section 4.3.1, the findings of Steinhauer and Ullman (2002) also provide a post-hoc interpretive framework for components not anticipated in our original predictions.

1.2. Present study

The present study has three goals. Firstly, to examine the effects of within-word processing of gender agreement violations. Secondly, to assess the effects of linear distance on gender agreement processing in short-distance dependencies using within-word (noun + suffix) and across-word (determiner + noun), while controlling for the structural distance and syntactic category variation not accounted for in previous research. Finally, this study aims to provide a comprehensive account of gender processing in Norwegian, establishing a baseline for further studies in L1 and L2-Norwegian processing.

Norwegian is an understudied language within the fields of ERP studies, and particularly with regards to morphosyntactic features. At the time of writing, two studies have examined number and gender agreement violations in this paradigm (Kubota et al., 2024; Sandstedt et al., 2025). Their results showed a P600 for gender agreement violations, lining up with findings from studies in other languages. However, their reported analyses were performed only in the late time window, i.e., 500–900 ms (Kubota et al., 2024) or 550–1150 ms (Sandstedt et al., 2025) post onset of target word, leaving a gap in the literature as to whether early negative effects found for Norwegian agreement violations.[2] Thus, there are, to our knowledge, no studies describing the full ERP response pattern in Norwegian. Providing a native-speaker baseline for future studies would pave the way for further exploration of gender processing in Norwegian, not only within the L1 context, but also for ERP studies on L2 learners of Norwegian.

For the purpose of this study, we will use three features which are marked for gender in Norwegian. Table 1 provides a simple overview of the grammatical gender system in Norwegian. Most Norwegian dialects traditionally used a three-gender system (masculine, feminine, neuter), but this is gradually being replaced by a two-gender system consisting of common gender and neuter (Busterud et al., 2019; Rodina and Westergaard, 2015, 2021).

Two features makes Norwegian ideal for investigating the effects of linear distance on within-phrase dependencies: the fact that definite articles are realized as suffixes attached to the nouns (e.g. *bil* 'car' – ***bil*en** 'car-**the**.$_{MASC}$'), and the double definiteness (**den** gule bil**en** '**the**.$_{MASC}$ yellow car-**the**.$_{MASC}$). The definite suffixes are the remains of an old postposed definiteness clitic (see Faarlund, 2009 for a detailed history). The determiner *den* in the double definiteness construction is required when an adjective precedes a noun marked for definiteness but conveys no additional syntactic information. Furthermore, both the determiner and the suffix convey information about the gender of the noun. This makes it possible to create gender violations within one single word (e.g. ****bilet*** 'car.$_{MASC}$-**the**.$_{NEUT}$') as well as across words, but within the same phrase (e.g., ****det*** *gule bilen* '**the**.$_{NEUT}$ yellow car-the.$_{MASC}$) by manipulating features which convey the same morphosyntactic information. This way, we eliminate confounds such as the ones O'Rourke and Van Petten (2011) encountered, where both syntactic category differences and structural distance may have influenced their results, and can look directly at the influence of linear distance on within-phrase processing.

To investigate linear distance effects, we compared processing of

Table 1
Overview of the Norwegian three-gender system.

Feature	Masculine	Feminine	Neuter
Indefinite article	**En** bil	**Ei** bok	**Et** hus
	A car	*A book*	*A house*
Definite suffix	Bil**en**	Bok**a**	Hus**et**
	Car-the	*Book-the*	*House-the*
Definite determiner	**Den** bilen	**Den** boka	**Det** huset
	That car-the	*That book-the*	*That house-the*
Adjective	En **stor** bil	Ei **stor** bok	Et **stort** hus
	A big car	*A big book*	*A big house*
Possessive	**Min** bil/bilen	**Mi** bok/boka **mi**	**Mitt** hus/huset
	min		**mitt**
	My car	*My book*	*My house*

[1] Note that the lack of a P600 in two studies (Penke et al., 1997; Weyerts et al., 1997) is likely related to the study design – see Rodriguez-Fornells and colleagues (2001) for a discussion on this.

[2] Note that the authors indicated in personal communication (J. Sandstedt, November 28, 2025) that any early negativities visible in the grand averages were negligible and therefore not subjected to formal analysis.

noun-suffix gender violations to that of determiner-noun gender violations. Additionally, we include a third condition with violation between the noun and a predicative adjective (e.g., *Bilen er gult** 'the car.MASC is yellow.NEUT). Due to linguistic constraints of Norwegian, it is not possible to test gender agreement across phrases while keeping the syntactic category of the target word constant.[3] However, given the paucity of experimental reports in this topic in Norwegian, we want to provide a nuanced overview of Norwegian agreement processing which may also serve as a baseline for future studies (e.g. in atypical L1 and typical L2-Norwegian acquisition). Thus, we decided to include this construction in our design, recognizing that any observed differences between determiner + noun and noun + predicative adjective processing may stem from either structural distance or the category of the target word. An overview of the three conditions with examples can be found in Table 2.The present study addresses the following three research question:

RQ 1: Do within-word gender violations elicit a (LAN +)P600 response pattern, indicating that they undergo the same morphosyntactic processing as across-word gender violations?

Considering that within-word processing of gender agreement has not yet been explored, our first research question concerns whether processing of gender violations between noun and suffix will elicit the same responses as have been found for other within-word manipulations and for gender violations across words. In the above section on within-word processing, it was established that the relatively few studies that have looked at violations within the same word have found regular number violations between noun and suffix to elicit a P600 preceded by a LAN response. This is in line with the general patterns of agreement processing across words, indicating that regular within-word morphosyntactic dependencies undergo the same processing as those occurring across separate words. Accordingly, we expect the within-word suffix violations in the present study to elicit a P600 response, indicating a process of reanalysis and repair. LAN responses consistently appear in other within-word manipulation studies and might therefore be expected in the present study, however, because the precise elicitors of the LAN remain unclear, this prediction is weaker.

Table 2
Stimuli examples.

Condition	Agreement	Violation
Suffix	Jon setter seg på **benken** i hagen *Jon sits down on bench.$_{MASC}$-the.$_{MASC}$ in garden-the*	Jon setter seg på ***benket** i hagen *Jon sits down on bench.$_{MASC}$-the.$_{NEUT}$ in garden-the*
Determiner	Kelneren på **den** dyre **kaféen** jobber hardt *The waiter in the.$_{MASC}$ expensive café-the.$_{MASC}$ works hard*	Kelneren på ***det** dyre **kaféen** jobber hardt *The waiter in the.NEUT expensive café-the.$_{MASC}$ works hard*
Adjective	Det hyggelige **bakeriet** er **travelt** om morgenen *The.$_{NEUT}$ nice bakery-the.$_{NEUT}$ is busy.$_{NEUT}$ in the morning*	Det hyggelige **bakeriet** er ***travel** om morgenen *The.$_{NEUT}$ nice bakery.$_{NEUT}$ is busy.$_{MASC}$ in the morning*
Fillers	Fuglen mister **et frø** på bakken *The bird drops a.INDEF seed. INDEF on the ground*	Fuglen mister ***et frøet** på bakken *The bird drops a.INDEF seed.DEF on the ground*

[3] While it is theoretically possible (e.g. using constructions such as a) *På loftet står **et skap fullt** av klær*; lit. 'In the attic there's a.$_{NEUT}$ cupboard.$_{NEUT}$ full.$_{NEUT}$ of clothes' vs. b) *I fjor var **loftet fullt** av rot*; lit. 'Last year was attic-the.$_{NEUT}$ full.$_{NEUT}$ of clutter'), there are relatively few adjectives that work for sentences such as a), meaning that there would likely be a lot of repetition in the stimuli. Further, it is unclear how we would process sentences like b), as the linear distance is affected by verb movement which creates a different structure for the across-phrase agreement. Future studies should look into the effects of verb movement on processing.

RQ 2: Does linear distance affect processing of gender violations when controlling for structural distance?

Our second research question is, given that the suffix violations follow the pattern outlined in RQ1, whether we can see an effect of linear distance in violations occurring within the same phrase when we compare constituents that are syntactically similar.

Repairs of violations on the suffix should be less effortful than for determiner-noun pairings. Firstly, because we would assume them to have a reduced memory cost compared to agreement violations across words, as there is no need to keep information related to gender agreement in memory. Secondly, as both constituents are immediately present, suffix violations should also have a reduced integration cost in terms of agreement processing as compared to across-word agreement. This means that on the one hand, if reduction of effort still to some extent outweighs the benefit of repair with determiner-noun pairings, we would expect P600 magnitudes for noun-suffix violations to be even larger. On the other hand, repair benefits may already be at ceiling for determiner-noun pairings, in which case the comparatively lower effort involved in processing suffix violations would not be expected to reduce the P600 amplitude any further. Alternatively, if the reduction of the P600 with distance is the result of a weakening of the links between the constituents, as proposed by Hagoort's unification model (Hagoort, 2013), we should see a larger P600 for violations on the suffix as compared to determiner violations.

