



HAL
open science

Electrophysiological correlates of second language processing

Jutta L. Mueller

► **To cite this version:**

Jutta L. Mueller. Electrophysiological correlates of second language processing. *Second Language Research*, 2005, 21 (2), pp.152-174. 10.1191/0267658305sr256oa . hal-00572089

HAL Id: hal-00572089

<https://hal.science/hal-00572089>

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.

Electrophysiological correlates of second language processing

Jutta L. Mueller *Max Planck Institute for Human Cognitive and Brain Sciences*

The aim of this article is to provide a selective review of event-related potential (ERP) research on second language processing. As ERPs have been used in the investigation of a variety of linguistic domains, the reported studies cover different paradigms assessing processing mechanisms in the second language at various levels, ranging from phoneme discrimination to complex sentence processing. Differences between ERP patterns of first language (L1) and second language (L2) speakers can help to specify and to test predictions derived from models of L2 processing or hypotheses concerning critical periods for some aspects of second language acquisition. The studies currently available suggest that ERPs are indeed sensitive to qualitative and quantitative differences in L2 speakers with regard to on-line processing.

I Introduction

Research on language processing has provided interesting insights into the human capacity to decode and integrate different types of linguistic information – such as prosodic, phonemic, semantic, syntactic and pragmatic information – needed in order to comprehend and produce an uncountable variety of meanings. While the nature and time course of the interplay of the different subsystems underlying language processing have not been clarified beyond dispute in first language (L1) processing, even less is known about how second language (L2) speakers process their first and second language. A basic question in second language research is how two or more languages are processed with respect to each other. It could be the case that the L2 forms a completely

separate system, relying on completely different processing mechanisms. However, it is equally conceivable that both L1 and L2 have access to either partially overlapping or identical processes. The study of similarities and dissimilarities of first and second language processing has yielded much evidence regarding possible intermediate and end-states in L2 acquisition and the influence of proficiency, as well as other factors on access and use of the L2. Many studies have reported persisting difficulties even in very advanced L2 speakers in the syntactic (Johnson and Newport, 1989; Weber-Fox and Neville, 1996) and phonological domain (Flege *et al.*, 1999). Yet, there are several studies on ultimate attainment in L2 acquisition that report cases of L2 speakers who are indistinguishable from L1 speakers on the basis of their linguistic behaviour in the above mentioned domains, and therefore seem to have reached native-like linguistic skills (Birdsong, 1992; Bongaerts, 1999). In the lexical domain the most extensively studied question concerns the degree of second language autonomy in lexical access and representation, depending on language context and proficiency level. A number of studies have shown that at least the initial stages of lexical processing in bilinguals are not driven by any one language, but there is also considerable evidence that this might be modified by linguistic and nonlinguistic context and proficiency level (for more detailed information, see Schreuder and Weltens, 1993).

Nonetheless it is important to bear in mind that the absence of differences in behavioural measures taken from L1 and L2 speakers does not necessarily mean that the underlying neural processing mechanisms are the same. Similarly, differences in behavioural measures, such as reaction times, are not necessarily the result of the involvement of different neuronal structures, even if they show qualitatively different patterns. Neurophysiological measures can add valuable information about timing and degree of activation of neural networks and therefore provide physiological correlates of behavioural differences. Therefore the question of convergence or divergence of L1 and L2 processing mechanisms is difficult to solve solely on the basis of behavioural data. Neurophysiological methods directly (Electroencephalography (EEG), Magnetoencephalography (MEG)) or indirectly (Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI)) measuring neural activity in the brain during language processing offer

a new window into the underlying structures and mechanisms of L2 processing. While methods tied to haemodynamic response such as PET or fMRI provide reliable information about the spatial location of the activated brain region with poor temporal resolution, electrophysiological measures (EEG, MEG) give a very precise real-time measure in the range of milliseconds with poor spatial accuracy. Depending on the research question each methodology can be fruitful in the study of language.

This article is restricted to the specific contributions of event-related potentials (ERPs) for research on L2 processing. Due to methodological constraints in the application of ERPs to the study of language production,¹ the studies reported in this review focus on L2 comprehension. After a short introduction to ERP methodology, selected studies of different domains of L2 processing research are reviewed and some possible future directions of research are pointed out.

II ERPs as a measure of language comprehension mechanisms

Language comprehension is characterized by its remarkable speed, despite the high complexity of the input. With a time resolution in the millisecond range, ERPs are a well-suited tool for the investigation of the nature and timing of processes in the short period between encoding of the acoustic signal and comprehension of the utterance. ERPs are small scalp-recorded voltage changes in the continuously measured electroencephalogram (EEG). They are precisely time-locked to an external event that is assumed to be related to the processing mechanism under study. In language research the external event usually comprises the presentation of a visual or auditory speech stimulus. As the signal-to-noise ratio in a single trial event is very small it is necessary to average EEG samples over many trials. The averaging procedure leads to an increase of the event related part of the signal and to a decrease of the part of the signal which is due to random variation (noise). The positive and negative deflections in the ERP, which are

¹ERPs cannot be recorded during natural speech production as articulatory movements cause muscularly induced artifacts in the signal. The same is true for eye movements and blinks, which should be avoided during the ERP epoch or mathematically corrected.

typically observed in relation to specific experimental conditions, are referred to as ERP components. ERP components can vary in amplitude, topography, latency, and polarity. Qualitative variations in ERPs of different experimental conditions (i.e., with distinct polarity and/or topography) can either stem from different neural generators or from differences with regard to the relative contribution of the generators (Rugg, 1999). ERPs can therefore help dissociate cognitive operations involved in different conditions on the neural level. However, it is important to keep in mind that similar ERP components do not allow for the conclusion that neural processes are identical. Similarities could also be due to the relative insensitivity of the EEG to a large part of brain activity. In order to produce measurable electrical fields on the scalp surface, nerve cells have to fire synchronously in very large groups and they must be configured in parallel in order to produce a so-called ‘open-field’. Due to these restrictions, most of the activity measured by ERPs stems from apical dendrites of pyramidal cells in the cortex. Activity from differently structured neuronal assemblies, as for example from many subcortical brain structures, therefore remains undetected (for more detailed information on ERPs, see Rugg and Coles, 1995; for methodological issues, see Picton, 2000). In the following, ERPs related to specific domains of language processing are introduced and L2 processing studies are reviewed in more detail. As ERP components have proven to be sensitive to different types of linguistic information, they provide information that can help disentangle different subprocesses in language comprehension. As stimuli were presented visually in the majority of studies, modality is mentioned only if it was auditory.

