PARIS-LODRON UNIVERSITY OF SALZBURG – DEPARTMENT OF LINGUISTICS

Processing a Second Language: Neurocognitive Correlates of Second Language Proficiency

Masterarbeit zur Erlangung des Grades “Master of Science (MSc)” an der Kultur- und Gesellschaftswissenschaftlichen Fakultät der Universität Salzburg

Dominik Freunberger

Betreuer: Prof. Dr. Dietmar Roehm

9/13/2011

Table of Contents

Abstract

In this event-related potential (ERP) experiment, we investigated the processing of semantically deviant versus well-formed sentences in a first versus a second language. One group were German native speakers who are learning English for 12.7 years on average, with an average age of acquisition of 9.9 years (L2 group); the second (control) group were native speakers of English (Native group). The semantic violation in this experiment was due to a semantic incongruity of the sentence-final verb and the preceding context, whereas the sentential context enabled a solid prediction about the incoming, sentence-final verb. Stimulus sentences were presented on a computer screen and participants had to judge whether the sentence was acceptable or not. For the processing of semantically deviant as compared to well-formed structures, we found a N400-effect for both groups, whereas the effect was most pronounced over central-right electrode sites in the L2 group and over posterior-right sites in the Native group. These results accord with findings from other studies, that also report of qualitatively similar ERP responses to semantic manipulations in native and non-native speakers of a language – even at an intermediate proficiency level (e.g. Hahne, 2001; Weber-Fox & Neville, 1996). In addition to the N400 as a response to deviant lexical elements, we found a P300 as a response to well-formed structures in the Native group but not in the L2 group. Results from Roehm et al. (2007) suggest that a P300 is engendered, when the (implicit) prediction about an incoming element is met. Since this was only the case in the Native group, we suggest that native and non-native speakers employ different parsing strategies.

1 Preface

Our everyday communicative needs require successful language comprehension. In turn, language comprehension depends on various distinct mechanisms that enable us to perceive, understand, and interpret our linguistic environment. The experimental investigation of these mechanisms underwent a major improvement when new, non-invasive methods were introduced into the field of psycholinguistics. In the last two decades more and more experiments focused on the investigation of second language acquisition and processing, since a multilingual environment becomes everyday routine for an increasingly number of people.

This work focuses on the investigation of second language processing using non-invasive methods that are based on electrophysiological brain activity. Therefore it is indispensable to shortly give a methodological introduction to the EEG and ERP technique (section 2), as well as an outline of the most important and prominent findings of language processing experiments (section 3). In section 4, two models of sentence processing will be introduced in order to sharpen the view for the (inter-) connections of the various results that are presented afore. The theoretical part of this thesis closes with a detailed review of experiments that concentrated on second language processing and second language learning (section 5). In the last sections (6-8) an experiment, where we investigated semantic processing in a first versus a second language, will be presented and discussed.

2 Methodological Introduction

The electroencephalography technique (EEG) has shown to be a suitable, non-invasive instrument for the examination of neurocognitive processes in the human brain. Event-Related Brain Potentials (ERPs), which are based on the human EEG, are currently the most widespread technique to investigate language processing. This section gives a brief introduction to the EEG technique, starting with some historical remarks (2.1), followed by ERPs in 2.2.

2.1 Electroencephalography – EEG

In the late nineteenth century Richard Caton (1875) took measurements of electrical activity in animals' brains. He discovered that electrical activity plays an essential role in the functioning of mammal brains. The first attempts to correlate cognitive processes to measureable changes in the human brain's electrical activity were made by Helmut Berger (1929). He described the so-called Berger effect, which was based on the observation of the suppression of a rhythmic electrical activity when mental activities (e.g. mental arithmetic) were performed. When uninjured participants were relaxed and/ or had their eyes closed the activity increased. This activity in the range from 8 – 10 Hz became known as the alpha rhythm, and the described effect as alpha blocking. Most essential to these findings is the correlation of electrical brain activity in healthy humans to distinct mental activities.

EEG continuously and instantaneously measures the cortex' electrical potentials using scalp electrodes that record potential differences relative to the reference electrode. These potential changes result primarily from "the summed postsynaptic activity of parallelly-oriented [sic] pyramidal cells perpendicular to the surface of the scalp" (Bornkessel-Schlesewsky & Schlesewsky, 2009:4). The electrical oscillations are a result of electrochemical transmissions between neurons in the postsynaptic cleft. Whenever a postsynaptic potential exceeds a certain threshold, the membrane potential abruptly changes from approximately -70mV to +30mV. For this moment, an electromagnetic dipole is generated and current flows between the two spatially separated poles with high or low depolarization, respectively (Kutas, Van Petten, & Kluender, 2006; Otten, & Rugg 2005).

Because the brain's activity is measured non-invasively from the surface of the scalp, the recorded signal does not necessarily reflect the actual cortical activity: The electrical signal might be modified by potential changes in different layers of the cortex. Therefore, the electrodes might even be incapable of measuring activity, given that signals cancel one another out, or the activity might be generated in neural populations that are not perpendicular to the scalp surface (Bornkessel-Schlesewsky & Schlesewsky, 2009). Further, the so-called "inverse problem" plays a crucial role in the interpretation of distributional properties. This source localization problem results from the non-uniqueness in the mapping from surface activity to underlying neural generators. The source of the activity in the cortex cannot be uniquely reconstructed from the measured surface distribution, though there are several mathematical approximations. For example, it does not necessarily mean that a neural generator is in occipital brain regions just because the measured activity is strongest over occipital electrode sites. If the source is known, the surface pattern can be anticipated; if it is unknown, there are always numerous possible solutions to the inverse problem, though physiological knowledge about brain structures obviates some of the potential sources (Bornkessel-Schlesewsky & Schlesewsky, 2009; Phillips, Mattout, Rugg, Maquet, & Friston, 2005).

2.2 Event-Related Brain Potentials – ERPs

ERPs are small potential changes in the brain's electrical activity that are time-locked to a certain internal or external event in order to examine the neural responses to a certain stimulus. Events of a certain type are usually also related to certain underlying cognitive processes (Kaan, 2007; Kutas, et al., 2006; Otten, & Rugg 2005). Here I will focus on language-related ERPs, that is, the event is usually a critical (part of a) word and/ or sentence. As the spontaneous electrical activity of the brain is approximately 10 – 100 μV and the stimulus-related changes range from 2 – 8 μV, an averaging procedure is required in order to extract the non-entropic activity from the entropic background activity. Averaging over a large number of trials of the same type increases stimulus-induced activity and reduces spontaneous background activity (Bornkessel-Schlesewsky & Schlesewsky, 2009; Luck, 2005b).