RQ 3: Do determiner-noun and noun-adjective gender agreement violations in Norwegian elicit (LAN +) P600 responses?

The purpose of the determiner-noun and noun-adjective conditions in isolation is to provide a baseline for Norwegian, and our research question connected to these two conditions is therefore whether we can replicate the previous findings in a Norwegian context. Based on previous research outlined earlier in the introduction, we expect both determiner-noun violations and adjective-noun violations to elicit P600 responses, possibly preceded by a LAN response.

2. Methods and analyses

2.1. Methods

2.1.1. Participants

41 native Norwegian speakers between the ages of 20–40 ($M = 25.4$, $SD = 4.3$) participated in the study, all right-handed and with normal or corrected-to-normal vision and no neurological diagnoses or language disorders. Three participants were excluded due to technical problems during data collection, and two additional ones were excluded because they turned out not to fulfil the requirements (one did not learn Norwegian from birth, the other had an unreported neurological diagnosis). The remaining 36 participants fulfilled all criteria and had a sufficient amount of useable data (<25% of trials rejected); thus, they were all included in the analysis.

2.1.2. EEG stimuli

For this study, three types of sentences were constructed varying in their type of gender violation: noun-suffix, determiner + noun, and noun + adjective. Only masculine and neuter nouns were used in order to control for the variation in the use of the feminine gender (see section on the Norwegian gender system above). Examples of the three conditions can be found in Table 2. The stimuli consisted of a total of 320 sentences, with 80 sentences per condition (40 congruent and 40 incongruent) and 80 filler sentences. All sentences were 6–8 words long, and the target words (nouns for suffix and determiner sentences, adjectives for adjective sentences) were always followed by two or three words to avoid wrap-up effects. Target words (underscored in Table 2) varied in length between 3 and 10 letters. For each condition, critical items were evenly distributed between masculine and neuter (i.e., 20 grammatical sentences had masculine target words, 20 had neuter target words, and the same for ungrammatical sentences). All target words

were matched on length and frequency both within (congruent vs. incongruent) and across the three sentence types. Frequencies for all words were retrieved from Norwegian Web As Corpus (NoWaC, Guevara, 2010), a corpus of 700 million Norwegian words collected from. no-domain sites in 2009 and 2010. For adjectives, separate frequencies were retrieved for masculine and neuter word forms respectively, and for nouns all forms of the word (indefinite and definite forms in both singular and plural) were added up into a total lemma frequency per word. Average frequencies for nouns in the suffix condition were 84.0/mill, for nouns in the determiner condition 104.85/mill and for adjectives in the adjective condition 79.93.

The fillers consisted of nouns with definiteness violations between article and noun, where congruent sentences had an indefinite article and an indefinite noun, whereas the incongruent sentences had an indefinite article and a definite noun. The nouns used for fillers were the same as in the suffix condition, but switched in grammaticality status, so that grammatical items in suffix condition were ungrammatical in the filler condition, and vice versa.[4]

For counterbalancing, two stimuli lists were created, using the same sentences but with the opposite congruency status – in other words, congruent items in list 1 were incongruent in list 2, and vice versa. Half of the participants were shown list 1, and the other half were shown list 2. Participants only saw one instance of each sentence.

The **suffix condition** consisted of 80 sentences with grammatical and ungrammatical noun-suffix constructions appearing as the fourth or fifth word. Ungrammatical items consisted of either a masculine noun with a neuter suffix (e.g., ***benket*** 'bench.$_{MASC}$ –the.$_{NEUT}$'), or neuter nouns with a masculine suffix (e.g., ***husen*** 'house.$_{NEUT}$-the.$_{MASC}$'). To minimize the risk of the target words in the suffix violation being interpreted as nonce words, they were embedded in high-constraint contexts to increase their predictability. Predictability ratings were collected for the sentences with the target words and 80 sentences with control words in low-constraint contexts from a separate group of 38 native speakers of Norwegian. The ratings were performed using a scale from 1 to 5, where 1 was not at all predictable and 5 was very predictable. The results of t-tests showed that the target words in the suffix sentences were significantly more predictable ($M = 4$, $SD = 1$) than the low-constraint control words ($M = 2$, $SD = 1$, $p < 0.01$).

The **determiner condition** contained 80 grammatical and ungrammatical determiner-noun pairings, which appeared as the third and fifth word in the sentence with an intervening adjective and were followed by two or three other words. Finally, the **adjective condition** consisted of congruent and incongruent adjective-noun pairings appearing as the third and fifth word, separated by the copula verb *er* ('is'). The adjectives used in this condition were also repeated in the determiner condition (where they are not critical items) and were counterbalanced so that they occurred once in congruent constructions in one sentence type, and once in incongruent in the other.

2.1.3. Procedure

Before coming to the lab, the participants were sent a link to a questionnaire mapping their language background, education level and whether they had history of neurological or language related diseases. All data collection was performed in the Socio-Cognitive Laboratory at the Department of Linguistic and Scandinavian studies at the University of Oslo. Once in the lab, participants filled in the consent form and took part in the EEG task. They were seated in a comfortable chair in front of a computer screen. The stimuli were presented in PsychoPy, version 2020.1.3 (Peirce et al., 2019) using a DELL U2412M screen with a resolution of 1920x1200 and a 59hz refresh rate at about 80 cm distance. Participants were shown how eyeblinks and other muscular artifacts affect the signal and were asked to try not to blink excessively during stimuli presentation. The EEG stimuli were presented in rapid serial visual presentation (RSVP), starting with a fixation cross displayed for 700 ms, followed by a 250 ms pause, and then each word displayed for 500 ms with a between-word interval of 250 ms.

Grammaticality judgement task: Each sentence was followed by a prompt to rate it as grammatically correct or incorrect by using the z and m keys on a standard keyboard, counterbalancing across participants which key indicated grammatical and which indicated ungrammatical. After every 15th sentence, the participants had the option to take a break for as long as they needed.

EEG signals were recorded using BrainVision Recorder, version 1.24.0101, with an ActiChAmp amplifier with a sampling rate of 500 Hz. Participants were fitted with a Brain Products ActiCap, with 33 active Ag/AgCl electrodes (Ground, FP1, FP2, F7, F8, F3, Fz, F4, FT9, FT10, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, TP9, TP10, P7, P3, Pz, P4, P8, O1, Oz, O2). Impedances were kept below 5kΩ. EEG signals were referenced online to Fz, and later re-referenced to average mastoids (TP9 + TP10). The EEG recording lasted for about 1 h, and the full session took about 2 h. Participants were compensated with a 150NOK (about 12€) gift card.

2.2. Analyses

2.2.1. EEG data preprocessing

The EEG data was preprocessed using EEGlab version 2024.2 (Delorme and Makeig, 2004) and ERPlab version 12 (Lopez-Calderon and Luck, 2014) in MATLAB version R2024b (The MathWorks Inc., 2024). The pipeline was an adaption of the one proposed in Luck (2022). We used bandpass filters of 0.1 and 30 Hz, with a 12 dB/oct roll-off. Eyeblinks with a probability of 0.9 or higher were removed using ICA and the data was segmented into epochs from −200 ms pre to 1000 ms post onset of target word with a 200 ms pre-stimulus baseline correction. Artifact detection was done using both a simple voltage threshold of −150/150 μV as well as a moving window peak-to-peak threshold of 150 μV. Epochs with artifacts exceeding the thresholds were excluded from further analyses (on average 2.1 trials, or 0.7%), as were epochs where the participant had given an incorrect response in the grammaticality judgement (on average 8.5 trials, or 2,7%). For averages per condition, refer to Table 3.

2.2.2. EEG data analysis

The EEG data were analyzed using a non-parametric cluster-based permutation test (Maris and Oostenveld, 2007). Cluster-based permutation testing is well-suited to this type of data, as the need for a priori choices regarding latency, electrode selection and polarity is minimized (Maris and Oostenveld, 2007; Mickan and Lemhöfer, 2020). This, combined with multiple comparison corrections, facilitates a data-driven analysis that is particularly relevant in this study given the

Table 3
Accuracy means (%) and standard deviations per condition and grammaticality status.