III ERPs in L2 lexical–semantic processing

The most extensively studied ERP component directly linked to language processing is the N400 (see Figure 1a). The N400 was first reported by Kutas and Hillyard (1980) who visually presented normal sentences (‘He spread the warm bread with butter’), sentences containing a semantic anomaly (‘He spread the warm bread with socks’) and sentences ending with a semantically congruent but physically deviating word, written in capitalized letters (‘He spread the warm bread with

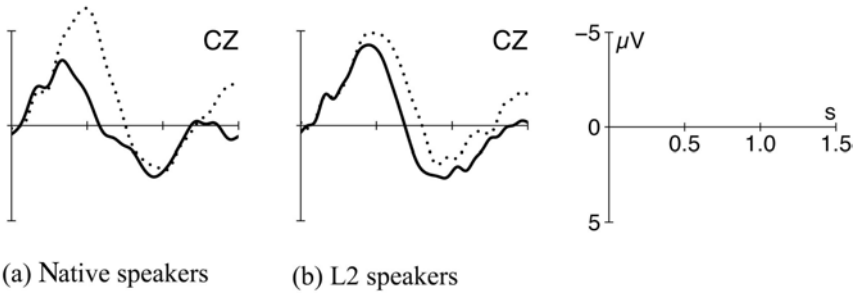


Figure 1 N400 in native speakers (a) and Russian L2 speakers (b) of German. Semantically anomalous words (dotted line) elicit a larger N400 at a central electrode position (CZ) than semantically normal words (solid line) (adapted from Hahne, 2001)

BUTTER’). They found an enhanced negativity peaking at around 400ms post-stimulus onset time-locked to the semantically incongruent word. The negativity was not present for the physical deviation. Here, a positivity (P300) was observed, a component which has been reported in many other studies in the case of infrequent and task-relevant attended stimuli (Donchin, 1981). The discovery of the N400 stimulated much experimental research regarding the specific nature of its underlying processes. Outright semantic violations are neither necessary nor sufficient for the occurrence of an N400. On the one hand it has been shown that the N400 decreases in amplitude or even vanishes when the same semantic violation is presented repeatedly (Besson *et al.*, 1992); on the other hand it is observed also for semantically correct words with a low ‘cloze probability’ that quantifies the expectancy of a specific word at a certain position in a sentence (Kutas and Hillyard, 1984). Moreover, presentation of nonwords leads to an increased N400 (Holcomb and Neville, 1990). In general it can be said that the N400 increases as a function of predictability of a word within its semantic context, which can range from a single word to general world knowledge (for review, see Kutas and Federmeier, 2000). The N400 has therefore been appreciated as a marker of the difficulty of semantic integration processes. In a series of studies the N400 has been used as a tool for investigating the nature of semantic processing mechanisms in second language learners with different ages of acquisition and proficiency levels. The studies referred to in this section make use

of the violation paradigm or the semantic priming paradigm. The violation paradigm tests for semantic integration processes at the sentence level while the semantic priming paradigm tests for lexical processes at the single word level.

First, some studies using semantic violation conditions are summarized. Ardal *et al.* (1990) visually presented normal and semantically anomalous sentences to fluent bilinguals in their L1 (French) and L2 (English) and to English monolinguals. An N400 was present in all cases, but it was delayed for the bilinguals' L2. Kutas and Kluender (1991) reported similar results, namely a delayed N400 with a reduced amplitude for the less fluent language of bilingual participants. Weber-Fox and Neville (1996) investigated the processing of semantic anomalies within sentences as well as syntactic violations in L2 learners who had acquired their L2 (English) at different ages (1–3, 4–6, 7–10, 11–13, >16 years). The N400 was found for monolinguals and all second language learner groups. However, bilinguals who learned their L2 after the age of 11 showed a delay of the peak latency of approximately 20 ms.

Hahne (2001) found an N400 for semantically anomalous words in Russian late L2 learners of German, however, with a reduced amplitude (cf. Figure 1b). Hahne (2001) attributes the difference in the N400 violation effect to an increased and delayed N400 for semantically normal words in the L2 speakers as she did not find any significant differences in the N400 for anomalous words between L1 and L2 speakers. In a parallel study with native Japanese late learners of German with a comparatively lower proficiency level, Hahne and Friederici (2001) found no significant difference with regard to the N400 between the Japanese L2 speakers of German and native speakers of German.

A paradigm that has been used in a variety of studies to uncover the nature of word-level lexical–semantic processes in the bilingual mental lexicon is the semantic priming paradigm. Semantic priming is described as the facilitation of a response to a target that is semantically related to a preceding prime. Numerous studies have reported semantic priming effects as indicated by response-time (RT) differences within languages in L1 (for review, see Neely, 1991) and L2 (e.g., French-Mestre and Prince, 1997), as well as across languages from L1 to L2 and from L2 to L1 (for review, see Kroll, 1993). In native speakers

semantic priming has also been observed to correlate with the N400 component. Semantically unrelated words give rise to an enhanced N400 compared to related words (e.g., Bentin *et al.*, 1985). Although it is a matter of debate whether the N400 observed in this context represents solely processes of post-lexical integration or also pre-lexical processes such as automatic spreading activation (for review, see Osterhout and Holcomb, 1995) it is interesting to investigate to what degree findings from L1 studies generalize to L2 studies. Furthermore, the priming paradigm also makes the examination of cross-linguistic lexical–semantic effects possible and can therefore be used to test hypotheses of language specificity with regard to lexical and conceptual processing.

