



**HAL**  
open science

## Second language processing: when are first and second languages processed similarly?

Laura Sabourin, Laurie A. Stowe

► **To cite this version:**

Laura Sabourin, Laurie A. Stowe. Second language processing: when are first and second languages processed similarly?. *Second Language Research*, 2008, 24 (3), pp.397-430. 10.1177/0267658308090186 . hal-00570749

**HAL Id: hal-00570749**

**<https://hal.science/hal-00570749v1>**

Submitted on 1 Mar 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Second language processing: when are first and second languages processed similarly?

**Laura Sabourin** and **Laurie A. Stowe** *School of Behavioral and Cognitive Neuroscience, University of Groningen*

Received March 2007; revised November 2007; accepted December 2007

In this article we investigate the effects of first language (L1) on second language (L2) neural processing for two grammatical constructions (verbal domain dependency and grammatical gender), focusing on the event-related potential P600 effect, which has been found in both L1 and L2 processing. Native Dutch speakers showed a P600 effect for both constructions tested. However, in L2 Dutch (with German or a Romance language as L1) a P600 effect only occurred if L1 and L2 were similar. German speakers show a P600 effect to both constructions. Romance speakers only show a P600 effect within the verbal domain. We interpret these findings as showing that with similar rule-governed processing routines in L1 and L2 (verbal domain processing for both German and Romance speakers), similar neural processing is possible in L1 and L2. However, lexically-driven constructions that are not the same in L1 and L2 (grammatical gender for Romance speakers) do not result in similar neural processing in L1 and L2 as measured by the P600 effect.

**Keywords:** L2 processing, ERP, P600, grammatical gender, L1 transfer, L1 interference

### I Introduction

Grammatical gender, also called noun class, is a lexical feature on nouns that is reflected at the level of syntax in terms of agreement both within the noun phrase and throughout the sentence. As outlined in the introduction to this special issue, the acquisition of a second language

Address for correspondence: Laura Sabourin, Linguistics Department, University of Ottawa, 70 Laurier Avenue E., Ottawa, ON K1N 6N5, Canada; email: lsabour3@uottawa.ca

(L2) grammatical gender system is very difficult. It is not only difficult for speakers of languages without grammatical gender (e.g. English) to acquire an L2 gender system, but acquiring L2 gender also seems to be difficult for speakers with gender in their first language (L1). However, L2 speakers do seem relatively proficient at knowing which gender class nouns belong to; they just have difficulty applying the correct gender rules. In order to investigate L2 grammatical gender comprehension we use event-related brain potentials (ERPs) to investigate the on-line processing of Dutch as an L2 by participants with either German or a Romance language as their L1.

Research on language processing has provided many insights into how the native language is processed. Language processing unfolds on-line at a very quick rate and any true attempt at fully understanding how language works must investigate the on-line time-course of different aspects of language. The electrophysiological measure of event-related brain potentials is widely used to address just these issues. ERPs provide a continuous on-line account of the brain's electrical activity, which is highly sensitive to manipulations of aspects of language processing difficulty (Osterhout and Holcomb, 1995). With this technique we have been able to considerably extend our knowledge about how and when various aspects of language input are processed in the brain. However, relatively little is known about L2 processing in the brain. This article explores the on-line neural processing of an L2 syntactic system.

The traditional way of examining L2 ability has been to investigate production or comprehension using off-line grammaticality judgement tasks. Both of these methods have drawbacks. In production tasks many things can affect performance other than linguistic knowledge (or lack thereof). For instance, working memory limitations may affect performance (Ardila, 2003), and such factors as attention and fatigue – which can greatly affect the performance of native speakers – are even more likely to affect L2 speakers. Using grammaticality judgement tasks to study comprehension provides an adequate measure of what L2 speakers know, but are limited in providing information on how grammatical information is used in comprehension. That is to say, it is not clear whether L2 participants perform this kind of task in the same way as native speakers do, which is necessary to demonstrate that they have

the same sort of knowledge of the language. This is where using the ERP methodology can help us gain insights into L2 processing that were not possible with more traditional methods for studying L2 (for reviews on the ERP methodology as used in L2 research, see Mueller, 2005; Stowe and Sabourin, 2005).

The experiments presented in this article investigate L2 processing by examining brain responses to syntactic violations as compared to the response to their grammatical counterparts. Two components in the ERP waveform have been shown to be associated with syntactic processing elicited by syntactic violations and increased processing difficulty in L1. These are the (early) left anterior negativity ((E)LAN; e.g. Gunter *et al.*, 2000) and the P600 (e.g. Osterhout and Holcomb, 1992) which is also referred to as the Syntactic Positive Shift (SPS) by some researchers (e.g. Hagoort *et al.*, 1993). The LAN is a negative shift in the waveform, which is maximum around 400 ms after the presentation of an ungrammatical word. This is in a latency range similar to the N400 component (Kutas and Hillyard, 1980), which is elicited by semantic processing difficulties but it has a more anterior scalp distribution with a left hemisphere bias (Kluender and Kutas, 1993; Friederici *et al.*, 1996). Scalp distribution, together with polarity, indicate whether two effects can or cannot be generated by the same population of neurons in the brain. LAN effects have been reported in response to violations of word-category constraints (Münte *et al.*, 1993; Friederici *et al.*, 1996). Similar early negativities have been found with number, case, gender and tense mismatches as well (Münte *et al.*, 1993). A LAN effect with a similar scalp distribution sometimes occurs in an earlier time window, between 125 and 180 ms (Neville *et al.*, 1991; Friederici, *et al.*, 1996), generally in response to unambiguous and easily identifiable syntactic-category violations (Gunter *et al.*, 1999).

The P600, which is the focus of the current study, is a positive deflection in the brain wave that has an onset at around 500 ms and reaches its maximum 600 ms after presentation of an ungrammatical or syntactically unexpected word. P600 effects have been reported in different languages for a range of syntactic violations, including phrase-structure violations (Neville *et al.*, 1991; Osterhout and Holcomb, 1992; Hagoort, *et al.*, 1993), subcategorization violations (Osterhout and Holcomb, 1992; Osterhout *et al.*, 1994; Ainsworth-Darnell *et al.*, 1998) and violations in

the agreement of number, gender and case (Hagoort, et. al., 1993; Münte, et. al., 1993; Osterhout, 1997; Coulson, King and Kutas, 1998). The P600 effect is also present for the less preferred reading of a syntactically ambiguous sentence (Osterhout *et al.*, 1994; Friederici *et al.*, 1996; Kaan and Swaab, 2003). It has also been reported for cases in which syntactic integration is difficult (for resolving *wh*-questions, see Phillips *et al.*, 2005; for agreement over a distance, see Kaan *et al.*, 2000). Although syntactic violations, syntactic preference violations and syntactic complexity elicit a P600 within approximately the same latency range, differences in their scalp distributions have been reported. Syntactic violations tend to show a posterior maximum while preference violations tend to be more frontally distributed (Hagoort, Brown and Osterhout, 1999; Friederici *et al.*, 2002; Kaan and Swaab, 2003), suggesting some difference in the brain sources underlying these effects. There is also some indication that the P600 effects found in sentence processing may not reflect syntactic processing *per se* but should be regarded more as indexing strategically controlled processes, such as attempting to reconstruct meaning from an ungrammatical or incoherent sentence (Gunter and Friederici, 1999; Hoeks *et al.*, 2004). Despite the somewhat indirect nature of the response, the P600 associated with ungrammaticality is useful for tracking the native-likeness of language processing since:

- it is very robust in native speakers unlike the earlier ERP responses; and
- it can only be elicited if the syntactic structure is being built in a fairly native-like way and thus fails in a native-like fashion.

In the past 10 years researchers have begun to investigate the extent to which these syntactic effects occur in L2 processing and what factors are most important in determining whether they occur, such as age of acquisition and proficiency. Weber-Fox and Neville (1996) investigated the role of experience in shaping brain organization for language in Chinese–English bilinguals. While native English speakers showed both a LAN and a P600 to word-category violations, L2 learners showed a different pattern depending on the age at which the L2 was learned. A native-like LAN and P600 were found for all participants who had acquired English before age 11 (although with some delays compared to native speakers), while participants who had learned

English at a later age showed an anterior negativity that was bilaterally distributed rather than unilaterally. Participants who learned English between the ages of 11 and 13 showed a delayed P600-like effect (starting at about 700 ms), while those who had acquired English later than age 16 showed no positivity at all within the time-window analysed.