Condition	Grammaticality	*M (%)*	*SD (%)*
Suffix	Grammatical	97.78	3.42
	Ungrammatical	97.71	2.34
Determiner	Grammatical	97.78	3.37
	Ungrammatical	98.40	2.88
Adjective	Grammatical	96.88	3.50
	Ungrammatical	93.40	6.79
Filler	Grammatical	97.99	2.52
	Ungrammatical	98.47	2.26

[4] An observant reviewer pointed out that using the same nouns in both the suffix and filler conditions with flipped grammaticality status might potentially lead participants to develop strategies for detecting ungrammaticalities. Given the large number of trials (320) and the fully randomized trial order, which ensures considerable variation in the distance between related items, we consider it unlikely that participants would notice or strategically exploit the repeated nouns. We nonetheless acknowledge this as a limitation of the design.

uncertainty surrounding the findings from the suffix condition (i.e., lack of previous research in general, and in Norwegian in particular) and lack of agreement on time window boundaries in the ERP literature. Analyses were performed in FieldTrip (Oostenveld et al., 2011) version 20240123 with MATLAB version R2024b (The MathWorks Inc., 2024). Tests were run over all electrode sites and across a time window of 0–1000 ms. Electrode neighbourhoods were defined by triangulation of the electrode layout using *ft_prepare_neighbours*. Clusters were formed from contiguous electrodes exceeding an individual sample threshold of $p < 0.05$ (two-tailed). The cluster-level test statistic was based on dependent-samples *t*-tests, and significance was evaluated against a permutation distribution of 1000 randomizations. The alpha level for the cluster-level inference was set at 0.025 to correct for two-tailed testing. We first ran separate tests comparing grammatical and ungrammatical sentences for each violation type. To directly compare suffix and determiner violations, we then computed difference waves (ungrammatical minus grammatical) for each condition and tested for the differences using the same procedure.

3. Results

3.1. Behavioural results

3.1.1. Grammaticality judgement

Accuracy was uniformly high across conditions (see Table 3). To examine whether accuracy differed across conditions, a binomial mixed model was fitted with violation type (suffix, determiner, adjective) and grammaticality (grammatical, ungrammatical) as fixed effects and participant as a random intercept. Behavioural analyses were run in R version 4.3.1 (R Core team, 2023) using RStudio version 2025.05.1 (Posit team, 2025) and the glmer function from the lme4 package (Bates et al., 2015). The packages car (Fox et al., 2026) and emmeans (Lenth et al., 2025) were used for post-hoc analyses. Model diagnostics confirmed adequate fit with no evidence of overdispersion (ratio = 0.92) or significant deviation from expected residual distributions (KS test: $p = 0.521$, dispersion test: $p = 0.734$, outlier test: $p = 1.00$). The model revealed a significant violation type × grammaticality interaction ($\chi^2(2) = 14.25$, $p < 0.001$), while the main effects of violation type and grammaticality were non-significant (both $ps > 0.18$). Estimated marginal means indicated that for grammatical trials, accuracy did not differ significantly between any of the three violation types (all $ps > 0.37$). For ungrammatical trials, accuracy for adjective violations was significantly lower than for both suffix (OR = 3.11, $p < 0.001$) and determiner violations (OR = 4.52, $p < 0.001$), while suffix and determiner did not differ from each other ($p = 0.505$). These results indicate that participants were less accurate at detecting adjective gender violations than violations in the other two conditions, while detection of suffix and determiner violations was comparable.

3.2. EEG results

3.2.1. Suffix violations

Grand average waveforms comparing sentences with and without suffix agreement violations can be seen in Fig. 1. Cluster-based permutation t-tests revealed a positive cluster ($p = 0.002$) from about 400 to 1000 ms,[5] across the whole scalp (involving all channels except FT9). Topographic plots (Fig. 1) show a positive deflection starting at right temporal sites and moving towards fronto-central, central and posterior areas. No negative clusters were found for suffix violations.

3.2.2. Determiner violations

Grand average waveforms and topographic plots for sentences with determiner violations are presented in Fig. 2. The cluster-based permutations *t*-test detected significant differences between the grammatical and ungrammatical sentences corresponding to a positive cluster ($p = 0.002$) from around 450 ms and throughout the first second after onset, and similarly to the suffix violations the cluster is widely distributed across all but three channels (Fp2, F8 and FT10). Topographic plots (Fig. 2) show that the positive deflection is centered over the midline at central and parietal sites and gradually expands over both hemispheres before shifting backwards toward parietal and occipital locations. Further, a negative cluster ($p = 0.03$) was found, starting at about 650 ms and continuing throughout the full length of the segment. The negative cluster includes electrodes Fp1, F3, F7, FT9, FC5, T7, T8, FT10, FC6, FC2, F4, F8, Fp2, and F2. Topographical plots of the cluster (Fig. 2) show a frontal negativity emerging in the left temporal region and spreading to cover all frontal and fronto-central sites.

3.2.3. Adjective violations

Cluster-based permutation analyses showed a significant difference ($p = 0.002$) between sentences with and without a noun-adjective violation, corresponding to a positive cluster with a duration from about 550 msec to about 1000 ms, and over all channels except Fp1, Fp2, FT9 and FT10. Topographic plots (Fig. 3) show a central and parietal positivity which gradually shifts to become more posterior. A negative cluster ($p = 0.04$) was also detected, starting at about 350 ms and ending at around 550 ms. The cluster comprises all channels except FT10. The topographical plots (Fig. 3) show a left anterior negativity which shifts toward central electrode sites before spreading bilaterally.

3.2.4. Comparing conditions

The results from the cluster-based analysis comparing the difference waves for suffix and determiner violations revealed significant differences between the suffix and determiner conditions (540–996 ms, $p = 0.002$), suggesting a greater P600 effect for suffix violations. Waveforms from parietal electrodes comparing all three conditions can be seen in Fig. 4.

4. Discussion

The aims of this study were three-fold; firstly, we examined processing of within-word gender violations between noun and suffix. Secondly, we investigated the effects of linear distance on within-phrase processing of gender violations by comparing noun-suffix to determiner-noun violations, while controlling for the structural distance. Thirdly, we provided a comprehensive picture of processing gender violations across different dependency structures in Norwegian, by including violations between noun and predicative adjective. The behavioural results show that the accuracy rates were lower for adjective violations than for suffix and determiner violations, a finding that is consistent with the ERP results which suggests that processing gender agreement in the adjective condition requires additional processing demands. Cluster-based permutation t-tests revealed posterior clusters for all three violation types, indicating a significant difference between grammatical and ungrammatical sentences corresponding to a P600 response. These findings suggest that noun-suffix violations go through the same processes of reanalysis as determiner-noun and noun-adjective violations. Further, comparisons between noun-suffix and determiner-noun showed that they differed in the P600 time window, with a larger P600 for noun-suffix violations. This is in line with the prediction that linear distance affects processing even at short distances, and that longer-distance dependencies result in smaller P600 responses. Additionally, negative clusters were found for determiner and adjective violations. In the following sections, the results will be discussed in detail and related to the existing literature.

[5] Following Sassenhagen & Draschkow (2019), who caution against reporting precise temporal boundaries for cluster-based permutation results, time windows are reported to the nearest 50–100 ms.

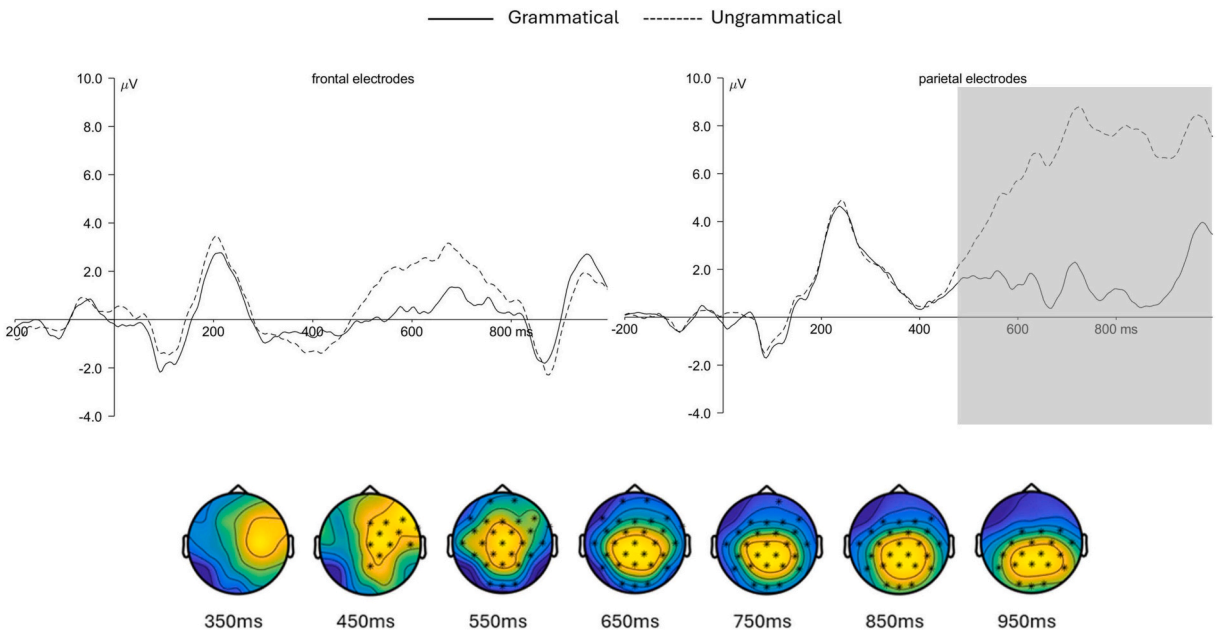


Fig. 1. Waveforms and topographical plots for suffix manipulations. Waveforms show average of F3, F7 and FC5 for frontal electrode sites, and average of P3, Pz and P4 for parietal sites. Positive is plotted up. Shaded area illustrates latency where cluster reached significance. The black asterisks in the topographical plots show electrode sites with significant difference between grammatical and ungrammatical sentences.

