

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://repository.ubn.ru.nl/handle/2066/128480>

Please be advised that this information was generated on 2021-06-13 and may be subject to change.

# Idiosyncratic Grammars: Syntactic Processing in Second Language Comprehension Uses Subjective Feature Representations

Kristin Lemhöfer<sup>1</sup>, Herbert Schriefers<sup>1</sup>, and Peter Indefrey<sup>1,2</sup>

## Abstract

■ Learning the syntax of a second language (L2) often represents a big challenge to L2 learners. Previous research on syntactic processing in L2 has mainly focused on how L2 speakers respond to “objective” syntactic violations, that is, phrases that are incorrect by native standards. In this study, we investigate how L2 learners, in particular those of less than near-native proficiency, process phrases that deviate from their own, “subjective,” and often incorrect syntactic representations, that is, whether they use these subjective and idiosyncratic representations during sentence comprehension. We study this within the domain of grammatical gender in a population of German learners of Dutch, for which systematic errors of grammatical gender are well documented. These L2 learners as well as a control group of Dutch native speakers

read Dutch sentences containing gender-marked determiner-noun phrases in which gender agreement was either (objectively) correct or incorrect. Furthermore, the noun targets were selected such that, in a high proportion of nouns, objective and subjective correctness would differ for German learners. The ERP results show a syntactic violation effect (P600) for objective gender agreement violations for native, but not for nonnative speakers. However, when the items were re-sorted for the L2 speakers according to subjective correctness (as assessed offline), the P600 effect emerged as well. Thus, rather than being insensitive to violations of gender agreement, L2 speakers are similarly sensitive as native speakers but base their sensitivity on their subjective—sometimes incorrect—representations. ■

## INTRODUCTION

In an increasingly “globalized” world, being able to speak several languages has become one of the standard requirements of modern life. As many foreign language learners only start to use that foreign language intensively when they are adults, most of them experience that it is hard to reach perfection in that new language: It is estimated that only 5–15% of adult second language (L2) learners reach native-like levels of proficiency (Birdsong, 2004). This holds particularly for domains like pronunciation and syntax, whereas semantic processing is often found to be comparable to that of natives from early on during L2 acquisition (Ojima, Nakata, & Kakigi, 2005; Duyck & Brysbaert, 2004; Stenberg, Johansson, & Rosen, 2004; Hahne & Friederici, 2001).

A controversial topic in the literature is whether or not native-like syntactic language processing is possible in adult L2 speakers and which factors determine the success of L2 acquisition (Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Gillon Dowens, Guo, Guo, Barber, & Carreiras, 2011; Clahsen & Felser, 2006; McDonald, 2000). Recently, this debate has focused on ERP data, as this method provides a direct measure of online language

processing (e.g., Kotz, 2009; Chen, Shu, Liu, Zhao, & Li, 2007; Rossi, Gugler, Friederici, & Hahne, 2006). The standard methodology in these studies is to have both native and L2 speakers (sometimes the latter split into “low” vs. “high” proficiency groups) as participants and to let these groups read or listen to sentences that contain syntactic violations. The question is whether the L2 group is sensitive to these violations at all, and if so, whether the ERP patterns of the L2 speakers resemble those of the native speakers (for overviews, see Kotz, 2009; Steinhauer, White, & Drury, 2009).

Within the domain of grammatical gender processing, which is the focus of this study, several EEG studies investigated whether nonnatives respond differently to violations of gender agreement than native speakers do. Gillon Dowens, Vergara, Barber, and Carreiras (2009) examined the processing of both gender and number agreement in native speakers of Spanish and late, but highly proficient L2 learners of Spanish with English as native language. The agreement violation was either located at the beginning or in the middle of the sentence. Overall, the same pattern of ERPs was found for the two location conditions and participant groups, namely, a LAN-type effect, a P600, and a very late negativity (starting at 1000 msec after noun onset). However, there were subtle differences in timing, latency, and scalp distribution between the participant groups depending on violation type

<sup>1</sup>Radboud University Nijmegen, <sup>2</sup>Heinrich Heine University Düsseldorf

and location, which were interpreted as evidence for small differences between native and nonnative processing. In a study with the same materials, but highly proficient Chinese L2 speakers of Spanish, Gillon Dowens et al. (2011) showed that the P600 effects of gender or number agreement violations also generalized to speakers with a native language that does not possess gender or number agreement. However, the LAN effects were missing for this group.

Foucart and Frenck-Mestre (2011) investigated the response of German learners of French to different instances of French gender agreement: determiner–noun agreement that is argued to be implemented similarly in German and French as well as adjective–noun agreement that differs from German. The results indicated that the L2 speakers, just like native speakers, displayed the expected P600 effects for determiner–noun agreement, whereas these were absent in the L2 speakers for violations of noun–adjective agreement. The authors interpreted their results in terms of differences in learnability of different syntactic constructions, depending on the overlap of L1 and L2. Note that this latter observation stands in contrast to Gillon Dowens et al.’s (2011) finding that speakers of a language without these features (Chinese) were nonetheless sensitive to gender and number agreement violations. Likewise, Gillon Dowens et al.’s (2009) first study demonstrated almost native-like effects of gender agreement in native speakers of English, a language without a grammatical gender system.

In summary, these studies suggest that violations of gender agreement are processed by L2 speakers in a way that is highly similar to that of native speakers, with some additional modulating influence of L1–L2 similarity with respect to how gender agreement is implemented. However, in all three studies, the L2 participants’ performance on gender assignment with respect to the experimental items was assessed offline after the experiment, resulting in very low error rates (below 6% in all three studies). Thus, the L2 speakers’ mastery of the gender system of their second language was very good. The low error rates on gender assignment might be because of the participants’ high level of proficiency (e.g., more than 12 years of immersion in a Spanish-speaking environment in the study by Gillon Dowens et al., 2009), the transparent nature of the grammatical gender system in romance languages given phonological gender cues of nouns, or other characteristics of the items (frequency, cognate status, etc.).

In contrast, L2 speakers at less advanced levels or with other language combinations often experience severe problems with certain syntactic aspects of L2, such as grammatical gender (e.g., Orgassa & Weerman, 2008; Dewaele & Véronique, 2001). For instance, in German learners of Dutch, Dutch nouns with incompatible gender between the two languages are likely to be assigned the wrong (but German-compatible) gender (Lemhöfer, Spalek, & Schriefers, 2008). Thus, incorrect representations of certain syntactic features might be common in

L2 speakers who are not as proficient and experienced as the populations investigated in some of the studies above. Furthermore, Lemhöfer, Schriefers, and Hanique (2010) found many of these incorrect representations to be stable across item repetitions, to be immune to feedback, and to be given with great subjective certainty. Possibly, such “stubborn,” incorrect representations are not just an aspect of early stages of L2 acquisition, but also of what is sometimes called “L2 fossilization” (e.g., White, 2003; Selinker, 1972), that is, the cessation of further learning in experienced L2 speakers.

Assuming, thus, that incorrect syntactic representations are a normal part of the second language system, the question arises whether and how these representations are used during syntactic processing. Will L2 speakers who frequently make gender errors use word gender at all during comprehension, or will they rather employ a “gender-free,” shallow way of processing in the sense of a “good enough” approach (Ferreira, 2003)? The latter might work especially well for gender agreement, which in Dutch is rarely necessary for comprehension and disambiguation. If, however, L2 speakers do use word gender during comprehension, which representations do they use? Will they base their syntactic processing on subjective representations, even the wrong ones, and if so, will they do it in the same way as natives base their processing on their (correct) representations? This is what this study sets out to investigate.

To our knowledge, this question has not yet been studied within the L2 syntactic sentence processing literature. Within the scope of the syntactic violation paradigm, three alternatives of how L2 speakers might respond to a syntactic violation have generally been considered until now: First, they might not respond to a violation at all (because of, e.g., shallow processing); second, they might respond to it exactly as native speakers do (which is, as we have seen, most likely in the case of high proficiency); and third, they might respond to it, but in a qualitatively different way from natives (e.g., by showing N400 effects rather than the usual “syntactic” ERP components; McLaughlin et al., 2010). Our considerations suggest a fourth possibility: An L2 speaker might respond to a violation as if it was correct and, the other way around, to a correct phrase as if it was a violation, provided that she has an incorrect subjective representation for the respective item. For example, if an L2 speaker of Dutch believes that the word *auto* (“car”) is neuter and hence takes the neuter definite determiner *het*, then the correct phrase *de auto* should be subjectively wrong to this speaker, and the incorrect phrase *\*het auto* should be processed as if it was just fine. This study will explore this fourth possibility that subjective, sometimes wrong representations are the basis of syntactic processing in L2 learners of intermediate proficiency. Furthermore, if these representations are indeed used, we will examine whether or not they are used in a similar way as (correct) representations are used by native speakers.

We study this issue not only because we aim to better understand syntactic L2 processing as compared with native processing but also because we are intrigued by the “stubbornness” of some L2 syntactic errors such as the word gender errors we previously observed in German learners of Dutch (Lemhöfer et al., 2008, 2010). How is it possible that these speakers, who are immersed in an L2 environment, adhere to errors such as *\*bet auto*, although they are surrounded by native speakers who always say *de auto*? One possible reason is that they do not notice the difference between the input they receive—like *de auto*—and their own errors any more (Schmidt, 1990). In that case, we should expect that the speakers do not respond to (either subjective or objective) violations of gender agreement between determiner and noun.