Kotz (2001) and Kotz and Elston-Guettler (2004) used categorical and associative semantic priming in L2 speakers to test predictions of the revised hierarchical model of the bilingual lexicon (Kroll and Stewart, 1994). This model assumes a weaker conceptual link for L2 words than for L1 words and thus predicts weaker categorical priming effects for L2 speakers as compared to L1 speakers. Kotz (2001) tested proficient early L2 learners (age of acquisition <4 years) of English, while Kotz and Elston-Guettler (2004) tested late L2 learners of English (age of acquisition >11 years) with a high or low level of proficiency. Both categorical (*heart–liver*) and associative priming (*heart–love*) are known to influence the N400 component in L1 speakers. Kotz (2001) reported native-like N400 patterns for categorical and associative priming in early proficient bilinguals, which indicates that word–word and word–concept links in L2 were equally strong in the group of L2 speakers as in natives. Kotz could therefore not confirm the predictions of the revised hierarchical model. Kotz and Elston-Guettler (2004) found differences between natives and L2 speakers with a high vs. low level of proficiency. L2 speakers showed an N400 effect for associative priming irrespective of proficiency level but no N400 effect for categorical priming. Associative priming led to a longer lasting negativity in the high proficiency group compared to the low proficiency group. The authors interpret this pattern of results as suggesting that the development of a direct link to the conceptual level in L2 is strongly influenced by age of acquisition. Associative priming, on the other hand, was influenced by proficiency level, which is also indicated by an

enhanced and longer lasting N400 effect for the high proficiency group (Kotz and Elston-Guettler, 2004). These two studies demonstrate how ERPs can be used to track down subtle differences in semantic processing of an L2 that otherwise might have remained undetected, since the reaction times in the above experiments did not yield any evidence for categorical priming effects for either L1 or L2 speakers.

Besides the investigation of lexical processing mechanisms within L2 or L1 the ERP priming paradigm has been used to investigate language-selective access and degree of top-down control of bilingual lexical processing when two languages possibly compete with each other. The study of interlingual homographs (orthographically identical word forms that have a different meaning in the L1 and L2) can provide information about the influence of a nontarget language on the target language. De Bruijn *et al.* (2001) tested effects of language context on semantic priming in the L2 (English) of Dutch native speakers. De Bruijn *et al.* visually presented triplets of words with the first word serving as language context (Dutch vs. English). The following two items functioned as prime and target, and could be semantically related or unrelated in the L2 (*house–angel–heaven* vs. *house–angel–bush*). As all primes were interlingual homographs (*angel* means ‘sting’ in Dutch) the possible influence of language context – set by the language activated by the first word on the relatedness effect – was put to test. Although semantic relatedness reduced the amplitude of the N400, no influence of the language context could be shown. The authors therefore argue in favour of language nonselective lexical access, as immediate Dutch language context did not inhibit the English meaning of the interlingual homograph. While nonselective access might account for the abovementioned findings within a bilingual language setting, a generalization to other language contexts seems premature. Taking Grosjean’s theory of ‘language mode’ into consideration (Grosjean, 1997), one might assume that selectivity of lexical access should vary depending on individual and situational factors.

In summary, the results from ERP studies in the lexical–semantic domain point to similarities between L1 and L2 speakers rather than to differences. The frequent observation of a latency delay or a reduced amplitude of the N400 effect (Ardal *et al.*, 1990; Kutas and Kluender, 1991; Hahne, 2001) suggests that differences between L1 and L2

processing are more quantitative than qualitative in nature. Differences might consist mainly in a slowdown or decrease in efficiency of semantic processing mechanisms. The most striking difference in the N400 between L1 and L2 processing is seen in the absence of the N400 for categorical priming in a late-learned L2 (Kotz and Elston-Guettler, 2004), which indicates that the use of categorical semantic information in L2 is indeed a process that is not easily established in late L2 learning.

IV ERPs in L2 morphological and syntactic processing

The discovery of an ERP component specific for semantic processing has promoted the search for ERP components related to syntactic processing. The pattern that emerged from studies in L1 syntactic processing, however, is less consistent than in the lexical–semantic domain. Many language comprehension studies using syntactic violation conditions have reported negativities that are more pronounced at anterior or left anterior electrodes occurring in the same time window as the N400 (300–500 ms). This component, often referred to as LAN (left anterior negativity), has been found frequently for morphosyntactic errors such as violations of tense, number or gender agreement (Weyerts *et al.*, 1997; Coulson *et al.*, 1998). It has therefore been interpreted as an index of morphosyntactic processing mechanisms (Friederici, 2002). Another view that has been proposed relates it to more general verbal working memory related processes (e.g., Coulson *et al.*, 1998). Some more recent studies helped to disentangle the two interpretations by demonstrating that working memory related negativities can be measured globally over the whole sentence, while morphosyntactic LAN effects are measurable only locally after the violated element of the sentence (e.g., King and Kutas, 1995). In word category violations a very early left anterior negativity (ELAN) with a latency between 100–300 ms has been found (Neville *et al.*, 1991; Friederici *et al.*, 1993; Hahne and Friederici, 1999) (see Figure 2a). The ELAN in response to a word that does not fit the anticipated word category has been interpreted to reflect rapid first-pass parsing processes. The parsing system here is assumed to retrieve only word category information in order to build an initial phrase structure (Friederici, 2002).

