

The Impact of Proficiency on Syntactic Second-language Processing of German and Italian: Evidence from Event-related Potentials

Sonja Rossi, Manfred F. Gugler, Angela D. Friederici, and Anja Hahne

Abstract

■ The present study investigated the role of proficiency in late second-language (L2) processing using comparable stimuli in German and Italian. Both sets of stimuli consisted of simple active sentences including a word category violation, a morphosyntactic agreement violation, or a combination of the two. Four experiments were conducted to study high- and low-proficiency L2 learners of German as well as high- and low-proficiency L2 learners of Italian. High-proficiency L2 learners in both languages showed the same event-related potential (ERP) components as native speakers for all syntactic violations. For the word category violation, they displayed an early anterior negativity (ELAN), an additional negativity reflecting reference-related processes, and a late P600 evidencing processes of reanalysis. For the processing of the morphosyntactic

error, an anterior negativity (LAN) and a P600 were observed, whereas for the combined violation, the same ERP components were found as in the pure category violation. In high-proficiency L2 learners, the timing of the processing steps was equivalent to that of native speakers, although some amplitude differences were present. Low-proficiency L2 learners, however, showed qualitative differences in the agreement violation characterized by the absence of the LAN and quantitative differences reflected in a delayed P600 in every violation condition. These findings emphasize that with a high proficiency, late L2 learners can indeed show native-like neural responses with the timing approximating that of native speakers. This challenges the idea that there are fundamental differences in language processing in the brain between natives and late L2 learners. ■

INTRODUCTION

Learning a second language (L2) is influenced by different factors that have varying impact on the different language components, such as syntax, semantics, phonology, and morphology, and that determine the nature and duration of the L2 acquisition process.

Two factors in particular have been discussed recently, the age of L2 acquisition and the level of proficiency of language use, the first of which has received more attention until now.

The factor age of acquisition (AoA) has been the focus of a number of studies. Several studies support the basic claim that the earlier the L2 is acquired, the better and faster a high level in the L2 is attained (Birdsong, 1999). Lenneberg (1967) assumes a critical period for language acquisition, whose termination (puberty) corresponds to the loss of adaptability and inability for reorganization in the brain. Bialystok and Miller (1999) showed that individuals who acquired the L2 before the age of 15 did better on a grammaticality judgment task than people who learned the L2 after the age of 15. The beneficial effect of an early AoA was also found in Flege,

Yeni-Komshian, and Liu (1999) regarding phonological aspects such as the accent in L2. Apart from behavioral studies, in recent years the development of various neurophysiological and imaging techniques such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), electroencephalography (EEG), and magnetoencephalography (MEG) have enhanced the understanding of brain mechanisms and allowed for detailed insight into language processes (for a review on imaging studies on bilingualism see Abutalebi, Cappa & Perani, 2001). Imaging studies investigating L2 processing have mostly concentrated on the question of whether the native language (L1) and L2 are localized in the same areas or if they are represented in different brain regions. Kim, Relkin, Lee, and Hirsch (1997) conducted an fMRI study using a production task and found overlapping activations in frontal regions for L1 and L2 in early L2 learners and different activations in late learners. This emphasized the potential importance of an early AoA in the cortical organization of an L2.

In a recent fMRI study (Perani et al., 2003), language exposure or usage in early bilinguals, apart from AoA, was suggested as important factors in L2 acquisition. The results showed a smaller activation in prefrontal regions in bilinguals with a high exposure to the L2 than in bilinguals with less L2 exposure. The lower prefrontal

Max Planck Institute for Human Cognitive and Brain Sciences, Germany

activation was seen in correlation with a lower activation threshold, indicating more automaticity and the need for less neural resources for learners with high L2 exposure.

Apart from imaging studies, a number of event-related potential (ERP) studies have been conducted to determine the temporal dynamics of language processing in L1 and L2. For natives, the temporal course of sentence comprehension evolves in three different processing stages (Friederici, 2002): (1) an initial autonomous phrase structure building process based on word category information (reflected by an early left anterior negativity [ELAN]), (2) a stage during which semantic (reflected by a centroparietal negativity [N400]) and morphosyntactic (reflected by an anterior negativity [LAN]) processing occurs, and (3) a final, more controlled stage in which different types of information are integrated and reanalyzed (reflected by a centroparietal positivity [P600]).

ERP studies on L2 learning have addressed the issue of whether the same processing steps are also observable in L2 learners and to what extent AoA plays a role. Weber-Fox and Neville (1996) investigated semantic and syntactic L2 processing during a reading comprehension task with ERPs. They investigated five groups of Chinese–English bilinguals with different L2 AoAs (1–3, 4–6, 7–10, 11–13, >16 years). The semantic violation, which in natives typically elicits an N400 associated with lexical–semantic integration processes (Kutas & Hillyard, 1980), showed an N400 effect in all bilingual groups; however, the N400 was delayed in bilinguals who acquired the L2 after the age of 11. ERPs for the syntactic violations in the L2 showed even more profound divergence due to late AoA. These effects were characterized by reduced asymmetries of syntactic negativities and the absence of the late positivity in phrase structure violations reported for bilinguals who were exposed to L2 after age 11. Weber-Fox and Neville (2001) found similar AoA dependencies in an ERP study investigating the processing of closed- and open-class words in Chinese–English bilinguals. Open-class words, which are more related to semantic aspects, elicited an N350 component, which did not differ across groups (monolinguals, and bilinguals who had acquired L2 between 1 and 3, 4 and 6, 7 and 10, 11 and 13, < 15 years). Closed-class words, on the other hand, showed latency delays of the N280 component when the L2 was learned after the age of 7. As closed-class words are related to grammatical information, these findings suggest that such processes are more sensitive to AoA.

Another crucial factor in L2 acquisition that has received less attention is the attained proficiency level. A PET study conducted by Perani et al. (1998) showed no differences in brain activation between the L1 and the L2 in late L2 learners with a high language level. Conversely, late learners with a poor L2 level showed a different activation in temporoparietal areas between L1 and L2

(Perani et al., 1996). This result suggests that the acquired proficiency level can play a more important role than L2 AoA. If this hypothesis is valid, the ERP differences between L1 and L2 observed in the study by Weber-Fox and Neville (1996) may be due to the low proficiency of their L2 participants. Recent ERP studies suggest that proficiency has to be considered as a factor influencing the L2 ERP pattern in semantics (Elston-Güttler, Paulmann, & Kotz, 2005; Kotz & Elston-Güttler, 2004) but particularly in syntax (Hahne, 2001; Hahne & Friederici, 2001). Hahne and Friederici (2001) investigated semantic and syntactic processing in Japanese native speakers who started learning German after puberty and who had not reached a very high L2 level. This group showed an N400 effect regarding the semantic condition, whereas in the phrase structure violation condition, they did not show any syntax-related ERP components, that is, neither an ELAN nor a P600 usually observed in native speakers (Hahne & Friederici, 1999, 2002). In another study, Hahne (2001) tested Russian native speakers who had learned German after the age of 10. Participants who had a higher level of L2 performance showed no ELAN, but a late positivity (P600) reflecting reanalysis processes in the syntactic condition. This effect was interpreted in terms of the higher L2 proficiency compared to the Hahne and Friederici study (2001). A recent ERP study (Mueller, Hahne, Fujii, & Friederici, 2005) in which syntactic processing was investigated in German natives who were trained in a miniature grammar of Japanese also found no ELAN but reported a P600 for word category violations. The question therefore arises as to whether high-proficiency late L2 learners can show automatic syntactic first-pass parsing processes apart from controlled processes such as the P600 (Hahne & Friederici, 1999). Friederici, Steinhauer, and Pfeifer (2002) studied syntactic processing in a miniature artificial grammar called Brocanto. They found that trained, high-proficiency participants in contrast to untrained people revealed a biphasic ERP pattern consisting of an early negativity and a P600 for phrase structure violations. These findings suggest that at least in L2 acquisition of a miniature language, a native-like ERP pattern can indeed be elicited.

In sum, we can state that in L2 studies with ERPs, early syntactic processing steps such as the ELAN have only been found in those learning an artificial miniature language. The present study aims to investigate the role of L2 proficiency in auditory sentence comprehension via ERPs in natural languages such as German and Italian. In order to test if late L2 learners with high proficiency show the same ERP components as native speakers and differ from those with low proficiency in terms of syntactic language processing mechanisms, we presented comparable sets of sentences in the active voice to two proficiency groups (high and low) in German and Italian. Three syntactic anomalies, namely, a word category, a morphosyntactic violation, and a combination

of the two, were included. Word category violations typically elicited an early (left) anterior negativity (ELAN) occurring before 300 msec after stimulus onset reflecting phrase structure building processes (Hahne & Friederici, 2002; Friederici, Pfeifer & Hahne, 1993; Neville, Nicol, Barss, Forster & Garrett, 1991) and a late centroparietal positivity (P600) occurring between 500 and 1000 msec after target onset reflecting processes of syntactic integration, reanalysis, and repair (Hahne & Friederici, 1999; Friederici et al., 1993; Osterhout & Holcomb, 1993). The early anterior negativity is assumed to be automatic in nature as it was found not influenced by probability variations or task manipulations. The P600, in contrast, represents a more controlled ERP component (Hahne & Friederici, 2002; Hahne & Friederici, 1999). Two recent studies with German (Rossi, Gugler, Hahne, & Friederici, 2005) and French (Isel, Hahne, & Friederici, 2004) testing word category violations in a prepositional phrase (PP) specifying the noun phrase have also reported a component in addition to the biphasic ELAN-P600 pattern, namely, a broadly distributed negativity preceding the P600. This ERP component was assumed to reflect unsatisfied reference processes due to the omission of a highly expected noun in the PP. Morphosyntactic violations, on the other hand, typically elicited an anterior negativity (LAN) appearing between 300 and 700 msec indicating the detection of the morphosyntactic agreement error, followed again by a P600 reflecting reanalysis processes (Coulson, King, & Kutas, 1998; Gunter, Stowe, & Mulder, 1997; Osterhout & Mobley, 1995; Friederici et al., 1993). Combined violations containing a word category violation usually elicit the same ERP pattern as the pure word category violation, indicating a primacy over semantics (Friederici, Gunter, Hahne, & Mauth, 2004; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001), argument structure (Frisch, Hahne, & Friederici, 2004), or morphosyntax (Rossi et al., 2005), at least concerning early processing steps.