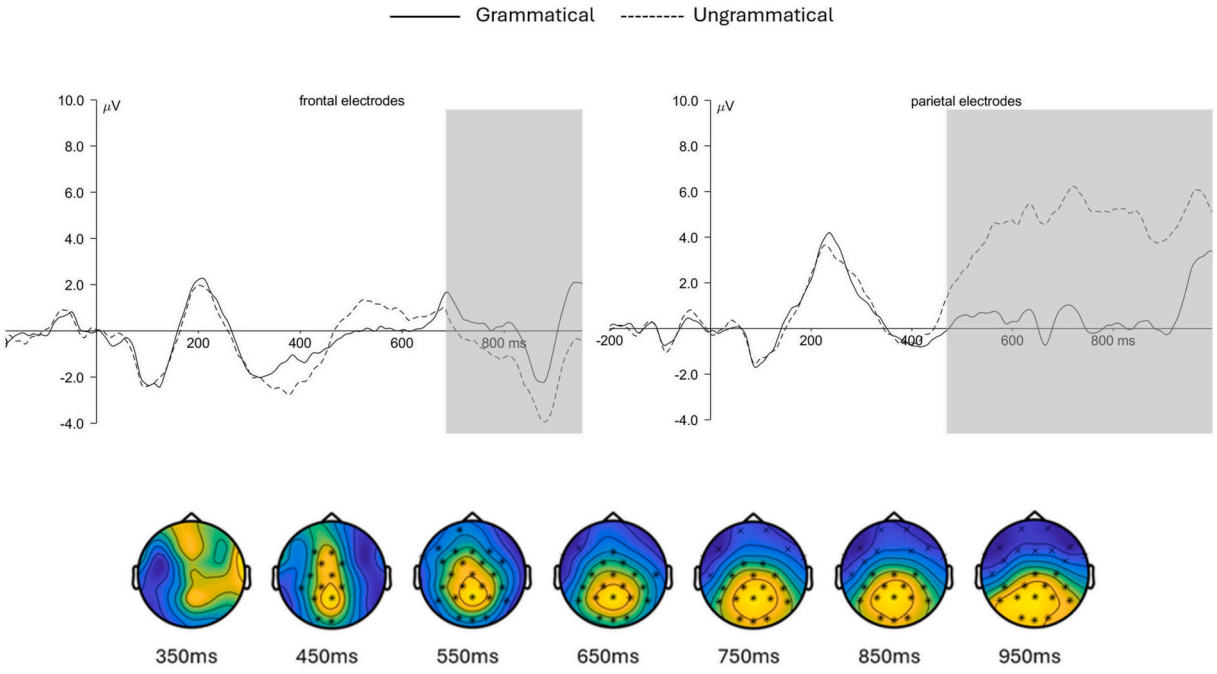


Fig. 2. Waveforms and topographical plots for the determiner condition. Waveforms show average of F3, F7 and FC5 for frontal electrode sites, and average of P3, Pz and P4 for parietal sites. Positive is plotted up. Shaded areas illustrate latencies where clusters reached significance. The black asterisks in the topographical plots show electrode sites with positive significant difference between grammatical and ungrammatical sentences, while the x-es indicate sites where the difference was negative.

4.1. Processing of within-word gender violations

In line with our predictions, the noun-suffix violations elicited a robust P600, starting out in the temporal areas of the right hemisphere before shifting towards central and parietal sites. This is in line with previous research investigating number violations between noun and suffix (Lück et al., 2006), and the P600, as well as the lack of an N400, suggests a morphosyntactic rather than a lexical processing of gender violations between noun and suffix, indicating that within-word violations undergo the same processes of reanalysis and repair as across-word violations.

Contrary to previous findings, no negativities were detected for within-word suffix gender violations in our study. Given the general unpredictability of the LAN response (see introduction), it is not straightforward to make assumptions as to exactly why we did not detect one for this condition. The two studies examining number violations

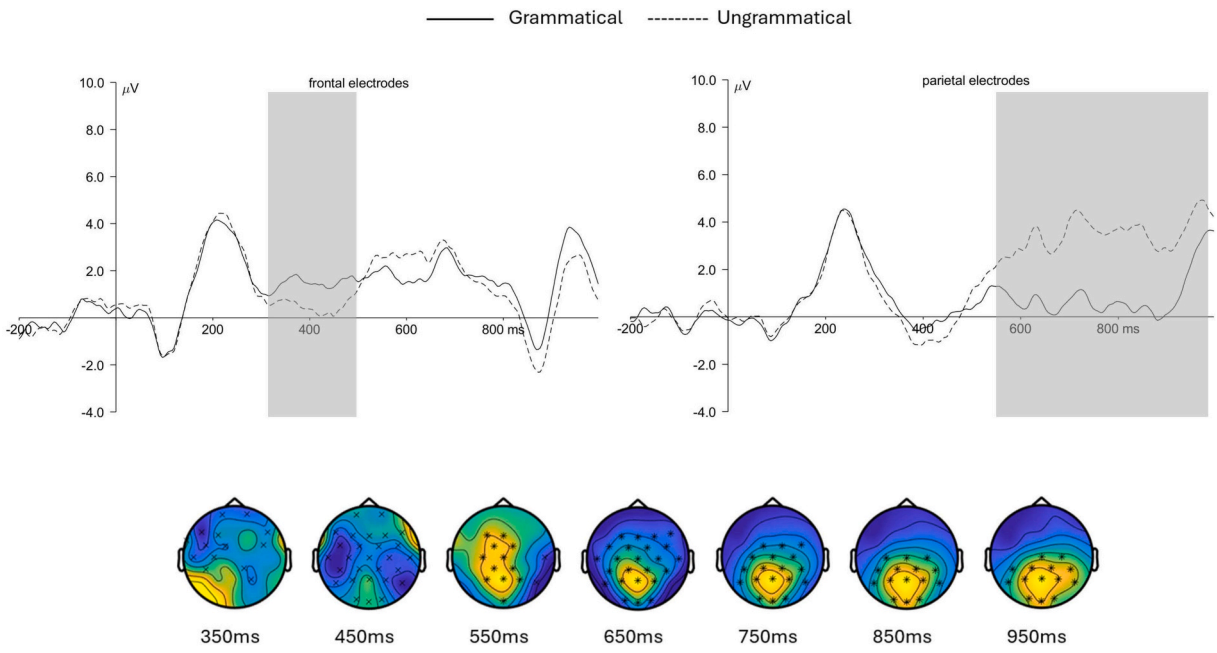


Fig. 3. Waveforms and topographical plots for the adjective condition. Waveforms show average of F3, F7 and FC5 for frontal electrode sites, and average of P3, Pz and P4 for parietal sites. Positive is plotted up. The shaded areas indicate latencies where clusters showed significant difference. The black dots in the topographical plots show electrode sites with positive significant difference between grammatical and ungrammatical sentences, while the x-es indicate sites where the difference was negative.