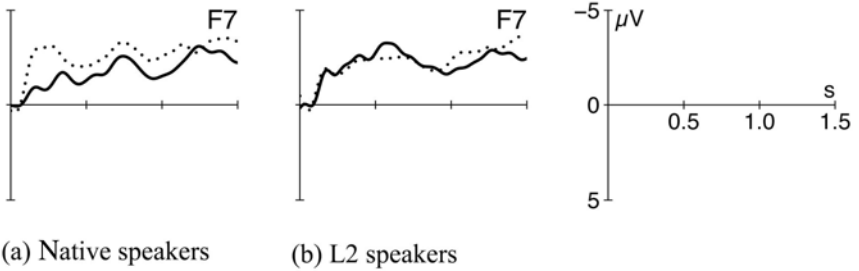


Figure 2 ERPs in response to syntactically correct and incorrect sentences in native speakers (a) and Russian L2 speakers (b) of German. Syntactically incorrect words (dotted line) elicit a larger ELAN at a left anterior electrode position (F7) than syntactically correct words only in German native speakers (solid line) (adapted from Hahne, 2001)

Another ERP component that has been reported frequently in the syntactic domain is the P600 or SPS (syntactic positive shift), a positive wave peaking at about 600ms and usually centro-parietally distributed (see Figure 3a). The P600 occurs for syntactic violations (Neville *et al.*, 1991; Hahne and Friederici, 1999) for syntactically complex structures (Kaan *et al.*, 2000) and for less preferred structures in ambiguous sentences (Osterhout *et al.*, 1994). Thus, it has been interpreted as reflecting processes of reanalysis and syntactic repair (Osterhout *et al.*, 1994; Friederici, 2002). In a more general way, the P600 is regarded to

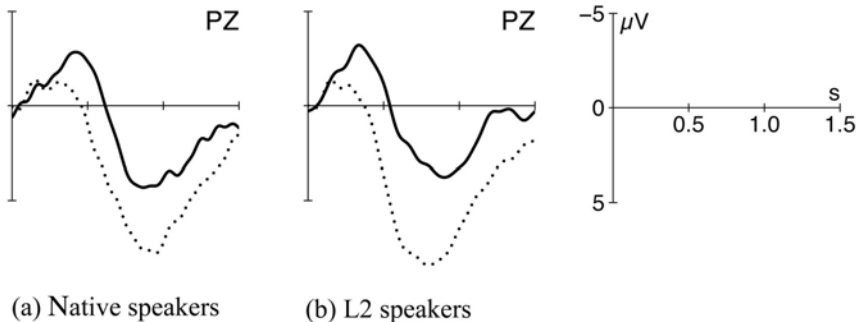


Figure 3 P600 in native speakers (a) and Russian L2 speakers (b) of German. Syntactically incorrect words (dotted line) elicit a larger P600 at a centro-parietal electrode position (PZ) than syntactically correct words (solid line) (adapted from Hahne, 2001)

indicate the difficulty of syntactic integration (Kaan *et al.*, 2000). Compared to the (E)LAN, the P600 seems to reflect a higher level of control, as it may be influenced by the percentage of incorrect items within the experiment (Hahne and Friederici, 1999).

The study by Weber-Fox and Neville (1996) was the first ERP study with L2 speakers focusing on syntactic processing. The study was designed to test Chinese–English bilinguals for critical period effects in the semantic and the syntactic domain. Participants in the study started using English at different ages (1–3, 4–6, 7–10, 11–13, >16 years of age). Phrases containing word category violations (*The scientist criticized Max's of proof the theorem* vs. *The scientist criticized Max's proof of the theorem*) led to both an early (N125) and a later left lateralized negativity (N300–500) in natives. Additionally, a P600 in the time window between 500–700 ms was found for L1 speakers. For the L2 speakers, the pattern differed from that of native speakers irrespective of age of acquisition. Yet, the syntactic ERPs were differentially affected across groups. Interestingly, the early left negativity (N125) was not observed in the L2 learners, except for the group with age of acquisition between 11–13, who showed a reversed left–right topographical distribution. In this group the N125 effect was larger over right scalp sites. The later negativity (N300–500) was present in all L2 groups, but bilaterally distributed if the age of acquisition was greater than 11 years. The P600 effect was similar to native speakers in the groups up to the age of 10 years. The ‘11–13’ years group displayed a delayed positivity starting at about 700 ms and in the group ‘>16’ years no positivity could be found in response to phrase structure violations (at least within the time window of 900 ms post stimulus onset the authors looked at). The authors argued in favour of a differential sensitivity of the syntactic processing system to age of acquisition effects. Neville and colleagues (Neville *et al.*, 1992; Weber-Fox and Neville, 2001) also compared the processing of open-class and closed-class words in L1 and L2 speakers. Open-class words are content words such as nouns or verb stems and closed-class words are grammatical function words. Open-class and closed-class words both elicited negativities in the ERPs, but with different temporal and topographical characteristics. While the negativity for open-class words was broadly distributed and peaked at 350 ms, the negativity for closed-class words was

stronger on left anterior electrode sites and peaked earlier (280ms). Neville *et al.* (1992), who tested native speakers of American Sign Language (ASL) in their L2 English, did not observe the N280 for closed-class words, while the N350 for open-class words was similar to English native speakers. Weber-Fox and Neville (2001) tested native Chinese L2 speakers with the same material. The negativities for both word classes was observed in the L2 speakers, but the peak latency of the N280 for closed-class words was delayed if the L2 was learnt after the age of 7 years. This was not the case for the N350 for open-class words. Weber-Fox and Neville (2001) noted that possible differences in the proficiency level between the ASL and the Chinese group might account for the observed differences in the N280. Nonetheless both studies indicate a differential susceptibility of function word processing to age of acquisition. However, the presence and topography of the left anterior negativity for the Chinese L2 group suggests that a similar network is involved in L1 and L2 function word processing even if a delay is observed in the L2 speakers.