Luck (2005a:131+) describes the averaging as follows: All *n* signals with the same or similar underlying cognitive activity are segmented synchronously by extracting the same time segments around the event. Then these *n* epochs are summed and the summation is divided by the number *n* of epochs. The background noise is characterized by a coincidental polarity and amplitude of its waveform, thus being minimized or deleted through averaging. Contrary to that, the polarity and amplitude of event specific activity is not random, therefore being enhanced by the averaging process. It should be clear that the higher the number *n* is the higher is the chance that the background noise is deleted completely.

Figure 1 depicts an illustration of how ERPs are retrieved from the scalp, before being amplified and averaged time-locked to the stimulus of interest to obtain the commonly used potential progressions, as displayed in the right panel. The yielded ERPs are classified with respect to the following dimensions (cf. Donchin, et al. 1978):

- *Polarity* gives information about whether the potential change is positive or negative relative to the control condition and/ or other conditions.
- *Latency* describes the time (in Milliseconds) from the onset of the critical stimulus to the onset of the observable potential change (*onset latency*) or to the peak of the potential change (*peak latency*).
- *Amplitude* refers to the effect "strength" by means of high or low voltage (in Microvolts). Typically high or low voltages are interpreted as increased or decreased neural activation, respectively, relative to the control condition.
- The *Topographic Distribution* of an effect is generally described using so-called "Regions of Interest" (ROIs), which are (pre-defined) groups of electrodes rather than single electrodes. For instance, an effect might be labeled "left-anterior" because the measured activity is strongest over left-anterior electrode sites (cf. section 2.1; inverse problem).

Components are often labeled according to their relative polarity ("N" for negativity, "P" for positivity, relative to control condition) and latency (e.g. "P300" refers to a positive deflection of the waveform 300ms after the onset of the critical stimulus; cf. Figure 1). Intermittently, peaks are labeled according to their order of appearance in the waveform (e.g. "P1" refers to the first positive peak). Further, distributional characteristics might be used, as in "LAN", which is a Left-Anterior Negativity, as well as functional interpretation, as in "SPS" (*Syntactic Positive Shift*).

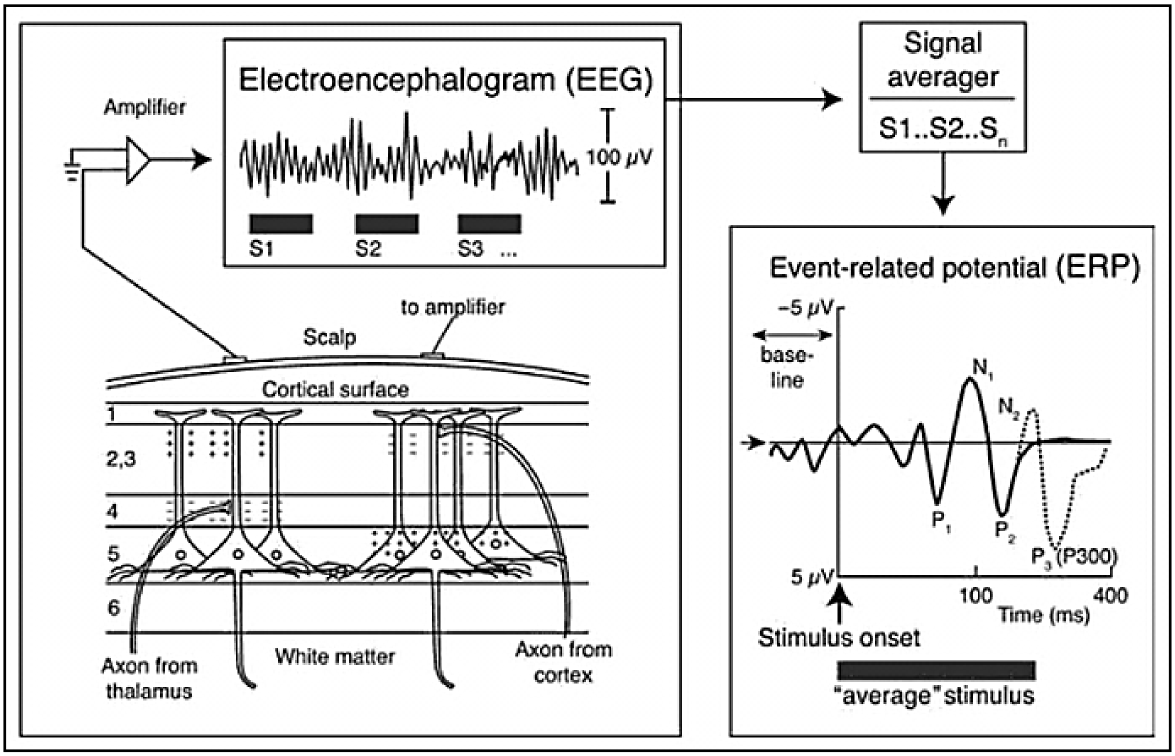

Figure 1: Schematic illustration of the retrieval of ERPs. The left panel shows the physiological basis of the human EEG: The signal is generated in neural populations that are perpendicular to the scalp surface, where the electrodes are mounted to detect the signal. Then the signal is amplified to yield the EEG; this in turn is segmented into same segments around the event (S1, S2, S2…). In the right panel the process of generating ERPs, namely time-locked averaging, is depicted. (Adapted from Bornkessel-Schlesewsky & Schlesewsky, 2009:6).

2.3 Interpreting ERPs

ERP components are typically associated with different functional interpretations. However, it is important to note that there is a key difference between an ERP-*component* and an ERP-*effect*: A component can occur alone, an ERP-effect is inherently relative. That means that an effect can only be observed if there is another (significantly) diverging waveform elicited by another condition, and/ or group of participants. In particular, a specific stimulus can give rise to a specific component, and another stimulus can engender qualitatively the same component but with a difference in its amplitude. This difference is referred to as an ERP-effect (Bornkessel-Schlesewsky & Schlesewsky, 2009).

The *polarity* of an effect is influenced by a number of non-neurophysiological aspects like the baseline used for comparison or the site of the reference electrode, as well as the orientation and location of the potential changes' intracerebral sources. The type of activation (inhibitory vs. excitatory) and the orientation of the synapses relative to the cell body also affect the polarity of an ERP effect. Therefore an effects' polarity can not be functionally interpreted in a sufficient way if there is no further detailed information about the underlying neural generators (Otten, & Rugg 2005).

An effects' *latency*, thus the time at which two waveforms diverge, indicates the time at which also the (assumed) cognitive processes deviate from one another. Nevertheless, this point in time does not inevitably mark the moment when the brain's response to different conditions starts to differ, for there is the chance that the activations diverge earlier in time than the electrodes are able to detect. Along with the onset latency, other parameters like rise time, time to peak, and effect duration can be of particular interest as well (Otten, & Rugg 2005).