In the present study, four ERP experiments were conducted directly comparing high- (Experiment 1) versus low-proficiency (Experiment 2) late L2 learners of German and high- (Experiment 3) versus low-proficiency (Experiment 4) late L2 learners of Italian. Although Germanic and Romance languages display some differences (e.g., Italian has a less rigid word order and allows a null-subject representation in contrast to German; for a review see Comrie, 1990), similar ERP effects are predicted for the two languages for the kind of violations tested here. Universal neural processing steps are supported by previous ERP studies on morphosyntactic violations, for instance, for which similar processing mechanisms have been found in the following languages: German (Gunter, Friederici, & Schriefers, 2000), English (Osterhout & Mobley, 1995), Dutch (Gunter et al., 1997), and Italian (De Vincenzi et al., 2003; Angrilli et al., 2002). We therefore presented two equivalent material sets, one in German and one in Italian, with

exactly the same structure, word counts, word order, and syntactic violations.

On the basis of the above-mentioned assumptions, we first expected no different ERP components between German and Italian syntactic processing at each of the proficiency levels. Second, concerning the word category violation, we expected the same ERP pattern (ELAN, additional negativity, P600) as observed in native speakers in both high-proficiency L2 groups with approximately the same latency. The ELAN, assumed to be automatic in nature, should be present in high-proficiency L2 learners given the fact that they should have already established an implicit representation of word category information due to the simple sentence structure "subject-verb-object" (S-V-O) realized in the active voice. The active voice is normally learned prior to passive voice in German and Italian language development and is therefore assumed to be processed more easily (Guasti, 2004). The simple S-V-O structure was chosen for the following reasons: (1) to enable equivalent German and Italian material to be created with no word order differences and (2) to ensure that low-proficiency L2 learners have no difficulties with the sentence constructions. In contrast, the ERP pattern elicited in the low-proficiency L2 group was expected to be affected by latency delays or amplitude variations. These differences would be due to the fact that although low-proficiency learners may be aware of these syntactic anomalies, they may have not internalized the rules and/or developed automatized processing routines to deal with them (Paradis, 2004; Ullman, 2001). Third, concerning the morphosyntactic agreement violation, the high-proficiency groups are expected to show the same processing steps, namely, a LAN and a P600, with similar latencies as native speakers. In contrast, low-proficiency learners should again show latency delays or amplitude variation effects. Fourth, the combined violation in high-proficiency learners is expected to resemble the pure word category violation normally obtained in natives (ELAN, additional negativity, P600). Low-proficiency learners, though, may be more disrupted by the presence of two syntactic anomalies. This may cause more processing problems, probably resulting in a more inconsistent ERP pattern. However, if low-proficiency learners also show a pattern similar to that found for the pure word category violation, this would indicate that even low-proficiency late learners follow the same processing steps for double violations as high-proficiency learners or native speakers.

METHODS

Participants

In total, 84 right-handed (assessed according to Oldfield, 1971) subjects took part in the present study. Fifteen participants, equally distributed across groups, were excluded

from statistical analyses because of too many artifacts. The 69 remaining subjects were 26 years of age, on average (range, 19–40 years). Almost all subjects were university students, they were paid for participation, and they had no known hearing deficits and had normal or corrected-to-normal vision. The participants were subdivided into four groups according to their L2 language and proficiency level. Because of lack of standardized proficiency tests equivalent for both German and Italian, the proficiency level was determined according to several behavioral measures such as the language-learning history, self-rating on linguistic proficiency, and performance on translation tests (see Table 1). On the basis of these measures, participants were divided into one of two groups and called high- or low-proficiency L2 learners.

High- versus low-proficiency L2 learners in both languages displayed clear differences regarding many behavioral aspects such as (1) the time spent in L2-speaking countries (low-proficiency learners spent less time there than high-proficiency learners); (2) the L2 learning period (high-proficiency learners had learned L2 for a longer period than low-proficiency L2 learners); (3) the L2 self-rating of listening, reading, speaking, and writing (high-proficiency learners rated themselves higher than low-proficiency learners); (4) the vocabulary translation test (low-proficiency learners made more percentages of errors when translating words from L2 into L1 and vice versa than high-proficiency learners), and (5) the preselection translation test (low-proficiency learners made more percentages of errors when translating from L2

into L1 than high-proficiency learners). In addition, the two groups clearly differed in their accuracy rates for the judgment task, recorded during the ERP session after each heard sentence (low-proficiency L2 learners made more errors than high-proficiency learners).

This division of subjects resulted in four experimental groups that make up the experiments in the present study:

Experiment 1 (high-proficiency L2 learners of German): 16 native Italian speakers (10 women) who had learned German after the age of 10 and had acquired a high L2 proficiency level.

Experiment 2 (low-proficiency L2 learners of German): 19 native Italian speakers (13 women) who had learned German after the age of 10 and had reached a low L2 proficiency level.

Experiment 3 (high-proficiency L2 learners of Italian): 16 native German speakers (12 women) who had learned Italian after the age of 10 and had acquired a high L2 proficiency level.

Experiment 4 (low-proficiency L2 learners of Italian): 18 native German speakers (10 women) who had learned Italian after the age of 10 and had reached a low L2 proficiency level.

Materials

Two sets of stimuli were used, one in German and an equivalent set in Italian (see Table 2). Both sets of

Table 1. Behavioral Proficiency Information

	<i>High-proficiency L2 Learners of German (n = 16)</i>	<i>Low-proficiency L2 Learners of German (n = 19)</i>	<i>High-proficiency L2 Learners of Italian (n = 16)</i>	<i>Low-proficiency L2 Learners of Italian (n = 18)</i>
Age (years)	27.3 ^a	23.1 ^{a,b}	27.5 ^a	24.8 ^{a,b}
Age of acquisition (years)	18.4	18.1	20.8	19.0
Time spent in L2-speaking countries (years)	3.8 ^{a,b}	0.5 ^{a,b}	1.0 ^{a,b}	0.2 ^{a,b}
L2 learning period (years)	8.1 ^{a,b}	3.5 ^{a,b}	4.8 ^{a,b}	2.1 ^{a,b}
L2 self-rating test (6-point scale)				
Listening	4.9 ^a	3.0 ^a	4.6 ^a	2.9 ^a
Reading	4.9 ^a	3.2 ^a	4.7 ^a	3.2 ^a
Speaking	4.4 ^a	3.1 ^{a,b}	4.3 ^a	2.7 ^{a,b}
Writing	4.1 ^a	3.1 ^a	3.8 ^a	2.9 ^a
Vocabulary translation test (errors in %)	8.2 ^a	34.0 ^{a,b}	12.4 ^a	53.4 ^{a,b}
Preselection translation test (errors in %)	7.8 ^a	76.1 ^a	11.1 ^a	70.0 ^a

^aSignificant difference between high- versus low-proficiency L2 learners in each language group.

^bSignificant difference between German versus Italian in each proficiency group.

Table 2. Items in the German and Italian Material

<i>Condition</i>		<i>German Example</i>	<i>Italian Example</i>
Correct	(60 items)	Der Junge im Kindergarten <u>singt</u> ein Lied. (The boy in-the kindergarten <u>sings</u> a song.)	Il signore nel bar <u>beve</u> un caffè. (The man in-the bar <u>drinks</u> a coffee.)
Category	(60 items)	Der Junge im <u>singt</u> ein Lied. (The boy in-the <u>sings</u> a song.)	Il signore nel <u>beve</u> un caffè. (The man in-the <u>drinks</u> a coffee.)
Agreement	(60 items)	Der Junge im Kindergarten <u>singst</u> ein Lied. (The boy in-the kindergarten <u>sing</u> a song.)	Il signore nel bar <u>bevo</u> un caffè. (The man in-the bar <u>drink</u> a coffee.)
Combined	(60 items)	Der Junge im <u>singst</u> ein Lied. (The boy in-the <u>sing</u> a song.)	Il signore nel <u>bevo</u> un caffè. (The man in-the <u>drink</u> a coffee.)
Filler 1	(60 items)	Der Junge im Chor singt eine Hymne. (The boy in-the choir sings a hymn.)	Il signore nel ristorante beve un tè. (The man in-the restaurant drinks a tea.)
Filler 2	(60 items)	Der Junge im Theater singt eine Melodie. (The boy in-the theatre sings a melody.)	Il signore nel locale beve un aperitivo. (The man in-the pub drinks an aperitif.)
Filler 3	(60 items)	Der Junge denkt: Du singst ein Lied. (The boy thinks: You sing a song.)	Il signore pensa: io bevo un caffè. (The man thinks: I drink a coffee.)

Literal translations are included in parentheses. The critical verbs are underlined. Correct, category, agreement, and combined indicate the experimental conditions.

materials were generated independently from each other and were not literal translations of each other to ensure a high degree of naturalness in each language. The level of difficulty and the specificity of the single words were controlled for, so that only simple words and words of intermediate difficulty were included into the materials. Each stimuli set in German and Italian consisted of four critical experimental conditions: *correct sentence*, *word category violation*, *morphosyntactic agreement violation*, and *combined violation*. In addition, three correct filler conditions were included. For each condition, 60 different sentences were generated, resulting in a total of 420 sentences per stimuli set. Both stimuli sets had the same amount of words per sentence per condition and the same word order. Critical sentences were all in the active voice in indicative present tense with the subject noun phrase including the PP followed by the verb and object.