The results of this study suggest that late acquirers of an L2 cannot process their L2 in a native-like way, while early acquirers do. However, age of acquisition and the level of proficiency both differed for the groups tested in this study, and it is not clear which is actually more important. Furthermore, the results are limited to a single L1 that differs considerably from the target L2. Conversely, these results suggest that early learners will show relatively native-like processing. Although processing is more native-like if the L2 was learned at an earlier age, Proverbio *et al.* (2002) – who compared early high proficient Italian–Slovenian bilinguals with Italian monolinguals – suggested that brain activation patterns for language processing in bilinguals may still differ from monolingual speakers due to changes in the functional organization of the brain as a direct result of acquiring more than one language.

Whether later learners show a P600 effect or not varies between studies. Hahne (2001) and Hahne and Friederici (2001) investigated phrase-structure processing in late L2 learners of German. The native German speakers in these studies showed both ELAN and P600 effects to unambiguous syntactic-category violations. Late Japanese learners of German, unlike the native speakers, did not show an ELAN (Hahne and Friederici, 2001). The Japanese learners also did not show a P600 effect to ungrammatical sentences relative to grammatically correct sentences. Rather, grammatical sentences showed a P600-like positivity when compared to the native speakers. Hahne and Friederici suggest that the greater positivity found in response to the grammatical sentences may be due to a need for greater recruitment of processing capacities even for correct sentences, and thus lead to a ceiling effect for the ungrammatical sentences.<sup>1</sup> In contrast to the Japanese learners, somewhat more proficient late Russian learners of German showed a

<sup>1</sup> Another interpretation of their results is a greater recruitment of memory resources for making judgements on ungrammatical sentences; see Sabourin and Stowe (2004) and the discussion presented in this article after Experiment 2.

P600 effect to the ungrammatical sentences, although like the Japanese learners they did not show an ELAN (Hahne, 2001). The results of this study show that later learners of an L2 can show at least a native-like P600, unlike the results of the Weber-Fox and Neville study, but they do not clarify under which circumstances this effect is seen. All three studies employed a syntactic-category violation, so it is not likely that the nature of the violation is the most important factor. However, this cannot be dismissed out of hand, since the target L2s were different. In the last two studies, the groups differed in proficiency as well as L1 so that it is not clear which of these factors may be more important.

A point that is also worth making at this juncture is that the majority of the studies that have examined L2 processing to date have used very similar syntactic violations: category violations. This has been partially rectified in a more recent study by Tokowicz and MacWhinney (2005), who investigated L1 effects on L2 processing by manipulating the L1 and L2 similarity of the syntactic construction being tested. They tested participants with L1 English and L2 Spanish on three different types of violations. One construction had similar syntax in the L1 and L2 (auxiliary omission), one had different syntax in L1 and L2 (number agreement on the determiner) and one construction that was unique to the L2 (gender agreement on the determiner). This design has the advantage that the general proficiency level of the participants is completely matched, so that differences must be due to differences in the way the various constructions are processed. Collapsing across construction type they found a significant P600 effect for the syntactic anomalies. However, when they examined the different constructions separately they found that only the similar and unique constructions resulted in a significant P600 effect. The authors conclude that the occurrence of implicit (unconscious and native-like) syntactic processing in L2 depends on the similarity between the L2 and L1, and dissimilarity can block appropriate processing. This idea is explored in more detail in the current study. A weakness of this study is that there is no direct comparison to a control L1 group; it is thus not clear that the constructions are totally comparable in L1 processing of Spanish and that the lack of the P600 is non-native-like. Given the consistency with which P600 effects are found in similar constructions, this is unlikely to explain the lack of an effect entirely, but the effect may be

smaller in some constructions and thus more prone to disappear in L2 processing noise, where it may not reach threshold.

To summarize, the relatively few ERP studies that have addressed on-line L2 processing, it seems that early (or extremely proficient) learners show more native-like processing, but completely native-like processing is generally not possible for late learners of a second language. Late L2 learners (the focus of the current study) typically do not show the early syntactic processing effects (the (E)LAN), but some do show the P600 effects (e.g. the Russian learners of German) and some do not (e.g. the Japanese learners of German). The presence of the P600 may depend on proficiency and/or similarity between L1 and L2.

In this article we investigate L2 processing to determine the circumstances under which the P600 is less likely to be native-like in later learners. Proficiency is a clear candidate to explain the inconsistencies in previous studies: participants have to have achieved a certain level of proficiency in order to show the P600. However, as shown in the study by Tokowicz and MacWhinney (2005), the degree of syntactic similarity between the L1 and L2 may also play an important role. The goal of the current experiments is to focus on the role of L1 and L2 similarity by comparing the processing of L2 Dutch in two groups: L1 speakers of German (a language very similar to Dutch) and L1 speakers of a Romance language (languages less similar to Dutch). A native Dutch control group is also tested. Additionally we consider two different constructions that vary in their degree of L1–L2 similarity for the Romance group but not for the German group. This allows us to address the issue of L1–L2 similarity while controlling completely for general proficiency level.

Like Tokowicz and MacWhinney (2005), we assume that the degree of similarity between L1 and L2 determines the extent to which L2 processing is native-like. We take a slightly different tack in addressing the issues of L1–L2 similarity on L2 processing. It is not really straightforward to define when L1 and L2 are similar. Obviously, two different languages are never identical. However, constructions can be similar in the sense that they involve the use of similar phrase-structure configurations. Second, they may invoke the use of similar grammatical features in a grammatical rule, with or without the same phrase-structure configuration. Third, similarity may involve similar grammatical distinctions

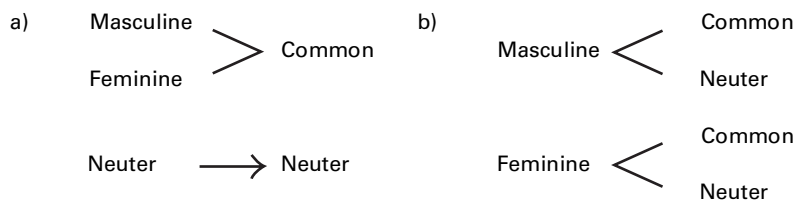


at the lexical level. It is not possible to disentangle all these levels in a single set of experiments. We chose here to focus on:

- the degree to which features that are employed in rules can be transferred; and
- the effect of (dis)similarity at the lexical level.

These two issues are addressed in separate experiments. In the first experiment ‘verbal domain dependency’ was investigated, while in the second experiment grammatical gender was tested. Specific examples of the constructions tested can be found in the materials section of each experiment. These constructions are of particular interest for an on-line processing study because in off-line studies speakers of L1s that differ from the L2 show relatively little difference in their knowledge of the grammatical features involved (Sabourin, 2003; Sabourin *et al.*, 2006). Languages typically discriminate between non-finite constructions by morphosyntactic features on the main verb and co-occurrence with specific auxiliary or modal verbs. The choice of the appropriate form is general and follows clear rules, although the actual verb form may be somewhat idiosyncratic. This gross distinction between the use of a past participle and the infinitive is present in Dutch, German and the Romance languages, although the details of the constructions vary. Thus, agreement within the verbal domain provides a case in which we can investigate L2 processing for a grammatical construction that uses similar syntactic features at the level of a general rule across the languages being investigated.

Grammatical gender, on the other hand, is a lexical property that is idiosyncratic for each noun, although the agreement processes that depend on it are regular. Although gender agreement is overtly present in the nominal agreement system of all languages being investigated, the Romance system is totally different at the lexical level, while Dutch and German express gender in a very similar fashion at the lexical level. Specifically, for a German speaker to learn Dutch gender assignment they can rely on a near perfect one-to-one correspondence (German neuter translates to Dutch neuter) or a many-to-one correspondence (German masculine and feminine translates to Dutch common gender). However, a speaker of French, Italian or Spanish would need to perform a one-to-many correspondence to learn the Dutch gender assignments. This is represented schematically in Figure 1.



**Figure 1** A schematic diagram of the gender assignment correspondences between (a) German and Dutch and (b) Romance and Dutch

Assuming that native-like processing depends on a very similar, grammatically predictable system, we predict that while both L2 groups will show a P600 within the verbal domain, only the German group will show a P600 effect for grammatical gender agreement.