With L2 grammatical gender as the focus of our study, we used the most simple type of phrase in which gender agreement surfaces in Dutch (or German), definite determiner–noun phrases. Definite determiners in Dutch are marked for gender, which can be common gender (definite determiner *de*) or neuter gender (definite determiner *het*). As already mentioned, German learners of Dutch have the tendency to map German masculine and feminine gender onto Dutch common gender, and German neuter gender onto Dutch neuter gender. This mapping tendency is particularly strong for cognates (i.e., words that are similar in form and meaning across languages [e.g., Dutch *bond*, German *Hund*, “dog”]). When such cognates do not follow the preferred mapping between German and Dutch gender, German learners are likely to adhere to incorrect but often stable gender representations for these words in Dutch (Lemhöfer et al., 2008, 2010). Thus, these words provide an ideal testing ground for the potential role of incorrect subjective representations in L2 processing: Although for gender-compatible cognates, the gender representations of nouns are highly likely to be correct (with error rates below 10% in Lemhöfer et al., for instance), the opposite is the case for gender-incompatible cognates, with previously observed error rates of up to 70%. In other words, subjective and objective correctness are likely to overlap for the first noun type, but to diverge for many items of the second type. Using these two word types will thus enable us to disentangle the roles of subjective and objective correctness.

German learners of Dutch as well as a control group of native speakers of Dutch read Dutch sentences containing target nouns preceded by their correct or their incorrect definite determiner while the EEG was recorded. The task was to read these sentences for comprehension (with occasional comprehension questions). Crucially, we categorized the items not only in the conventional way—according to predefined categories differing in objective correctness (correct vs. incorrect gender)—but also in a way that is new to this field of research, namely, according to each participant’s individual subjective representations. That is, after the EEG experiment, we asked the

L2 learners in an offline questionnaire to indicate the correct gender for each target; subsequently, we used these offline data to re-sort the items per participant according to the participant’s response. Similar methodologies—the categorization of items based on participant behavior rather than on predefined categories—are well known from memory research, where remembered items are contrasted with not remembered ones (e.g., Brodeur et al., 2011), or from the error processing literature, where trials with correct and incorrect responses are compared (e.g., Maier, Yeung, & Steinhauser, 2011).

We included a control group of Dutch native speakers to determine which ERP components should serve as “markers” for the successful (native-like) processing of agreement. Previous ERP studies on the processing of gender or number agreement violations in native speakers of various languages have almost invariably shown P600 effects, at least when the violation concerned two adjacent words in the sentence (O’Rourke & Van Petten, 2011; Barber & Carreiras, 2005; Wicha, Moreno, & Kutas, 2004; Hagoort, 2003; Münte, Szentkuti, Wieringa, Matzke, & Johannes, 1997). However, the studies differ in whether they also observed an early, anterior, and often left-lateralized negativity (LAN; O’Rourke & Van Petten, 2011; Barber & Carreiras, 2005; Gunter, Friederici, & Schriefers, 2000) or not (Foucart & Frenck-Mestre, 2011; Hagoort, 2003, for the middle sentence position; Martin-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006; Wicha et al., 2004). Whereas the P600 component is generally assumed to reflect strategic processes of syntactic reanalysis and repair, the LAN is probably indicative of a more automatic process of morphosyntactic reference computation (e.g., O’Rourke & Van Petten, 2011; Rossi et al., 2006; Friederici, 2002). Still, it is currently unclear why this LAN effect is absent in many syntactic agreement studies, both in L1 and L2 speakers. It has been argued that a LAN effect in L2 speakers arises only at the highest level of (native-like) proficiency, whereas P600 effects occur earlier during L2 acquisition (Steinhauer et al., 2009). However, not all available results point in this direction (e.g., Ojima et al., 2005, found only the early effect, but not a P600 in high-proficiency speakers processing subject–verb agreement in English). Furthermore, this claim is of course based on the assumption that the used manipulation does cause LAN effects in native speakers, which, as we have just argued, is not always the case. We therefore included a native control group to be able to compare our L2 speakers’ performance to that of native speakers on the exact same material and experimental procedure.

Because of the possibility that our L2 speakers might not respond to gender agreement manipulations at all, we included a second manipulation of determiner–noun agreement, namely, number agreement. This type of agreement is implemented in the same way in Dutch and German and is usually unproblematic for German learners of Dutch; violations of number agreement should be highly salient to our L2 population. Including this violation

type hence enabled us to test whether L2 speakers would respond to grammatical violations at all (or rather engage in an entirely “grammar-free” processing mode).

## METHODS

### Participants

Twenty-one right-handed native speakers of Dutch and 29 German learners of Dutch, all students at Radboud University Nijmegen, took part in the experiment for course credit or payment (€10/hr). Data from seven L2 learners were excluded from analyses because they had not noticed any grammatical errors in the sentences at all, indicating extremely low proficiency and/or grammatical awareness in Dutch, such that a group of 22 speakers remained. The number of women was 14 (native speakers) and 17 (L2 learners). All reported to have normal or corrected-to-normal vision and not to be dyslexic. The mean age was 23.6 years (native speakers) and 23.2 years (L2 learners).

All of the native speakers had experience with other foreign languages, especially English, but also French or German. Likewise, the L2 learners reported speaking other foreign languages beside Dutch, in particular English, but also French, Italian, and Spanish. Four participants stated using English more often than Dutch. None of the other (gender-marking) languages were spoken more often or more proficiently than Dutch. All other results from a language background questionnaire given to the L2 learners are summarized in Table 1.

### Materials

#### Gender Agreement Condition

Given that the largest differences in error rates for gender assignment in German learners of Dutch have been

demonstrated for gender-compatible versus -incompatible cognates, we used only cognates as target nouns, with “cognates” defined as translation equivalents with obvious common etymological roots and high form overlap between Dutch and German (e.g., *schip<sub>D</sub>*–*Schiff<sub>G</sub>* “ship,” *spek<sub>D</sub>*–*Speck<sub>G</sub>* “bacon,” but also many identical cognates like *strand/Strand* “beach,” *radio/Radio* “radio”). Note that because of the close relation between Dutch and German, cognates form a large part of the vocabulary, such that the large proportion of cognates in the materials was not extraordinary.

Forty cognate nouns with compatible gender and 40 cognates with incompatible gender between Dutch and German were selected from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). In analogy with our previous studies, Dutch neuter nouns were categorized as “gender-compatible” when they were neuter in German, too, and Dutch common gender nouns when they were either feminine or masculine in German. No targets with transparent grammatical gender (e.g., nouns with suffixes indicating grammatical gender or nouns with natural gender) were included. Within each of the two groups of 40 nouns, 20 nouns were *de*-words in Dutch, that is, of common gender, whereas the other 20 nouns were of neuter gender, taking the definite determiner *het*. The target nouns were between three and nine letters long (mean = 5.4) and had an average frequency of 55 occurrences per million in Dutch according to the CELEX database.

Because each participant would see each noun twice (once with the correct, once with the incorrect determiner), two sentence frames per target noun were constructed. The critical determiner and noun never occurred in sentence-initial or sentence-final position. The occurrence of correct and incorrect determiners was then counterbalanced across these sentence frames (see “list construction” below for more details). The structure

**Table 1.** Results from the Language Background Questionnaire Given to German Learners of Dutch

	Mean	SD	Range
Age of first contact with Dutch (years)	19.2	3.4	6–23
Years of experience with Dutch	4.8	6.0	1–25
Self-ratings <sup>a</sup>			
How often do you read Dutch literature?	5.1	1.9	1–7
How often do you speak Dutch?	5.6	1.2	3–7
How often do you listen to Dutch radio/watch Dutch TV?	3.8	1.7	1–6
Self-rated reading experience in Dutch	5.0	1.3	1–7
Self-rated writing experience in Dutch	4.7	1.2	2–7
Self-rated speaking experience in Dutch	5.4	1.0	4–7
Mean Dutch experience (mean of previous 3)	5.0	0.9	2.7–7.0

<sup>a</sup>Self-ratings were given on a scale from 1 (*low/rarely*) to 7 (*high/very often*).



**Table 2.** Examples for Sentences with English Translations in Gender and Number Agreement Conditions

Condition	Example Sentence
Gender agreement: correct	<i>De oude man wachtte op <u>bet</u><sub>neu</sub> <u>einde</u><sub>neu</sub> van zijn leven.</i> (The old man waited for the <u>end</u> of his life.)
Gender agreement: incorrect	<i>Het volk verlangt naar *<u>de</u><sub>com</sub> <u>einde</u><sub>neu</sub> van de dictatuur.</i> (The people longed for the <u>end</u> of the dictatorship.)
Number agreement: correct	<i>Een rondreis langs <u>de</u><sub>pl</sub> <u>dorpen</u><sub>pl</sub> van het eiland duurt twee dagen.</i> (A round trip of the <u>villages</u> of the island takes two days.)
Number agreement: incorrect	<i>De geschiedenis van *<u>bet</u><sub>sing</sub> <u>dorpen</u><sub>pl</sub> wordt in het gemeentearchief gedocumenteerd.</i> (The history of the <u>villages</u> is documented in the municipal archive.)