As mentioned above, three critical violation and three filler conditions were included in addition to the correct control condition. In the *category violation*, the verb directly followed the preposition so that the noun expected in the PP was absent. This absence created a word category violation. The prepositions used in the experiment were merged forms consisting of a preposition and a determiner (e.g., *im* = *in dem*—in the).

In the *agreement violation*, the verb was violated morphosyntactically. The correct third person singular inflection was altered and replaced with the incorrect second person singular in German and with the first person singular in Italian. It was not possible to use the same violating person in both languages because the

first person singular in German and the second person singular in Italian would have created correct sentences in the present subjunctive.

Last, the *combined violation* contained both types of error, the category and the agreement error.

Two filler conditions were also constructed to balance the amount of correct and syntactically incorrect sentences with the same sentence structure. A third filler condition was included to balance the morphosyntactic error by using either the first or second person singular in a correct sentence.

The sentences of each stimuli set were spoken by two female native speakers in a soundproof booth and recorded digitally with 16 bits. Because the reading of sentences without a noun after the preposition (as was the case for the phrase structure and combined violation) may alter the prosody of the sentence, a complete PP including a filler word after the preposition was integrated for the recording and edited out from the digital speech file afterward. To ensure the same transition in test and control sentences from the last phoneme of the preposition to the first phoneme of the verb, this additional noun was constructed such that the first phoneme of the filler word corresponded to the first phoneme of the following verb and the last phoneme of the filler word corresponded to the last phoneme of the preposition (cf. Hahne & Friederici, 1999).

The material was pseudorandomized and resulted in 21 different randomizations, which enabled the presentation of one randomization list for each person in each experiment.

Procedure

To control for proficiency in the L2, each participant had to pass a preselection translation test before being allowed to take part. On the phone, we read out 10 German or Italian sentences with the same structure as the experimental trials but composed of quite selective words. Potential subjects were asked to translate these sentences from the L2 to the L1. To belong to the high-proficiency group, participants had to make less than 20% translation errors, and to be accepted into the low-proficiency group, they had to perform 50% correct translations at most.

Before the experiment, the participants were instructed to minimize eye and body movements during the acoustic presentation of the sentences. At the beginning of the experiments, the participants were presented with 21 training trials. After the EEG experiment, the participants performed a vocabulary translation test in which they had to translate lists containing the 60 verbs and the 60 subject nouns from the German and Italian experiments each from the L2 to the L1 and vice versa. This also provided an additional measure of L2 proficiency. Finally, a questionnaire concerning demographic, medical, and general information about the age of L2 acquisition, language-learning history, self-rating scales on linguistic proficiency in both languages, and other speech-relevant information was filled out by each participant (see Table 1).

During the experiment, participants listened to the L2 sentences in a soundproof booth with dimmed lighting and were seated in a comfortable chair 1 m in front of a computer monitor. The presentation of the trials was as follows. First, a fixation star appeared on the monitor for 500 msec, and then the acoustic presentation of the sentence followed via loud speakers for approximately 3500 msec as subjects fixated on the star on the screen to help reduce eye movements. After the auditory presentation, the fixation star was still visible for 1500 msec, then a response sign appeared for 2000 msec to which participants were instructed to perform a grammaticality judgment task. That is, participants were asked to press one of two buttons on a response pad (left and right button for correct sentences were counterbalanced across participants) to judge the grammatical correctness of the sentence. An intertrial interval of 1000 msec followed before the next trial started.

ERP Recording

The EEG was recorded with 23 Ag/AgCl electrodes placed in an elastic cap (EASY CAP, GmbH, Herrsching, Germany) at the following positions: F3, F4, FC3, FC4, F7, F8, FT7, FT8, I_z, C3, C4, T7, T8, Cz, CP5, CP6, P3, P4, P7, P8, Pz, O1, and O2 (nomenclature based on Sharbrough et al., 1991). The vertical electrooculogram (VEOG) was recorded from two electrodes placed above and be-

low the right eye, and the horizontal electrooculogram (HEOG) was recorded from two electrodes at the outer canthus of each eye. The EEG recording was referenced online to the left mastoid and rereferenced off-line to averaged mastoids. Electrode impedance was kept below 3 k Ω , and the EEG signal was digitized with 250 Hz and amplified within a band pass from DC to 70 Hz.

Data Analysis

Behavioral Data

Behavioral data were the accuracy rates (in percent) during the grammaticality judgment task. For each language, a repeated measures ANOVA was computed with the factors Condition (correct, category, agreement, and combined) and Group (high- vs. low-proficiency L2 learners). If the ANOVA revealed a significant main effect of condition ($p < .05$), subsequent paired t tests between the single conditions were computed. The significance level was adjusted according to the false discovery rate method (Benjamini & Hochberg, 1995).

ERP Data

For ERP data analysis for all four experiments, only correctly answered trials in the judgment task were included. Trials affected with artifacts were excluded. If participants had many eye movement artifacts, their trials were corrected with an EOG algorithm. In Experiment 1 (high-proficiency L2 learners of German), 11.2% of trials were rejected, on average (correct, 11.3%; category, 11%; agreement, 12.1%; combined, 10.4%); in Experiment 2 (low-proficiency L2 learners of German), 8.5% of trials were excluded (correct, 8.9%; category, 8.3%; agreement, 8.2%; combined, 8.7%); in Experiment 3 (high-proficiency L2 learners of Italian), 11.8% of trials were excluded (correct, 10.4%; category, 12.2%; agreement, 12.5%; combined, 12%); and in Experiment 4 (low-proficiency L2 learners of Italian), 13.8% of trials were discarded (correct, 13.8%; category, 13.3%; agreement, 13.6%; combined, 14.4%).

ERPs were computed for each participant and each experimental condition and afterward, for each group. A 1500-msec time range after onset of the verb was averaged with a 100-msec poststimulus onset baseline. This baseline was chosen because different word categories preceded the critical verb in the different conditions, a preposition in the category and combined conditions, and a noun in the correct and agreement conditions (cf. Hahne, 2001; Hahne & Friederici, 1999, 2001, 2002). To ensure that in the poststimulus onset baseline time window, no differences due to the different preceding word classes were present, and to justify the application of this 100-msec poststimulus onset baseline, an ANOVA was performed with the factor Condition (correct, category, agreement, combined) in

the 0- to 100-msec time window in each experiment. In all four experiments, the main effect of Condition did not reach significance: Experiment 1, $F(3,45) < 1$; Experiment 2, $F(3,54) = 1.38, p > .10$; Experiment 3, $F(3,45) = 1.79, p > .10$; Experiment 4, $F(3,51) < 1$.

The statistical ERP analyses on the mean amplitudes were conducted separately for the three violation conditions, each compared to the correct condition. The following consecutive time windows were chosen according to the literature and visual inspection of the present data: agreement violation, 500–650, 650–900, and 900–1500 msec; category and combined violation, 100–250, 250–650, 650–1100, and 1100–1500 msec. The following electrodes entered statistical analyses and were subdivided into four regions of interest (ROIs): left anterior (F7, F3, FT7, FC3), right anterior (F8, F4, FT8, FC4), left posterior (CP5, P7, P3, O1), and right posterior (CP6, P8, P4, O2). For the main analyses, repeated measures ANOVAs were computed for each time window with the factors Condition (correct vs. incorrect), Region (anterior vs. posterior), and Hemisphere (left vs. right). To capture differences between high- and low-proficiency L2 learners, the between-subjects factor Group (high- vs. low-proficiency L2 learners) was included. The analyses were conducted separately for each language (German or Italian), that is, Experiments 1 and 2 and Experiments 3 and 4 were compared. All statistical analyses followed a hierarchical schema (cf. Mueller et al., 2005; Rossi et al., 2005; Frisch et al., 2004; Hahne, 2001; Hahne & Friederici, 1999, 2001). Whenever the main analyses revealed a significant interaction between either Region or Hemisphere with the factor Condition at $p < .05$, further ANOVAs were calculated on the next level. Whenever an interaction with both factors Condition and Group reached significance, further step-down ANOVAs in each group were performed. ROI analyses were calculated whenever both factors Region and Hemisphere interacted with the factor Condition alone, or with both Condition and Group. A correction according to Greenhouse and Geisser (1959) was applied and reported here as the corrected significance. Results of the main analyses (involving Condition, Region, Hemisphere, and Group) are reported in tables, whereas analyses on hierarchically lower levels are mentioned in the text.

Additionally, the peak latency of the difference between each incorrect and the correct condition was calculated for the 650- to 1500-msec time window to further determine the latency differences of the P600 component. A t test was performed afterward between the two groups (high- vs. low-proficiency L2 learners) of each language.

Additional ERP Analyses

Apart from the comparison between high- and low-proficiency L2 learners of German or Italian, a compar-

ison between native German speakers from another study (Rossi et al., 2005) and high-proficiency L2 learners of German from the present study (Experiment 1) was performed to determine to what extent these two groups differ or overlap in their processing steps. For this purpose, 30 right-handed native German speakers (22 women; average age 26 years, range 18–36 years) who had listened to the same German material used in the present study were included in these additional analyses. The same statistical analyses procedure (main analyses followed by licensed lower level analyses) was applied as for the comparisons between high- and low-proficiency L2 learners. The time windows were the following: agreement, 500–650 and 650–1500 msec; category and combined, 100–250, 250–650, and 650–1500 msec. The whole P600 650- to 1500-msec time window was chosen for the comparison between native German speakers and high-proficiency L2 learners of German because high-proficiency L2 learners of German displayed reliable positivity effects in both time windows analyzed (650–900 and 900–1500 msec for the agreement; 650–1100 and 1100–1500 msec for the category and combined condition).