Under the assumption that stimulus-locked activity is stable across epochs, the *amplitude* of an effect can be interpreted in terms of activation strength of the neural generator, and hence the strength of the cognitive demands needed for the processing of the stimulus under examination. Single-trial and/ or single participant analyses should be avoided because of the low signal-to-noise ratio. Amplitude differences might reflect differences in the strength of the postsynaptic potentials, the activation of more or fewer neurons in a population, or less synchrony in the time course of the activated neurons (Kutas & Federmeier, 2011). It should be noted, however, that amplitude differences can occur even if there is no activation difference, and vice versa. Moreover, there are well-known repetition effects, that is, an effect's amplitude decreases after participants are exposed to the same type of stimulus over a longer period (Otten, & Rugg 2005).

As noted above, interpreting differences in the *scalp distributions* of ERP effects is inherently problematic in case of ERPs being the only data available. The measured activity might result from a single cell population or from multiple cell populations concurrently active; what makes it difficult to identify the actual underlying generator. In addition, different distributions might not only reflect distinct neural sources; but also different contributions of several neural regions, in activity strength as well as in their time course. That is to say, if one finds a topographic difference between two conditions there might either be two separate cortical sources involved, or there are networks with partially (but not completely) overlapping concurrently active brain regions (Otten, & Rugg 2005).

Before turning to language related ERP components, a short note on obtaining *no* significant differences between conditions (‘null results’) in an experiment: If one does not find a significant difference in amplitude, latency, or surface distribution between two conditions, this does not inevitably mean that there are no differences. The waveforms may have been analyzed and/or quantified in an inappropriate way. Further, null results can occur due to statistical deficits, viz. there might be a lack of statistical power in the experiment. Finally, ERPs do not reflect all of the brain’s activity; the activation differences of neural populations must show the features described above (simultaneously active, parallel-oriented, perpendicular to the scalp surface) (Bornkessel-Schlesewsky & Schlesewsky, 2009; Otten, & Rugg 2005).

At this point, the distinction between ‘endogenous’ and ‘exogenous’ components should be noted, whereas the former are more informative for psycholinguistic purposes: Exogenous components are supposed to be relatively invulnerable by an individuals’ vigilance or attention; they appear approximately < 200 ms post stimulus onset. Endogenous components are relative invariantly elicited by a stimulus, yet they are affected by several factors, like expectancy, task demands, strategies, and so on (Kutas, et al., 2006).

3 ERPs in the Study of Language

This section will draw an outline of the most prominent language related ERP components (ELAN in 3.1; LAN in 3.1; N400 in 3.3, P600/ SPS in 3.4, and further language related components in section 3.5) by giving a short characterization of their discovery and the experimental conditions that elicit the respective component. Each subdivision closes with a short functional characterization, whereupon it has to be noted that these descriptions will be slightly modified throughout the remainder of this thesis. Unless otherwise stated, ERP measurements were calculated on the underlined word(s) in the examples. Table 2 gives a summary of the domains of occurrence of language-related ERP components.

3.1 ELAN

Neville et al (1991) observed an early left-anterior negativity (ELAN), in the time window from 150 – 200 ms, when participants were confronted with word category violations, as shown in (1). Further, these constructions do not merely violate word category restrictions, but also the initially built phrase structure. Friederici, Pfeifer, and Hahne (1993) replicated these findings for German, as in example (2), Isel, Hahne, Maess, and Friederici (2007) for French. They interpreted this negativity as a neurophysiological reflection when highly automatic phrase structure building processes cannot proceed with the assumed structure. In addition to the early negativity, they report of late positivity (a P600/ SPS; see section 3.4), probably reflecting the repair of the violated phrase structure, or the attempt to integrate the erroneous input into a meaningful interpretation (see also Friederici, 2002; or section 4). Yamada and Neville (2007) found an ELAN even after reading nonsensical sentences (‘Jabberwocky’), as shown in (3). Hitherto, the ELAN was observed after the

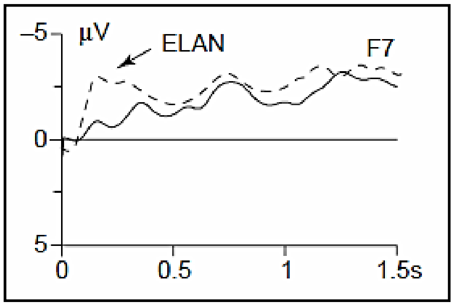

Figure 2 shows the grand average waveform for electrode F7 after processing phrase structure violations, as in ‘**Die Bluse wurde am gebügelt*’ (“*The blouse was on ironed”). Negativity is plotted upwards. (adapted from Friederici, 2002:81).

processing of ungrammatical structures only, whereas other components have also been elicited by grammatically well-formed structures (Kutas, et al., 2006).

(1) *The man admired Don’s of sketch of the landscape.

(2) *Das Parkett wurde im gebohnert.

*The parquet was in the$_{DAT}$ polished.

(3) *Minno can kogg the mibe with her that nove.

The ELAN has shown to be highly sensitive to the degree of language proficiency and/ or the age of acquisition of a second language (see section 5). Pakulak and Neville (2009) showed that the ELAN is even amenable to proficiency differences in monolinguals, as reflected by a more focal negativity (temporarily and spatially) in higher proficient speakers as compared to lower proficient speakers. Hahne, Eckstein, & Friederici (2004) investigated the development of semantic and syntactic processing in children with an age ranging from 6 to 13 years. For the syntactic condition they used phrase structure manipulations like in (2). They observed an adult-like ELAN + P600-pattern in the group of 13 year old children only. For the younger groups, the authors report of a later sustained negativity and a P600, but no negativity in the time window of the ELAN. They conclude that the highly automatic phrase structure building processes the ELAN reflects are not yet developed in an adult-like way in young children.

The early left anterior negativity – if at all – has been taken as to reflect early automatic processes of local phrase structure building, moreover, assigning word category information to the initially built phrase structure. Because this early negativity has only been observed in a very restricted set of constructions, namely the ones shown in (1-3), it is still controversial what the ELAN actually reflects and if it is qualitatively distinct from the left-anterior negativity (LAN) (Kutas, et al., 2006). See the end of section 3.2 for another view on the ELAN/ LAN debate.

3.2 LAN

Kutas and Hillyard (1983) investigated the processing of incongruous number marking in English sentences. They observed a left-anterior negativity peaking between 300 and 500 ms after the erroneous word was encountered. Friederici, Pfeifer, and Hahne (1993) found a similar pattern[1] for incorrectly inflected verb forms (4), suggesting that the LAN reflects morphosyntactic processing. Coulson, King, and Kutas (1998) report of an anterior negativity after participants read sentences with deviant pronoun case inflection, as shown in (5) and (6). Gunter, Friederici, and Schriefers (2000) found a LAN for gender violations, as in (7).

(4) *Das Parkett wurde bohnere.
    The parquet was polish.
(5) *The plane took we to paradise and back
(6) *Ray fell down and skinned he knee.
(7) *Sie bereiste den Land auf einem kräftigen Kamel.
    *She traveled the$_{masc}$ land$_{neuter}$ on a strong camel.