Target nouns are underlined. Assignment of correct and incorrect determiners to sentences was counterbalanced across experimental lists. neu = neuter gender; com = common gender; pl = plural; sing = singular.

of these sentences was, at least up to the critical noun, as similar as possible within sentence pairs. A cloze test was run on an independent sample of 11 native speakers of Dutch who did not take part in the actual experiment. For this cloze test, the experimental sentences were truncated just before the critical noun phrase. Cloze probability for the target noun was below 0.1 for all sentences (mean = 0.005). The use of additional gender-incompatible cognates in other parts of the sentences than the critical noun phrase was avoided. The sentences were between 6 and 15 words long (mean = 10.2); the target position was between the second and the eleventh word (mean = 5.3). Examples for sentences in gender and number agreement conditions are given in Table 2.

#### Number Agreement Condition

Besides the determiner–noun phrases with correct versus incorrect gender agreement, we included a second type of violation that we expected to be especially salient to German learners of Dutch. In both German and Dutch, there is one single, gender-unmarked definite determiner for plural nouns (*de* in Dutch, *die* in German). Including violations of Dutch number agreement (e.g., \**bet paarden*, the<sub>sing</sub> horses) thus served as a sort of baseline, providing information on whether the L2 learners were sensitive to any sort of agreement violation at all.

Number agreement violations were constructed by combining the plural of a neuter gender noun with its singular determiner *bet*. To this end, 32 neuter cognate nouns were additionally selected from the CELEX database. All these 32 nouns were gender-compatible with German to avoid confusion as to the nouns' gender. The mean Dutch frequency of these nouns according to CELEX was 117 o.p.m., and their mean length was 5.5 letters.

For each number target, two sentence frames were constructed in the same way as in the gender condition. Each participant received each item once in a sentence with the correct plural determiner *de* and once in another sentence with the incorrect (singular) determiner *bet*. Sentences were between 7 and 14 words long (mean = 10.8); the

position of the target noun was between the third and the eighth word (mean = 5.2). Examples for the sentences are given in Table 2.

#### Filler Sentences and Comprehension Questions

To compensate for the exclusive use of *bet*-words as targets in the number agreement condition, we added 32 filler sentences containing plurals of *de*-words. The target nouns in these fillers also comprised noncognates. Length and kind of the sentences were comparable to the experimental ones.

We refrained from using grammaticality judgments along with the EEG measurement to avoid an unnatural, “grammar-focused” mode of processing. Rather, participants were instructed to read the sentences for meaning. To test for sentence comprehension and to keep participants attentive, about 10% of the sentences in the experiment (25 of 256 sentences) were followed by a yes/no comprehension question. Only grammatically correct sentences were followed by comprehension questions. For example, following the sentence “It is not safe to drink water from the river,” the question “Can you get sick from the water of the river?” was presented. Half of the questions required a “yes” answer, and the others a “no”.

#### List Construction

We counterbalanced with which determiner (correct or incorrect) and with which sentence frame a target occurred first by constructing four pseudorandomized experimental lists. Participants were randomly assigned to one of the four lists. List construction had been such that both the first and the second half of the experiment each comprised the full set of targets, that is, all targets appeared for the first time in the first and for the second time in the second half. All lists met the following randomization restrictions: (a) no more than three subsequent correct or incorrect sentences followed each other, (b) there were no more than three subsequent sentences with the same critical

determiner (*de* or *bet*), and (c) no two sentences were followed by questions in immediate succession.

The total of 256 sentences ( $2 \times 80 = 160$  gender agreement sentences,  $2 \times 32 = 64$  number agreement sentences, and 32 filler sentences) was presented in six blocks of 44 sentences, with breaks in between the blocks. The first one or two sentences of each block were additional dummy sentences to allow for the EEG signal to settle down after the break. The EEG experiment was preceded by a brief practice session of 10 sentences that were similar to the experimental materials (e.g., there were determiner errors in four of these sentences).

## Procedure

Participants were tested individually. They sat in front of a computer screen and a button box in a dimly lit cabin. The instruction was to read the presented Dutch sentences for comprehension, such that they could answer the occasional questions by pressing one of the two buttons (right for “yes,” left for “no”). Participants were also asked to try not to blink during sentence presentation.

Sentences were presented word-by-word in the center of the screen in black 24 pt Arial letters on a light gray background. Before the beginning of each sentence, a fixation cross appeared in the same central location of the screen for 500 msec. After a blank of 250 msec, the first word appeared. Each word stayed on the screen for 500 msec, followed by a blank screen for 300 msec.<sup>1</sup> The interval between the last word of a sentence and the fixation before the next was 1500 msec. Questions were presented as a whole after the last word of a sentence and remained on the screen until a response was given (or otherwise until a deadline of 10 sec had passed, which however never happened).

After the EEG experiment, participants were asked whether they noticed anything unusual about the sentences and, more concretely, whether the sentences had all been correct. Participants who did not notice any grammatical mistakes in the sentences (none of the native, but seven nonnative participants) were removed from the data set.

After removal of the EEG cap and additional electrodes, the L2 learners were given an offline questionnaire listing all target nouns in random order. They were asked to write down the correct singular definite determiner in front of each noun and to rate the certainty of their response on a 4-point scale. The questionnaire also briefly tested their knowledge of the plural determiner by asking participants to write down the plural forms of six given singular det + N phrases, half of them containing *de*-words and the other half *bet*-words. Finally, they filled in the language background questionnaire summarized in Table 1.

The complete experimental session took about 1.5–2 hr for Dutch and 2–2.5 hr for German participants.

## EEG Recording

The EEG was recorded using an elastic cap containing 27 passive tin electrodes (Electro-Cap International, Eaton, OH). The positions of electrodes are shown in Figure 1. Electrodes were also placed on both mastoids and on the forehead (between both eyes). The left mastoid electrode served as reference (and was later re-referenced to the average of right and left mastoid), the forehead electrode as the ground. Impedances for EEG electrodes were below 3 k $\Omega$ . The EOG was measured by two horizontal electrodes placed at the outer side of both eyes and two vertical electrodes above and below the right eye. Impedances for EOG electrodes were below 5 k $\Omega$ . The EEG and EOG signal was amplified (time constant = 8 sec, bandpass = 0.05–30 Hz) and sampled with a frequency of 500 Hz.

## EEG Data Analysis

The EEG and EOG signals were segmented into epochs from 100 msec before until 1000 msec after onset of each critical noun. The baseline was corrected based on the average EEG activity in the 100-msec interval before target onset as a reference. Blink detection and ocular correction were applied using the Gratton and Coles algorithm as implemented in Brain Vision Analyzer Version 1.05 (Gratton, Coles, & Donchin, 1983). Trials with amplitudes below  $-100 \mu\text{V}$  or above  $+100 \mu\text{V}$  in one or more EEG electrodes were removed semiautomatically, that is, after inspection (2.7% of critical trials).<sup>2</sup>

Because of the different time windows of effects for native and nonnative participants (see below), the two

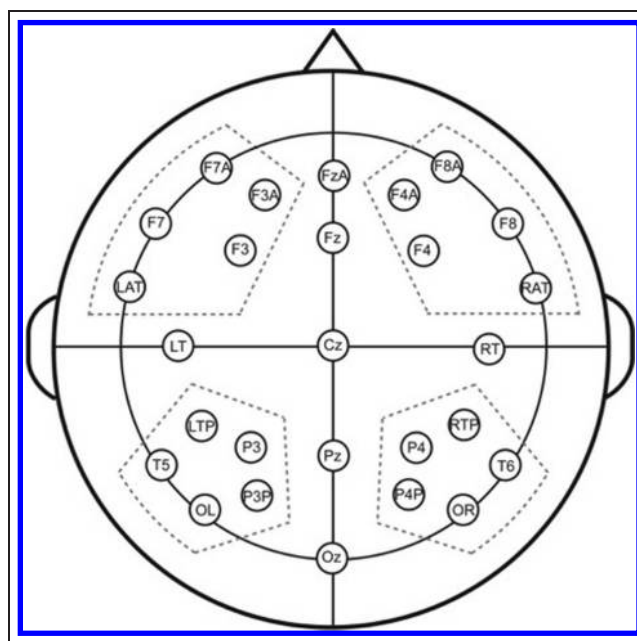


Figure 1. Positions of the electrodes on the EEG cap.

groups were analyzed separately. Data from the lateral electrodes (those included in dashed lines in Figure 1) were collapsed into quadrants and analyzed using repeated-measures ANOVA, with the factors Hemisphere (right vs. left), Region (anterior vs. posterior), and Correctness (correct vs. incorrect). The analysis of midline electrodes was dropped because its results were highly similar to those for lateral sites (apart from reduced statistical power because of the smaller number of electrodes). For the lateral analysis, significant interactions including the Correctness factor were followed up by planned simple effect ANOVAs to explore the nature of the effects. For all analyses, we report only significant effects concerning the experimental factor Correctness.

Because differences in the timing of ERP effects can be expected between native and nonnative speakers, we first identified the relevant time windows for each participant group by visual inspection and a time-course analysis involving  $t$  tests ( $p = .1$ ) in consecutive time windows of 50 msec for the four lateral quadrants. Intervals in which at least two adjacent 50 msec windows showed effects of Correctness of the same direction in the same quadrant were selected for further analysis.

## RESULTS

### Behavioral Results

The mean percentage of errors in the content questions was 2.7% for native speakers of Dutch (range = 0–8%,  $SD = 3.3\%$ ) and 4.4% (range = 0–16%,  $SD = 4.7\%$ ) for the L2 learners (in both cases, for those participants who entered the final analyses).