RESULTS

Experiments 1 and 2: High- versus Low-proficiency L2 Learners of German

Behavioral Data

High-proficiency L2 learners of German performed the judgment task accurately (93.9%, on average; correct, 96.4%; category, 94.5%; agreement, 87.9%; combined, 96.9%). Low-proficiency L2 learners of German performed the judgment task with an accuracy rate of 86.2%, on average (correct, 90%; category, 81%; agreement, 81.4%; combined, 92.5%). A Condition by Group comparison revealed a significant main effect of Condition, $F(3,99) = 13.36, p < .001$, and a reliable main effect of Group, $F(1,33) = 14.38, p = .001$, but no Condition \times Group interaction. The Group main effect indicated that low-proficiency L2 learners of German performed the judgment task worse than did high-proficiency learners. Subsequent paired t tests (adjusted at $\alpha = .042$, according to Benjamini & Hochberg, 1995) concerning the factor Condition revealed that more errors were made in the category and agreement conditions in contrast to the correct and combined conditions.

ERP Data

The ERP data for the single incorrect conditions versus the correct are displayed in Figure 1A and B. The plots reported are filtered with an 8-Hz low-pass filter for presentation purposes only. The results of the main analyses are displayed in Table 3.

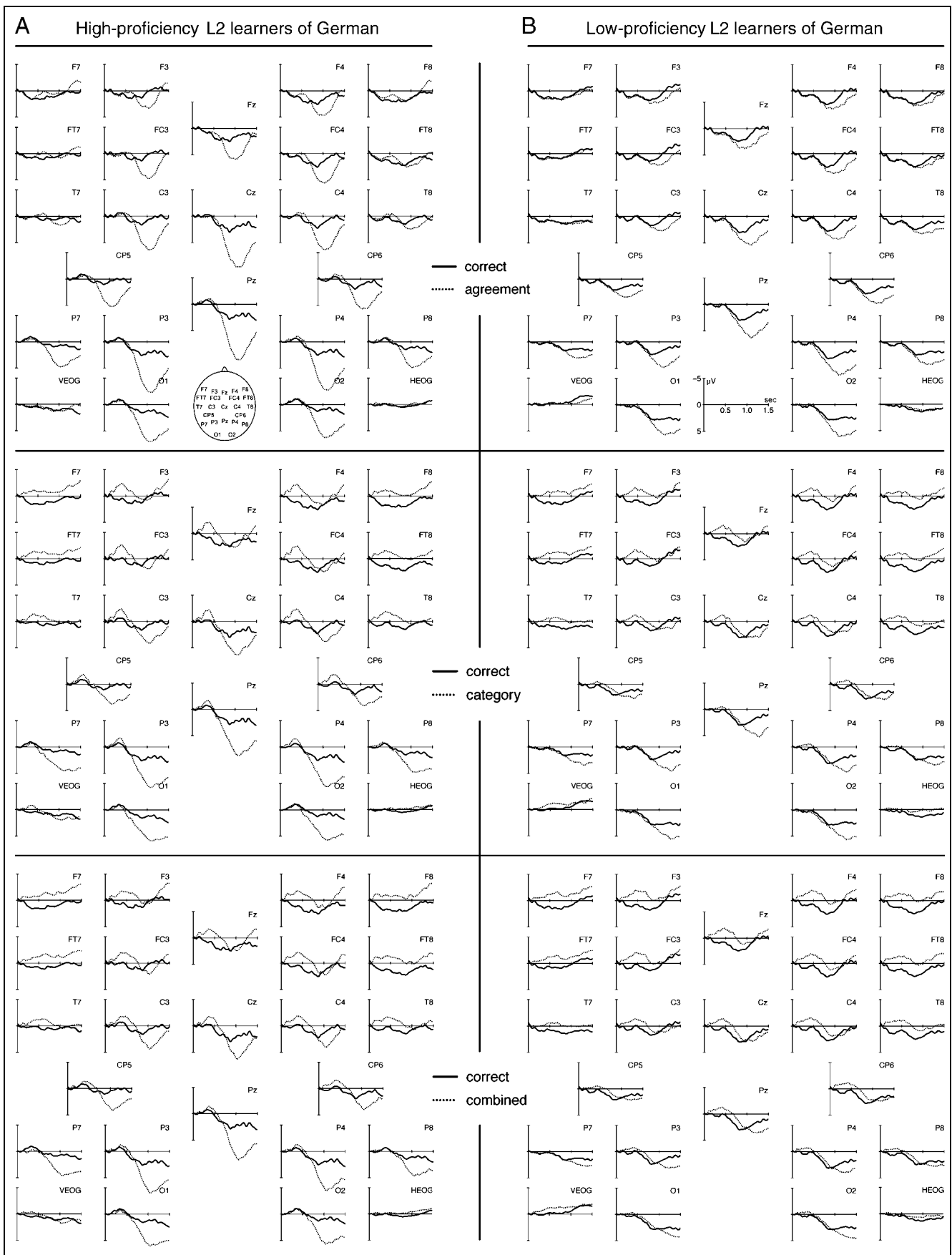


Figure 1. Grand-average ERPs. (A) ERPs for the incorrect (category, agreement, and combined) versus the correct condition from verb onset up to 1500 msec for 16 high-proficiency L2 learners of German (Experiment 1). (B) ERPs for 19 low-proficiency L2 learners of German (Experiment 2). Negative voltage is plotted upward.

Table 3. ANOVAs of ERP Data for Each Experimental Violation versus the Correct Condition in Experiments 1 and 2

Source	df	Experiment 1 vs. 2: High vs. Low-proficiency L2 Learners of German											
		Agreement vs. Correct				Category vs. Correct				Combined vs. Correct			
		500-650 msec	650-900 msec	900-1500 msec		100-250 msec	250-650 msec	650-1100 msec	1100-1500 msec	100-250 msec	250-650 msec	650-1100 msec	1100-1500 msec
Cond	1,33	2.29	7.86**	77.05***	8.42**	33.54***	<1	10.31**	27.89***	44.53***	1.42	1.65	
Cond × Group	1,33	3.72	1.38	3.16	<1	<1	7.06*	1.26	<1	<1	7.12*	1.25	
Cond × Reg	1,33	7.86**	13.66***	59.77***	18.99***	87.81***	107.40***	145.80***	26.38***	73.56***	82.25***	103.26***	
Cond × Reg × Group	1,33	3.47	7.14*	12.74***	<1	2.97	9.37**	11.15**	1.98	1.75	12.45***	12.43***	
Cond × Hem	1,33	<1	<1	1.96	14.15***	3.30	2.78	<1	29.17***	4.13*	3.18	1.40	
Cond × Hem × Group	1,33	1.99	1.59	1.11	<1	<1	<1	<1	1.18	<1	<1	<1	
Cond × Reg × Hem	1,33	<1	1.69	<1	<1	<1	<1	<1	<1	<1	<1	<1	
Cond × Reg × Hem × Group	1,33	1.23	6.48*	<1	<1	<1	<1	<1	2.42	1.03	<1	<1	

The *F* values are reported. Cond = Condition; Reg = Region; Hem = Hemisphere.

**p* < .05.

***p* < .01.

****p* < .001.

The morphosyntactic agreement violation displays a LAN effect for high-proficiency L2 learners of German, whereas no such negativity can be observed for low-proficiency learners. A P600 component is present in both groups. However, the effect differs across groups with respect to the latency as it starts much earlier (from 650 msec on) in high-proficiency learners than in low-proficiency L2 learners (from 900 msec on). In addition, the amplitude of the P600 is reduced in low-proficiency L2 learners in contrast to high-proficiency learners. Concerning the category and the combined violation, both high- and low-proficiency L2 learners show an ELAN effect between 100 and 250 msec and an additional negativity following this. As in the agreement violation, high-proficiency L2 learners start with the P600 from 650 msec on, whereas low-proficiency learners have a delayed and reduced P600 (from 1100 msec on).

Agreement violation. Between 500 and 650 msec, there was a reliable interaction between Condition and Region. As the ERPs concerning the anterior negativity (Figure 1A and B) displayed clear differences between the two groups (low-proficiency L2 learners of German do not show any tendency toward a negativity effect), the marginal (*p* < .1) interactions of Condition × Group and Condition × Region × Group were taken into consideration as well. Subsequent region analyses in each group revealed a negativity effect in anterior regions for high-proficiency L2 learners of German, $F(1,15) = 7.45$, $p < .05$, but no reliable effect for low-proficiency L2 learners. Between 650 and 900 msec, a main effect of Condition and the significant interactions Condition × Region, Condition × Region × Group, and Condition × Region × Hemisphere × Group were found. The four-way interaction reflected a positivity effect in the left posterior, $F(1,15) = 14.29$, $p < .01$, and right posterior, $F(1,15) = 8.28$, $p < .05$, quadrants for high-proficiency L2 learners of German and no reliable effect for low-proficiency learners. Between 900 and 1500 msec, a reliable main effect of Condition and the interactions Condition × Region and Condition × Region × Group reached significance. The three-way interaction indicated positivity effects for both groups, although high-proficiency learners showed a larger P600 in posterior regions, $F(1,15) = 93.23$, $p < .001$, than low-proficiency learners, $F(1,18) = 38.40$, $p < .001$. In addition, low-proficiency learners showed a negativity effect in anterior regions, $F(1,18) = 11.34$, $p < .01$. The peak latency analysis in the 650- to 1500-msec time window revealed a reliable difference between high- and low-proficiency L2 learners of German, $t(33) = -2.51$, $p < .05$, indicating a later P600 peak in low- (1195 msec) than in high-proficiency L2 learners (1098 msec).