Hahne (2001) and Hahne and Friederici (2001) investigated phrase structure processing in late L2 learners of German. Similar to Weber-Fox and Neville (1996) they used word category violations and compared correct sentences (*Der König wurde ermordet*. 'The king was killed.') to syntactically incorrect sentences (*Der König wurde im ermordet*. 'The king was in the killed'). During auditory sentence processing native speakers of German show an early left anterior negativity (ELAN) followed by a P600 for syntactically incorrect sentences (see Figures 2a and 3a). Hahne and Friederici (2001) reported that native Japanese late L2 learners of German showed neither of the two ERP effects in response to word category violations, but instead showed a greater P600 for the correct sentences as compared to native speakers. To explain the absence of the P600 effect the authors suggest that the L2 speakers might already have to recruit an upper level of processing capacities for the processing of correct sentences, possibly leading to a kind of 'ceiling effect'. Alternatively, one could conclude that Japanese participants recruit first-pass and second-pass parsing processes that are not reflected in any ERP effects. Clearly, in the majority of cases syntactically incorrect sentences were correctly judged as incorrect (76.9%) by the Japanese participants, indicating that

the incorrect syntactic structure was identified and processed in some way despite the absence of a P600 modulation.

In contrast to Japanese speakers, native Russian L2 learners of German showed a P600 for syntactically incorrect sentences with a slightly delayed peak latency (Hahne, 2001). As was the case in the Japanese group, no ELAN effect was observed for the Russian group (see Figures 2b and 3b). These studies indicate that, at least for the emergence of the P600, proficiency might play a crucial role. As reflected in the ratio of correct grammaticality judgements in the syntactic condition (92.9%) the Russian group was more proficient than the Japanese group (Hahne, 2001). One additional factor directly or indirectly influencing the ERP differences between the two groups might be related to the presence of similar syntactic structures in the L2 learners' L1. While syntactic structures like those used in the German test sentences (prepositional phrases) are familiar to native Russian speakers, they are unknown to native speakers of Japanese. Nonetheless, there was no indication for the availability of relatively automatic syntactic processes as reflected in the ELAN for either group.

To address the issue of how different native languages influence the learnability of specific syntactic phenomena in the L2, Sabourin (2003) compared the processing of finiteness, subject–verb agreement and grammatical gender violations in native Dutch speakers and German, English and Romance L2 speakers of Dutch. For her native participants Sabourin found a P600 for each condition. The L2 speakers displayed a P600 for the finiteness violation, where incorrect infinitives occurred instead of finite verbs. For subject–verb agreement violations a P600 was found to be similar to natives in the German group, but for the English and the Romance group it was found to be less widely distributed. The most interesting finding of Sabourin's study is the observation that for the gender agreement violation the only group that showed a P600 effect at all was the group with the most similar gender system in the native language, which was the German group. The author suggests that native-like processing mechanisms, as reflected in the P600, are only acquired if the respective grammatical feature is similar in the native language. However, it cannot be ruled out that these ERP differences resulted from subtle differences in proficiency rather than from transfer of processing mechanisms directly, as the German group

performed better than the other learner groups in the grammaticality judgement during the experiment. A further problem of Sabourin's study might be the unequal sample size across groups (Dutch: 29; German: 19; English: 17; Romance: 12), which might have hampered the statistical power in the L2 groups. Nonetheless, this study demonstrates how ERPs can be used to test predictions concerning possible transfer processes from L1 to L2.

In a recent study Hahne *et al.* (2003) investigated the processing of regular and irregular participles and plurals in late L2 learners of German. In native speakers of German it has been found that incorrect applications of regular and irregular suffixes lead to differential ERP effects, namely a LAN (and a P600) for the incorrect use of regular suffixes, and an N400 for incorrect use of irregular suffixes (Weyerts *et al.*, 1997). Since the LAN can be interpreted as indicating morphosyntactic processes and the N400 as indicating lexical–semantic processes, these findings suggest a dual-route model that assumes rule-based decomposition for regular words and lexical storage for irregular words (Pinker, 1991). Hahne *et al.* found that participles and plurals with incorrect irregular suffixes both elicited an N400 in L2 speakers. Whereas incorrect regular suffixes led to a LAN and a P600 effect in the participle system, only a P600 was found for plurals with incorrect regular suffixes. The authors argue that, as for native speakers, two processing routes are available to L2 speakers. Hahne *et al.* point to the increased difficulty for L2 speakers of German in processing plurals vs. participles to explain the absence of the LAN for incorrectly applied regular plural suffixes. The development of relatively automatic morphosyntactic processes therefore seems possible but may be influenced by the complexity of the morphosyntactic domain under study.

The correlation between proficiency and certain syntactic ERP effects is also underlined by a study conducted by Friederici *et al.* (2002) on auditory processing of an artificially constructed miniature language (Brocanto). Due to the small size of Brocanto (14 lexical elements), the participants, whose native language was German, were able to achieve a high level of performance in the production and comprehension of Brocanto sentences. After learning, participants were presented correct Brocanto sentences and sentences containing word category violations. A biphasic ERP pattern consisting of an early

anterior negativity and a P600 was found in response to violations of phrase structure rules even for structures that were totally new to the learners and could not be transferred from the L1. This finding is in concord with the assumption that proficiency plays a major role in the development of neural correlates of syntactic processing. However, it is not clear if mechanisms underlying artificial language processing are qualitatively the same as processes underlying natural language processing. A first study by Mueller *et al.* (2004) using a miniature language paradigm with naturally spoken Japanese sentences revealed differences between Japanese native speakers and learners of the miniature Japanese in syntactic ERP patterns. Participants reached a high proficiency level, but did not display an anteriorly focused negativity as was seen for word category violations in Japanese participants. However, a P600 was observed for the same conditions in both groups of participants (natives and learners).