In the offline gender questionnaire that was given to the L2 learners only, the mean error rate for the included participants was 33.1% (range = 15–48%,  $SD = 7.1\%$ ), with most errors given to nouns that were gender-incompatible between Dutch and German (mean 59.0%) and relatively few errors on gender-compatible ones (mean 7.1%). All German participants produced the correct plural definite determiner in Dutch.

### ERP Results

The data of two participants per group had to be discarded because of a high percentage of artifacts or, in one case, technical failure, such that the final number of included participants was 19 for the native and 20 for the L2 learner group.

We will start with the description of results for the condition that served as a sort of control condition, the number agreement condition.

#### *Number Violation Condition, Native Speakers*

The ERP waves for native speakers in the number violation condition are shown in the top panel of Figure 2.

Visual inspection and the time-course analysis showed that there were two large time windows in which effects occurred: a window from 100 to 550 msec after noun onset where a negativity for incorrect sentences relative to correct ones was observed at anterior lateral sites and a window from 700 to 1000 msec where a positivity for incorrect sentences relative to correct ones occurred at posterior and a negativity at anterior sites.

For the 100–550 msec time window, there was a negativity that was present primarily at left anterior and, with marginal significance, also at right anterior sites (main effect Correctness:  $F(1, 18) = 3.21, p = .09$ ; Correctness  $\times$  Region [anterior vs. posterior]:  $F(1, 18) = 7.04, p = .016$ ; Correctness  $\times$  Region  $\times$  Hemisphere:  $F(1, 18) = 4.33, p = .052$ ; analyses of Correctness effects for all four quadrants following this latter interaction: left anterior:  $F(1, 18) = 11.24, p = .004$ ; right anterior:  $F(1, 18) = 4.07, p = .059$ ; left and right posterior: both  $F < 1$ ). This corresponds to the so-called LAN effect reported in the literature, although our effect is earlier and more sustained than in many other studies.

For the 700–1000 msec time window, the analyses revealed a negativity at anterior sites with a simultaneous positivity in posterior regions (interaction Region  $\times$  Correctness:  $F(1, 18) = 16.10, p = .001$ ; main effect Correctness, anterior region:  $F(1, 18) = 5.14, p = .036$ ; posterior:  $F(1, 18) = 5.68, p = .028$ ). The latter effect thus represents the classic P600 component in response to syntactic violations. None of the other effects involving Correctness were significant (all  $ps > .25$ ).

#### *Number Violation Condition, L2 Learners*

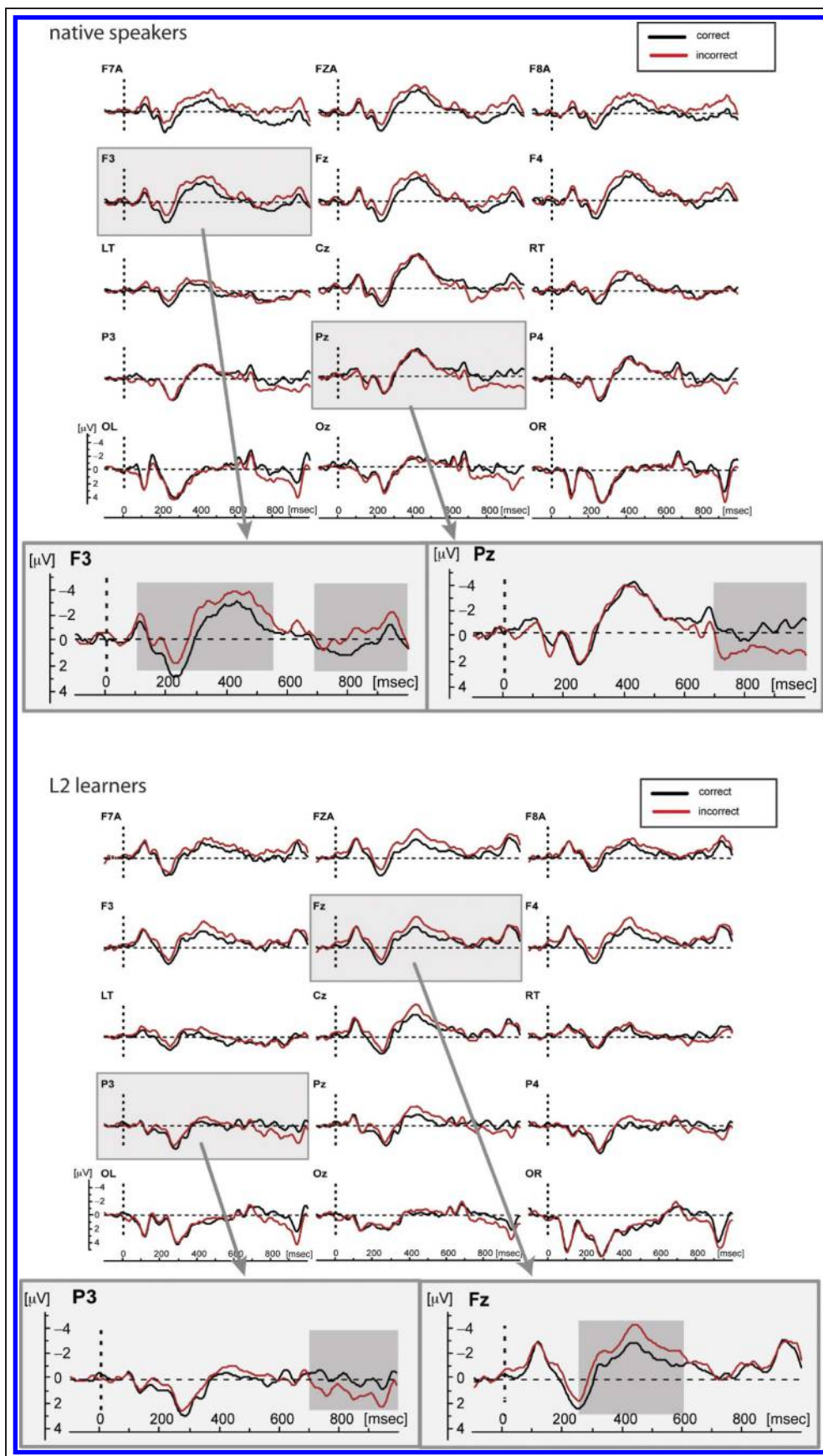
The grand-averaged waveforms of the L2 learners in the number violation condition are shown in the bottom panel of Figure 2. Visual inspection and the time-course analysis in L2 learners resulted in a similar picture as for the native speakers: There was an early window (between 250 and 600 msec after target noun onset) characterized by a negativity for incorrect relative to correct sentences. Furthermore, there was a later window starting at 700 msec and lasting until the end of the epoch, with a posterior positivity for agreement violations.

In the early window (250–600 msec), there was a marginally significant, negative-going main effect of Correctness,  $F(1, 19) = 4.06, p = .058$ , which was—somewhat different from the anterior effect in native speakers—broadly distributed, as there were no interactions with Region or Hemisphere (all  $ps > .30$ ).

The analysis for the 700–1000 msec time window showed a significant interaction of Region and Correctness,  $F(1, 19) = 9.04, p = .007$ , which was because of a positivity for incorrect sentences that occurred only in the posterior,  $F(1, 19) = 7.77, p = .012$ , but not in the anterior region ( $F < 1$ ). All other effects of Correctness were *ns* ( $p > .28$ ).



**Figure 2.** Grand-averaged ERP waveforms for the critical noun in the number agreement condition for native speakers (top) and L2 learners (bottom), for all midline and a subset of lateral electrodes. One representative electrode for each of the two observed effects, the early anterior negativity and the P600, is enlarged, and the time windows of analysis (100–550 msec and 700–1000 msec for native speakers and 250–600 msec and 700–1000 msec for L2 learners) are marked in gray for these electrodes.



### Common Analyses with Group as a Factor

Although a common analysis of both groups was complicated by the different time windows (in case of the early effect), we used the overlapping windows in which both groups had shown effects (250–550 msec and 700–1000 msec) to assess whether differences in the effects between the participant groups were statistically reliable. We conducted the same ANOVAs as before, but added Group as a factor. Although the effects of Correctness and/or Region  $\times$  Correctness were found again, there was no significant effect involving Group and Correctness in either of the two windows (250–550 msec: all  $ps > .12$ ; 700–1000 msec: all  $ps > .13$ ). Thus, statistically, the amplitudes and scalp distributions of the effects did not significantly differ from another (however, there was a latency difference of the early effect which is hard to test statistically).