Category violation. In the 100- to 250-msec time window, the main effect of Condition and both two-way

interactions Condition \times Region and Condition \times Hemisphere reached significance. Subsequent region and hemisphere analyses revealed reliable negativity effects in the anterior region, $F(1,33) = 24.74, p < .001$, and in the right hemisphere, $F(1,33) = 13.75, p < .001$. No interaction including the factor Group reached significance. Between 250 and 650 msec, a main effect of Condition and a Condition \times Region interaction was present. The interaction resulted in a reliable negativity effect in the anterior region, $F(1,33) = 96.59, p < .001$. Again, no interactions with the factor Group resulted from the main analyses. Between 650 and 1100 msec, reliable interactions Condition \times Group, Condition \times Region, and Condition \times Region \times Group were present. The interaction Condition \times Region \times Group revealed a negativity effect in the anterior, $F(1,15) = 5.56, p < .05$, and a positivity effect in the posterior region, $F(1,15) = 38.14, p < .001$, in high-proficiency L2 learners of German. Low-proficiency L2 learners, in contrast, only showed a negativity effect at anterior sites, $F(1,18) = 11.03, p < .01$, but no posterior P600 in this time window. Between 1100 and 1500 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Group were reliable. Subsequent region analyses in each group showed a larger positivity effect in the posterior region in high-proficiency L2 learners, $F(1,15) = 100.56, p < .001$, than in low-proficiency learners, $F(1,18) = 34.67, p < .001$. In addition, high-proficiency learners showed a negativity effect in anterior regions, $F(1,15) = 8.02, p < .05$. The peak latency analysis in the 650- to 1500-msec time window revealed a reliable difference between high- and low-proficiency L2 learners of German, $t(33) = -2.57, p < .05$. This indicates an earlier P600 peak in high- (1126 msec) than in low-proficiency L2 learners (1225 msec).

Combined violation. In the 100- to 250-msec time window, the main effect of Condition and both two-way interactions Condition \times Region and Condition \times Hemisphere reached significance. Subsequent region and hemisphere analyses revealed reliable negativity effects in the anterior region, $F(1,33) = 67.34, p < .001$, and in the left, $F(1,33) = 9.96, p < .01$, and right hemisphere, $F(1,33) = 45.75, p < .001$. No interaction including the factor Group reached significance. Between 250 and 650 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Hemisphere reached reliability. The interaction Condition \times Region resulted in an anterior negativity, $F(1,33) = 114.50, p < .001$, whereas Condition \times Hemisphere revealed a reliable negativity effect in the left, $F(1,33) = 30.97, p < .001$, and the right hemisphere, $F(1,33) = 48.49, p < .001$. Again, no interactions with the factor Group resulted from the main analyses. Between 650 and 1100 msec, reliable interactions Condition \times Group, Condition \times Region, and Condition \times

Region \times Group were found. The interaction Condition \times Region \times Group revealed a negativity effect in the anterior, $F(1,15) = 9.61, p < .01$, and a positivity effect in the posterior region, $F(1,15) = 25.36, p < .001$, in high-proficiency L2 learners of German but only a negativity effect at anterior sites in low-proficiency L2 learners, $F(1,18) = 18.71, p < .001$. Between 1100 and 1500 msec, the interactions Condition \times Region and Condition \times Region \times Group were reliable. Subsequent region analyses in each group showed a larger positivity effect in the posterior region in high-proficiency L2 learners, $F(1,15) = 38.09, p < .001$, than in low-proficiency learners, $F(1,18) = 6.07, p < .05$. In addition, both groups displayed a negativity effect in anterior regions: high-proficiency L2 learners, $F(1,15) = 6.92, p < .05$; low-proficiency L2 learners, $F(1,18) = 5.90, p < .05$. The peak latency analysis in the 650- to 1500-msec time window revealed a reliable difference between high- and low-proficiency L2 learners of German, $t(33) = -3.58, p < .001$, indicating a later P600 peak in low- (1273 msec) than in high-proficiency L2 learners (1127 msec).

Experiments 3 and 4: High- versus Low-proficiency L2 Learners of Italian

Behavioral Data

High-proficiency L2 learners of Italian showed high accuracy rates in the judgment task (96.6%, on average; correct, 97.6%; category, 97.3%; agreement, 92.6%; combined, 99%). Low-proficiency L2 learners of Italian performed the judgment task accurately as well (86.6%, on average; correct, 91.2%; category, 85.1%; agreement, 78.8%; combined, 91.4%). An ANOVA including the factors Condition and Group revealed a significant main effect of Condition, $F(3,96) = 15.76, p < .001$, and a reliable main effect of Group, $F(1,32) = 15.19, p < .001$. The Group effect suggests that low-proficiency L2 learners of Italian made higher percentages of errors while performing the judgment task than high-proficiency L2 learners. Subsequent paired t tests (adjusted at $\alpha = .042$, according to Benjamini & Hochberg, 1995) concerning the factor Condition revealed that more errors were made in the agreement condition compared to the correct, category, and combined conditions, and fewer errors were made in the combined compared to the category condition.

ERP Data

The ERP data for the single incorrect conditions versus the correct are displayed in Figure 2A and B. The plots reported are filtered with an 8-Hz low-pass filter for presentation purposes only. The results of the main analyses are displayed in Table 4.

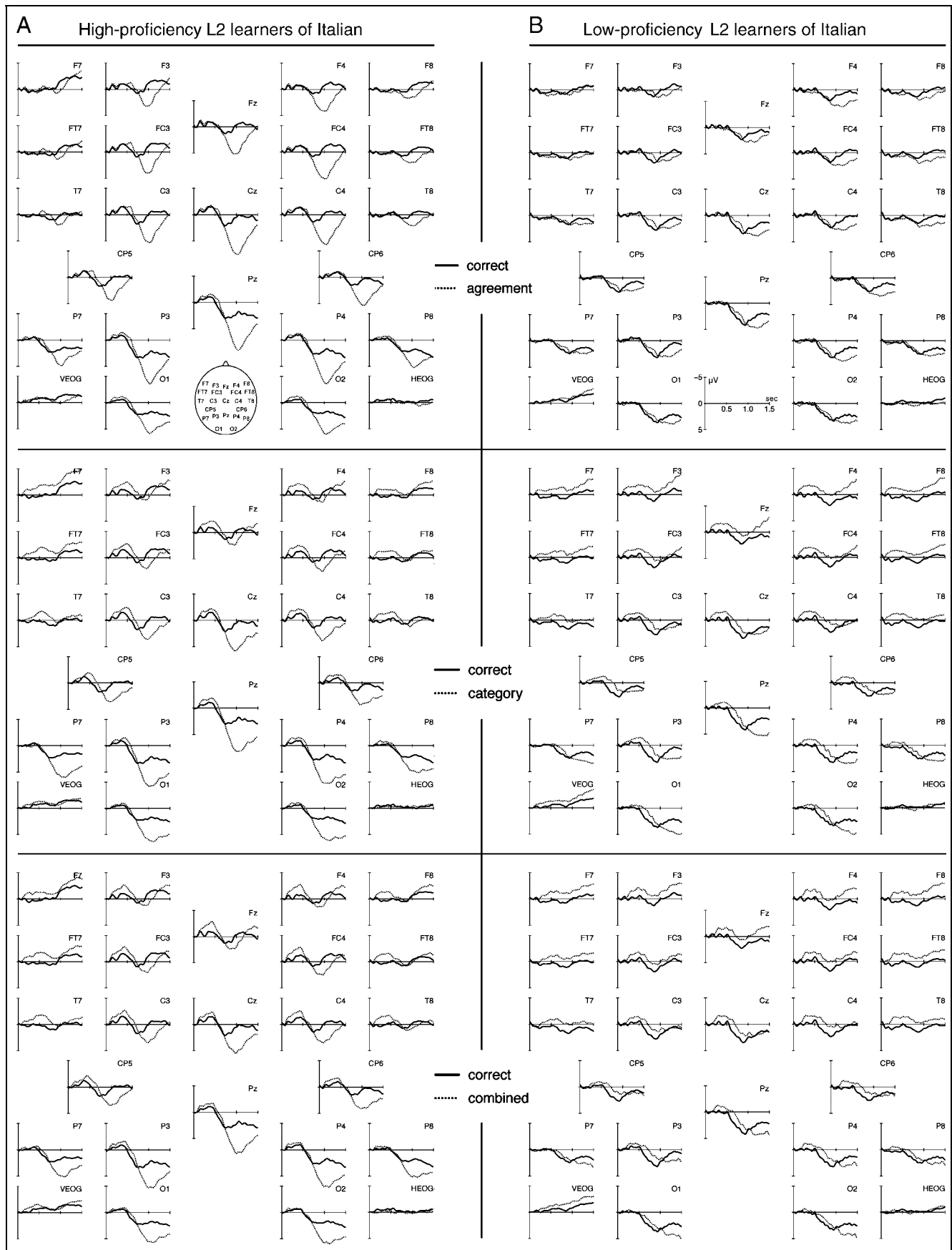


Figure 2. Grand-average ERPs. (A) ERPs for the incorrect (category, agreement, and combined) versus the correct condition from verb onset up to 1500 msec for 16 high-proficiency L2 learners of Italian (Experiment 3). (B) ERPs for 18 low-proficiency L2 learners of Italian (Experiment 4). Negative voltage is plotted upward.