In sum, the picture regarding syntactic ERP effects in L2 learners is twofold. On the one hand, some striking similarities to native speakers were discovered, notably the P600 (Weber-Fox and Neville, 1996; Hahne 2001; Hahne *et al.*, 2003; Sabourin, 2003), which suggests that more controlled syntactic parsing processes can at least in principle be acquired in a number of grammatical domains. On the other hand, the processes underlying the ELAN or LAN effect seem to be comparatively difficult to acquire if the L2 is learnt late (Hahne 2001; Hahne and Friederici, 2001; Weber-Fox and Neville, 1996). Studies using (artificial) miniature languages that ensure a very high proficiency level of the trained participants (Friederici *et al.*, 2002; Mueller *et al.*, 2004), as well as studies of near-native L2 speakers seem to be a promising area of research that could help specify conditions necessary for the development of specific syntactic ERP components.

V ERPs in L2 speech sound categorization and segmentation

There is a limited number of studies on L2 speakers that investigate comprehension processes preceding lexical–semantic and syntactic analysis. Before the listener can assign structure and meaning to a given speech stream he or she must distinguish relevant phonemes accurately and detect the onset and offset of words. An ERP component that is

sensitive to detection of changes in the auditory modality is the mismatch negativity (MMN), peaking at 100–250 ms post stimulus onset. It is measurable when physically ‘deviating’ stimuli are presented randomly in sequences of very frequent ‘standard’ stimuli. Elicitation of the MMN does not require participants to listen actively to the stimulus, and has been interpreted as an index of pre-attentive processes (for review, see Näätänen, 2001). Its independence from conscious attention makes it suitable to study language processing mechanisms even in young children. It has been shown that the MMN is not only sensitive to acoustic stimulus properties, but also to categorical distinctions between speech sounds. Näätänen *et al.* (1997) have shown that the amplitude of the MMN increases if standard and deviant stimuli are phoneme prototypes of the native language. The language specific MMN is probably based on the formation of permanent memory traces for the native language phonemes within the first year of life (Cheour *et al.*, 1998). For second language learners the acquisition of appropriate distinctions between L2 phonemes is crucial. Cheour and colleagues (Cheour *et al.*, 2002; Shestakova *et al.*, 2003) conducted a longitudinal study on second language learning in 3–6-year-old children who learned French in a school-like setting. They showed that within two months training 3–6-year-old Finnish children developed a full MMN response to vowel categories specific to French, which was not present in children without teaching. When looking at late L2 learners, Winkler *et al.* (1999) reported a native-like MMN in response to a L2 specific phoneme distinction. Proficient Hungarian late second language learners of Finnish displayed the MMN for a vowel distinction that is relevant in Finnish but not in Hungarian. The MMN for this distinction was not present in Hungarian speakers with no command of Finnish (Winkler *et al.*, 1999).

These findings show that the categorization processes reflected in the MMN can be learnt during and after childhood and are not necessarily hampered by late age of acquisition (>12 years). It is questionable, however, whether this holds true for all possible phoneme distinctions across languages. Results of a study conducted by Nenonen *et al.* (2003) give rise to a less optimistic view. In their study the processing of duration in speech and nonspeech stimuli was tested in native Finnish children and Russian children who arrived in Finland after the age of 4 years. In the Finnish language long and short vowels distinguish

between the meaning of words and are therefore to be considered different phonemes (e.g., /tuli/ 'fire' vs. /tu:li/ 'wind'), which is not the case in Russian. Nenonen *et al.* presented the long and short syllable /ka:/ vs. /ka/ and a long and short complex tone as standard and deviant in blocks. The duration difference in both conditions was 50ms. Whereas the MMN response to the deviant tone did not differ between native and L2 speakers, the MMN related to phoneme distinction was decreased in amplitude in the L2 speakers. The authors take this difference as evidence that acoustic change detection is comparable in the L1 and the L2 speakers but that 'fine tuning in the processing of speech–sound duration may be inhibited at the pre-attentive level' (Nenonen, 2003: 494) in the L2. This finding is of particular interest as the participants tested can be considered early L2 learners. Studies currently available seem to suggest that phonemic features differ with regard to their learnability in L2 acquisition and that there might be differential age and proficiency constraints for different distinctive features.

Recently an ERP component has been reported that possibly indicates speech segmentation processes independent of acoustic cues (Sanders *et al.*, 2002; Sanders and Neville, 2003). As languages use different rhythmical cues for segmenting speech – such as syllables, morae and stress – it has been questioned if L2 speakers with a rhythmically different L1 use the same rhythmic segmentation cues as native speakers. Efficient speech segmentation could be a prerequisite for the acquisition of native-like efficiency with regard to syntactic and semantic processing. Sanders *et al.* (2003) reported an enhanced N100 (a negativity 100ms post stimulus onset) for word initial syllables compared to physically matched word medial syllables for native speakers of English. A similar N100 effect has been observed comparing stressed and unstressed syllables. Neither effect has been observed for late Japanese L2 learners of English (age of acquisition >12 years). The authors concluded that the L2 speakers neither segmented speech in a native-like manner nor did they use stress information (which is carried by the parameters loudness, duration and pitch) in the way natives do. As the native language of the group investigated by Sanders *et al.* is considered to be mora-timed, it relies probably less on syllable and stress information as other languages. The absence of the N100 effects could then be a result of the delay in language learning itself, or could

reflect the Japanese participants' use of L1 processing strategies that differ from those used by native speakers of English.