### Summary of Number Violation Condition

Both L2 learners and native speakers of Dutch were sensitive to the number violation condition and displayed largely similar ERP patterns: The native speakers showed a

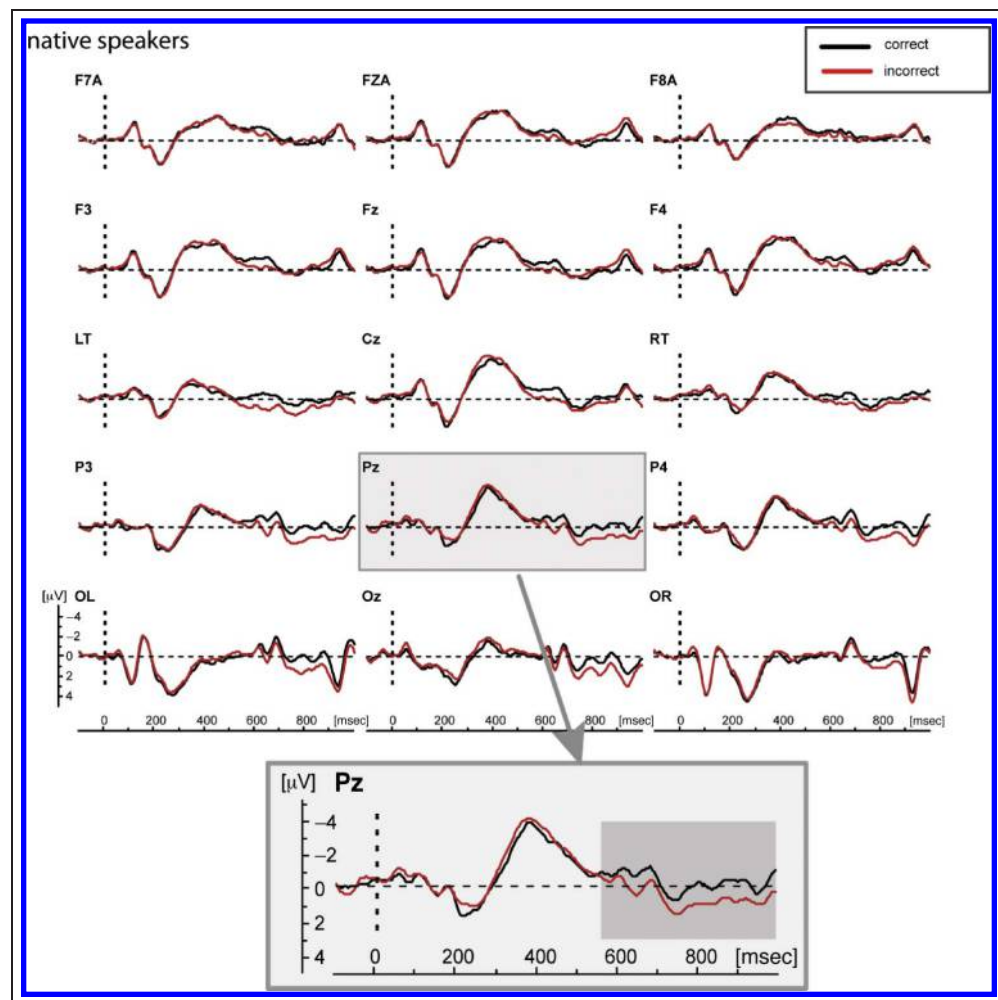
long-lasting anterior negativity between 100 and 550 and again between 700 and 1000 msec after onset of the critical plural noun as well as a posterior positivity starting at 700 msec. Given its polarity and scalp distribution, the latter component can be seen as the classical P600. In the L2 learners of Dutch, we observed a similar, but only marginally significant negativity, which however was delayed in comparison with the native group (250–600 msec), more broadly distributed, and which was less sustained than that in the native speakers. However, in a common analysis of the 250–550 msec window, these differences did not become significant in the form of an interaction with Group.

The positivity in the later window (the P600) was highly similar in both groups, with the same scalp distribution (across the posterior region) and latency (starting at 700 msec). Again, the common analyses did not reveal any interactions of Group with the experimental factor Correctness.

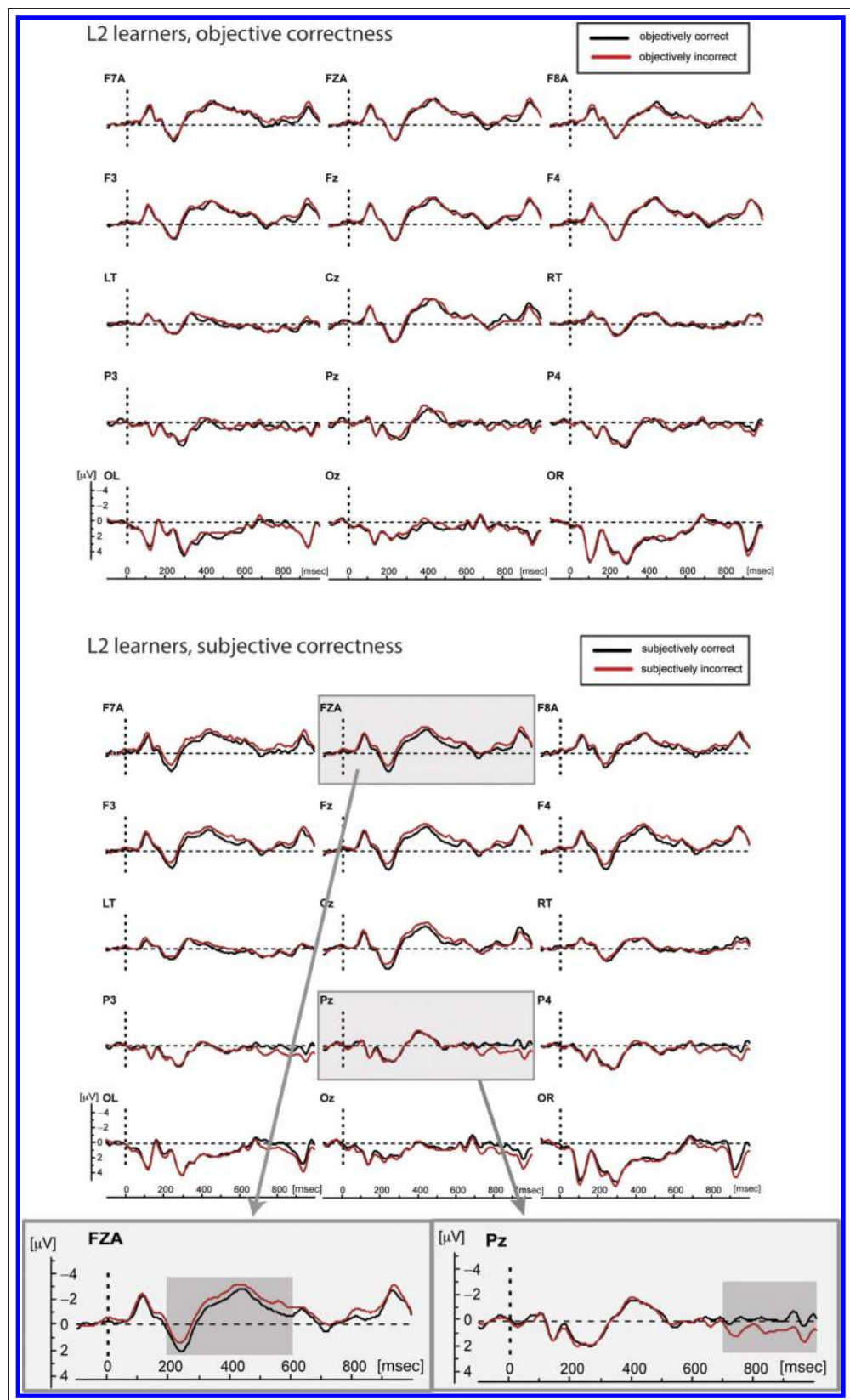
### Gender Violation Condition, Native Speakers

The ERP waves for native speakers in the gender violation condition are depicted in Figure 3.

**Figure 3.** Grand-averaged ERP waveforms for the critical noun in the gender agreement condition for native speakers, for all midline and a subset of lateral electrodes. A representative electrode (Pz) for the observed effect, the P600, is enlarged, and the time window of analysis (550–1000 msec) is marked in gray for this electrode.



**Figure 4.** Grand-averaged ERP waveforms for the critical noun in the gender agreement condition for objective correctness (top) and subjective correctness (bottom) for L2 learners, for all midline and a subset of lateral electrodes. There were no effects of objective correctness. In contrast, subjective correctness resulted in an early anterior negativity and a P600; one representative electrode for each of the two observed effects is enlarged, and the time windows of analysis (200–600 msec and 700–1000 msec) are marked in gray for these electrodes.



Visual inspection and the time-course analysis for native speakers revealed significant positive-going deflections for incorrect sentences relative to correct ones between 550 and 1000 msec after onset of a target noun,

primarily at posterior sites. This was confirmed by the statistical analysis (interaction Region  $\times$  Correctness:  $F(1, 18) = 6.17, p = .023$ ; Correctness, anterior region:  $p > .65$ , posterior region:  $F(1, 18) = 6.54, p = .020$ ;



all other interactions involving Correctness were *ns*, all  $ps > .13$ ).

Thus, in the gender violation condition, there was no early anterior negativity, but instead a P600 with a shorter latency (starting at 550 rather than at 700 msec) than in the number violation condition before.

### Gender Violation Condition, L2 Learners

We first analyzed the gender violation condition for L2 learners according to objective correctness, that is, we conducted the same analysis as for native speakers. The ERP waveforms of this contrast are shown in the top panel of Figure 4. Neither visual inspection nor the time-course analysis showed any ERP effects in at least two adjacent 50-msec windows.

Subsequently, we based our analyses on the subjective correctness of a given trial for a given participant. To this end, trials were re-sorted depending on each participant's responses in the offline determiner questionnaire, administered after the EEG experiment (see Table 3 for an example). As mentioned above, there was an average error rate of 33.1% in the questionnaire. For these incorrectly answered items—that is, nouns that were assigned the incorrect determiner—subjective correctness was reversed compared with objective correctness. Note that this procedure only exchanged the assignments of trials to the gender-correct and gender-incorrect conditions but did not change the number of trials in each condition. If, for example, a participant judged *\*bet radio* to be correct, the trial in which this article–noun combination was presented would be categorized as subjectively correct, whereas the trial containing the combination *de radio* would be categorized as subjectively incorrect for this participant.