Table 4. ANOVAs of ERP Data for Each Experimental Violation versus the Correct Condition in Experiments 3 and 4

Source	df	Experiment 3 vs. 4: High vs. Low-proficiency L2 Learners of Italian											
		Agreement vs. Correct				Category vs. Correct				Combined vs. Correct			
		500-650 msec	650-900 msec	900-1500 msec	100-250 msec	250-650 msec	650-1100 msec	1100-1500 msec	100-250 msec	250-650 msec	650-1100 msec	1100-1500 msec	
Cond	1,32	<1	<1	49.71***	11.98**	17.54***	<1	3.18	5.06*	18.35***	<1	1.64	
Cond × Group	1,32	1.04	4.41*	6.68*	<1	<1	9.44**	3.38	1.30	<1	13.64***	3.43	
Cond × Reg	1,32	3.64	<1	1.16	7.19*	4.68*	12.90***	50.81***	2.08	8.57**	7.12*	35.01***	
Cond × Reg × Group	1,32	<1	<1	1.47	<1	<1	3.08	<1	<1	<1	1.29	<1	
Cond × Hem	1,32	3.02	6.14*	4.69*	3.32	<1	<1	<1	3.01	<1	<1	<1	
Cond × Hem × Group	1,32	4.55*	<1	<1	<1	<1	<1	<1	1.41	3.82	1.03	<1	
Cond × Reg × Hem	1,32	1.06	3.49	5.63*	<1	1.00	<1	<1	<1	<1	<1	<1	
Cond × Reg × Hem × Group	1,32	<1	<1	<1	<1	<1	2.08	1.27	<1	<1	<1	<1	

The *F* values are reported. Cond = Condition; Reg = Region; Hem = Hemisphere.

**p* < .05.

***p* < .01.

****p* < .001.

Qualitative differences between high- and low-proficiency L2 learners of Italian are visible, especially relative to the LAN effect in the agreement condition. High-proficiency L2 learners show a left-lateralized LAN effect between 500 and 650 msec. Low-proficiency L2 learners do not show any reliable negativity effect, neither between 500 and 650 msec nor in a delayed time window between 650 and 900 msec (a tendency toward a negativity is visible in this latter time window in Figure 2B, but the effect does not reach significance). The P600 effect starts earlier and has a larger amplitude in high- than in low-proficiency L2 learners of Italian. Concerning the category and the combined violations, both groups have an ELAN effect between 100 and 250 msec and an additional negativity following this. Substantial differences again concern the timing of the P600 effect: high-proficiency L2 learners begin reanalysis processes earlier (from 650 msec on) than low-proficiency L2 learners of Italian (from 1100 msec on).

Agreement violation. In the 500- to 650-msec time window, only a reliable Condition × Hemisphere × Group interaction reached significance. Subsequent hemisphere analyses in each group revealed a left-lateralized negativity effect in high-proficiency L2 learners of Italian, $F(1,15) = 5.21, p < .05$, whereas low-proficiency L2 learners did not show any reliable effect. Between 650 and 900 msec, the interactions Condition × Group and Condition × Hemisphere reached significance. The Condition × Hemisphere interaction did not show any reliable effect in the left and right hemispheres. The resolution of the Condition × Group interaction showed marginal effects in high-, $F(1,15) = 3.90, p < .10$, and no reliable effect in low-proficiency L2 learners, $F(1,17) = 1.15, p > .10$. The marginal effect in high-proficiency learners indicates a tendency toward a positivity in this time range. Between 900 and 1500 msec, the main effect of Condition and the interactions Condition × Group, Condition × Hemisphere, and Condition × Hemisphere × Region were found. The three-way interaction resulted in reliable positivity effects in all quadrants: left anterior, $F(1,32) = 10.98, p < .01$; right anterior, $F(1,32) = 24.90, p < .001$; left posterior, $F(1,32) = 34.97, p < .001$; right posterior, $F(1,32) = 37.14, p < .001$. The resolution of Condition × Group reflected a larger positivity effect for high-proficiency, $F(1,15) = 43.55, p < .001$, L2 learners of Italian than for low-proficiency learners, $F(1,17) = 10.66, p < .01$. The peak latency analysis between 650 and 1500 msec showed a marginally significant difference between high- and low-proficiency L2 learners of Italian, $t(32) = -1.69, p = .10$. The P600 peak occurred at 1115 msec in high- and at 1185 msec in low-proficiency L2 learners of Italian.

Category violation. In the 100- to 250-msec time window, a main effect of Condition and an interaction

Condition \times Region were found. Subsequent region analyses revealed a reliable negativity effect in the anterior region, $F(1,32) = 15.84, p < .001$. No interactions with the factor Group were present. Between 250 and 650 msec, a main effect of Condition and a Condition \times Region interaction were present. Region analyses revealed a larger negativity effect in anterior, $F(1,32) = 16.17, p < .001$, than in posterior areas, $F(1,32) = 9.54, p < .01$. Between 650 and 1100 msec, the main analyses revealed reliable interactions Condition \times Group and Condition \times Region. The latter interaction revealed no significant effects in both regions, whereas the interaction Condition \times Group revealed a positivity effect for high-proficiency learners, $F(1,15) = 4.20, p < .05$, and a negativity effect in low-proficiency L2 learners, $F(1,17) = 5.31, p < .05$. Between 1100 and 1500 msec, the interaction Condition \times Region was reliable and resulted in a negativity effect in anterior, $F(1,32) = 5.44, p < .05$, and a positivity effect in posterior, $F(1,32) = 42.94, p < .001$, regions. The peak latency analysis in the 650- to 1500-msec time window revealed a reliable difference between high- and low-proficiency L2 learners of Italian, $t(32) = -2.29, p < .05$, indicating an earlier P600 peak in high- (1138 msec) than in low-proficiency L2 learners (1246 msec).

Combined violation. Between 100 and 250 msec, only the main effect of Condition reached significance. Between 250 and 650 msec, the main effect of Condition and the interaction Condition \times Region were present. Subsequent region analyses showed a larger negativity effect in anterior, $F(1,32) = 22.90, p < .001$, than in posterior areas, $F(1,32) = 5.72, p < .01$. In the time range 650–1100 msec, the main analyses revealed reliable interactions Condition \times Group and Condition \times Region. The resolution of Condition \times Region resulted in no effects at both regions. The interaction Condition \times Group showed a positivity effect for high-proficiency L2 learners of Italian, $F(1,15) = 6.83, p < .05$, and a negativity effect for low-proficiency learners, $F(1,17) = 6.95, p < .05$. Between 1100 and 1500 msec, only a Condition \times Region interaction was found, resulting in a significant negativity effect in anterior, $F(1,32) = 5.71, p < .05$, and a positivity effect in posterior areas, $F(1,32) = 26.23, p < .001$. The peak latency analysis between 650 and 1500 msec showed a reliable difference between high- and low-proficiency L2 learners of Italian, $t(32) = -2.09, p < .05$, indicating a later P600 peak in low- (1228 msec) than in high-proficiency L2 learners (1142 msec).

Native German Speakers versus High-proficiency L2 Learners of German

The ERP data for the single incorrect conditions versus the correct condition in native German speakers are displayed in Figure 3. The plots reported are filtered

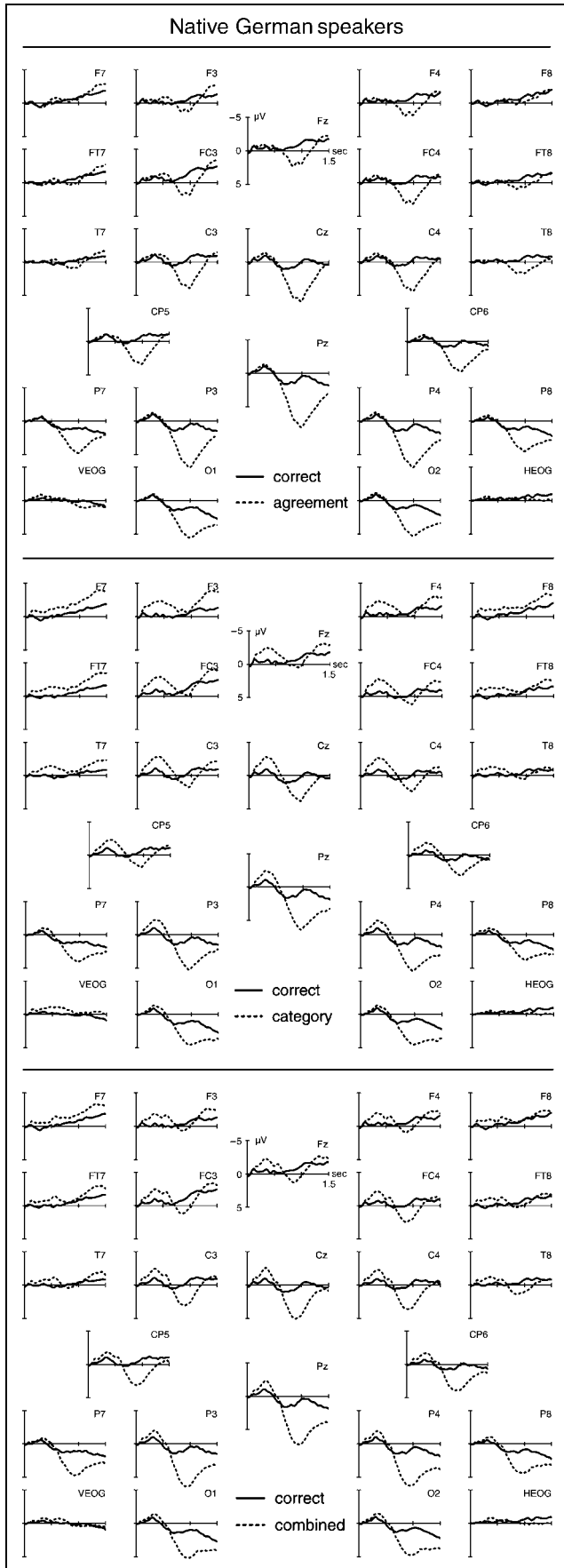
with an 8-Hz low-pass filter for presentation purposes only. The results of the main analyses are displayed in Table 5.

ERP Analyses

Agreement violation. Between 500 and 650 msec, the main analyses revealed a main effect of Condition and a reliable interaction Condition \times Region. The resolution of the interaction reflected a negativity effect in anterior regions, $F(1,44) = 10.50, p < .01$. Between 650 and 1500 msec, the main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Hemisphere reached reliability. The three-way interaction resulted in positivity effects in three out of four quadrants: right anterior, $F(1,44) = 9.36, p < .01$; left posterior, $F(1,44) = 160.34, p < .001$; right posterior, $F(1,44) = 165.50, p < .001$. No interaction including the factor Group was found.