Münte, Schlitz, and Kutas (1998) found a longer lasting left anterior negativity, hence assuming that there is more than one LAN. This so-called 'sustained' LAN seems to reflect general additional processing costs due to a higher working memory load. They observed a left anterior negativity, beginning 300ms after the first word was encountered, for 'before' as compared to 'after' in sentences like shown in (8). Negativity further increased with the ongoing sentence. 'Before' requires more additional discourse-level computations because the default order of chronological events is disturbed. ERP plots of this study are shown in Figure 3. The processing of wh-dependencies also leads to left-anterior negativities: Kluender and Kutas (1993) report of a LAN at 'who' (9) as compared to 'that' (10), that is, at the moved *wh*-element. Further, they also found a potential difference directly after the gap, namely on 'into'. The first LAN probably reflects the identification of a filler that has

[1] The negativity that is reported here was most *pronounced* over left-anterior sites, though the overall activity was broadly distributed.

to be kept in memory, the second negativity might reflect the integration of the critical element into its' base position, for this reason, working memory load is increased.

(8) After/ Before the scientist submitted the paper, the journal changed its policy.

(9) What did he wonder that he could coerce her into …

(10) What did he wonder who$_i$ he could coerce ___$_i$ into …

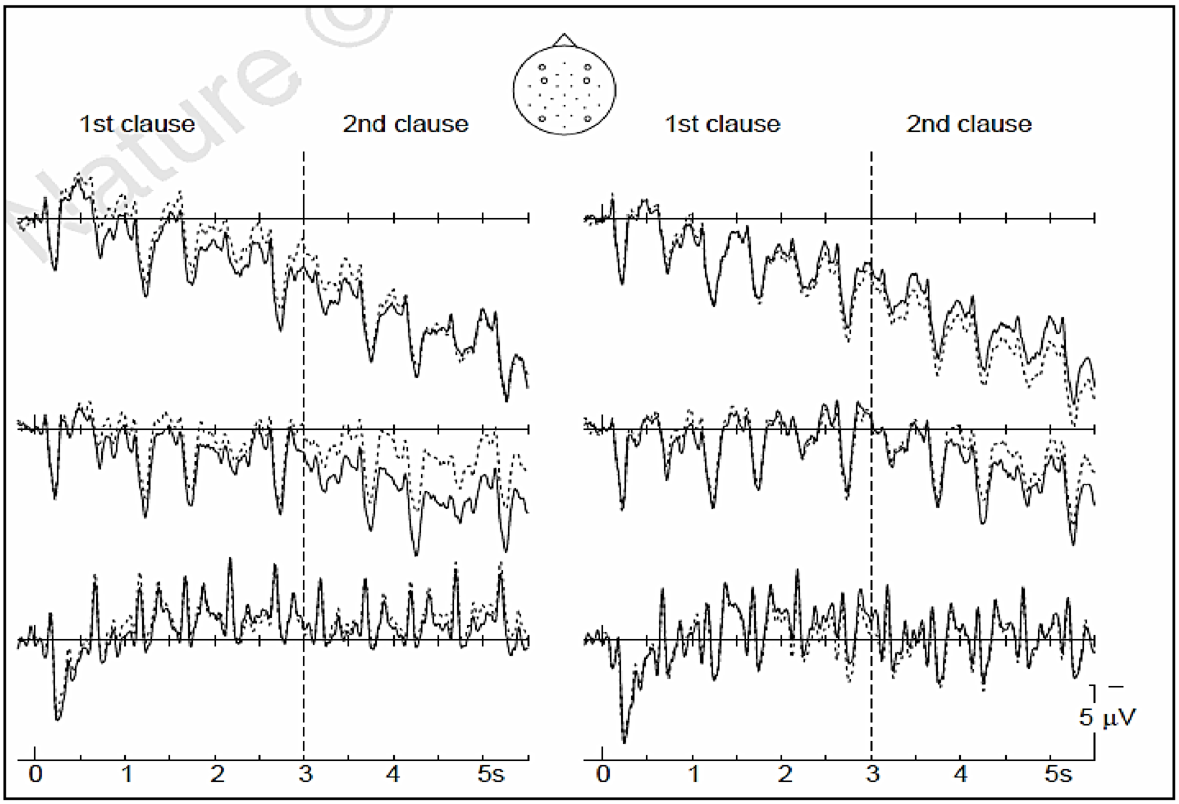

Figure 3 shows the grand average waveforms from frontal, prefrontal, and occipital sites. The dotted line represents the 'before' sentence, the solid line the 'after' sentence. The vertical dashed line illustrates the onset of the embedded clause ('that' and 'who', respectively). Negativity is plotted upwards. (adapted from Munte, et al., 1998:71)

The LAN and the N400 have similar latencies, but the assumed underlying functional properties (on the base of the conditions that elicit these components) seem to be different. Hence, there is consensus regarding the assumption that these components reflect different underlying mechanisms, namely thematic role assignment (LAN) and the processing of semantic relations (N400); further, the LANs distribution differs from the N400s (Friederici, 2002). Additionally, the LAN clearly is attributed to syntactic processing, but as it can be observed even in grammatical constructions, some argue that it might reflect general aspects of increased working memory load. Sometimes the LAN and the (late) ELAN share the time window of their appearance, thus they are hard to distinguish using only temporal aspects. Rossi et al. (2005) report of an ELAN (+ LAN + P600) for phrase structure violations, and double violations (phrase structure and congruency), and of a LAN (+ P600) for bare congruency violations, therefore concluding that word category information and morphosyntactic information are processed quite independently.

Though this could be regarded as evidence for the functional difference of the LAN and the ELAN, some argue that the timing differences of these two components are due to experimental conditions: For example, Hagoort (2003) proposes a model where semantic and (morpho-) syntactic information are processed parallel as soon as they are available. Hagoort assumes that the difference in the timing of the ELAN and the LAN is due to differences in the online availability of word category and morphosyntactic information, but not due to a basic functional difference. For instance, when word category is encoded by a prefix (like 'GEbohnert', in (2)), the information is available earlier (especially in the auditory modality), thus leading to a negativity with an early onset; when morphosyntactic deviations are encoded by a suffix, what is later available in time, the negativity's latency is increased. Thereby it follows that a component occurs earlier than another one because of the availability of the necessary information, but this does not imply that the earlier information is needed for the computation of the latter. This leads to the conclusion that thematic (and also semantic) integration is computed independently of syntactic information (in contrast to syntax-first models), as soon as the respective information is available[2].

[2] Hagoort reports, for instance, of N400s preceding ELANs when the onset of the ERP-measurement is calculated on the 'category violation point' (that is, when it is clear that, e.g. a word is identified as a noun instead of a verb) rather than on the onset of the word. Hagoort therefore concludes that '[…] this is the clearest evidence so far for the claim that semantic binding can start before word category information is provided.' (2003:22/23).