The ERP waveforms for subjectively correct and incorrect phrases are depicted in the bottom panel of Figure 4. Visual inspection and the time-course analysis revealed consistent patterns of significant effects between 200 and 600 msec and between 700 and 1000 msec after noun onset. In the early time window between 200 and 600 msec, we again observed an anterior negativity (Region  $\times$  Correctness:  $F(1, 19) = 6.10, p = .023$ ; Correctness, anterior region:  $F(1, 19) = 9.42, p = .006$ ; posterior region:  $F < 1$ ; all other effects involving Correctness were *ns*, all  $ps > .17$ ).

The opposite pattern of results was obtained in the time window between 700 and 1000 msec, where we again obtained a P600, that is, a positivity in the posterior region (Region  $\times$  Correctness:  $F(1, 19) = 19.91, p = .000$ , Correctness, anterior region: no effects, all  $ps > .14$ , posterior region:  $F(1, 19) = 8.42, p = .009$ ; no other Correctness effects in the main analysis were significant, all  $ps > .30$ ).

### Common Analyses with Group as a Factor

Like in the number condition, we aimed at a statistical test of the observed differences between L1 and L2 speakers, especially in terms of the absence of the early effect in the native speakers. We therefore conducted common analyses with Group as a factor again and for the two time windows: the early window (200–600 msec), where at least the L2 speakers had shown effects of subjective correctness, and the 700–1000 msec window, where both groups had shown a P600. Note that these analyses involved the objective correctness factor for the native group, but subjective correctness in the case of L2 speakers, because we wanted to compare how L2 speakers use their subjective representations with how native speakers use their objectively correct ones.

In the first window (200–600 msec), there was indeed a significant interaction of Group, Region, and Correctness,  $F(1, 37) = 6.24, p = .017$ . Follow-up analysis confirmed that only L2 speakers displayed an effect of Correctness (see separate analysis above), whereas all Correctness effects were *ns* for native speakers (all  $ps > .30$ ). Thus, the difference of early effects between the two groups was statistically reliable. In contrast, the effects in the later window (700–1000 msec) were statistically indistinguishable between the two groups (all  $ps$  involving Group and Correctness  $> .10$ ).

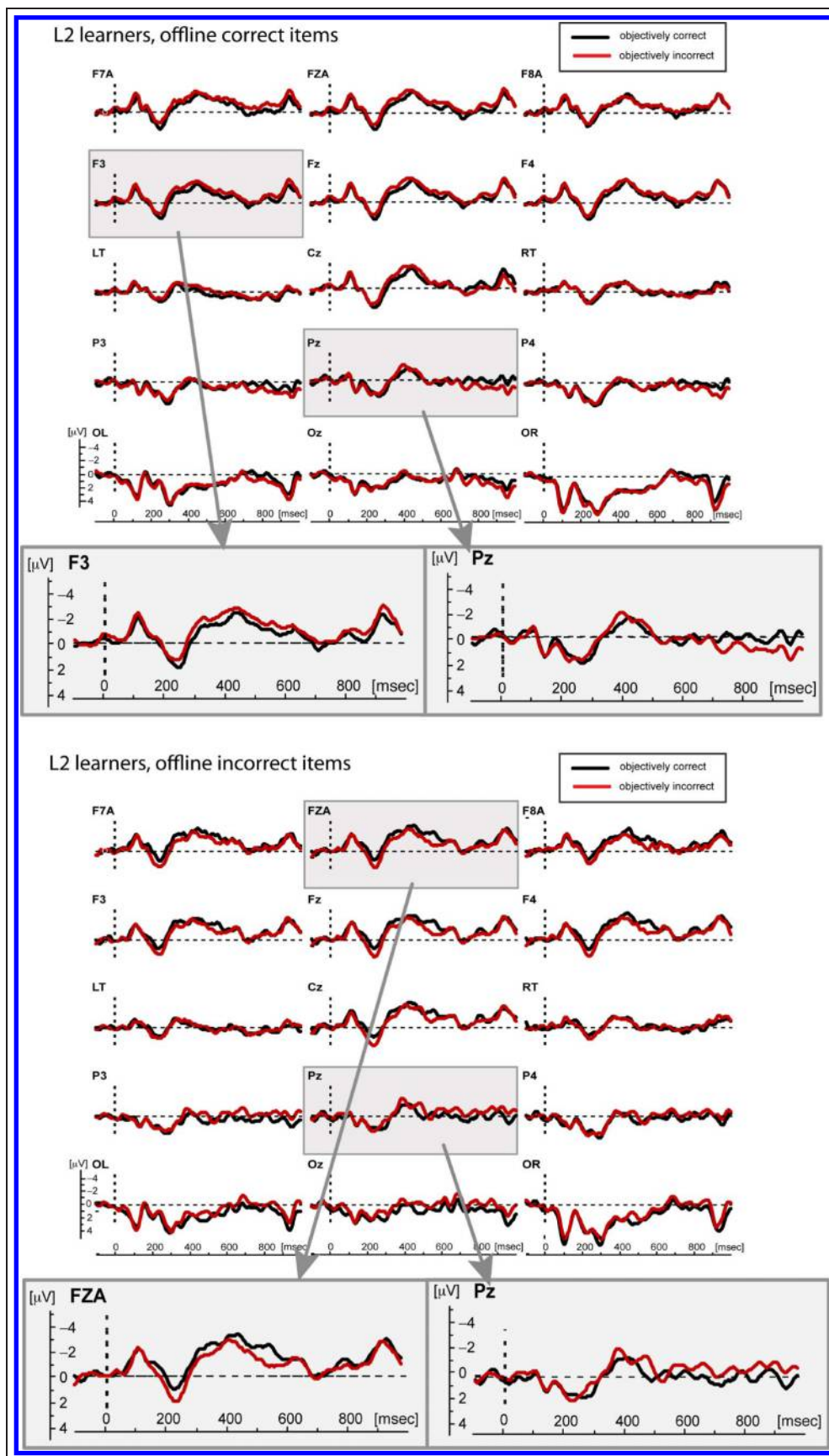
### L2 Speakers: Separate Analysis for Correctly and Incorrectly Represented Items

To get a better picture of what exactly caused the effects of objective correctness to disappear in the L2 learners and to be able to compare the effects to those reported in the literature, we examined the data more closely. For each participant, we split the items into those that were assigned the correct gender in the offline task (offline

**Table 3.** Examples for Subjective Correctness Values for Experimental Phrases Based on Offline Determiner Responses

Phrase in Experiment	Objective Correctness of Phrase	Example for Offline Determiner Response	Subjective Correctness of Phrase
<i>de tomaat</i> (the tomato)	correct	de tomaat	correct
<i>bet tomaat</i> (*the tomato)	incorrect		incorrect
<i>de radio</i> (the radio)	correct	*het radio	incorrect
<i>bet radio</i> (*the radio)	incorrect		correct

**Figure 5.** Grand-averaged ERP waveforms for the critical noun in the gender agreement condition for L2 learners, calculated separately for offline correct items (top) and offline incorrect items (bottom). Black lines represent objectively correct trials; red lines represent objectively incorrect trials. It can be seen that the effects reverse for offline incorrect items (bottom), because for these items, subjective correctness runs counter to objective correctness. Two representative electrodes are enlarged; again, an early anterior negativity and a later positivity can be observed. Because we did not run any statistical analyses on these data, we did not mark any time windows.





correct items) and those that were not (offline incorrect items). Offline correct items can be assumed to possess correct gender representations in an L2 speaker's lexicon; hence, for these items, objective and subjective correctness converge. The top panel of Figure 5 shows the comparison of (objectively and subjectively) correct versus incorrect trials for these items. Note that this comparison, which includes about two thirds of the items, corresponds to the "standard" analysis in those L2 studies where incorrect offline responses (e.g., on grammatical judgments) are excluded from analysis (Foucart & Frenck-Mestre, 2011; Gillon Dowens et al., 2009, 2011). It is clearly visible from the figure that the same effects are observed as we reported earlier for "subjective correctness" in general, that is, an early (from around 200 msec) anterior negativity and a later posterior positivity (around 700–1000 msec). However, whether these effects arise from objective or subjective correctness cannot be decided here, as the two are fully overlapping.

In contrast to other studies where the remaining data are typically excluded, a substantial number of offline incorrect items (33%) allowed us to analyze the data for these items as well, that is, the cases for which objective and subjective correctness diverge. Because they were assigned the wrong gender offline (and thus probably possess an incorrect gender representation), what is perceived as subjectively correct is actually objectively incorrect, and the other way around. Although this data subset should be treated with some caution, because it contains, on average, only half of the items of the set of offline correct items above (and is therefore a lot noisier), the bottom panel of Figure 5 gives an indication of what is happening. In the figure, the color coding is based on objective correctness (black lines show objectively correct trials, whereas red lines represent objectively incorrect ones). It can be seen that the effects reverse, namely, objectively incorrect trials now show (trends toward) an early anterior positivity and a later posterior negativity.<sup>3</sup> This data pattern makes sense when recalling that, for these items, subjective correctness is reversed relative to objective correctness: What is observed are the same LAN and P600 effects as before based on subjective correctness. Thus, these data provide further evidence for subjective, and not objective, correctness driving syntactic agreement processing in L2 speakers.