## II Experiment 1

In this first experiment the L2 processing of dependencies within the verbal domain is investigated. In the L1s of the L2 participants being investigated, this dependency surfaces in the inflection on the main verb. Both German and the Romance languages, like Dutch, use a finite participial morphological form of the verb in past constructions and an infinitive morphological form in the non-finite construction; see the examples in (1) and (2):

- |       |   |        |
|-------|---|--------|
| 1) a. | Ik heb in Groningen <i>gewoond</i> .            | Dutch  |
|       | I have in Groningen lived <sub>past.part.</sub> |        |
|       | I have lived in Groningen                       |        |
| b.    | Ich habe in Groningen <i>gewohnt</i> .          | German |
| c.    | J'ai <i>habité</i> en Groningen.                | French |
| 2) a. | Ik wilde in Groningen <i>wonen</i> .            | Dutch  |
|       | I wanted in Groningen to live <sub>inf.</sub>   |        |
|       | I wanted to live in Groningen.                  |        |
| b.    | Ich wollte in Groningen <i>wohnen</i> .         | German |
| c.    | J'ai voulu <i>habiter</i> en Groningen.         | French |

This allows for the investigation of L2 processing of structures that function grammatically in the same way in both the L1 and L2 of the learners. However the morphological form that reflects this dependency does differ across these languages; furthermore, the surface word order positions (i.e. phrase-structure configuration) of the main verb and the auxiliary verb – which indicates whether a participle or non-finite form

is grammatical – differ between Romance and Dutch. See, for example, the French sentences above where the word order is different but the participle (1c) and infinitive forms (2c) are parallel to the distinction in German and Dutch. Despite the word order differences, it is expected that Romance and German speakers who are reasonably proficient in Dutch will show a P600 effect.

### *1 Methods*

*a Participants:* In total 60 participants were tested for this study. Twenty-nine were native speakers of Dutch, 19 were native speakers of German and 12 were native speakers of a Romance language. All participants were right-handed, had corrected to normal vision and were neurologically unimpaired. All L2 participants were late learners, having acquired Dutch at age 14 or later. Of these 60 participants 45 were used in the final analyses presented below. The 15 participants not analysed were excluded either due to excessive artefacts from eye blinks and/or muscle movements, or because they scored below 85% accuracy on agreement within the verb domain.<sup>2</sup> Of the 45 participants, 23 were native speakers of Dutch, 14 were late German learners of Dutch and 8 were late Romance learners of Dutch. In view of the goal of only testing relatively advanced L2 speakers, participants were required to have a high level of proficiency. All L2 participants filled out a language background questionnaire that included questions about their use of Dutch as well as their L1. Participants were also asked if they had learned other languages. Only participants living in the Netherlands for at least three years were considered. This was to ensure that all participants had some opportunity to learn Dutch in a natural setting. Information about length of exposure to Dutch for the participants used in the final analyses can be found in Table 1.

*b Materials:* Forty sentences were used as part of this experiment. Each sentence had a grammatical (the correct past participle form of the verb) and ungrammatical (the incorrect infinitive form of the verb) version. Two versions of the test were made such that sentences that were

<sup>2</sup> Only participants with at least 12 trials per condition (in both experiments) remaining after artefact rejection were used in final statistical analyses.

**Table 1** Participant information and accuracy scores for Experiments 1 and 2. The final column represents the scores on assigning gender to the nouns used in Experiment 2

First language	Exposure to Dutch: age range (average in years in brackets)	Accuracy: Experiment 1: average as a percentage (range in brackets)	Accuracy: Experiment 2: average as a percentage (range in brackets)	Accuracy: gender assignment: average as a percentage (range in brackets)
Dutch ( $n = 23$ )	n/a	97.2 (95–100)	94.3 (90–99)	n/a
German ( $n = 14$ )	3–32 (9.8)	92.0 (89–95)	82.1 (77–88)	93.0 (80–98)
Romance ( $n = 8$ )	3–23 (11)	90.7 (87–95)	59.4 (52–67)	78.0 (60–80)

grammatical in one version were ungrammatical in the other and vice versa; participants saw an equal number of grammatical and ungrammatical target sentences and version was counterbalanced across participants. An example of the sentences used can be seen below.

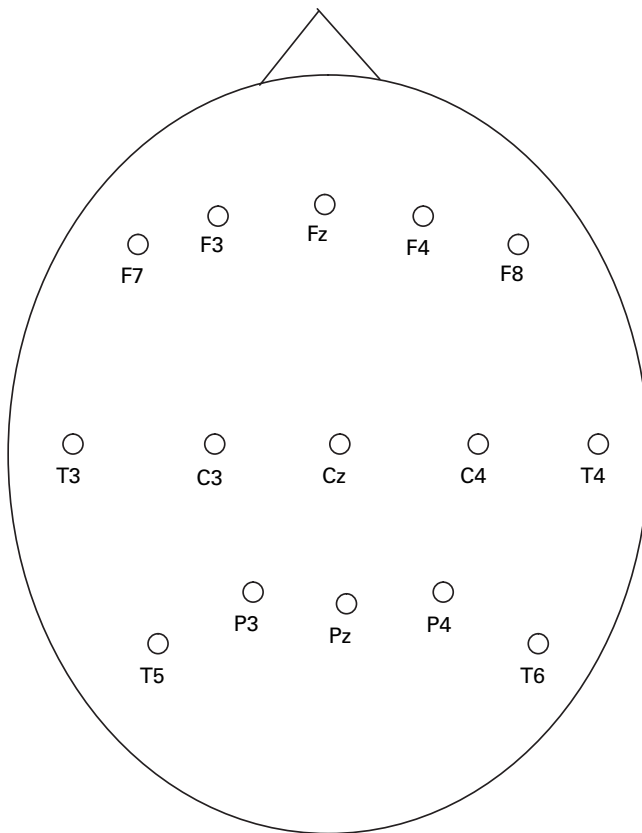
- 3) Ik heb in Groningen gewoond/\*wonen.  
 I have in Groningen lived<sub>past/part.</sub>/\*to live<sub>inf.</sub>  
 I have lived in Groningen.

For this experiment, all items had the critical word (the word at which the sentence became ungrammatical in the ungrammatical version of the sentence) at the end of the sentence. All words used in the sentences were of a fairly high frequency to ensure familiarity to all the L2 speakers. The log frequency of each item (determined through the CELEX database; Burnage, 1990) was between 1.11 and 2.98 (average 1.8). These 40 sentences were embedded in a 280 sentence grammaticality judgement paradigm. Sentences analysed in Experiments 1 and 2 were both presented randomly interspersed among other sentence types.<sup>3</sup>

*c Procedure:* During ERP measurements, participants were seated in a dimly lit sound-attenuated room, facing a computer monitor. Sentences were presented word by word in the middle of the screen; each word was presented for 250 ms, followed by a blank screen for 250 ms after which the next word would appear. Each sentence was preceded by a centrally-presented asterisk to alert participants that a new sentence was about to begin. After each sentence, a delay screen was displayed, followed by a screen prompting participants to make a grammaticality judgement by pressing one of two buttons. The delay was included to prevent motor artefacts from obscuring the response to the final target word. Between sentences participants were given two seconds in which they were allowed to blink. Participants were asked to try their best not to blink during presentation of the sentences. The experimental session started with a practice session so that participants could get used to the word-by-word presentation of the sentences and to practise not blinking during the sentence trials. The experiment lasted approximately one hour. Participants were given breaks during the test session as needed.

<sup>3</sup> Other sentence types presented were canonical and violations of subject–verb agreement, indefinite noun and adjective agreement, and noun–relative-pronoun agreement.

*d EEG recording and analysis:* Electroencephalogram (EEG) activity was recorded by means of tin electrodes mounted in an elastic cap (Electro-Cap International). Fifteen electrode sites, based on the international 10–20 system, were used for analyses (F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4 and T6). See Figure 2 for a diagram of the electrode positions used in the current study. All electrodes were referenced to averaged mastoids. Both horizontal and vertical electrooculograms (EOGs) were measured. Electrode impedances were kept below 5 k $\Omega$ . EEG and EOG signals were sampled at 1000 Hz, amplified and digitally filtered with a cut-off frequency of 30 Hz; effective sample frequency was 100 Hz.