3.3 N400

This broad negativity with an onset around 250 ms and peak around approximately 400 ms post stimulus onset was first described by Kutas and Hillyard (1980): They compared sentences with a semantically inappropriate sentence final word, as in (11) with congruent sentences (12) and found the waveforms to diverge significantly from one another. Astonishingly, they did not observe an (expected) P3b-effect[3], but moreover, large centro-parietal negativity. When they compared the same material to sentences containing a physically unexpected sentence closure (semantically congruent, as in *I shaved off my mustache and BEARD*) they did not find an N400 effect; thereby they could tell that the N400 did not just indicate *any* manipulation using words. Further they found that the N400 not only obtains for semantic violations; but for all sentence final words, with its amplitude being enhanced for unexpected words as compared to expected words, as measured by offline cloze-probability judgments (Kutas & Hillyard, 1984). The N400 does not only occur after semantic anomalies, but also after the processing of contextually meaningful words; thus the N400 might reflect the brains normal response to meaningful sentences.

(11) He planted string beans in his <u>car</u>.

(12) I shaved off my mustache and <u>beard</u>.

To date; several experimental conditions have been identified to obtain N400-effects. For instance, word lists containing a semantic incongruent item, or lexical priming paradigms with target words violating semantic expectation, whereby the N400 amplitude is increased for unrelated items as compared to related items. Moreover, there is an N400 effect after the presentation of content words vs. function words, concrete vs. abstract words, and high vs. low frequent words (Brown & Hagoort, 1999; among others). Further, several input modalities (e.g. written and spoken words, sign language, pseudowords), as well as primed

[3] The P3b was one of the first components classified; it occurs central-parietal after the presentation of improbable events as compared to probable event. The less probable an event is the higher is the amplitude of the P3b; importantly it must be noted, that this is only the case, for instance, in an odd-ball paradigm, where the task demands to detect deviant stimuli!

(vs. not primed) line-drawings, pictures (e.g. Barrett & Rugg, 1990), and faces have yielded robust N400 effects (Kutas & Federmeier, 2000; 2011). Brown and Hagoort (1999) report that the N400 is most pronounced over centro-parietal electrode sites, whereas the effect is more right-lateralized in the visual than in the auditory modality. Rugg (1990) found the N400 to be sensitive to word frequencies, and especially to repetition of these, whereas the repetition effects seems to superpose frequency effects. Swaab, Camblin and Gordon (2004) found an inversed repetition effect: The smaller the lag between the two elements (name-name, name-pronoun) the smaller was the N400's amplitude. This might be due to an advantage of referential linking when the antecedence is closer to the pronoun/ name. Further, they obtained results that can be viewed as evidence for the phenomena of the 'repeated name penalty'; that is, there are increased processing demands (as reflected in an increased N400-amplitude) when coreferential linking is made via repetition of the name rather than with a pronoun (with an additional influence of discourse prominence).

Roehm et al. (2007) obtained N400s after processing antonymous vs. related vs. unrelated words in the construction viewed in (13). The amplitude of the effect varied as a function of the words' relations to each other, whereas the unrelated words yielded higher amplitudes as compared to related words. Further, they found a P300 for the related condition in the sentential context, as well as a P300 as a correlate of different parsing strategies in the word-pair context. See the discussion at the end of this thesis (section 8) for more on this topic. The ERP-plots of this study are shown in Figure 4.

(13) Das Gegenteil von schwarz ist weiß/ gelb/ nett.
The opposite of black is white/ yellow/ nice.

The occurrence of the N400 is not restricted to the word-level; Van Berkum, Hagoort and Brown (1999) report of a broad negativity peaking approximately 400ms after the processing of a discourse-anomalous word. The authors interpret these findings as evidence that our language processing system integrates words in the same manner on sentence-level as on discourse-level. Burkhardt (2006) observed a N400 when meaning is integrated into world-knowledge via inferential linking, and Burkhardt and Roehm (2007) report of the elicitation of a N400 when a depended element is referentially linked to a prior more or less plausible antecedence, with the amplitude being decreased or increased, respectively.

When an anaphoric element initiates a search process for an appropriate antecedence via lexical-semantic property-checking of potential antecedences, cognitive demands are reflected in a N400 as well (Filik, Sanford, & Leuthold, 2008).

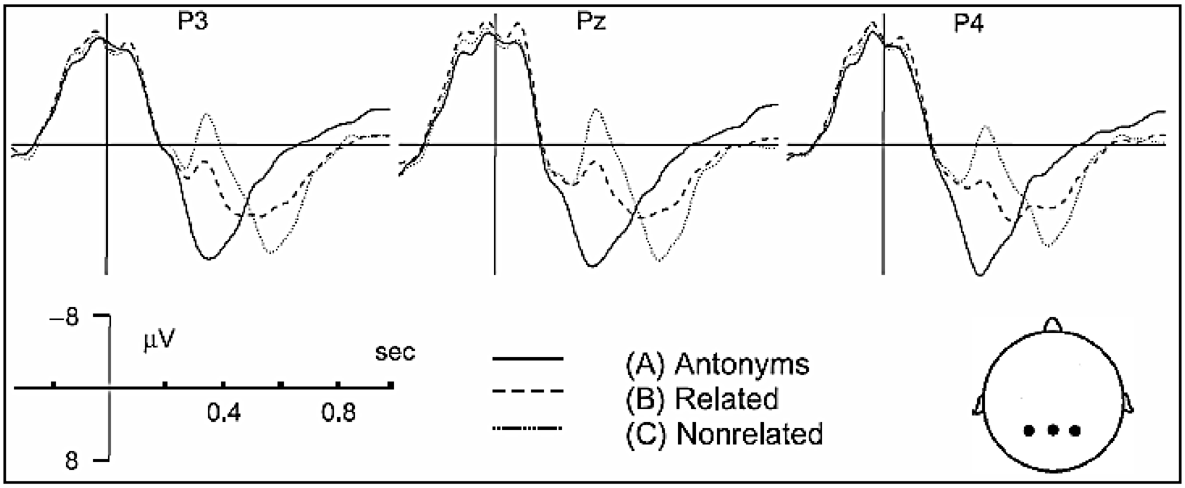

Figure 4 shows the grand average waveforms for the conditions antonymous, related and unrelated. Negativity is plotted upwards (adapted from Roehm, et al., 2007:1264).

Van Petten and Kutas (1991) found the N400 amplitude to be amenable to distance differences between two elements; however, the closer the two to-be-linked elements are (whether in word lists or discourse contexts), the higher is the amplitude of the N400, and vice versa (see Swaab, et al., 2004, for oppositional results in discourse processing). Further, a N400 can be elicited by violating long-term semantic relations, as shown in (14), yielding the lowest amplitude for the expected 'palms', a higher amplitude for the more implausible 'pines', and the strongest effect for the totally unexpected 'tulips' (Federmeier & Kutas, 1999).