### *Summary of Gender Violation Condition*

The data of the gender violation condition showed a posterior P600 for native speakers. The latency of this effect was shorter than for the number violation, starting at 550 rather than at 700 msec after onset of the critical noun. Furthermore, the early (and sustained) negativity found for number violations was absent for gender violations in native speakers.

L2 learners showed no effects of objective correctness at all. However, after re-sorting trials according to subjective correctness, subjective violations of word gender resulted in a similar biphasic pattern as the one we had already observed in the number violation condition: an anterior negativity between 200 and 600 msec and a posterior-located P600 between 700 and 1000 msec. Additional analyses for correctly and incorrectly represented items (based on offline gender assignments) showed that the standard effects of objective correctness reversed in the case of incorrectly represented nouns, which explains why collapsing across these two item types had resulted in null effects for objective correctness before.

## **DISCUSSION**

This study investigated the issue of syntactic processing by second language learners from a new point of view, namely, that of the role of subjective, idiosyncratic, and often incorrect representations. To this end, we used violations of gender agreement between determiner and nouns, as processed by German learners of Dutch (and native speakers of Dutch as a control group), informed by previous results showing robust and systematic gender errors in this population (Lemhöfer et al., 2008, 2010). Previous EEG studies of L2 gender processing have focused on high-proficiency populations with very low error rates concerning gender, whereas we chose to look at an L2 learner group of intermediate proficiency, with high error rates for "gender-difficult" nouns.

We used both the conventional, objective correctness definition as well as a new subjective correctness definition that expressed for each determiner–noun phrase and each L2 participant separately whether a phrase was subjectively correct or incorrect. Before we turn to the results of the gender agreement conditions, we will briefly discuss the results of the additional number agreement manipulation.

### **Number Agreement Condition**

To test whether our population of L2 learners was sensitive to determiner–noun agreement violations at all, we included a number agreement condition. Violations of number agreement should be especially obvious to German learners of Dutch, because the two languages are highly similar with respect to the use of plural determiners. Indeed, as expected, both the native speakers and the L2 learners responded similarly to violations of number agreement between determiners and nouns. Both groups showed an early, anterior, negative effect (in the literature often referred to as LAN), followed by a posterior P600. The onset of the LAN effect was delayed by about 150 msec in the L2 group, something that is frequently observed for ERP effects in L2 learners, although usually for later effects like N400 and P600 (e.g., Rossi et al., 2006; Weber-Fox & Neville, 1996). The

early negativity was descriptively less anteriorly distributed in L2 speakers than in the native speakers, something that has been observed before for L2 speakers of high, but not yet native-like proficiency (see Steinhauer et al., 2009, for an overview).

In contrast to these slight differences concerning the early component, the P600 occurred with a similar latency, amplitude, and distribution in both groups. Note that often, smaller or delayed P600 effects have been found for L2 learners of less than native-like proficiency (Steinhauer et al., 2009). The fact that our two groups displayed almost identical P600 effects is thus in line with our expectation that German learners of Dutch would process number agreement in a very similar way to native speakers of Dutch because of the high overlap between German and Dutch with respect to this feature.

### **Gender Agreement Condition and the Role of Incorrect Subjective Representations**

The crucial comparison in our study concerned gender agreement. Our results show that our L2 learners had syntactic gender representations that differed from those of native speakers in one third of the cognate nouns of our stimulus set. As evident from our ERP results, L2 learners do use these subjective, sometimes incorrect representations for syntactic processing.

More precisely, the following pattern of results was obtained. When analyzing the ERP data from the gender agreement condition in the “conventional” way, that is, comparing objectively correct determiner–noun phrases to objectively incorrect ones, the group of L2 learners did not show any effects. In contrast, for the same materials, the native speakers showed the most commonly reported ERP component for grammatical violations, the P600. Thus, on the basis of these data, one might have concluded that the L2 learners were insensitive to grammatical gender information. This would be consistent with the idea of more “shallow,” that is, less syntactically driven processing (cf. the “good enough” account by Ferreira, 2003) in L2 compared with native speakers (see Clahsen & Felser, 2006) and with behavioral studies suggesting that L2 speakers may not make use of word gender (Lew-Williams & Fernald, 2010; Scherag, Demuth, Rösler, Neville, & Röder, 2004; Guillelmon & Grosjean, 2001). However, the recategorization of trials in terms of subjective correctness for the L2 learners showed a different picture: Subjectively unexpected determiners triggered an ERP response on the subsequent noun. Most importantly, this ERP response included the P600 effect, the effect that was also shown by the native speakers as well as by both native speakers and L2 learners in the number agreement condition.

A set of additional analyses illustrated in more detail how the null effect for objective correctness had come about. For offline correct items, that is, nouns for which subjective and objective correctness were identical, the

standard violation response (here, an early negativity and a P600) was obtained. Crucially, though, the effect of objective correctness was reversed when looking at incorrectly represented nouns (the offline incorrect items): These trials showed an early positivity and a late negativity for objectively incorrect trials. In other words, objectively incorrect trials behaved as correct ones and vice versa, which corresponds exactly to how these items were represented subjectively. Thus, it is subjective correctness that determined the ERP effects. As a consequence, when collapsing across these two item types in the analysis of objective correctness, the effects cancelled each other out.

The pattern of effects that was shown by the L2 speakers for subjective correctness was similar to the one obtained in the number agreement condition: An early, anterior negativity was followed by a P600. In contrast, in the native speakers, there was no early negativity and a P600 with a shorter latency (550 msec) than in all other conditions of this study (700 msec). The fact that the L2 speakers, but not the native speakers, showed a LAN effect for gender agreement violations is in contradiction with Steinhauer et al.’s (2009) observation that a missing LAN has often been found in L2 learners who have not (yet) reached near-native proficiency. This observation makes sense when considering that the LAN is usually thought to reflect automatic processes that might be missing in an L2 acquired fairly late in life. In contrast, the P600 is believed to indicate more strategically driven processes of repair and reanalysis (O’Rourke & Van Petten, 2011; Rossi et al., 2006; Friederici, 2002), which might be more readily available to late L2 learners.

At this point, we are uncertain why the native speakers showed a LAN only for number, but not for gender agreement, whereas L2 speakers showed it in both cases. One possible explanation is that there might be differences in Dutch as opposed to German gender agreement processing—note that Sabourin and Stowe (2008) and Hagoort (2003) also did not observe a LAN for gender agreement violations between determiner and nouns (in midsentence position) in Dutch native speakers, whereas Gunter et al. (2000) did observe a LAN for similar gender violations in Germans. If this was the case, our German participants might have transferred their German “processing mode” to Dutch. However, this is a speculative explanation, as there are not enough ERP studies on native German and Dutch gender agreement to support this claim. It should however be noted that when exactly LAN effects do or do not occur in syntactic agreement violations, both in L1 and L2, is not yet understood at all (for an overview, see Bornkessel-Schlesewsky & Schlesewsky, 2009, pp. 117–123). A second possible reason for the observed difference between L1 and L2 speakers concerns the fact that, in contrast to our number agreement condition, the exact composition of trials within the gender conditions was different between groups, because we compared effects of objective correctness for the native group with those of subjective correctness for L2 speakers.

Thus, in summary, the central finding in our data is that both native speakers and L2 learners were sensitive to deviations from their own representations, as indexed by a P600. For native speakers, these deviations concern the objectively correct, canonical gender of nouns, and for the L2 learners, these deviations concern the subjective, sometimes incorrect gender. The two groups differed with respect to the occurrence of a LAN effect, which unexpectedly was present for L2 speakers only. This is especially remarkable, as the native speakers did show this LAN effect for violations of number agreement. Note, however, that both obtained patterns, the LAN + P600 combination and the P600 only, are patterns that have been observed for agreement violations in native speakers before. Thus, the L2 speakers did not show an ERP pattern that is atypical of native speakers (as in, e.g., McLaughlin et al., 2010; Hahne & Friederici, 2001), although it differed from the specific ERP pattern of our native group to a larger degree than was the case for the number agreement condition.

One additional interesting question concerning subjective L2 representations is in how far they are truly idiosyncratic (i.e., varying from speaker to speaker) or whether they are rather uniform for a given group of L2 speakers. If the latter holds true, then re-sorting the trials per every individual participant might not be necessary; it would suffice to determine which items are generally most affected by incorrect representations. To clarify this point, we looked more closely at the data of the offline gender assignments for the gender-incompatible items. Because of L1 transfer, these items had a much higher probability of errors (59% on average) than the compatible ones (7%). To assess how uniform the tendency toward errors was within this category, we computed the mean of all correlations between any two L2 participants for the offline responses to these items, which was only small to moderate and nonsignificant ( $r = .21$ ). Thus, which specific nouns within the “difficult” category would be incorrectly represented varied greatly between our German learners, justifying the idea that the subjective representations are “idiosyncratic.”

### **Implications for L2 Syntactic Processing Research**

Our results have important general implications for research on L2 syntactic processing. They illustrate in detail that an apparent null effect of (objective) syntactic correctness was in fact the result of opposite effects from correctly and incorrectly represented items. That is, what would subsequently be interpreted as a lack of sensitivity to the investigated syntactic feature in L2 speakers might in fact be a disguised sensitivity to subjective correctness.