Category violation. In the time range 100–250 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Hemisphere were present. Region analyses showed a larger negativity effect in the anterior, $F(1,44) = 36.91, p < .001$, than in the posterior region, $F(1,44) = 6.23, p < .05$. Hemisphere analyses revealed negativity effects in both the left, $F(1,44) = 15.52, p < .001$, and the right hemisphere, $F(1,44) = 30.34, p < .001$. No interactions including the factor Group were present. Between 250 and 650 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Group reached significance. The three-way interaction reflected a negativity in anterior, $F(1,29) = 24.50, p < .001$, and posterior regions, $F(1,29) = 10.70, p < .01$, in native German speakers and a larger negativity in anterior areas in high-proficiency L2 learners of German, $F(1,15) = 45.52, p < .001$. Between 650 and 1500 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Group were found. The three-way interaction resulted in a negativity effect in anterior [native German speakers, $F(1,29) = 6.54, p < .05$; high-proficiency L2 learners of German, $F(1,15) = 8.56, p < .01$] and a positivity effect in posterior regions for both groups [native German speakers, $F(1,29) = 36.09, p < .001$; high-proficiency L2 learners of German, $F(1,15) = 86.73, p < .001$].

Combined violation. In the time range 100–250 msec, a main effect of Condition and the interactions Condition \times Region, Condition \times Region \times Group, Condition \times Hemisphere, and Condition \times Hemisphere \times Group were present. Region analyses showed a negativity effect in anterior, $F(1,29) = 14.08, p < .001$, and posterior areas, $F(1,29) = 7.72, p < .01$, for native



German speakers, whereas hemisphere analyses in this group revealed a negativity in the left, $F(1,29) = 13.62$, $p < .001$, and right hemisphere, $F(1,29) = 12.30$, $p < .01$. For high-proficiency L2 learners of German, region analyses showed a negativity in anterior regions, $F(1,15) = 37.85$, $p < .001$, whereas hemisphere analyses showed a negativity effect in the right hemisphere, $F(1,15) = 18.83$, $p < .001$. Between 250 and 650 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Group reached significance. The three-way interaction reflected a negativity in anterior, $F(1,29) = 14.86$, $p < .001$, and posterior regions, $F(1,29) = 5.18$, $p < .05$, in native German speakers and a larger negativity in anterior areas, $F(1,15) = 51.73$, $p < .001$, in high-proficiency L2 learners of German. Between 650 and 1500 msec, a main effect of Condition and the interactions Condition \times Region and Condition \times Region \times Group were found. The three-way interaction resulted in a positivity effect in posterior regions for both native German speakers, $F(1,29) = 79.05$, $p < .001$, and for high-proficiency L2 learners of German, $F(1,15) = 47.83$, $p < .001$. In addition, high-proficiency L2 learners showed a negativity effect in anterior regions, $F(1,15) = 9.51$, $p < .01$.

DISCUSSION

The present ERP study investigated the role of L2 proficiency in late L2 learners of German and Italian. The age of L2 acquisition did not differ, either between high- versus low-proficiency L2 learners or between German and Italian learners. Two equivalent stimuli sets in German and Italian containing three syntactic violations (word category, agreement, and a combination of both errors) were acoustically presented. The experiments provided direct comparisons between high- and low-proficiency groups in several syntactic violation conditions, all of which are discussed in detail below.

Word Category Violation

Overall, the ERP pattern of components (ELAN, additional negativity, P600) shown by high-proficiency L2 learners of German and Italian was comparable to what has been found previously in native German speakers (Rossi et al., 2005; Hahne & Friederici, 1999, 2002; Friederici et al., 1993). This indicates that with high proficiency, the same processing steps can be attained as native speakers, namely, (1) the process of phrase

Figure 3. Grand-average ERPs for 30 native German speakers for the incorrect (category, agreement, and combined) versus the correct condition from verb onset up to 1500 msec. Negative voltage is plotted upward. Modified from Rossi et al. (2005) with permission from Elsevier.

Table 5. ANOVAs of ERP Data for Native German Speakers versus High-proficiency L2 Learners of German

<i>Native German Speakers vs. High-proficiency L2 Learners of German</i>									
<i>Source</i>	<i>df</i>	<i>Agreement vs. Correct</i>		<i>Category vs. Correct</i>			<i>Combined vs. Correct</i>		
		<i>500–650 msec</i>	<i>650–1500 msec</i>	<i>100–250 msec</i>	<i>250–650 msec</i>	<i>650–1500 msec</i>	<i>100–250 msec</i>	<i>250–650 msec</i>	<i>650–1500 msec</i>
Cond	1,44	4.90*	81.19***	25.16***	34.71***	11.20**	24.91***	35.38***	15.07***
Cond × Group	1,44	1.19	1.18	<1	<1	<1	<1	1.50	1.84
Cond × Reg	1,44	10.04**	95.73***	19.92***	47.87***	141.62***	19.86***	22.62***	116.80***
Cond × Reg × Group	1,44	1.38	2.62	<1	15.75***	5.71*	6.00*	8.55**	6.05*
Cond × Hem	1,44	<1	1.49	7.75**	<1	<1	10.51**	1.21	<1
Cond × Hem × Group	1,44	2.13	1.75	3.69	2.88	3.07	9.32**	2.12	3.71
Cond × Reg × Hem	1,44	3.40	4.33*	<1	<1	<1	1.58	<1	1.62
Cond × Reg × Hem × Group	1,44	2.80	<1	<1	<1	<1	<1	1.14	1.04

The *F* values are reported. Cond = Condition; Reg = Region; Hem = Hemisphere.

**p* < .05.

***p* < .01.

****p* < .001.

structure building reflected by the early anterior negativity between 100 and 250 msec; (2) the processing of reference-related information reflected by the additional negativity between 250 and 650 msec; and (3) reanalysis processes reflected by the late positivity between 650 and 1500 msec. These findings underline that even late L2 learners are able to attain native-like language processing steps of sentence comprehension. Even early automatic processes such as that reflected by the early anterior negativity can be achieved in L2 learners with increasing proficiency.

Low-proficiency L2 learners of German and Italian also displayed a three-component ERP pattern, but with some quantitative differences in contrast to high-proficiency L2 learners. In particular, the P600 was delayed in low-proficiency L2 learners of both German and Italian. This component started at 1100 msec in low-proficiency learners in contrast to the much earlier onset (from 650 msec on) in high-proficiency learners. It should be noted that the P600 in L2 learners was also reported to be delayed in Hahne (2001). Apart from the latency delay, low-proficiency L2 learners of German also displayed a clear amplitude reduction. The latency delay and amplitude reduction indicate some reanalysis problems in low-proficiency learners, perhaps because they need more time to initiate the reanalysis of the sentence and to try to integrate the syntactically incongruent word. This seems plausible under the assumption that low-proficiency L2 learners cannot use enough resources to reanalyze the sentence and, thus, take more time to integrate all syntactic features.

Surprisingly, an ELAN effect was also found for low-proficiency L2 learners, irrespective of L2 language. This can be attributed to the active sentence structure of the materials used in the present study, in contrast to other L2 studies that used passive sentences (Hahne, 2001; Hahne & Friederici, 2001) and obtained no early negativity. The development of first-language (L1) acquisition may provide some hints in explaining this difference. In L1 acquisition, passive voice is learned after active voice and in general, passive sentences are more complex and usually require additional auxiliary verbs and moving elements to another position (Guasti, 2004). This difference in complexity in passive versus active sentences was also found in ERP studies with children. Although the ELAN with the same latency and distribution as in adults was not present until the age of 13 when word category violations in passive constructions were presented acoustically (Hahne, Eckstein, & Friederici, 2004), a left anterior negativity similar to that of adults was observed in children between 31 and 34 months when word category violations embedded in active sentences were presented (Oberecker, Friedrich, & Friederici, 2005). Moreover, the presentation setting in the present study, which contained syntactic violations and no semantic ones, may have facilitated participants to concentrate on syntactic processing of a restricted amount of syntactic rules. The relatively high accuracy rates during the grammaticality judgment task in low-proficiency L2 learners of both languages (low-proficiency L2 learners of German made 81% correct judgments concerning the category

violation and low-proficiency L2 learners of Italian judged 85.1% correct) provide additional evidence that the low-proficiency L2 learners in the present study realized the word category violation quite well, despite their low proficiency as defined by other behavioral measures (listed in Table 3). This suggests that low-proficiency learners recognized word category information relevant for the early syntactic ERP component quite well.

In L2 studies up until now, this early syntax-related component has only been observed in high-proficiency L2 learners in an ERP study investigating an artificial miniature grammar (Friederici et al., 2002). Kubota, Ferrari, and Roberts (2004) showed a similar early syntactic magnetic component in an MEG study in response to NP raising phrase structure violations in English. The present study provides ERP evidence that this early anterior negativity can also be elicited for phrase structure violations by low-proficiency L2 learners both of German as well as of Italian in simple active voice sentences.

To investigate whether the ERP components relative to word category violations in high-proficiency L2 learners coincide in time with the processing steps observed in native speakers, a direct comparison was performed between native German speakers from another study (Rossi et al., 2005) who listened to the same German material as used in the present study and high-proficiency L2 learners of German in Experiment 1. All ERP components occurred in the same time ranges. However, amplitude variations were present concerning both the additional negativity and the P600, but not for the ELAN. L2 learners, despite their high proficiency, showed larger amplitudes than native speakers. This signals an increase in the resources necessary to process reference-related information and to syntactically reanalyze sentences in an L2.