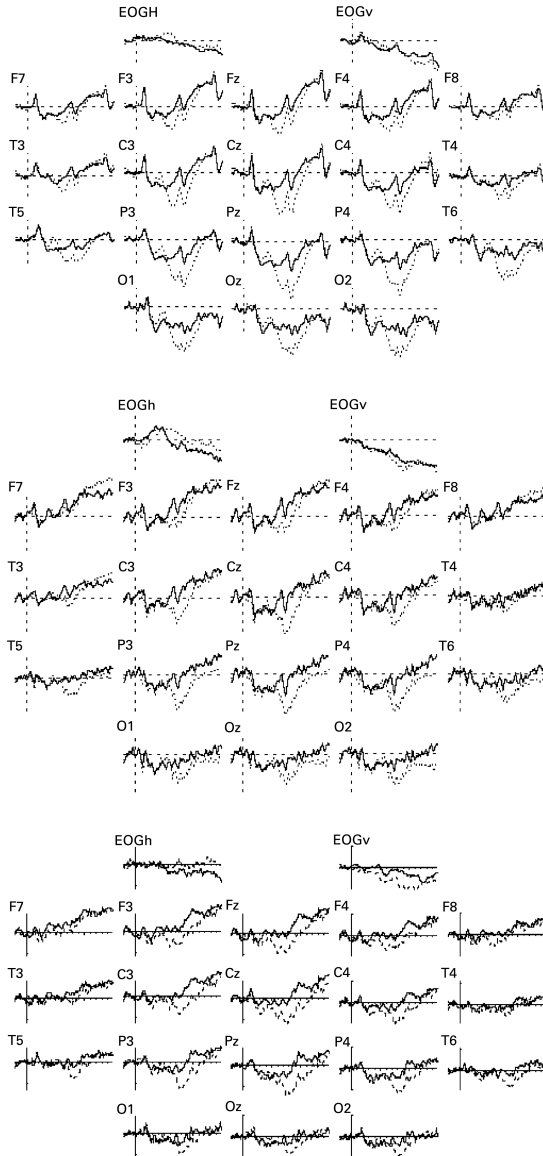
**Figure 2** Approximate electrode locations for Experiment 1 and 2. At the top is the nose. F represents electrodes placed over frontal sites, T represents temporal sites, P represents parietal sites and C represents central sites. Even numbers represent electrodes on the right hemisphere and odd numbers represent electrodes placed over the left hemisphere. Z represents electrodes placed on the midline.

The ERP analyses examined the average mean amplitudes of the ERP waveform from two time-windows in which the P600 normally occurs in native speakers. Only mean amplitudes to sentences that were correctly judged during the session were included in analyses. Since it has been noted that L2 speakers may have a delayed response, the first interval chosen represents a P600 time-window in the L1 time range, and the second interval is later to investigate possible delays in L2 processing. Time-window one (TW1) was from 511 ms to 720 ms and time-window two (TW2) was from 721 ms to 900 ms. For each TW we performed 4-way mixed ANOVAs. The factors analysed were sentence type (grammatical vs. violation), scalp distribution (anterior vs. central vs. posterior) and laterality (lateral left vs. medial left vs. midline vs. medial right vs. lateral right) as the within-participants factors. The factor language (Dutch vs. German vs. Romance) was the between-participants factor. Greenhouse–Geisser corrections are used when there is more than one degree of freedom in the numerator.

## 2 Results

*a Behavioural results:* The average accuracy per group per condition is shown in Table 1. A 2-way ANOVA with grammaticality as the within-participants effect and L1 as a between-participants effect was performed on the behavioural data. The interaction between L1 and grammaticality shows a trend toward significance ( $F(2,42) = 2.911$ ,  $p = .065$ ). A significant main effect of L1 is found ( $F(2,42) = 6.177$ ,  $p = .004$ ). Tukey HSD *post hoc* tests show that this main effect is due to the native speaker group performing significantly better than both L2 groups (Dutch vs. German:  $p = .02$  and Dutch vs. Romance:  $p = .017$ ). No significant differences were found between the German and Romance speakers ( $p = .868$ ). On the whole, the L2 groups performed pretty well even with the word-by-word presentation in this task, showing that they are aware of the correct structure of these sentences.

*b ERP data:* The grand-average waveforms for each group can be seen in Figure 3. Inspection of the figure shows that the native speakers exhibit clear effects of grammaticality at virtually all electrodes, although it is smaller at the frontal and lateral electrodes. The L2 groups show similar effects, with a similar distribution.



**Figure 3** The grand-average waveforms for each of the language groups for Experiment 1. The waves on the top are for the Dutch native speakers, the middle waves are for the German learners of Dutch, and the bottom waves are for the Romance learners of Dutch. The solid line represents the grammatical form of the sentence while the violation is seen in the dashed line. The scale shown here is  $\pm 5$ mV, with negative plotted up. Each wave shows the entire waveform analyzed with a 200 ms baseline and lasting to 1500 ms.



In the first ANOVA investigating the P600 effect in TW1 a main effect of sentence type was found ( $F(1,41) = 5.03, p > .05$ ), indicating that the violations were more positive than the grammatical sentences in this TW. Not all electrodes show the grammaticality effect equally strongly: a significant interaction between laterality and sentence type was also found ( $F(4,164) = 5.08, p < .01$ ), indicating that the positivity for violations was greater over medial and midline electrode sites than over lateral sites. A trend towards a significant interaction between sentence type and language was also found ( $F(2,41) = 2.86, p = .069$ ), reflecting the fact that Dutch native speakers tended to show a larger P600 effect in this TW than the German and Romance speakers. This trend was further investigated by looking at the language groups individually for this TW. In the separate 3-way ANOVAs the Dutch group showed a significant main effect of sentence type in this TW ( $F(1,21) = 20.05, p < .001$ ) and a significant interaction between sentence type and laterality ( $F(4,84) = 7.192, p = .001$ ), indicating that the effect is largest over midline sites and smallest over right lateral sites. Neither the German group ( $F(1,13) = .04, p = .837$ ) nor the Romance group ( $F(1,7) = .877, p = .38$ ) taken separately showed a significant main effect of sentence type in this TW. However, critically, the German group did show a significant 3-way interaction ( $F(8,104) = 3.671, p < .05$ ), indicating that the violations were significantly more positive at the medial posterior electrode sites. The Romance group also showed a trend toward a significant 2-way interaction between laterality and sentence type ( $F(4,28) = 3.515, p = .069$ ) as the P600 effect was largest over medial and right lateral electrode sites.

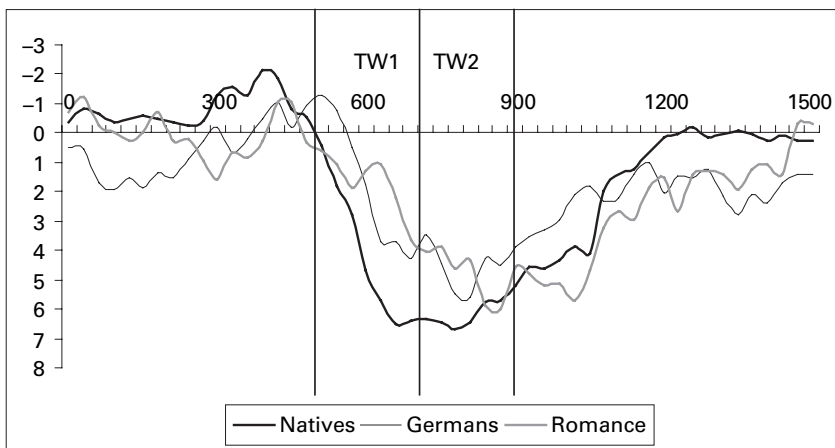
In the ANOVA for TW2 no interactions or trends toward an interaction with language were found. A significant interaction between scalp distribution and sentence type was found ( $F(2,84) = 25.996, p < .001$ ), indicating that the P600 effect was found over central and posterior electrode sites, which given the lack of interaction appears to be approximately equally strong for all groups. This was confirmed by significant effects in each of the language groups analysed separately (all  $F_s > 4.5, p < .05$ ). A significant interaction between laterality and sentence type was also found ( $F(4,168) = 5.322, p = .005$ ), indicating that the P600 effect was larger over medial and midline electrode sites. In the individual group analyses however, this effect was only confirmed for the native speakers ( $F(4,84) = 4.79, p = .011$ ).

### 3 Discussion

The statistical analyses confirms that all groups in this study showed a robust P600 effect, a positivity that is maximal over posterior and medial electrodes for the ungrammatical sentences, particularly in TW2. The trend towards a significant interaction between language and sentence type in TW1 together with less reliable effects in the L2 groups suggests that the L2 groups' P600 was somewhat slower to reach its peak and is smaller or more restricted in its distribution. Figure 4 shows difference waves (ungrammatical–grammatical conditions), illustrating the time course of the P600 effect during the two TWs at electrode Pz for all three language groups. It can be seen that the P600 reached maximum for the native speakers during TW1 and continued at maximum into TW2 while the peak for both L2 groups is in TW2.

The P600 for the L2 speakers shows only a quantitative difference (in latency and amplitude) from the Dutch native speakers and not a qualitative difference. This holds even though the L2 speakers performed less accurately on the sentence-final grammaticality judgement task. Thus it appears that:

- the late learners in this study can show a native-like P600; and
- native-like knowledge (in terms of accuracy measures under time pressure) is not necessary for native-like processing to be seen as measured by the P600 effect.