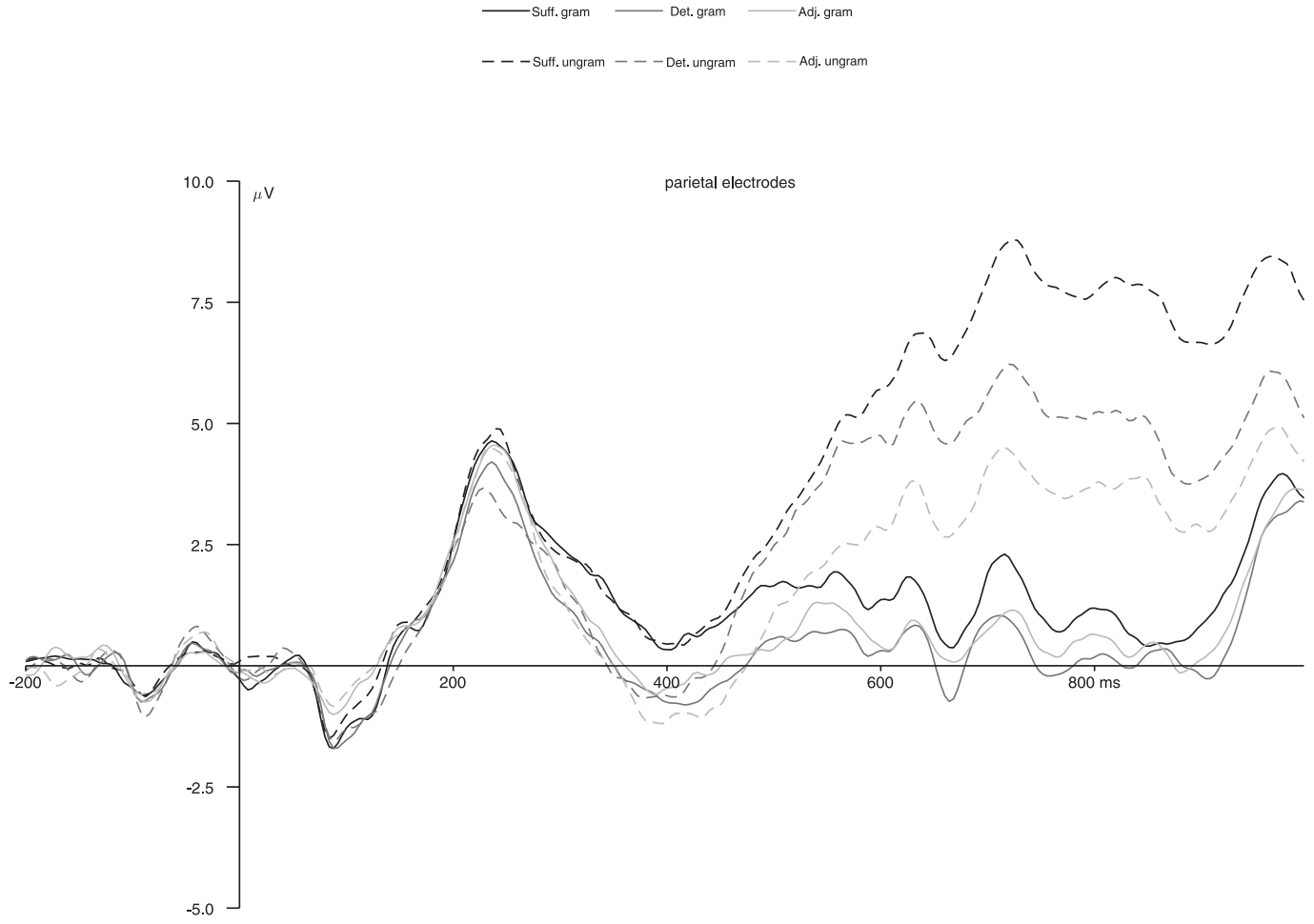


Fig. 4. Overlay of waveforms for all conditions. Waveforms show average of P3, Pz and P4. Positive is plotted up.

between noun and suffix in German (Lück et al., 2006; Weyerts et al., 1997) both elicited LAN responses despite notable methodological differences, including the sentential position of the violation, the choice of reference electrode(s), and the mode of stimuli presentation (written versus auditory). It therefore seems unlikely that such methodological dissimilarities can account for the absence of a negativity in the present study. A more plausible explanation is that LAN elicitation is sensitive to the linguistic feature under investigation (number versus gender), language-specific properties (German versus Norwegian), or individual factors known to modulate its occurrence, such as working memory capacity (Caffarra et al., 2019; Vos et al., 2001). Future research should keep working on assessing exactly what these factors are, and to what extent they interact in order to reliably elicit LAN responses.

4.2. Effects of linear distance

In order to detail how linear distance impacts processing of within-phrase violations, it is first necessary to describe the response found in determiner-noun violations. As with the suffix violations, and in line with our expectations and previous results, the gender agreement violations between determiner and noun elicited a clear P600 starting over the midline of central and parietal regions before spreading to both hemispheres over posterior sites. The P600 is accompanied by a late frontal negativity, which appears as widespread across both hemispheres. We will return to discuss the late negativity in section 4.1.3.

The pairwise comparisons between the suffix and determiner conditions showed that violations between noun and suffix elicited a higher P600 amplitude than violations between determiner and noun. Thus, the present results show processing differences between gendered definiteness markers which are similar in terms of the information they convey but differ in the placement and distance to the noun they are linked to, while controlling for the structural distance. This finding lends support to the idea that a longer linear distance leads to a reduced P600, even when comparing within-word to within-phrase violations. Our results demonstrate that if the distance effect is a result of a cost-benefit tradeoff, as proposed by e.g. Alemán Bañón and colleagues (2012), the cost reduction from processing between-word to within-word is sufficient to induce more processes of reanalysis and repair for suffix violations than for determiner violations, resulting in a larger P600 for the former. Alternatively, if the increase in distance weakens the links between the agreeing constituents (as proposed by Hagoort, 2013), making the violation less salient, this effect is measurable even at very short distances.

It appears likely that both cost-benefit and violation salience are at play in agreement violations across distance. Firstly, the weakened links of Hagoort's unification model (2013) appear to be a reasonable explanation for the P600 decline at shorter distances. Based on this account, the P600 magnitude would be seen as an index of violation salience, in the sense that it reflects the explicitness of the error rather than – or, perhaps more accurately, in addition to – the repair effort. This suggests that a mismatch between noun and suffix would be more salient and thus detected more readily than a violation separated by a longer string of words or phrases. A question which may tip the scales in favor of this approach is that if the parser is a *least-effort device* that follows a *minimal everything principle* (Fodor, 1998; Fodor and Inoue, 1998; Fodor and Inoue, 2000), can a cost/benefit account explain the continuum from short- to long-distance dependencies? Empirical results show a gradual decline in the P600: strong for within-word violations, smaller for across-word/within-phrase and across-phrase violations, and absent for long-distance cases such as across relative clauses (e.g., O'Rourke and Van Petten, 2011). If the parser can often accept sentences without repairing incongruities – and already gives up repairs at the across-word level under a cost/benefit tradeoff – why does it ever attempt repair at all, given its putative least-effort strategy?

However, violation salience does not provide a satisfactory explanation for the results found by O'Rourke and Van Petten (2011), where agreeing constituents separated by a relative clause did not elicit a P600 at all. Under Hagoort's account, the lack of a P600 would indicate that the links between the constituents are weakened to the extent that the violation was not detected, however, O'Rourke and Van Petten's results also showed that the distance did not affect the LAN response which was similar across all their conditions. This finding indicates that the violation was detected, but not subsequently repaired, indicating a "good enough" processing strategy (Christianson et al., 2001; Ferreira et al., 2002). Based on these arguments, it appears likely that the amplitude reduction seen with increasing linear distance in the current study is related to both a reduction in the strength of the links between the dependencies, and a cost/benefit trade-off – at least for longer-distance dependencies.

It should be noted that the order of the dependencies differed between the conditions in the current study. The noun was used as the target word for both conditions, but for the determiner condition the violating feature preceded the noun, whereas in the suffix condition, the violation was realized on the noun ending. This could potentially also have impacted the processing, although it is unknown to what extent. Previous findings have shown that prenominal and postnominal adjective placement in French impacts processing, so that noun + adjective elicited a larger P600 than adjective + noun (Foucart and Frenck-Mestre, 2012). However, this finding is likely related to postnominal adjective placement being more canonical than prenominal adjectives, and a similar canonicity is not likely to have impacted the present results. How the order of dependencies impacts processing is an interesting question which should be further explored in future studies.

4.3. P600, LAN, and late frontal Negativity: Interpreting Norwegian gender violations

An overarching goal of this study was to provide a comprehensive account of gender processing in Norwegian for use as a baseline in later studies. As determiner + noun and noun + adjective are among the most explored dependencies in other languages, typically leading to P600 responses, and, albeit less reliably, LAN responses, both structures were included in order to replicate these findings for Norwegian. Both conditions elicited strong P600 responses, indicative of reanalysis and repair processes. While the accuracy scores for suffix and determiner violations were almost at ceiling, the scores for adjective violations were slightly lower. This behavioural result matches previous findings (e.g., Kaan, 2002), and lends support to the idea that increased structural distance also increases processing difficulty.