(14) They wanted to make the hotel look more like a tropical resort. So along the driveway they planted rows of palms/ pines/ tulips.

Bornkessel-Schlesewsky and Schlesewsky (2009:10) subsume the classic functional interpretation of the N400 as "lexico-semantic processing [and] integration into a meaningful context.” The N400 has hitherto been viewed as indexing bare semantic-lexically processes, but more recent findings let suggest that this ‘classic’ interpretation (for a review, see Kutas & Federmeier, 2000; 2011; Kutas, et al., 2006) is a misbelief, regarding manipulations that are clearly not in the lexico-semantic domain and yet yielding N400s[4]; some of these findings are discussed here.

For example, Weckerly and Kutas (1999) found a N400 after the processing of inanimate subjects (15) as compared to animate subjects (16) in object relative clauses. Surely, these results can be interpreted in terms of lexico-sematic integration, but one cannot disclaim that this violation operates (at least) at the interface of syntax and semantics. Furthermore, there are several studies investigating deviant word orders and yielding N400s; Bornkessel et al. (2004) obtained a N400 effect after the processing of sentence-final dative-active-verbs following dative-object-subject (DAT-OS) word order (17) as compared to subject-dative-object (DAT-SO; (18)). Contrary to that, they found a P600 (see 2) for the same word order variation using accusative objects, hence, the authors interpret these findings as evidence for the N400 being evoked when thematic roles have to be re-assigned (as it is the case for dative objects)[5]. Similar findings for the processing of word order variations are reported by Bornkessel, Schlesewsky, and Friederici (2002) using ‘scrambled’ structures (permutation in the German Middlefield). Frisch and Schlesewsky (2001) found an N400 for the processing of case conflicts using double-case-violations (19). Frisch (2000) also used double-case-violations, but with the verb intervening the arguments, as in (20). He also found a N400 with additional late positivity for the NOM-NOM-sentence as compared to the NOM-ACC-constructions.

(15) The novelist that the movie…

(16) The movie that the novelist…

[4] Note that the experimental environments mentioned above (e.g. picture or music processing) can also be seen as evidence that the N400 does not reflect pure lexico-semantic processes, as it is hard to define lexico-semantic properties in pictures or music.

[5] Further, they found a N400 after the processing of dative-object-experiencer-verbs following OS vs. SO constructions.

(17) …dass Maria Sängerinnen folgen.

…that $\text{Maria}_{NOM/ACC/DAT.SG}$ $\text{singers}_{NOM/ACC/DAT.PL}$ follow_{PL}.

'…that singers follow Maria.'

(18) …dass Maria Sängerinnen folgt.

…that $\text{Maria}_{NOM/ACC/DAT.SG}$ $\text{singers}_{NOM/ACC/DAT.PL}$ follow_{SG}.

'…that Maria follows singers.'

(19) …welcher Angler der Jäger gelobt hat.

…which angler_{NOM} the hunter_{NOM} praised has.

(20) Welcher Kommissar lobte *der Detektiv/ den Detektiv im Radio?

Which commissar_{NOM} praised *the detective_{NOM}/ the detective_{ACC} in the radio?

Regarding these results, it is quite clear that the classic interpretation of the N400 as a reflection of pure lexical-semantic processes is hard to pursue any longer. The N400 reflects much more fine-grained processes, and a one-to-one-mapping from the component to the functional interpretation fails. The N400 surely reflects processing mechanisms in the lexical-semantic domain, but one cannot disclaim that this negativity probably is the electrophysiological manifestation of mechanisms in other (at least overlapping) domains, just as the interface between syntax and semantics (thematic role assignment).

3.4 P600/ SPS

The P600 is a positive deflection of the ERP-waveform starting at approximately 400 – 550 ms post stimulus onset, peaking around 600 ms post stimulus onset and ending approximately 800 – 1000 ms after the presentation of the critical stimulus. The P600 is strongly correlated with syntactic processing, namely syntactic repair, reanalysis, reintegration, increase in discourse complexity, and general increased processing costs.

Osterhout and Holcomb (1992) presented syntactically anomalous sentences (21) and obtained positivity with a latency of 600 ms for the deviant sentence. Hence it was clear that the P600 is a reflection of the processing of syntactic violations. Further, they also found a P600 in grammatical sentences, but with a non-preferential reading; these so-called 'garden path sentences' are structures, where at a certain point the initial analysis cannot be upheld any longer due to a local ambiguity (as measured on 'to' in the 'persuade' version of sentence (21)). Hagoort, Brown, and Groothusen (1993) labeled this positivity as 'syntactic positive shift' (SPS) according to its functional characterization. They concluded that the SPS is the measureable surface activity when the parser faces the impossibility of assigning an incoming word to the preferred sentence structure (as it is the case in garden path sentences), therefore, the parser has to 'shift' the built structure (see also van Berkum, Brown, & Hagoort, 1999).

(21) The broker hoped/ persuaded to sell the stock was sent to jail.

Friederici, Pfeifer, and Hahne (1993) investigated the processing of morphosyntactic anomalous sentences using an incorrect form of a sentence final verb (4); the stimuli were presented auditory. The authors report of a late positivity in the time range from 600 – 1200 ms post stimulus onset over parietal and posterior sites after (in addition to an anterior negativity around 400 ms). Moreover, phrase structure violations (2) also elicited a P600, though preceded by an early left-anterior negativity (ELAN). The occurrence of the P600 is not restricted to (morpho-) syntactic violations and garden path sentences; Kaan, Harris, Gibson, and Holcomb (2000) constructed grammatical non-garden path sentences, as shown in (22). They found a more pronounced positivity for the 'who'- version as compared to 'whether', hence assuming that the P600 reflects integration difficulties. In the sentence with 'who' the thematic role assignment takes place at the verb 'imitate', thus leading to increased processing costs as compared to 'whether'. Additionally, integration is even more impeded as 'who' and the verb 'imitate' are separated by several words. ERP plots of this study are shown in Figure 5.

(22) Emily wondered whether/ who the performer in the concert had imitated for the audience's amusement.