**Figure 4** The ERP difference waveform at electrode Pz for Experiment 1 is shown for all three language groups

This confirms that late L2 processing can be relatively native-like for constructions that are relatively similar between L1 and L2 and that are expressed in a relatively similar way in German and the Romance languages as in Dutch. Note that it is not necessary that the rule apply to an identical phrase-structure configuration. While German has a surface word order that is very similar to Dutch in this construction, Romance languages do not.

The results of this experiment are consistent with the hypothesis that L2 speakers can achieve native-like processing, at least in cases where they can make some use of L1 processing routines in their L2. This seems to be the case at least where the languages are similar and the process is independent of lexical limitations at the level of the abstract features involved (although not at the level of morphological instantiation). Furthermore, they suggest that the languages do not have to instantiate the rule in a completely similar way since the Romance word order is different. The degree of similarity between L1 and L2 that is necessary for native-like processing however remains unclear. In the next experiment we examine the processing of a construction in which lexical grammatical features are necessary for rule use.

### **III Experiment 2**

In the second experiment, carried out in the same group of participants, the L2 processing of gender agreement within the nominal domain is investigated. In Experiment 1, both the German and Romance speakers show late ERP effects that are similar to those of native speakers, suggesting native-like processing is possible at least when the languages are similar and the constructions are rule governed. Furthermore, given this syntactic context, the participial or infinitival form must always occur, although the precise morphological expression of the form may be idiosyncratic for a given lexical item. However, agreement within the nominal domain is different. Nouns in languages like Dutch, German and Romance languages carry intrinsic grammatical gender, which is not overtly marked on the noun. For speakers of English this gender assignment often seems arbitrary, and it may require actual memorization of lexical items in order to assign the correct gender to nouns.

Speakers of other gender languages such as German and Romance languages have a basis from within their own language to help them

with the second-language acquisition of gender. In fact, both German and Romance speakers, and even English speakers, can have the knowledge necessary to assign nouns to the correct Dutch gender category at a very high level of accuracy (Sabourin, 2003; Sabourin *et al.*, 2006). The issue addressed here is whether the existence of the gender distinction in the L1 is sufficient to allow for native-like processing in L2, or whether this can only occur for a system that is lexically very similar to the system in L1, so that L1 processing can be transferred to L2.

The instantiation of the Dutch gender agreement system requires Romance speakers, but not German speakers, to learn the gender of each noun on a word-by-word basis. The German system, although it differs from the Dutch system in that it contains three genders, is similar in terms of assignment and agreement of gender; most German nouns and their Dutch equivalents have the same gender (masculine and feminine in German are mapped onto the common gender in Dutch and German neuter is mapped onto Dutch neuter; Figure 1), and the elements that require gender agreement are very similar in the two languages (for a review of the Dutch and German gender systems and their similarities, see Sabourin, 2003). For this reason, it may be possible for German speakers to transfer processing routines as well as an initial lexical-gender assignment from the L1 to the L2. However the Romance system is completely different at the lexical level. If a completely cognate system is necessary for native-like processing, then the Romance group should not show a P600 effect in their L2 Dutch.

### 1 Methods

*a Participants:* The same participants as in Experiment 1 were used in this experiment.

*b Materials:* Forty sentences were tested as part of this experiment. Each sentence had a grammatical and ungrammatical version. Examples of the sentences tested can be seen in (4–5) below.

- 4) Het/\*De kleine kind probeerde voor het eerst te lopen.  
 The<sub>neut/\*com</sub> small child<sub>neut</sub> tried for the first to walk.  
 The small child tried to walk for the first time.
- 5) Hij komt eraan met de/\*het verse koffie.  
 He comes to with the<sub>\*neut/com</sub> fresh coffee<sub>com</sub>.  
 He is coming with the fresh coffee.

The ungrammatical versions of the sentences in this experiment all used the incorrect form of the definite determiner. Half the sentences contained *het* and half *de* as the correct determiner. An adjective (inflected in the same manner for either common or neuter gender) was always present between the determiner and noun, which means that the participants could not use mere linear co-occurrence frequency to see that the sequence is correct. Although the determiner used in the sentences is incorrect, the sentences only become ungrammatical at the noun. Thus, ERPs were measured from the critical word, the noun. Fairly high frequency words were used in all sentences to ensure that they were familiar to all the L2 speakers. Lexical items were selected from the CELEX database (Burnage, 1990) and had log frequencies above 1.11.<sup>4</sup>

Forty sentences were used as part of this experiment. Two versions of the experimental list were created, with correct and incorrect forms of each sentence appearing on different versions; an equal number of sentences in each condition appeared in both versions. Sentences of each condition were evenly spread across the lists interspersed among sentences with other structures.

*c Procedure:* The same grammaticality judgement procedure was used as for Experiment 1. One extra task for participants was included. After the ERP session, participants were asked to assign gender to all critical nouns used in the experiment. The EEG recording procedure and analyses, with one exception, was also the same. For Experiment 2, the mean amplitudes used for the native speakers and the German learners of Dutch represent correct judgements only (as in Experiment 1); however, the mean amplitudes used for the Romance learners came from all sentences in order to increase the power.<sup>5</sup>

## 2 Results

*a Behavioural results: grammaticality judgement:* A 2-way ANOVA with grammaticality as the within-participants factor and L1 as a

<sup>4</sup> Out of the 40 nouns tested, six were of non-congruent gender between Dutch and German. This was done to test a further question concerning the effect of congruency. No differences in ERP pattern were found when looking at responses to all nouns compared to just the responses to congruent gender items.

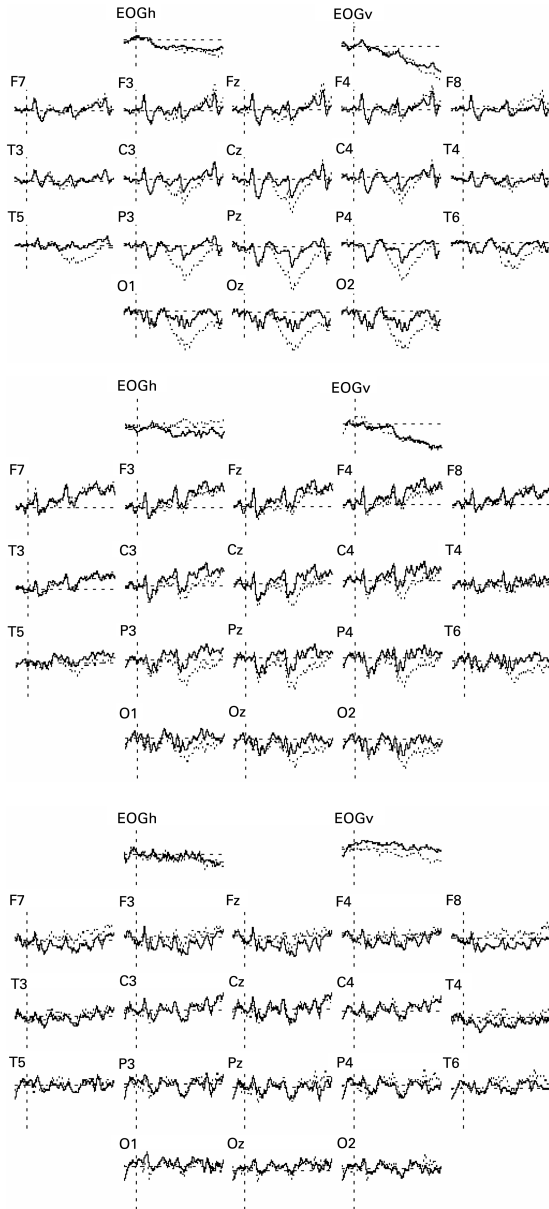
<sup>5</sup> Other than an increase in statistical power, using all sentences does not change the statistical results for this group.

between-participants factor was performed on the behavioural data. The group scores can be seen in Table 1. A significant interaction between grammaticality and L1 was found ( $F(2,42) = 14.739$ ,  $p < .001$ ). This interaction reflects the fact that the native speakers, though not performing perfectly, do very well on all sentences (94.5 % correct on grammatical sentences and 94.1% correct on the ungrammatical sentences), while the L2 groups showed a strong effect of grammaticality, tending to miss the ungrammatical sentences. *Post hoc* Tukey HSD tests show that the native speakers performed significantly better than both the German ( $p = .003$ ) and Romance ( $p < .001$ ) speakers. The German group performed significantly better than the Romance group ( $p < .001$ ). As can be seen by the scores this task was considerably more difficult for both L2 groups than the same task within the verbal domain on which they showed similar levels of proficiency.<sup>6</sup>

*b Behavioural results: gender assignment:* The results of the gender assignment task are shown in Table 1. Results show a much higher degree of accuracy in this off-line task, demonstrating that the L2 speakers were mainly familiar with the gender class of the nouns used in this experiment. However, even in this task, the same pattern was seen with the native Dutch speakers performing significantly better than the L2 German group and both outperforming the L2 Romance group (in all cases  $p < .01$ ).