In addition to the P600, the determiner condition elicited a late frontal negativity from around 660 ms. This frontal negativity was not part of our expectations, however, it is not a novel finding, as similar negativities have been found in several studies. For example, in studies on Italian (Caffarra et al., 2015) and Spanish (Caffarra and Barber, 2015), Caffarra and colleagues found that article-noun gender violations involving nouns with a transparent gender marking elicited a negative response from 350-750 ms over frontal medial electrode sites. Similar, but frontal negativities, have also been found in other studies (Alemán Bañón and Rothman, 2016; Blais et al., 2024; Lau et al., 2013; Otten and Van Berkum, 2008; Royle et al., 2021; Sabourin and Stowe, 2004, 2008; Zawiszewski et al., 2016).

There are two main ideas that have been hypothesized to be the cause of this late frontal negativity: an added working memory (WM) cost, or a phase reversal of the P600, that is, the negative end of the P600 dipoles. Anterior negativities have also been attributed to phonological mismatch processing (e.g., Royle et al., 2021), however, this account is unlikely to apply to the current study that used written stimuli.

There are (at least) three ways in which WM has been suggested to be the cause of a late frontal negativity. Sabourin and Stowe found late anterior negativities when participants were presented with subject-verb number agreement violations (2004) and with determiner-noun gender violations (2008). They proposed that the findings reflected

the need to maintain information in WM in order to resolve the ungrammaticality, with sustained frontal negativities reflecting the use of memory resources (Ruchkin et al., 1990; Honda et al., 1996). Relatedly, the negativity has been attributed to the added load of having to keep information in WM in order to perform grammaticality judgements (Alemán Bañón and Rothman, 2016). However, neither of these accounts provide any explanations as to why the negativity only appeared for the determiner condition in the present results. It would be reasonable to assume that the adjective condition would impose at least the same WM load as the determiner condition, and further, it seems unlikely that the grammaticality judgement task should invoke this response with one sentence type only. A more likely explanation for the negativity in the determiner condition only, might stem from a failure to fulfill a relatively strong prediction (as compared to the other two conditions), requiring WM costs to revise the association between the cue and the target word (Alemán Bañón and Rothman, 2016; Lau et al., 2013; Otten and Van Berkum, 2008). A phrase beginning with *den* + adjective will always be completed by a noun corresponding in gender to the determiner. The same predictability is not present for the suffix condition (where the noun and suffix appear together) or the adjective condition (*caféen er* 'the café is' could just as well be completed by e.g. a preposition phrase: *på hjørnet*, 'on the corner').

The suggestion that the negativity could reflect the phase reversal of the P600 relates to the topographical distribution of the responses, and entails that the orientation of the dipoles created by the postsynaptic potentials was such that positive end of the dipoles was captured by the posterior electrodes, while the negative end was captured by the frontal electrodes (Barber and Carreiras, 2005; Zawiszewski et al., 2016). It is important to note, however, that late frontal negativities have been found in studies where the phase-reversal account seems less plausible. For example, Alemán Bañón and Rothman (2016), found larger P600s for singular + plural errors than for plural + singular errors, from which it should follow that the frontal negativity for singular + plural errors would also be larger if caused by a phase reversal – however, this was not the case, which the authors took to indicate that the frontal negativity is separate from the P600. Further evidence against a phase reversal account comes from task-manipulation studies, showing that removing an explicit grammaticality judgment task eliminates or substantially reduces the P600 while the anterior negativity persists (Royle et al., 2013). The dissociation between the two components under these conditions suggests that the anterior negativity reflects a functionally distinct process rather than being an artifact of the P600. Thus, the occurrence of the late frontal negativity and its possible relation to working memory is something which should be examined further in future research.

In the adjective condition, the P600 was preceded by a negativity starting at 328 ms and initially appearing as most prominent in left anterior sites. However, around 400 ms it spreads further back and covers central and centro-parietal sites across both hemispheres. This is not the usual topography of the LAN response, which is typically constricted to left anterior regions. It is possible that this extended distribution is a result of a combined LAN and N400, as has been found in other studies (e.g., Blais et al., 2024; Steinhauer and Ullman, 2002). Steinhauer and Ullman found that irregular verbs elicited N400s, indicative of a reliance on lexico-semantic processing, whereas regular verbs elicited a LAN in male participants and an N400 followed by a LAN in female participants, suggesting that there are individual differences in the reliance on rule-based morphosyntactic processing versus memory-based lexico-semantic processing. Similarly, Blais and colleagues found that the same adjective violations that elicited a LAN-P600 combination in adults, led to an N400-P600 in adolescents. This indicates that morphosyntactic processing might undergo a maturation process. Given the ambiguity of the topography, we cannot rule out that this response reflects a superposition of LAN and N400 components, as individual participants may vary in their reliance on morphosyntactic versus lexico-semantic processing strategies. Nevertheless, we interpret this negativity primarily in light of the LAN literature, while acknowledging that the atypical distribution warrants some interpretive caution.

The adjective condition was the only one to elicit an early negativity in the LAN window. As previously noted, there is still a significant amount of uncertainty surrounding the occurrence of the LAN, as previous findings to little or no degree reveal any clear patterns. The following section will discuss two factors which could have contributed to the emergence of the early negativity in the present study: the number of morphosyntactic markers in the sentence and working memory load.

One of the proposed factors which elicit the LAN response is the amount of morphosyntactic markings and the expectations arising from these (Friederici, 2002; Molinaro et al., 2011), which would provide a plausible explanation for the present results. Recall that we had sentences such as *Jon sitter på benk**en** i hagen* ('Jon sits on bench-**the** in the garden'), with one morphosyntactic marker indicating the gender of the target noun; *Kelneren på **den** dyre kafe**en** jobber hardt* ('The waiter in **the** expensive café-**the** works hard'), with two morphosyntactic markers indicating the gender of the target noun; and ***Det** hyggelige bakeri**et** er travelt om morgenen* ('**The** nice bakery-**the** is busy-$_{\text{NEUT}}$ in the morning'), with three morphosyntactic markers indicating the gender of the target noun. If LAN elicitation is indeed driven at least in part by the cumulative number of morphosyntactic markings, this could provide an explanation for why the LAN was observed in the adjective condition specifically, as this condition contained the greatest number of gender-marking morphosyntactic cues (three) compared to the determiner (two) and suffix (one) conditions. In further support of cue-accumulation driving the LAN is the fact that Sandstedt and colleagues (2025) also used predicative adjective gender violations in their study, but without the additional determiner (e.g., *huset hun likte var dyrt å fikse opp*, 'house-the she liked was expensive-NEUT to fix up'), meaning that their sentences contained only two gender-marking cues: the definiteness suffix and the adjective itself. While their results focus on the P600 time window, earlier time windows were considered but excluded because any negativity visible in the grand averages was too small and ambiguous in nature to justify formal analysis (see footnote on p. 10). The P600 they reported, however, had a similar timing and amplitude as the one found for adjective violations in the present study, with both starting around 550 ms and with a voltage peak of about 4 μV. This could indicate that the presence of the additional gender cue in the present study is what has triggered the LAN response, meanwhile, the extra marking does not seem to have affected the P600. It should be noted, however, that the LAN response has also been found with fewer cues in several other studies (e.g., Caffarra and Barber, 2015), and has failed to appear with the same amount of cues in other studies (e.g., Alemán Bañón et al., 2012), meaning that our results are at best evidence in support of one of several factors driving the LAN. The relevance of the number of cues should be explored further in future studies.

As noted in the introduction, the emergence of the LAN has also been connected to an increased working memory load. Considering the assumption that the cost of processing increases with distance, partly due to the cost of keeping information in WM, this could explain the occurrence of a LAN in this condition, as it is the only one in the present study to go across phrases. If this is the case, it is worth noting that the linear distance in the adjective condition is equal to that of the determiner condition, implying that the structural distance is adding more to WM cost than the linear distance. However, previous findings do not indicate that the LAN response is sensitive to distance. For example, in the study by O'Rourke and Van Petten (2011), distance affected the P600 only, whereas the LAN response remained consistent across conditions. Similarly, Alemán Bañón and colleagues (2012) did not find a LAN for either of their conditions, regardless of increased structural distance.

Concluding from this, it would appear more likely that the LAN emerged due to the number of morphosyntactic markers than the distance between the constituents, but we cannot rule out that the latter also contributed to the response. More research, conducted in a variety

of languages and using several different features and syntactic structures, is needed in order to learn more about the LAN response and when we can expect it to appear.

4.4. Overall summary

The present study demonstrates that within-word gender violations between noun and suffix are processed similarly to other agreement structures. Furthermore, we also found an effect of linear distance on within-phrase dependencies, possibly due to a combination of cost/benefit calculation and higher salience. This indicates that P600 amplitude is sensitive to linear distance even in very short-distance dependencies. Additionally, this study provides the first comprehensive investigation of gender processing in Norwegian in determiner-noun and noun-adjective dependencies, as well as the previously unexplored noun-suffix structure. Our results are in line with former research from other languages, with robust P600s for all three conditions, and a LAN effect for noun-adjective violations which we attributed to the increased number of gender cues for this condition. An unpredicted finding was the late anterior negativity for the determiner condition, which may be a result of a phase reversal or higher WM cost as a result of failed predictions. Overall, these findings indicate that Norwegian gender agreement engages the same reanalysis mechanisms observed in other languages, while being sensitive to linear distance and cue salience, and they motivate further work to pinpoint how distance, cue density, and prediction interact during agreement processing.