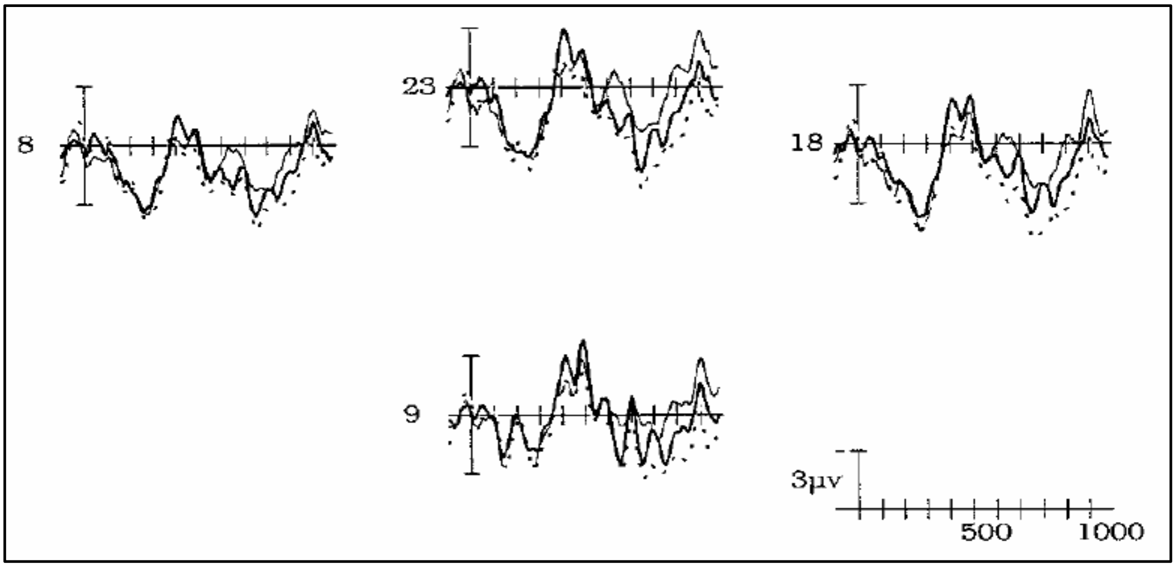

Figure 5 shows the waveforms for sentences like in (22) time locked to the verb. The thin solid line depicts the 'whether' condition, and the thick solid line the 'who' condition. The results of the third condition (dotted line; 'which') are not discussed here. Negativity is plotted upwards (adapted from Kaan, et al., 2000:170).

Mecklinger, Schriefers, Steinhauer, and Friederici (1995) found an early positivity (P345; see also 3.5) after the processing oft the sentence final verb in constructions containing only case ambiguous feminine nouns. The disambiguation towards an object- or subject relative clause reading is delivered by the verb. Note, however, that this early positivity was only found in the 'fast comprehenders' group (with reaction times (RTs) < 800 ms). The authors suggest that this P345 reflects a fast revision of the initial parse, though leaving hierarchical phrase structure untouched. These results (among others) initiated a debate whether there is more than one P600. Hagoort, Brown, and Osterhout (1999) assumed a more frontal P600 when the preferred reading of ambiguous sentences has to be revised towards a non-preferred reading, and a more posterior P600 for the repair of ungrammatical sentences. Evidence for this assumption comes from a study by Kaan and Swaab (2003): They found a posterior P600 for revision and repair of ungrammaticality, and a frontal P600 for the resolution of ambiguities and when discourse complexity increased (see also Burkhardt, 2006; 2007). Friederici, Hahne, and Saddy (2002) manipulated grammaticality and complexity and observed a N400 + a centro-parietal P600 for the grammaticality violation, and an early positivity + a fronto-central P600 for the complexity condition, hence assuming that there are different neural contributors to the repair-P600 and the complexity-P600.

Besides P600s being elicited by syntactic complexity, ungrammaticality, or discourse complexity, there are also studies reporting of P600 after the processing of semantically anomalous sentences, the so-called 'semantic P600'. For example, Kim and Osterhout (2005) presented sentences like in shown (23) and found a P600 on the verb 'devouring' as compared to the control sentence in (24). In this case, the verb clearly does not violate syntactic structures, and/ or morphosyntactic rules, but rather thematic constraints, namely that the verb 'devouring' requires (or at least prefers) an animate agent. Similar to these findings, Kuperberg et al. (2003) also report of a P600 after reading sentences like (26), compared to the control condition (25). These results suggest, that when the thematic structure of a verb is violated, a robust P600 can be observed. Kuperberg at al. further found the amplitude of the positivity to vary as a function of semantic relatedness (or 'thematic fit'). For a review of 'unusual' P600s, see Kolk and Chwilla (2007).

(23) The hearty meal was devouring…

(24) The hearty meal was devoured…

(25) For breakfast the boys would only eat toast and jam.

(26) For breakfast the eggs would only eat toast and jam.

To sum up, the P600 reflects much more than just simple reanalysis after encountering ungrammatical structures. As this late positivity often is observed in combination with (early) left-anterior negativities, is appears to reflect the process of repairing deviances that are recognized earlier (as reflected by an (E)LAN).

3.5 Further Language-Related ERP-Components

The so-called *mismatch negativity* (MMN) is a negative deflection with a peak around 150 – 200 ms. The MMN is the brains response to changes in the auditory input, moreover, whenever a certain threshold (e.g. in the voice onset time) is exceeded. For example, when a listener hears similar phonemes that are intervened by odd phonemes (like /p/ /p/ /p/ /b/ /p/), a MMN is engendered.

The *scrambling negativity* can be observed when scrambled structures (viz. clause-medial word order variations), like in (27), are processed. Bornkessel, Schlesewsky, and Friederici (2002) assume that this negativity arises because of a mismatch between the preferred order (Subject-Object, in German) and the actual order. Bornkessel-Schlesewsky and Schlesewsky (2006) propose that this negativity is different from the well-established N400 and LAN, because, functionally speaking, it is somewhat between the domains associated with the respective components.

(27) Gestern hat den Jäger der Gärtner beruhigt.
Yesterday has the hunter$_{OBJ}$ the gardener$_{SUBJ}$ calmed
'Yesterday, the gardener calmed the hunter.'

The *P345* is an early parietal positivity that is elicited when grammatical functions need to be reanalyzed, as it is the case for relative sentences like in (28) (Mecklinger, et al., 1995).

(28) Das ist die Journalistin, die die Sekretärinnen informiert haben.
This is the journalist, who$_{AMB.SG}$ the secretaries$_{AMB.PL}$ informed have$_{PL}$
"This is the journalist whom the secretaries informed."

The disambiguation point is at the clause final auxiliary 'haben', which forces the reader towards a unpreferred object-first reading, thus, the initially assigned grammatical functions need to be reanalyzed, as reflected in the P345. This component has only been observed in relative clauses, like in the example shown above. For embedded *wh*-questions (which are assumed to be structural identical), different ERP-patterns are found (Bornkessel, Fiebach, & Friederici, 2004).

Component		Domains of occurrence											
		Lexical factors	Morpho-syntactic factors	Word-level composition	Constituent structure	Sentence-level composition					Well-formedness	Complexity	Working memory
						Gramm. relations	Thematic roles	Linking	Word order	Semantic interpretation			
NEG	ELAN				√								
	LAN		√	√				√					
	sLAN								√				√
	SCR NEG								√				
	N400	√		√			√	√	√		√		
POS	P345					√	√						
	P600		√	√	√	√			√	√	√	√	√

Table 1 shows a summary of language related ERP-components and the domains where they have been observed (from Bornkessel-Schlesewsky & Schlesewsky, 2009:296).