It is important to note, however, that a major impact of incorrect subjective representations is not to be expected for all L2 speaker populations. In fact, as proficiency approaches the near-native level, subjective correctness will

converge with objective correctness, just as is the case for native speakers. Indeed, previous ERP studies that reported effects of what we call objective correctness of grammatical gender (Foucart & Frenck-Mestre, 2011; Gillon Dowens et al., 2009, 2011) used high-proficient L2 speakers with almost native-like mastery of the L2 gender system (less than 6% errors) as their participant population, such that incorrect subjective representations of grammatical gender did presumably not play an important role. Our data complement their findings with the observation that, in a population of L2 learners who are still in the process, rather than at the end, of acquiring their second language, gender agreement processing can be largely based on idiosyncratic and partially incorrect subjective representations.

Besides grammatical gender processing, other domains of L2 syntactic processing may also be affected by the occurrence of incorrect syntactic representations. The degree to which these representations might play a major role for a given set of materials and syntactic constructions is, plausibly, indicated by error rates in the often-used grammaticality judgment task. Although these error rates are usually negligible in native speakers, L2 speakers are often much worse at judging whether a sentence is grammatical or not, raising the possibility of incorrect grammatical representations. For example, in a study by Kotz, Holcomb, and Osterhout (2008), early Spanish L2 speakers of English displayed accuracy rates in grammaticality judgments of only 56% on long sentences containing reduced relative clauses. As in other L2 studies (Pakulak & Neville, 2011; Chen et al., 2007; Ojima et al., 2005; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996), sentences with correct and incorrect grammaticality judgments were grouped together in the analyses in that study. In our view, collapsing across sentences that are perceived as correct and as incorrect into the same conditions might lead to inconsistent results, in particular when they are compared with those of native speakers (for whom subjective and objective correctness converge). Thus, the important role of subjective correctness in L2 syntactic processing shown by the present results suggests that ERP data of L2 speakers should not be averaged across stimuli that are perceived as correct and incorrect.

Some of the studies that chose not to exclude incorrectly judged sentences from their analysis assume that ERP data in L2 syntactic processing reflect some “implicit” syntactic knowledge even when “explicit” knowledge is absent, for example, when judgment performance is at or near chance level (Kotz et al., 2008; Tokowicz & MacWhinney, 2005). In contrast, a number of recent studies on L2 syntactic acquisition rather suggest a parallel course of behavioral and neural correlates of syntactic discrimination ability (Davidson & Indefrey, 2009, 2011). Although we do not want to deny that a certain “gut feeling” might contribute to syntactic processing in a second language, these and our results clearly show that the role of implicit knowledge is fairly limited when compared

with that of subjective, but explicit, representations. In our data, the decisive criterion was whether a phrase matched a participant's explicit expectation or not, rather than its objective correctness. We therefore claim that there is, in general, a tight coupling—rather than a dissociation—between L2 learner's behavioral responses and their ERP patterns in syntactic processing.

In our case, the source of subjective, incorrect L2 gender representations is presumably incorrect L1–L2 transfer (as indicated by especially high error rates for gender-incompatible nouns between Dutch and German). However, this does not mean that the effects we observed are restricted to the present population and materials. First, L1–L2 transfer effects are a frequent, if not the most frequent, source of L2 grammatical errors and L1–L2 processing differences also in syntactic domains other than word gender (e.g., Antón-Méndez, 2011; Jegerski, VanPatten, & Keating, 2011; Ionin & Montrul, 2010; Hertel, 2003; Montrul, 1999). Hence, there are numerous syntactic phenomena that might be affected by incorrect syntactic representations in L2 originating from L1 influences. Second, incorrect representations also occur in the absence of direct L1–L2 conflicts. English learners of gender languages, for example, develop incorrect gender representations despite the fact that the grammatical gender feature is absent in their L1. Although, strictly speaking, our data leave open the possibility that incorrect representations that are not because of L1 transfer are too weak to affect L2 syntactic processing, it seems important for future research on any kind of L2 syntactic processing to consider and control for the previously ignored possibility of these idiosyncratic, partially incorrect representations in L2 speakers. To avoid the unnatural task of grammaticality judgments during the EEG measurement, this is best accomplished by an additional offline assessment of each participant's correctness perception of each experimental item. These offline data can then be used to either exclude “misrepresented” items (if not too numerous) or to recategorize them in a way demonstrated in this study. Only then will it be “fair” to compare L1 and L2 groups directly with respect to their sensitivity to syntactic information.

To conclude, this study shows that intermediate-level second language learners use their partially incorrect, idiosyncratic second language grammar during syntactic processing and that they use it in a similar way as native speakers use their correct representations.

### Acknowledgments

This research was made possible by a “veni” grant from the Netherlands Organization for Scientific Research (NWO) to Kristin Lemhöfer (dossier no. 016.084.015). We thank Julia Lennertz and Yvonne Maas for their great help in the EEG laboratory and Hubert Voogd and Jos Wittebrood from the DCC technical group for their technical support.

Reprint requests should be sent to Kristin Lemhöfer, Donders Institute for Brain, Cognition, and Behaviour—Centre for Cognition,

Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands, or via e-mail: k.lemhofer@donders.ru.nl.

### Notes

1. This comparably slow pacing was used, first, because L2 speakers are somewhat slower readers and, second, because Dutch (compound) words can be very long (13–15 letters is not unusual).
2. We removed trials with one or more peaks above or below the mentioned limits; trials in which the high amplitude was merely because of a gradual shift of the signal were not removed.
3. We refrained from statistically analyzing these two separate comparisons because of the unequal number of trials and a generally reduced statistical power relative to the full analysis of subjective or objective correctness.

### REFERENCES

- Antón-Méndez, I. (2011). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, *14*, 318–331.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (Release 2) [CD-ROM]*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*, 137–153.
- Birdsong, D. (2004). Second language acquisition and ultimate attainment. In A. Davies & C. Elder (Eds.), *Handbook of applied linguistics* (pp. 82–105). London: Blackwell.
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2009). *Processing syntax and morphology: A neurocognitive perspective*. Oxford, UK: University Press.
- Brodeur, M. B., Debrulle, J. B., Renoult, L., Prevost, M., Dionne-Dostie, E., Buchy, L., et al. (2011). The influence of contour fragmentation on recognition memory: An event-related potential study. *Brain and Cognition*, *76*, 115–122.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*, 161–174.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.
- Davidson, D. J., & Indefrey, P. (2009). Plasticity of grammatical recursion in German learners of Dutch. *Language and Cognitive Processes*, *24*, 1335–1369.
- Davidson, D. J., & Indefrey, P. (2011). Error-related activity and correlates of grammatical plasticity. *Frontiers in Psychology*, *2*, 219.
- Dewaele, J.-M., & Véronique, D. (2001). Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition*, *4*, 275–297.
- Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 889–906.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*, 379–399.



- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78–84.
- Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish—Evidence from event related potentials. *Neuropsychologia*, 49, 1651–1659.
- Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2009). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22, 1870–1887.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, 55, 468–484.
- Guillemon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory and Cognition*, 29, 503–511.
- Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, 12, 556–568.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–899.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- Hertel, T. J. (2003). Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research*, 19, 273–304.
- Ionin, T., & Montrul, S. (2010). The role of L1 transfer in the interpretation of articles with definite plurals in L2 English. *Language Learning*, 60, 877–925.
- Jegerski, J., VanPatten, B., & Keating, G. D. (2011). Cross-linguistic variation and the acquisition of pronominal reference in L2 Spanish. *Second Language Research*, 27, 481–507.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, 109, 68–74.
- Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and nonnative readers of English. *Acta Psychologica*, 128, 514–527.
- Lemhöfer, K., Schriefers, H., & Hanique, I. (2010). Native language effects in learning second-language grammatical gender: A training study. *Acta Psychologica*, 135, 150–158.
- Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, 59, 312–330.
- Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and nonnative Spanish speakers. *Journal of Memory and Language*, 63, 447–464.
- Maier, M. E., Yeung, N., & Steinhauser, M. (2011). Error-related brain activity and adjustments of selective attention following errors. *Neuroimage*, 56, 2339–2347.
- Martin-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research*, 1093, 178–189.
- McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, 21, 395–423.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., et al. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, 60, 123–150.
- Montrul, S. (1999). Causative errors with unaccusative verbs in L2 Spanish. *Second Language Research*, 15, 191–219.
- Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, 24, 933–947.
- Münste, T. F., Szentkuti, A., Wieringa, B. M., Matzke, M., & Johannes, S. (1997). Human brain potentials to reading syntactic errors in sentences of different complexity. *Neuroscience Letters*, 235, 105–108.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, 17, 1212–1228.
- Orgassa, A., & Weerman, F. (2008). Dutch gender in specific language impairment and second language acquisition. *Second Language Research*, 24, 333–364.
- O'Rourke, P. L., & Van Petten, C. (2011). Morphological agreement at a distance: Dissociation between early and late components of the event-related brain potential. *Brain Research*, 1392, 62–79.
- Pakulak, E., & Neville, H. J. (2011). Maturation constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, 23, 2752–2765.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, 18, 2030–2048.
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, 24, 397–430.
- Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93, B97–B108.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129–158.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 10, 209–231.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, 25, 13–41.
- Stenberg, G., Johansson, M., & Rosen, I. (2004). Using a second language: Semantic priming effects on the N400. *Psychophysiology*, 41, S92.
- Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173–204.
- Weber-Fox, C., & Neville, H. J. (1996). Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256.
- White, L. (2003). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, 6, 129–141.
- 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. *Journal of Cognitive Neuroscience*, 16, 1272–1288.