CRediT authorship contribution statement

Sarah Cameron: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Natalia Kartushina:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Björn Lundquist:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Sendy Caffarra:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

S. Caffarra received funding from PRIN PNRR P2022SMEJW, supported by the European Union – Next Generation EU. The authors gratefully acknowledge equipment, resources and assistance provided by MultiLing's Socio-Cognitive Laboratory. We thank Béibhinn Maire Reardon and Eva Van Kampen for help with data collection, and all participants for their time and effort. We would also like to thank our two anonymous reviewers for their valuable input and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brainres.2026.150390.

References

Alemán Bañón, J., Rothman, J., 2016. The role of morphological markedness in the processing of number and gender agreement in Spanish: an event-related potential investigation. Language, Cognition and Neuroscience 31 (10), 1273–1298. https://doi.org/10.1080/23273798.2016.1218032.

Alemán Bañón, J., 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 Res. 1456, 49–63. https://doi.org/10.1016/j.brainres.2012.03.057.

Bambini, V., Canal, P., Breimaier, F., Meo, D., Pescarini, D., Loporcaro, M., 2021. Capturing language change through EEG: Weaker P600 for a fading gender value in a southern Italo-Romance dialect. J. Neurolinguistics 59, 101004. https://doi.org/10.1016/j.jneuroling.2021.101004.

Barber, H., Carreiras, M., 2005. Grammatical gender and number agreement in Spanish: an ERP comparison. J. Cogn. Neurosci. 17 (1), 137–153. https://doi.org/10.1162/0898929052880101.

Barber, H., Salillas, E., Carreiras, M., 2004. Gender or genders agreement? In *The On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond* (pp. 309–327). Scopus. https://doi.org/10.4324/9780203509050.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

Beatty-Martínez, A.L., Bruni, M.R., Bajo, M.T., Dussias, P.E., 2021. Brain potentials reveal differential processing of masculine and feminine grammatical gender in native Spanish speakers. Psychophysiology 58 (3). https://doi.org/10.1111/psyp.13737.

ERP Evidence for Early and Late Grammar Consolidation of Gender Agreement in French-Speaking Adolescents Vol. 1, 2024, 73–86.

Busterud, G., Lohndal, T., Rodina, Y., Westergaard, M., 2019. The loss of feminine gender in Norwegian: A dialect comparison. J. Comparat. Germanic Linguist. 22 (2), 141–167. https://doi.org/10.1007/s10828-019-09108-7.

Caffarra, S., 2023. The Impact of Sentence Context on the Morphological Processing of Single Words: Electrophysiological evidence. *Linguistic Morphology in the Mind and Brain*. Routledge.

Caffarra, S., Barber, H.A., 2015. Does the ending matter? The role of gender-to-ending consistency in sentence reading. Brain Res. 1605, 83–92. https://doi.org/10.1016/j.brainres.2015.02.018.

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 (8), 1019–1030. https://doi.org/10.1111/psyp.12429.

Caffarra, S., Mendoza, M., Davidson, D., 2019. Is the LAN effect in morphosyntactic processing an ERP artifact? Brain Lang. 191, 9–16. https://doi.org/10.1016/j.bandl.2019.01.003.

Christianson, K., Hollingworth, A., Halliwell, J.F., Ferreira, F., 2001. Thematic roles assigned along the garden path linger. Cogn. Psychol. 42 (4), 368–407. https://doi.org/10.1006/cogp.2001.0752.

Ciaccio, L.A., Bürki, A., Clahsen, H., 2023. Inter-individual variability in morphological processing: an ERP study on German plurals. J. Neurolinguistics 67, 101138. https://doi.org/10.1016/j.jneuroling.2023.101138.

Delorme, A., Makeig, S., 2004. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134 (1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

De Resende, N.C.A., Mota, M.B., Seuren, P., 2019. The processing of grammatical gender agreement in Brazilian Portuguese: ERP evidence in favor of a single route. J. Psycholinguist. Res. 48 (1), 181–198. https://doi.org/10.1007/s10936-018-9598-z.

Faarlund, J.T., 2009. On the history of definiteness marking in Scandinavian. Journal of Linguistics 45 (3), 617–639.

Ferreira, F., Bailey, K.G.D., Ferraro, V., 2002. Good-enough representations in language comprehension. Curr. Dir. Psychol. Sci. 11 (1), 11–15. https://doi.org/10.1111/1467-8721.00158.

Fiebach, C.J., Schlesewsky, M., Friederici, A.D., 2002. Separating syntactic memory costs and syntactic integration costs during parsing: the processing of German WH-questions. J. Mem. Lang. 47 (2), 250–272. https://doi.org/10.1016/S0749-596X(02)00004-9.

Fodor, J., 1998. Learning to parse? J. Psycholinguist. Res. 27, 285–319. https://doi.org/10.1023/A:1023258301588.

Fodor, J.D., Inoue, A., 1998. Attach anyway. In: Fodor, J.D., Ferreira, F. (Eds.), *Reanalysis in Sentence Processing*. Springer, Netherlands, pp. 101–141. https://doi.org/10.1007/978-94-015-9070-9_4.

Fodor, J., Inoue, A., 2000. Garden Path Re-Analysis: Attach (Anyway) and Revision as Last Resort, pp. 21–61. https://doi.org/10.1007/978-94-011-3949-6_2.

Foucart, A., Frenck-Mestre, C., 2011. Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. Biling. Lang. Congn. 14 (3), 379–399. https://doi.org/10.1017/S136672891000012X.

Foucart, A., Frenck-Mestre, C., 2012. Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. J. Mem. Lang. 66 (1), 226–248. https://doi.org/10.1016/j.jml.2011.07.007.

Fox, J., Weisberg, S., Price, B., Adler, D., Bates, D., Baud-Bovy, G., Bolker, B., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., Graves, S., Heiberger, R., Krivitsky, P., Laboissiere, R., Maechler, M., Monette, G., Murdoch, D., Nilsson, H., … R-Core, 2026. *car: Companion to Applied Regression* (Version 3.1-5) [Computer software]. https://cran.r-project.org/web/packages/car/index.html.

Friederici, A.D., 2002. Towards a neural basis of auditory sentence processing. Trends Cogn. Sci. 6 (2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8.

Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91 (4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011.

Friederici, A.D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. Trends Cogn. Sci. 16 (5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001.

Gascon, A., Lebel, V., Royle, P., Drury, J. E., Steinhauer, K., 2011, April. *Task effects in event related potentials for semantics and gender in French.* [Conference paper] 10th Symposium of Psycholinguistics, Donostía-San Sebastian, Spain. https://www.