In summary, the processing of word category information shows no qualitative differences between high- and low-proficiency L2 learners of German and Italian as the same ERP components (ELAN, additional negativity, P600) were present in the L2 as in the L1. There were, however, quantitative differences, reflected in particular by a clear latency delay of the P600 in low-proficiency L2 learners of German and Italian. The temporal similarity of the ERP components in high-proficiency L2 learners and native speakers underlines the importance of high proficiency in achieving native-like processing mechanisms and for approaching native-like timing. However, amplitude differences remain with respect to some processing steps. The presence of an ELAN component in low-proficiency learners indicates that in the context of simple active sentences, early phrase structure building processes can occur, even when high proficiency has not been obtained. Lastly, the same ERP components in the same latency ranges were found in both German and Italian for each proficiency group, respectively. This provides evidence for universal syntactic comprehen-

sion processing mechanisms, at least within the family of Indo-European languages.

Morphosyntactic Agreement Violation

High-proficiency L2 learners of German and Italian showed the same ERP components as native speakers (Rossi et al., 2005; De Vincenzi et al., 2003; Friederici et al., 1993), namely, a LAN reflecting the detection of the morphosyntactic agreement error and a P600 reflecting reanalysis processes. A qualitative difference arose in low-proficiency L2 learners of German and Italian who failed to show a LAN. Visual inspection of the agreement violation in low-proficiency L2 learners of Italian (Figure 2B) reveals a tendency toward a negativity in a delayed time window (650–900 msec) prior to the P600, but this effect did not reach significance. In both low-proficiency groups, the P600 was delayed (from 900 msec on), as was the case in category violations. These latency delays may reflect processing problems when initiating the different processing steps, in this case the process of reanalysis. Both low-proficiency groups also showed a clear amplitude reduction of the P600, indicating that they could not fully establish reanalysis processes.

Concerning the absence of the LAN effect in low-proficiency L2 learners, another factor contributing to this result might be considered. In L1 acquisition, word category information in terms of knowing what is a noun, a verb, and so forth is acquired quite early. First words, which are mostly nouns, are produced by the child at about 12 months of age. Between 18 to 24 months, two-word utterances follow, mostly consisting of a noun and a verb without any morphology. Morphological features arise at about 2 years of age and consolidate until the age of 5 and beyond (Guasti, 2004). Parallel development might be observed in late L2 acquisition. In an initial stage, category information is acquired, and in following stages, the whole variety of morphological features follows (Hawkins, 2001). Because morphosyntax seems to be acquired later than word category information, it is plausible to assume that low-proficiency L2 learners have more problems in processing morphosyntactic aspects than they do in processing word category information. This would explain why the low-proficiency participants did not show reliable LAN effects, but they did show a reliable ELAN while processing word category violations embedded in simple sentences.

The comparison between native German speakers and high-proficiency L2 learners of German revealed no temporal or amplitude differences across groups, either for the LAN or for the P600 component. The relatively late occurrence of the LAN component in these two groups (500–650 msec) might be due to the acoustic presentation of the sentences in contrast to many other studies in the visual modality (Coulson et al., 1998; Gunter et al., 1997; Osterhout & Mobley, 1995) in which the LAN occurred from 300 msec on. Because the mean length of

the German verbs used in the agreement violation of the present study was 629 msec, and the morphosyntactic error was encoded in the last two phonemes of the verb, it is possible that the detection of the error, and the resulting LAN effect, were somewhat delayed.

Combined Violation

In all groups, the combined violation resembled the pure word category violation, giving rise to the same ERP components in the same time windows (ELAN, additional negativity, P600). High-proficiency L2 learners of both German and Italian showed these components in the same time windows as the category violation. This was also the case for both low-proficiency groups, indicating the same latency effects as in the pure category violation. Again, the P600 in low-proficiency learners was delayed as compared to high-proficiency learners. Analogous to the word category violation, low-proficiency L2 learners of German also showed a reduced P600 amplitude. These differences reflect the same reanalysis problems in the combined violation as in the category violation. Overall, the findings concerning the combined violation are in accord with previous L1 studies that report comparable effects for pure word category violations and a combined violation including a word category violation (Friederici et al., 2004; Frisch et al., 2004; Hahne & Friederici, 2002). Thus, no additivity or interaction effects between both anomalies are reported in the combined violation, at least concerning early processes. This reflects a primacy of phrase structure building processes not only over semantics (Friederici et al., 2004; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001), but also over other syntactic features such as morphosyntactic subject–verb agreement processes (Rossi et al., 2005).

The comparison between native German speakers and high-proficiency L2 learners of German revealed no timing differences, but variations in amplitude, especially visible in the additional negativity. As in the category violation, this component had a larger amplitude in L2 learners in contrast to native speakers, indicating more processing problems. Unlike the category violation, the P600 amplitude in the combined condition was comparable across groups. This might result from the fact that a double-error condition is more salient than a single-error condition and, thus, L2 learners might reanalyze a combined violation by using resources comparable to those of native speakers. However, this explanation remains speculative, as further research is needed to investigate the fine-grained mechanisms engaged while processing double violations.

Conclusion

It is still an open debate whether only an early AoA provides the necessary prerequisite for reaching native-

like L2 processing or if this can even be fully achieved after childhood. The present study demonstrates that despite a late age of L2 acquisition, high L2 proficiency learners can show native-like brain processing mechanisms during sentence comprehension; moreover, they approach native-like timing of these processes. Although the AoA was kept constant over both groups, some qualitative and quantitative ERP differences were present between high- and low-proficiency L2 learners. High-proficiency L2 learners of German and Italian who learned L2 after the age of 10 displayed the same ERP components as previously observed in other studies with native speakers concerning all syntactic violations presented. Low-proficiency L2 learners, however, displayed some qualitative differences in the agreement violation, where no reliable LAN effect was found. In addition, some quantitative differences between high- and low-proficiency L2 learners were observed in the agreement, word category, and combined violations. This was visible in a delayed and reduced P600 effect in low-proficiency L2 learners, indicating more uncertainty and problems during syntactic reanalysis. In word category violations, the presence of an ELAN in both high- and low-proficiency learners especially indicates that initial phrase structure building processes can even take place in an L2 in a context of natural languages if sentences are simple and embedded into an active structure. In accordance with previous findings, the combined violation showed the same ERP pattern as the pure category violation in all groups. In general, this study also provides evidence for universal syntactic processing mechanisms, at least within Indo-European languages, as the very same processing steps were observed in both German and Italian.

Acknowledgments

This research was supported by the Max Planck Society and grants by the University of Salzburg awarded to S. R. We thank three anonymous reviewers for helpful comments and Kerrie Elston-Güttler for native English proofreading.

Reprint requests should be sent to Sonja Rossi, Max Planck Institute for Human Cognitive and Brain Sciences, PO Box 500 355, 04303 Leipzig, Germany, or via e-mail: rossi@cbs.mpg.de.

REFERENCES

- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4, 179–190.
- Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccirelli, L., et al. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: An event-related potential study. *Neuroscience Letters*, 322, 5–8.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289–300.

- Bialystok, E., & Miller, B. (1999). The problem of age in second-language acquisition: Influences from language, structure, and task. *Bilingualism: Language and Cognition*, 2, 127–145.
- Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Erlbaum.
- Comrie, B. (1990). *The world's major languages*. Oxford: Oxford University Press.
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13, 21–58.
- De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., et al. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language*, 85, 280–296.
- Elston-Güttler, K. E., Paulmann, S., & Kotz, S. A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience*, 17, 1593–1610.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78–104.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science*, 6, 78–84.
- Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *NeuroReport*, 15, 165–169.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1, 183–192.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, U.S.A.*, 99, 529–534.
- Frisch, S., Hahne, A., & Friederici, A. D. (2004). Word category and verb-argument structure information in the dynamics of parsing. *Cognition*, 91, 191–219.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Guasti, M. T. (2004). *Language acquisition: The growth of grammar*. Cambridge, MA: Bradford.
- 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.
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34, 660–676.
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30, 251–266.
- Hahne, A., Eckstein, K., & Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *Journal of Cognitive Neuroscience*, 16, 1302–1318.
- Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 194–205.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13, 339–356.
- Hahne, A., & Jescheniak, J. D. (2001). What's left if the jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research*, 11, 199–212.
- Hawkins, R. (2001). *Second language syntax: A generative introduction*. Oxford: Blackwell.
- Isel, F., Hahne, A., & Friederici, A. D. (2004). *Neurochronometry of the syntax-semantic interplay: ERP evidence from spoken French relative clauses*. Poster presented at the AMLaP, Aix-en-Provence, France.
- Kim, K. H. S., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, 388, 171–174.
- Kotz, S. A., & Elston-Güttler, K. E. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, 17, 215–235.
- Kubota, M., Ferrari, P., & Roberts, T. P. L. (2004). Human neuronal encoding of English syntactic violations as revealed by both L1 and L2 speakers. *Neuroscience Letters*, 368, 235–240.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and non-native speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17, 1229–1244.
- Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151–165.
- Oberecker, R., Friedrich, M., & Friederici, A. D. (2005). Neural correlates of syntactic processing in two-year-olds. *Journal of Cognitive Neuroscience*, 17, 1667–1678.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97–113.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes*, 8, 413–437.
- Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34, 739–773.
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins.
- Perani, D., Abutalebi, J., Paulesu, E., Brambati, S., Scifo, P., Cappa, S. F., et al. (2003). The role of age of acquisition and language usage in early, high-proficient bilinguals: An fMRI study during verbal fluency. *Human Brain Mapping*, 19, 170–182.
- Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *NeuroReport*, 7, 2439–2444.
- Perani, D., Paulesu, E., Sebastian Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, 121, 1841–1852.

- Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, *384*, 228–233.
- Sharbrough, F., Chatrian, G. E., Lesser, R. P., Lüders, H., Nuwer, M., & Picton, T. W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200–202.
- Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*, 105–122.
- Weber-Fox, C. M., & 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.
- Weber-Fox, C. M., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research*, *44*, 1338–1353.