In sum, the above reviewed studies suggest that some new phoneme distinctions – as for example new vowel categories – can be acquired in L2 acquisition (Cheour *et al.*, 1998; Winkler *et al.*, 1999; Shestakova *et al.*, 2003) while other distinctions – such as distinctions based solely on the duration parameter (Nenonen *et al.*, 2003) – might be more difficult to establish in L2 acquisition. The first ERP study of the N100 for speech segmentation in L2 speakers (Sanders and Neville, 2003) suggests that native-like speech segmentation mechanisms cannot be used by late L2 learners, at least in those cases where the L1–L2 pairing precludes the transfer of segmentation strategies from the learners' L1.

VI Conclusions and outlook

In the last decades ERPs have become increasingly important as a tool for measuring language-related brain processes. ERP components that are thought to reflect specific functions in native language comprehension can be used as an objective measure of these functions in different groups of language users, such as L2 speakers of various proficiency levels and age of acquisition groups. Theories – such as the revised hierarchical model (Kroll and Stewart, 1994), which predicts problems with categorical-semantic processing in (less advanced) L2 speakers, or the declarative/procedural model (Ullman, 2001), which hypothesizes an increased reliance on lexical storage based processes as compared to automatized syntactic processes for L2 speakers – can be put to test in ERP studies.

From the available studies the developmental order of ERP markers of language processing in L2 and the degree of variability of developmental trajectories can also be speculated upon. It is probable that highly automatic syntactic processes, as indexed by the ELAN or LAN, are not available until more controlled syntactic and semantic processes, as indicated by the P600 and the N400, are established, if they are acquired at all by late L2 learners. It is equally conceivable that similar interdependencies exist with respect to the presence of ERPs in segmental processing or phonological discrimination.

Given an adequate interpretation of their functional characteristics, ERPs can be regarded as a powerful instrument in diagnosing of areas of

difficulty for L2 learners and in identifying domains in which native-like language comprehension processes are not normally attained. Moreover, ERP patterns observed in L2 speakers can also help to further clarify the general nature of the respective ERP components in L1 speakers. The case of the ELAN component, which has not yet been reported for L2 speakers, underlines that the development of the underlying process requires either a lot of time and practice or specific biological prerequisites available only in early childhood. In contrast to the difficulty in the acquisition of rather automatic syntactic processing mechanisms, the P600, which reflects more controlled syntactic processes, has been found in advanced L2 speakers of several languages. The available studies suggest that the P600 is strongly affected by the actual proficiency level of the L2 speakers (Hahne, 2001; Hahne and Friederici, 2001; Sabourin, 2003). Thus, the prerequisites for the acquisition of the process indexed by the P600 seem to be principally available to late L2 learners.

To summarize, ERP studies of second language processing have shed light on processing differences between native and L2 speakers. Factors such as age of acquisition or proficiency level seem to play a crucial role in determining similarity or dissimilarity of ERP patterns in native and L2 speakers. Careful control of the variables that might influence L2 processing – such as learning environment, amount of L1 and L2 use, or learning motivation – might help to further elucidate the nature of differences between L1 and L2 speakers with respect to ERPs. The specification of the prerequisites for the development of ‘nativelike’ ERP patterns appears to be a promising field for future investigations.

Acknowledgements

For helpful comments and suggestions on previous versions of this article, I would like to thank Claudia Felser, Angela D. Friederici, Anja K. Ischebeck, Sonja A. Kotz, Shirley-Ann Rüschemeyer and an anonymous *Second Language Research* reviewer.

VII References

- Ardal, S., Donald, M.W., Meuter, R., Muldrew, S. and Luce, M.** 1990: Brain responses to semantic incongruity in bilinguals. *Brain and Language* 39, 187–205.