researchgate.net/publication/261946604_Task_effects_in_event_related_potentials_for_semantics_and_gender_in_French.
Gibson, E., 1998. Linguistic complexity: Locality of syntactic dependencies. Cognition 68 (1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1.
Gibson, E., 2000. The dependency locality theory: a distance-based theory of linguistic complexity. Image, Language, Brain 2000, 95–126.
Guevara, E., 2010. NoWaC: A large web-based corpus for Norwegian, pp. 1–7.
Gunter, T.C., Stowe, L.A., Mulder, G., 1997. When syntax meets semantics. Psychophysiology 34 (6), 660–676. https://doi.org/10.1111/j.1469-8986.1997.tb02142.x.
Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: erps reveal early autonomy and late interaction. J. Cogn. Neurosci. 12 (4), 556–568. https://doi.org/10.1162/089892900562336.
Hagoort, P., 2003. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. J. Cogn. Neurosci. 15 (6), 883–899. https://doi.org/10.1162/089892903322370807.
Hagoort, P., 2013. MUC (memory, Unification, Control) and beyond. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00416.
Hammer, A., Jansma, B.M., Lamers, M., Münte, T.F., 2008. Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. Brain Res. 1230, 177–191. https://doi.org/10.1016/j.brainres.2008.07.004.
Honda, M., Barrett, G., Yoshimura, N., Ikeda, A., Nagamine, T., Shibasaki, H., 1996. Event-related potentials during paired associate memory paradigm. Electroencephalogr. Clin. Neurophysiol. 100, 419–439.
Kaan, E., 2002. Investigating the Effects of Distance and Number Interference in Processing Subject-Verb Dependencies: An ERP Study. *Journal of Psycholinguistic Research*, *31*(2), 165–193. https://doi.org/10.1023/A:1014978917769.
Kluender, R., Kutas, M., 1993. Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. J. Cogn. Neurosci. 5 (2), 196–214. https://doi.org/10.1162/jocn.1993.5.2.196.
Kolk, H.H.J., Chwilla, D.J., van Herten, M., Oor, P.J.W., 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. Brain Lang. 85 (1), 1–36. https://doi.org/10.1016/s0093-934x(02)00548-5.
Kubota, M., Alonso, J.G., Anderssen, M., Jensen, I.N., Luque, A., Soares, S.M.P., Rothman, J., 2024. Bilectal exposure modulates neural signatures to conflicting grammatical properties: Norway as a natural laboratory. Lang. Learn. 74 (2), 436–467. https://doi.org/10.1111/lang.12608.
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*(Volume 62, 2011), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.
Lau, E.F., Holcomb, P.J., Kuperberg, G.R., 2013. Dissociating N400 effects of prediction from association in single-word contexts. J. Cogn. Neurosci. 25 (3), 484–502. https://doi.org/10.1162/jocn_a_00328.
Lenth, R. V., Piaskowski, J., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Hervé, M., Jung, M., Love, J., Miguez, F., Riebl, H., Singmann, H., 2025. *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 2.0.1) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html.
Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: An open-source toolbox for the analysis of event-related potentials. Front. Human Neurosci. 8, 213.
Luck, S. (2022). Book: Applied Event-Related Potential Data Analysis (Luck). Retrieved September 9, 2022, from https://socialsci.libretexts.org/Bookshelves/Psychology/Book%3A_Applied_Event-Related_Potential_Data_Analysis_(Luck).
Lück, M., Hahne, A., Clahsen, H., 2006. Brain potentials to morphologically complex words during listening. Brain Res. 1077 (1), 144–152. https://doi.org/10.1016/j.brainres.2006.01.030.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.
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 Res. 1093 (1), 178–189. https://doi.org/10.1016/j.brainres.2006.03.094.
Mickan, A., Lemhöfer, K., 2020. Tracking syntactic conflict between languages over the course of L2 acquisition: a cross-sectional event-related potential study. J. Cogn. Neurosci. 32 (5), 822–846. https://doi.org/10.1162/jocn_a_01528.
Molinaro, N., Vespignani, F., Job, R., 2008. A deeper reanalysis of a superficial feature: an ERP study on agreement violations. Brain Res. 1228, 161–176. https://doi.org/10.1016/j.brainres.2008.06.064.
Molinaro, N., Barber, H.A., Carreiras, M., 2011. Grammatical agreement processing in reading: ERP findings and future directions. Cortex 47 (8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019.
Münte, T.F., Schiltz, K., Kutas, M., 1998. When temporal terms belie conceptual order. Nature 395 (6697), 71–73. https://doi.org/10.1038/25731.
Nevins, A., Dillon, B., Malhotra, S., Phillips, C., 2007. The role of feature-number and feature-type in processing Hindi verb agreement violations. Brain Res. 1164, 81–94. https://doi.org/10.1016/j.brainres.2007.05.058.
Oostenveld, R., Fries, P., Maris, E., Schoffelen, JM (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience, Volume 2011 (2011), Article ID 156869, doi:10.1155/2011/156869.
Osterhout, L., Mobley, L.A., 1995. Event-related brain potentials elicited by failure to agree. J. Mem. Lang. 34 (6), 739–773. https://doi.org/10.1006/jmla.1995.1033.
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 *The On-line Study of Sentence Comprehension*. Psychology Press.
Otten, M., Van Berkum, J.J.A., 2008. Discourse-based word anticipation during language processing: prediction or priming? Discourse Process. 45 (6), 464–496. https://doi.org/10.1080/01638530802356463.
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 Res. 1392, 62–79. https://doi.org/10.1016/j.brainres.2011.03.071.
Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., 2019. PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods 51 (1), 195–203. https://doi.org/10.3758/s13428-018-01193-y.
Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T.F., Clahsen, H., 1997. How the brain processes complex words: an event-related potential study of German verb inflections. Brain Res. Cogn. Brain Res. 6 (1), 37–52. https://doi.org/10.1016/s0926-6410(97)00012-8.
Rodina, Y., Westergaard, M., 2015. Grammatical Gender in Norwegian: Language Acquisition and Language Change. J. Germanic Linguist. 27 (2), 145–187. https://doi.org/10.1017/S1470542714000245.
Rodriguez-Fornells, A., Clahsen, H., Lleó, C., Zaake, W., Münte, T.F., 2001. Event-related brain responses to morphological violations in Catalan. Cogn. Brain Res. 11 (1), 47–58. https://doi.org/10.1016/S0926-6410(00)00063-X.
Royle, P., Steinhauer, K., 2023. Neural Correlates of Morphology Computation and Representation. In: Grimaldi, M., Brattico, E., Shtyrov, Y. (Eds.), *Language Electrified: Principles, Methods, and Future Perspectives of Investigation*. Springer, US, pp. 447–503. https://doi.org/10.1007/978-1-0716-3263-5_14.
Royle, P., Drury, J.E., Steinhauer, K., 2013. ERPs and task effects in the auditory processing of gender agreement and semantics in French. The Mental Lexicon: Special Issue on the Neural Correlates of Lexical Processing 8 (2), 216–244. https://doi.org/10.1075/ml.8.2.05roy.
Royle, P., Courteau, É., Brucher, A., Misirliyan, C., Steinhauer, K., 2021. October). Heterogeneous ERP patterns for different error types in French complex noun phrases. Slide Slam *Society for the Neurobiology of Language*.
Ruchkin, D.S., Johnson, R., Canoune, H., Ritter, W., 1990. Short-term memory storage retention: an event-related brain potential study. Electroencephalogr. Clin. Neurophysiol. 76, 419–439.
Sabourin, L., Stowe, L., 2004. Memory effects in syntactic ERP tasks. Brain Cogn. 55 (2), 392–395. https://doi.org/10.1016/j.bandc.2004.02.056.
Sabourin, L., Stowe, L.A., 2008. Second language processing: when are first and second languages processed similarly? Second. Lang. Res. 24 (3), 397–430. https://doi.org/10.1177/0267658308090186.
Sandstedt, J., Kubota, M., Anderssen, M., Darby, J.A., Helset, S., Tavakoli, E., Vangsnes, Ø.A., Rothman, J., 2025. Bidialectal language representation and processing: evidence from Norwegian ERPs. J. Mem. Lang. 140, 104557. https://doi.org/10.1016/j.jml.2024.104557.
Sassenhagen, J., Draschkow, D., 2019. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology 56, e13335. https://doi.org/10.1111/psyp.13335.
Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language, 83*(9), 62–65. https://www.mcgill.ca/neurocoglab/files/neurocoglab/Steinhauer2002b_A.pdf.
Swaab, T.Y., Ledoux, K., Camblin, C.C., Boudewyn, M.A., 2011, December 15.. *Language-Related ERP Components*. The Oxford Handbook of Event-Related Potential Components. https://doi.org/10.1093/oxfordhb/9780195374148.013.0197.
Tanner, D., Van Hell, J.G., 2014. ERPs reveal individual differences in morphosyntactic processing. Neuropsychologia 56, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002.
The MathWorks Inc. (2024). MATLAB. https://www.mathworks.com.
van de Meerendonk, N., Kolk, H.H.J., Vissers, C.T.W.M., Chwilla, D.J., 2010. Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. J. Cogn. Neurosci. 22 (1), 67–82. https://doi.org/10.1162/jocn.2008.21170.
Vos, S.H., Gunter, T.C., Kolk, H.H.J., Mulder, G., 2001. Working memory constraints on syntactic processing: an electrophysiological investigation. Psychophysiology 38 (1), 41–63. https://doi.org/10.1111/1469-8986.3810041.
Weyerts, H., Penke, M., Dohrn, U., Clahsen, H., Münte, T.F., 1997. Brain potentials indicate differences between regular and irregular German plurals. Neuroreport 8 (4), 957–962. https://doi.org/10.1097/00001756-199703030-00028.
Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. J. Cogn. Neurosci. 16 (7), 1272–1288. https://doi.org/10.1162/0898929041920487.
Zawiszewski, A., Santesteban, M., Laka, I., 2016. Phi-features reloaded: an event-related potential study on person and number agreement processing. Appl. Psycholinguist. 37 (3), 601–626. https://doi.org/10.1017/S014271641500017X.
Zeller, J., Bylund, E., Lewis, A.G., 2022. The parser consults the lexicon in spite of transparent gender marking: EEG evidence from noun class agreement processing in Zulu. Cognition 226, 105148. https://doi.org/10.1016/j.cognition.2022.105148.