*c ERP results:* Grand-averaged waveforms for each language group can be seen in Figure 5. As in Experiment 1, the native speakers show a widely distributed positivity, which is largest at posterior electrode sites. However, in addition to a P600 effect, the native speakers show a late negativity, which is most prominent at right frontal electrodes. The German group shows a positivity with a similar scalp distribution but no frontal negativity, while the Romance group fail to show a positivity but do exhibit an anterior negativity with a similar distribution to the one seen in the native speakers, but which is apparent from an earlier latency.

<sup>6</sup> Since the two groups did not differ on the grammaticality judgement task within the verbal domain, we regard them as being matched on overall proficiency in their L2. It seems likely that the differences in proficiency for gender are due to L1 differences.



**Figure 5** The grand-average waveforms for each of the language groups for Experiment 2. The waves on the top are for the Dutch native speakers, the middle waves are for the German learners of Dutch and the bottom waves are for the Romance learners of Dutch. The solid line represents the grammatical form of the sentence while the violation is seen in the dashed line. The scale shown here is  $\pm 5\text{mV}$ , with negative plotted up. Each wave shows the entire waveform analyzed with a 200 ms baseline and lasting to 1500 ms

The data for this experiment were analysed initially using the same statistical design and two TWs as were used in Experiment 1 looking at the P600 effect. Results for the P600 TWs are also discussed in the same manner as for Experiment 1 above. The frontal negativity is analysed in an exploratory fashion looking at consecutive chunks of 30 ms TWs to determine whether this effect is significant for the native speakers and whether a similar pattern can be statistically supported in the Romance group.

In the ANOVA investigating the P600 effect in TW1 (511–720 ms) three significant interactions were found. First, there was a significant interaction between sentence type and language ( $F(2,41) = 3.664, p = .034$ ). This interaction is explored below in separate ANOVAs for each language group. There was also a significant interaction between scalp distribution and sentence type ( $F(2,82) = 4.196, p = .043$ ), which indicates that the P600 effect is largest over central and posterior regions. A significant interaction between laterality and sentence type was also found ( $F(4,164) = 9.03, p < .001$ ), indicating that the P600 effect is only present over medial and midline sites. In summary, the effect was similar in scalp distribution to the P600 found in Experiment 1, but the effect was not equivalent for all three groups of participants.

In the separate 3-way ANOVAs the Dutch group showed a significant main effect of sentence type in this early TW ( $F(1,21) = 7.994, p = .01$ ). The Dutch group also showed a significant interaction between sentence type and laterality ( $F(4,84) = 10.977, p < .001$ ), indicating that the P600 effect is largest over midline sites and smallest over right lateral sites. The Dutch also showed a trend towards a significant interaction between scalp distribution and sentence type ( $F(2,42) = 3.059, p = .091$ ), which indicates a tendency for the P600 effect to be larger over more posterior electrode sites. The German group also showed a significant main effect of sentence type ( $F(1,13) = 6.907, p = .021$ ). No significant interactions were found for the German group. No significant effects were found for the Romance group (all  $p$  values  $> .18$ ). This is consistent with the effects seen in the figures.

In the overall ANOVA for TW2 a near significant interaction between language and sentence type was found ( $F(2,41) = 3.151, p = .053$ ). This is also explored in separate ANOVAs for each language group. A significant interaction between scalp distribution and sentence



type was also found ( $F(2,82) = 19.086, p < .001$ ), indicating that the P600 effect was larger over the more posterior electrodes and not present at all over frontal electrodes. A significant interaction was also found between laterality and sentence type ( $F(4,164) = 5.946, p = .004$ ), indicating that the P600 effect was present over medial and midline sites and not over more lateral sites.

In separate 3-way ANOVAs the Dutch group showed a significant main effect of sentence type in this later TW ( $F(1,21) = 8.37, p = .009$ ). The Dutch group also showed a significant interaction between sentence type and laterality ( $F(4,84) = 7.235, p = .002$ ), indicating that the effect was largest over midline and medial sites as well as a significant interaction between sentence type and scalp distribution ( $F(2,42) = 21.185, p = .002$ ), indicating that the P600 effect was present over posterior and central sites and not present over frontal electrode sites. The German group showed a trend towards a significant main effect of sentence type ( $F(1,13) = 4.426, p = .055$ ) and a significant interaction between scalp distribution and sentence type in this later time window ( $F(2,26) = 8.675, p = .009$ ), indicating a P600 effect on central and posterior electrodes. The German group also showed a trend towards a significant interaction between laterality and sentence type ( $F(4,52) = 3.226, p = .059$ ), showing a tendency towards a larger P600 effect over medial and midline electrode sites and no positivity over left lateral sites. Again, no significant effects or interactions were found for the Romance group (all  $p$  values  $> .12$ ).

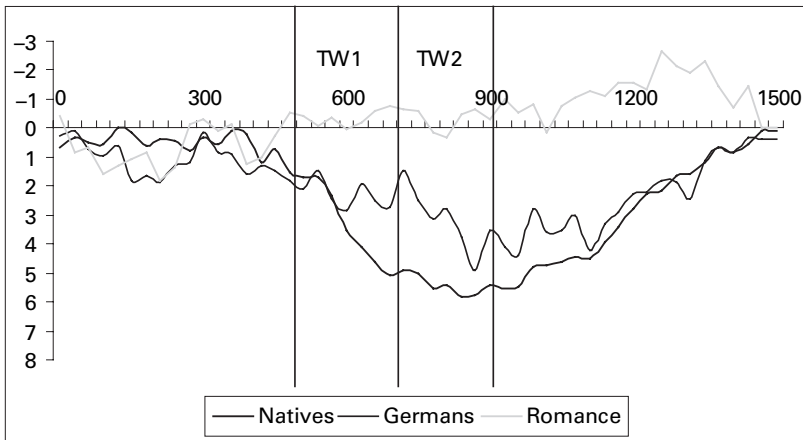
In the exploratory analysis looking at the frontal negativity to the gender agreement violations found in the Dutch and Romance speakers' waveforms, ANOVAs looking at the effect of sentence type were performed on separate 30 ms TWs from 0 to 1500 ms to identify in which intervals, if any, this effect was significant. Effects are only reported as significant if more than three consecutive TWs show significant effects at  $p < .05$ ; this is to counteract the tendency to find significance in non-orthogonal multiple comparisons. These analyses were only performed on the native speaker and Romance groups; the goal was to determine whether the Romance pattern was possibly related to the response found in the native group. For the native speakers, this analysis resulted in significant effects at electrode Fz, where there was a significant negativity from 600 to 720 ms, and at electrode F8 where

there was a significant negativity from 960 to 1500 ms. In the Romance speakers the electrodes F7, F3, Fz and F8 all showed a significant early negativity from 270 to 420 ms. For electrodes F7 and F8 the significant negativity continued until 1500 ms.

### 3 Discussion

For Experiment 2, as in Experiment 1, an overall P600 response was found. Unlike Experiment 1, though, the L2 groups showed quite different ERP effects. The German participants showed a clear P600 effect (although it was attenuated relative to the native speakers and peaked later), while there was no evidence whatsoever for a P600 effect in the Romance data. This can be seen clearly in Figure 6, which shows the difference waves (ungrammatical–grammatical) at electrode Pz (where the P600 effect tends to be maximal) for each language group.