- Bentin, S., McCarthy, G. and Wood, C.C.** 1985: Event-related potentials associated with semantic priming. *Electroencephalography and Clinical Neurophysiology* 60, 343–55.
- Besson, M., Kutas, M. and Van Petten, C.** 1992: An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience* 4, 132–49.
- Birdsong, D.** 1992: Ultimate attainment in second language acquisition. *Language* 68, 706–55.
- Bongaerts, T.** 1999: Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In Birdsong, D., editor, *Second language acquisition and the critical period hypothesis*, Mahwah, NJ: Erlbaum, 133–59.
- Cheour, M., Shestakova, A., Alku, P., Ceponiene, R. and Näätänen, R.** 2002: Mismatch negativity shows that 3-6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters* 325, 187–90.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K. and Näätänen, R.** 1998: Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience* 1, 351–53.
- Coulson, S., King, J.W. and Kutas, M.** 1998: Expect the unexpected: event-related brain response to morphosyntactic violations. *Language and Cognitive Processes* 13, 21–58.
- De Bruijn, Ellen R.A., Dijkstra, T., Chwilla, D.J. and Schriefers, H.J.** 2001: Language context effects on interlingual homograph recognition: evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition* 4, 155–168.
- Donchin, E.** 1981: Surprise . . . surprise? *Psychophysiology* 18, 493–13.
- Flege, J.E., Yeni-Komshian, G. and Liu, H.** 1999: Age constraints on second language acquisition. *Journal of Memory and Language*, 41, 78–104.
- Frenck-Mestre, C. and Prince, P.** 1997: Second language autonomy. *Journal of Memory and Language* 37, 481–501.
- Friederici, A.D.** 2002: Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences* 6, 78–84.
- Friederici, A.D., Hahne, A. and Pfeifer, E.** 1993: Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research* 1, 183–92.
- 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 Science* 99, 529–34.
- Grosjean, F.** 1997: Processing mixed language: issues, findings, and models. In De Groot, A.M.B. and Kroll, J.F., editors, *Tutorials in bilingualism: psycholinguistic perspectives*, Mahwah, NJ: Erlbaum, 225–54.
- 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.** 1999: Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience* 11, 194–205.
- 2001: Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition* 4, 123–41.
- Hahne, A., Mueller, J.L. and Clahsen, H.** 2003: Second language learners' processing of inflected words: behavioral and ERP evidence for storage and decomposition. *Essex research reports in linguistics* 45, 1–43.
- Holcomb, P.J. and Neville, H.J.** 1990: Semantic priming in visual and auditory lexical decision: a between modality comparison. *Language and Cognitive Processes* 5, 281–312.
- Johnson, J.S. and Newport, E.L.** 1989: Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21, 60–99.
- Kaan, E., Harris, A., Gibson, E. and Holcomb, P.J.** 2000: The P600 as an index of integration difficulty. *Language and Cognitive Processes* 15, 159–201.
- King, J.W. and Kutas, M.** 1995: Who did what and when – using word-level and clause-level ERPs to monitor working-memory usage in reading. *Journal of Cognitive Neuroscience* 7, 376–95.
- Kotz, S.A.** 2001: Neurolinguistic evidence for bilingual language representation: a comparison of reaction times and event-related brain potentials. *Bilingualism: Language and Cognition* 4, 143–54
- Kotz, S. A. and Elston-Guettler, K.** 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, 215–35.
- Kroll, J.** 1993: Accessing conceptual representations for words in a second language. In Schreuder, R. and Weltens, B., editors, *The bilingual lexicon*. Amsterdam: John Benjamins, 53–81.
- Kroll, J.F. and Stewart, E.** 1994: Category interference in translation and picture naming: evidence for asymmetric connection between bilingual memory representations. *Journal of Memory and Language* 33, 149–74.
- Kutas, M. and Federmeier, K.D.** 2000: Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Neuroscience* 4, 463–70.
- Kutas, M. and Hillyard, S.A.** 1980: Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–05.
- 1984: Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307, 161–63.
- Kutas, M. and Kluender, R.** 1991: What is who violating? A reconsideration of linguistic violations in the light of event-related potentials. In Heinze, H.J., Münte, T.F. and Mangun, G.R., editors, *Cognitive electrophysiology: basic and clinical applications*, Boston: Birkhauser, 183–210.

- Mueller, J.L., Hahne, A., Fujii, Y. and Friederici, A.D.** 2004: Electrophysiological correlates of syntactic and thematic processing in a miniature language: how different are learners from natives? 11th Annual Meeting of the Cognitive Neuroscience Society, San Francisco, CA, USA, April 2004. *Journal of Cognitive Neuroscience*, Supplement, 36.
- Näätänen, R.** 2001: The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 8, 1–21.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huottilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, A., Luuk, A., Allik, J., Sinkkonen, J. and Alho, K.** 1997: Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385, 432–34.
- Neely, J.H.** 1991: Semantic priming effects in visual word recognition: a selective review of current findings and theories. In Besner, D. and Humphreys, G.W., editors, *Basic processes in reading: Visual word recognition*, Hillsdale, NJ: Erlbaum, 264–337.
- Nenonen, S., Shestakova, A., Houttilainen, M. and Näätänen, R.** 2003: Linguistic relevance of duration within native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research* 16, 492–95.
- Neville, H.J., Mills, D.L. and Lawson, D.S.** 1992: Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex* 2, 244–58.
- Neville, H.J., Nicol, J.L., Barss, A., Forster, K.I. 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. and Holcomb, P.J.** 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*, New York: Oxford University Press, 171–209.
- Osterhout L., Holcomb, P.J. and Swinney, D.A.** 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.
- Picton, T.W., Bentin, S., Berg, P., Donchin, E., Hillyard, S.A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D. and Taylor, M.J.** 2000: Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–52.
- Pinker, S.** 1991: Rules of language. *Science* 253, 530–35.
- Rugg, M.D.** 1999: Functional neuroimaging in cognitive neuroscience. In Brown, C.M. and Hagoort, P., editors, *The neurocognition of language*, New York: Oxford University Press, 15–30.

- Rugg, M.D.** and **Coles, M.G.H.** 1995: *Electrophysiology of mind: event-related brain potentials and cognition*, New York: Oxford University Press, 171–79.
- Sabourin, L.** 2003: Grammatical gender and second language processing: an ERP study. Unpublished doctoral dissertation, University of Groningen.
- Sanders, L.** and **Neville, H. J.** 2003: An ERP study of continuous speech processing II: segmentation, semantics and syntax in nonnative speakers. *Cognitive Brain Research* 15, 214–27.
- Sanders, L.D., Newport, E.L.** and **Neville, H.J.** 2002: Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience* 5, 700–03.
- Schreuder, R.** and **Weltens, B.** 1993: *The bilingual lexicon*. Amsterdam: John Benjamins.
- Shestakova, A., Huotilainen, M., Ceponiene, R.** and **Cheour, M.** 2003: Event-related potentials associated with second language learning in children. *Clinical Neurophysiology* 114, 1507–12.
- 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, 105–22.
- Weber-Fox, C.M.** and **Neville, H.J.** 1996: Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8, 231–56.
- 2001: Sensitive periods differentiate processing of open- and closed-class words: an ERP study of bilinguals. *Journal of Speech, Language and Hearing Research* 44, 1338–53.
- Weyerts, H., Penke, M., Dohrn, U., Clahsen, H.** and **Münte, T.F.** 1997: Brain potentials indicate differences between regular and irregular German plurals. *NeuroReport* 8, 957–62.
- Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R.J.** and **Näätänen, R.** 1999: Brain responses reveal the learning of foreign language phonemes. *Psychophysiology* 36, 638–42.