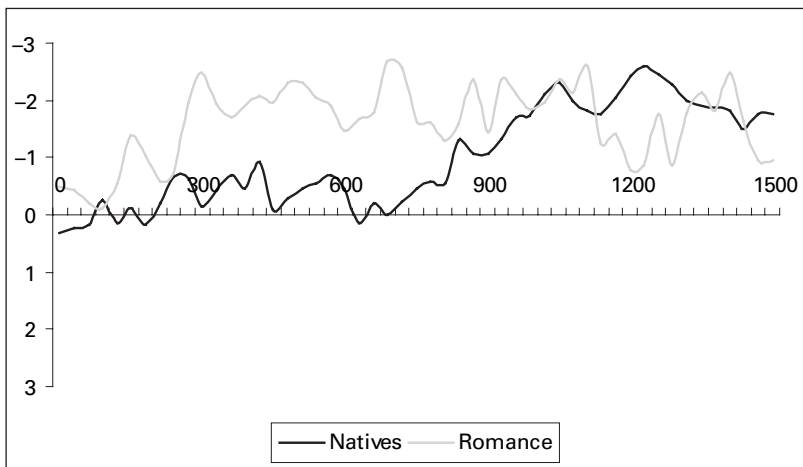
In this construction we find clear effects, which appear to be due primarily to differences in similarity between L1 and L2. As can be seen from the results of Experiment 1 (which was carried out over the same participants), both groups had achieved a reasonable general level of proficiency in their L2, which was sufficient to elicit a native-like P600 and accurate grammaticality judgements. The L2 groups performed at a similar level when dealing with a construction that was compatible



**Figure 6** The ERP difference waveform at electrode Pz for Experiment 2 is shown for all three language groups

between L1 and L2. The difference between L2 groups found in Experiment 2 suggests that in the case of gender, it is not sufficient to have gender in the L1, but that the systems must be very similar to that of the L1, down to the lexical level, in order for the processes eliciting the P600 effect to be employed in the L2. The behavioural results are consistent with this conclusion. Although off-line grammaticality judgement and gender identification tasks suggest little difference between these two groups (Sabourin, 2003; Sabourin *et al.*, 2006), as soon as L2 speakers were forced to respond more quickly, both groups tended to make more errors, and this was more noticeable for native speakers of an L1 whose gender system is less like that of Dutch, the Romance group.

The main goal of the current experiment was to investigate the aspects of syntactic processing reflected in the P600. However, the native speakers also showed a second response to gender violations that is less familiar, a frontal negativity. The Romance speakers, although they showed no detectable P600 effect, did show a frontal negativity for these gender violations. This can be seen in Figure 7, which shows the difference waves for the Dutch and Romance groups at electrode F8, the electrode where the later negativity is most noticeable in native speakers.



**Figure 7** The ERP difference waveform at electrode F8 for Experiment 2 is shown for the Dutch and Romance groups

This frontal negativity suggests that the Romance speakers are sensitive to the ungrammaticality despite their lack of a P600. It is clearly different from the native response in time course and extent: the earliest effect shown by the native speakers (600–720 ms after onset of the violation) was limited to the front central electrode (Fz) followed by a much later effect limited to the right frontal electrode (F8), while the Romance speakers showed effects from a much earlier TW (starting 270 ms after presentation of the violation) with a broader frontal distribution. Although the effect is somewhat different across the two groups, the frontal scalp distribution suggests that this negativity may reflect one aspect of the native-like processing of this type of construction.

We have interpreted this as a negative deflection of the waveform to ungrammatical sentences. Conversely, this negativity could also be considered a positivity for the grammatical sentences. This is the analysis given by Hahne (2001) and Hahne and Friederici (2001) for similar frontal effects found in their experiments in response to categorical ungrammaticalities. This interpretation is at least partly based on the fact that they do not see a frontal negativity in the native response to categorical ambiguities, which is not true in the current study. They suggest that this effect may reflect a P600 effect due to effortful processing that is stronger for the grammatical sentences, possibly because the L2 speakers see a chance of success at interpreting these sentences. The effects in the current study seem to be too early for such an explanation. Additionally, the distribution is not the same as that found for the P600 in this group in Experiment 1. For these reasons, we interpret this finding as a processing negativity, similar to the effect found for native speakers in the later time window.

The overall pattern of results suggests that native speakers make use of several kinds of neurological resources to deal with ungrammatical sentences, including the neural substrates that generate the P600 and the frontal negativity. These appear to differ in nature sufficiently that one (the P600) is more suitable to be used by L2 speakers with a cognate system, while the other (the frontal negative wave) is more suitable to be used by an L2 group, who are proficient enough to recognize that there is an error, but cannot easily make use of native processing routines to deal with the ungrammaticality.

It is commonly proposed that the P600 reflects an attempt to revise or repair the structure or meaning of a complex or ungrammatical sequence, during sentence comprehension (Kaan and Swaab, 2003). These processes make use of syntactic information. To be able to invoke these processes, the syntactic process must be native-like to some extent. Without being able to make efficient use of lexical gender, this process appears to be unavailable to the Romance speakers.

The frontal negativity seen in the native speakers in this study, we suggest, can best be interpreted as a memory resource that is used when attempting to maintain information in the hopes that a resolution for the ungrammaticality will be encountered later in the input (see also Sabourin and Stowe, 2004). This is primarily based on evidence that long-lasting frontal negativities occur in a number of circumstances from various domains that seem to require memory resources (Ruchkin *et al.*, 1990; Honda *et al.*, 1996). In native speakers, such a resource is most likely to be employed when reanalysis fails; this effect will, therefore, occur relatively late in native speakers. A working memory account seems compatible with the early use of this resource by L2 speakers who have some awareness of the ungrammaticality but are not able to use native processing routines to deal with the ungrammaticality. However, it is not clear under this account why the German L2 group does not show this effect. This account is speculative, but interesting, and deserves further research.

These results also highlight the extra insight that can be gained by the use of neural measurements of language processing in L2. By using ERPs we are able to learn more about the differences between off-line and on-line processing. An off-line study of Romance speakers' knowledge of gender in their Dutch L2 suggested that they were fairly proficient at gender-processing, as proficient as the German L1 group (Sabourin *et al.*, 2006). The on-line grammaticality judgements collected in this study suggest that they are incapable of using this information under time pressure, since they performed not much above chance level. The neural measurements nevertheless suggest that the information is available and accessed, but that it cannot be used in the same way that native speakers use it. The German group on the other hand, can transfer L1 processing routines, which allows more native-like reanalysis processes.

#### IV General discussion and conclusions

The focus of the two experiments reported here was on the effects of L1 similarity on transfer in L2 processing. This was investigated by comparing the P600 effects found in L1 and L2 processing. Previous research had suggested that the P600 might be lacking for relatively late learners (Weber-Fox and Neville, 1996) or for L2 speakers with low proficiency (Hahne, 2001). However, these studies did not manipulate the type of construction or similarity between L1 and L2. Although in both studies late learners did not show the P600, both studies were carried out with groups whose L1 was not particularly similar to the target language to be learned (Chinese learners of English and Japanese learners of German). Additionally, the proficiency level of these groups was clearly lower than that of earlier learners (Weber-Fox and Neville, 1996) or of learners of the same age group with a more similar language (Russian learners of German; Hahne and Friederici, 2001). The results reported by Tokowicz and MacWhinney (2005) suggest that similarity between L1 and L2 plays a central role in the degree to which L2 processing is native-like. We confirm this point in the current study.

The findings of Experiment 1 in the current study demonstrate that some degree of native-like processing is possible for late learners. However, L2 processing is slower (shown by delayed onset and/or delayed maximum of the P600 effect).<sup>7</sup> This relatively native-like use of the P600 seems to be limited to processing of aspects of the syntax that are quite similar between L1 and L2, as seen in the differences between the results of Experiments 1 and 2. For the verb domain condition, which is similar in the L1s of both groups, a clear but delayed P600 maximum is found. This result was found for the Romance group despite the fact that there is a difference in surface word order between L1 and L2. However, for gender processing – for which the degree of similarity between the target language and the L1s of these groups differs – there is less evidence for native-like processing. Only the German group shows a P600 effect. For German speakers, the determiner choice in Dutch is very similar to that in their own language down to the lexical level. Simply

<sup>7</sup> We note here that the Romance speakers do show an early negative response to the gender violations but, since we are uncertain as to the linguistic significance of this effect, if any, we have chosen to focus on the language related P600 component.

having a gender system in the L1, like the Romance speakers, does not necessarily support native-like syntactic reanalysis and repair routines in L2. Nevertheless, it leads to enough sensitivity to gender violations to evoke a processing response, seen as a frontal negativity, which we suggest may reflect allocation of memory resources.

We interpret the results of Experiment 2 as demonstrating differences in processing efficiency that primarily result from L1 influence, rather than differences in overall proficiency in these participants. This is because, although the Romance speakers performed less well in the nominal domain than the Germans in the on-line behavioural test, both groups were equivalent on agreement in the verbal domain and showed similar native-like processing on this aspect of grammar for which the L1s do not differ. In addition to this, the Romance speakers do show a significantly different ERP pattern to the ungrammatical vs. the grammatical sentences within the nominal domain, although it is not the same as the response seen for the German learners of Dutch. The ERP pattern found in the Romance speaker data is similar to an anterior frontal effect found in the native speaker data, although it is started at an earlier time point. This suggests that the ability to use particular grammatical aspects of the L2 in a native-like way differ depending on the L1.

Up until now we have not discussed the process underlying L1 effects. Much of the literature on L1 effects has concentrated on the presence or absence of a particular abstract grammatical feature, like gender or verbal dependency, in the L1. However, the difference between the results for the verbal and nominal domains suggests that this may not be the correct way of looking at the issue. Rather, taken together, the results suggest that transfer of processing routines rather than grammatical representations from L1 to L2 may explain much of the results that have been obtained to date. If late L2 learners typically transfer processing routines, these will be more or less successful depending on the degree of similarity between the languages. Thus, in the case of similar constructions (processing in the verbal dependency in our case) transfer will succeed (as the processing routines are similar across the different languages), while for dissimilar constructions (e.g. Romance speakers learning the Dutch grammatical gender system) or for lexical instantiations of features, transfer will fail (as the processing routines are not similar). This perspective is consistent with our own

results and is largely compatible with the results obtained by Tokowicz and MacWhinney (2005). We discussed three circumstances in which transfer will succeed. In domains where agreement is completely rule governed, transfer can take place on the basis of an abstract feature despite surface differences, such as the place of the non-finite verb form in the sentence. On the other hand, in gender agreement – in which features must be lexically specified and are essential for the operation of a grammatical agreement rule – it appears that having a similar distinction in the native language is not enough. In German it is not only the case that the feature gender is present; the agreement between determiner and noun works in a congruent fashion in Dutch and German down to the lexical level (see Figure 1).

When there is not sufficient congruence, transfer will fail. In Romance languages, although the feature gender is present, lexical-gender assignment and agreement does not work the same way as in Dutch. Similarly, Tokowicz and MacWhinney (2005) showed interference effects when L1 and L2 are dissimilar. Totally unsuccessful routines are likely to be abandoned or modified, but the degree of success with which this can occur determines the degree of native-likeness that can be attained in the L2. It is presumably more likely to occur with quite general rules that are clearly identifiable in the input. Gender agreement in Spanish correlates strongly with the final vowel of the noun and has been shown to be relatively successfully learned by English speakers (Tokowicz and MacWhinney, 2005), although there are lexically marked exceptions that are likely more difficult for the L2 processor. We suggest that only in the case of transfer can the use of purely lexical properties – like gender assignment in Dutch – result in native-like processing by late acquirers of a language, at least those at the level of proficiency tested in the current experiment. However, clear qualitative distinctions can be made between the native and L2 processing. It would be interesting to see if these differences remain operative even at a level of near-native proficiency (for behavioural evidence of L1 processing effects in highly proficient L2 learners, see Hopp, 2006).

Taken together the ERP results reported here suggest that the P600 component can indeed be found in late L2 processing, suggesting that some aspect of native-like processing is possible for this group. They also provide a potential explanation for why not all studies find this effect; the



presence of the P600 appears to depend largely on the syntactic construction being investigated and on the degree of similarity between L1 and L2. Constructions that are similar in the L1 and L2 and are rule governed result in proficient late L2 learners showing the native-like P600 effect: L1 and L2 processing routines are similar. Constructions that are not grammatically similar or depend on lexically specific features that are rarely the same in the L1 and L2 are unlikely to result in a P600 effect in the L2 processing: L1 and L2 processing routines are not similar.

## V References

- Ainsworth-Darnell, K., Shulman, H. and Boland, J.** 1998: Dissociating brain responses to syntactic and semantic anomalies: evidence from event-related potentials. *Journal of Memory and Language* 38, 112–30.
- Ardila, A.** 2003: Language representation and working memory with bilinguals. *Journal of Communication Disorders* 36, 233–40.
- Burnage, G.** 1990: *A guide for users*. Nijmegen: CELEX Centre for Lexical Information.
- Coulson, S., King, J. and Kutas, M.** 1998: Expect the unexpected: event-related brain response to morphosyntactic violations. *Language and Cognitive Processes* 13, 21–59.
- Friederici, A.D., Hahne, A. and Mecklinger, A.** 1996: Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22, 1219–48.
- Friederici, A.D., Steinhauer, K. and Pfeifer, E.** 2002: Brain signatures of artificial language processing: evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences* 99, 529–34.
- Gunter, T.C. and Friederici, A.D.** 1999: Concerning the automaticity of syntactic processing. *Psychophysiology* 36, 126–37.
- Gunter, T.C., Friederici, A.D. and Hahne, A.** 1999: Brain responses during sentence reading: visual input affects central processes. *Neuroreport* 10, 3175–78.
- Gunter, T.C., Friederici, A.D. and Schriefers, H.** 2000: Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience* 12, 556–68.
- Hagoort, P., Brown, C. and Groothusen, J.** 1993: The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes* 8, 439–83.
- Hagoort, P., Brown, C. and Osterhout, L.** 1999: The neurocognition of syntactic processing. In Brown, C. and Hagoort, P., editors, *The*

- neurocognition of language*. Oxford: Oxford University Press, 273–316.
- Hahne, A.** 2001: What's different in second language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research* 30, 251–66.
- Hahne, A.** and **Friederici, A.D.** 2001: Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition* 4, 123–41.
- Hoeks, J.C.J., Stowe, L.A.** and **Doedens, G.** 2004: Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research* 19, 123–41.
- Honda, M., Barrett, G., Yoshimura, N., Ikeda, A., Nagamine, T.** and **Shibasaki, H.** 1996: Event-related potentials during paired associate memory paradigm. *Electroencephalography and Clinical Neurophysiology* 100, 419–39.
- Hopp, H.** 2006: Subject–object ambiguities in L2 parsing. In Bamman, D., Magnitskaia, T. and Zaller, C., editors, *Proceedings of the 30th annual Boston University conference on language development*. Boston, MA: Cascadilla Press, 281–92.
- Kaan, E., Harris, A., Gibson, E.** and **Holcomb, P.** 2000: The P600 as a measure of syntactic integration difficulty. *Language and Cognitive Processes* 15, 159–201.
- Kaan, E.** and **Swaab, T.Y.** 2003: Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *Journal of Cognitive Neuroscience* 15, 98–110.
- Kluender, R.** and **Kutas, M.** 1993: Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8, 573–633.
- Kutas, M.** and **Hillyard, S.** 1980: Reading between the lines: event-related brain potentials during natural sentence processing. *Brain and Language* 11, 354–73.
- Münste, T., Heinze, H.** and **Magnum, G.** 1993: Dissociation of brain activity related to syntactic and semantic aspects of language. *Journal of Cognitive Neuroscience* 5, 335–44.
- Mueller, J.L.** 2005: Electrophysiological correlates of second language processing. *Second Language Research* 21, 152–74.
- Neville, H.J., Nicol, J., Barss, A., Forster, K.** and **Garrett, M.F.** 1991: Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3, 151–65.
- Osterhout, L.** 1997: On the brain response to syntactic anomalies: manipulations of word position and word class reveal individual differences. *Brain and Language* 59, 494–522.
- Osterhout, L.** and **Holcomb, P.** 1992: Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31, 785–806.

- 1995: Event-related potentials and language comprehension. In Rugg, M.D. and Coles, M.G.H., editors, *Electrophysiology of mind: event-related brain potentials and cognition*. Oxford: Oxford University Press, 171–215.
- Osterhout, L., Holcomb, P. and Swinney, D.** 1994: Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory and Cognition* 20, 786–803.
- Phillips, C., Kazanina, N. and Abada, S.H.** 2005: ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22, 407–28.
- Proverbio, A.M., Čok, B. and Zani, A.** 2002: Electrophysiological measures of language processing in bilinguals. *Journal of Cognitive Neuroscience* 14, 994–1017.
- Ruchkin, D.S., Johnson, R., Canoune, H. and Ritter, W.** 1990: Short-term memory storage retention: an event-related brain potential study. *Electroencephalography and Clinical Neurophysiology* 76, 419–39.
- Sabourin, L.** 2003: *Grammatical gender and second language processing*. Groningen Dissertations in Linguistics, volume 42. Groningen: CLCG.
- Sabourin, L. and Stowe, L.A.** 2004: Memory effects in syntactic ERP tasks. *Brain and Cognition* 55, 392–95.
- Sabourin, L., Stowe, L.A. and de Haan, G.J.** 2006: Transfer effects in learning an L2 grammatical gender system. *Second Language Research*, 27, 1–29.
- Stowe, L.A. and Sabourin, L.** 2005: Imaging the processing of a second language: effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching* 43, 329–53.
- Tokowicz, N. and 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.M. and Neville, H.J.** 1996: Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8, 231–56.