4 Neurocognitive Models of Sentence Processing

Friederici (1995; 2002:79) proposes the following three stages of (auditory) sentence processing, given the assumption that the components described above are hierarchically mapped to each other. The primary acoustic analysis and identification of the phonemes takes place in phase 0 (< 100 ms); violation in this stage elicits a negativity around 100 ms post stimulus onset. In phase 1 (100 – 300 ms), word category information is encountered and the initial sentence structure is formed, autonomously and independently of contextual or semantic influences. Word category and/ or local phrase structure violation leads to an ELAN. In phase 2 (300 – 500 ms), following the identification of lemma and morphological information, semantic and morphosyntactic information is integrated and semantic and thematic roles are assigned. When thematic role integration fails, a LAN is engendered. When semantic integration is disturbed, a N400 can be observed. In phase 3, syntactic integration is processed; if there are difficulties with the integration of the different information types, reanalysis and repair processes are initiated, as reflected by a P600. Friederici assumes that processes of phrase structure building are highly automatic (phase 1), the latter semantic and syntactic processes might also interact and seem to be more accessible by conscious control (see also Hahne & Friederici, 1999).

Bornkessel and Schlesewsky (2009; 2006) present an extension of Friederici's neurocognitive model of sentence comprehension: The *extended argument dependency model* (eADM). This model also assumes that the comprehension process can be divided in hierarchically ordered stages. In the eADM, word category information, as encoded by phrase structure, and relational structure are processed separately from one another. The relational structure encodes the arguments' relations to each other, and to the verb, respectively. The arguments are ranked hierarchically with respect to the so-called 'prominence information' (see also Bornkessel-Schlesewsky & Schlesewsky, 2009). The prominence scales can vary across languages, though there is a general underlying ranking, with respect to morphosyntactic information (nominative > accusative/ ergative > nominative), argument order (argument 1 > argument 2), animacy (animate > inanimate), definiteness (definite > indefinite/ specific > unspecific), and person ($1^{st}/2^{nd}$ person > 3^{rd} Person). Prominent arguments (according to the hierarchy) are rather interpreted as Actors, less prominent arguments as Undergoers. In stage 1 in the eADM, word category information is encountered and the basic constituents are structured; in stage 2, prominence is computed, arguments are linked, and lexical/ associative semantic properties are processed (interacting with discourse environment); in stage 3 the different information types are mapped and integrated.

5 Processing a Second Language

As shown above, several ERP components have been identified and associated with underlying functional characterizations, whereas nowadays functional explanations are, of course, far beyond the (somewhat simplified) classical interpretations. Therefore, a one-to-one mapping from component to function is not always given. Nevertheless, these components are an appropriate instrument for investigating second language processing, because they can at least provide an answer to the question, whether second language processing is qualitatively and/ or quantitatively same or different from L1 processing (in terms of ERP measurements). To be more precise, are different linguistic aspects, like semantics and syntax, learned and processed in the same manner in natives and second language learners? In addition to ERPs, behavioral data (e.g. acceptability and/ or grammaticality judgments; response times) is available in order to examine the conscious performance of L2-learners as compared to native speakers. As will be clear after section 6, it makes sense to subdivide this section into semantic (5.1) and syntactic processing (5.2)[6], as these two aspects differ from each other with respect to several parameters. Further, alternations in AoA, and/ or

[6] For a recent, extensive review on ERP studies investigating bilingualism (including areas that are not discussed in this thesis, e.g. phonetic/ phonological aspects, language switching, or inhibitory effects), see Moreno, Rodríguez-Fornells, and Laine (2008).

proficiency, type of language instruction, or experimental settings affect these aspects differently, as will be shown. Each section presents ERP-studies that investigated L2-processing compared to L1-processing (within or across groups[7]), and present data from longitudinal studies to outline changes in ERPs and behavioral data during L2-learning (5.3). Age of Acquisition (AoA), Years of L2-Experience, and L2-proficiency are discussed at the respective studies. Before turning to ERP and behavioral evidence, a rough introduction to language learning and processing theories is needed.

Since Lenneberg (1967) postulated the Critical Period Hypothesis (CPH), there is an ongoing debate whether second language acquisition is constrained by the same mechanisms that regulate the effectiveness of language exposure during first language acquisition. The CPH proposes the existence of a critical period due to maturational processes (from childhood to adolescence) during which neuroplasticity decreases, and hence the effortlessness of learning languages decreases as well[8]. This assumption relies (among other insights) on different restitution states after brain lesions in children and adults, whereas Lenneberg reports of functional hemispheric compensation mechanisms in children, but not in adults. Interesting (but nevertheless tragic) evidence comes from the 'wild child' Genie (Curtiss, 1977), who acquired the first language after the age of 13. On the lexical-semantic level, she reached a very high level of proficiency (at least near-native like), whereas the performance on the (morpho-) syntactic level remained rudimentary. Consequentially, these insights lead to the assumption that different aspects of language (e.g. phonology, semantics, and syntax) are affected differently by maturational processes and therefore are bound to different sensitive periods.

As will be clear after the results from ERP studies investigating second language processing, there is still no consensus whether late second language learners are able to reach

[7] Picton et al. (2000) point out that marginal across-group differences should be treated with care, as these differences in amplitude, latencies, and distribution might interact with other experimental manipulations than intended and controlled.

[8] In terms of Universal Grammar (UG) based assumptions, some argue that the accessibility to the UG is lost after a critical period (cf. Steinhauer, White, & Drury, 2009).

native-like language competence, and moreover, whether the same mechanisms are involved in the acquisition of a first and a second language. Further, it remains an open question how factors like age of acquisition, years of language experience, and ultimate language proficiency contribute to the processing nature of a second language as compared to a native language.

5.1 Semantic Processing of a Second Language

Ardal et al. (1990) were one of the first to examine lexical-semantic processing in bilinguals' first and second languages, compared to monolinguals; their participants had to perform a lexical decision task while their EEG was recorded. For both groups and all languages they found a N400 after reading anomalous sentences. It should be noted that participants in the L2 group were highly fluent in both, their L1 and L2. However, the mean latency of the N400 was shortest in the monolinguals group, the first language of the bilingual group next, and finally the second language of the bilingual group (approximately +40 ms relative to the learner's L1). There was also a difference in the behavioral data; bilinguals had longer reaction times (RTs) in both, their first and second language, compared to the monolingual group. According to the authors, the longer N400-latencies in bilinguals are due to a larger mental lexicon, thus, bilinguals need more time to integrate the incoming word into the prior context; this might also explain the RT differences. The negativities they observed differed not only in their latency but also in their distribution; whereas the N400 was stronger over electrode sites on the left hemisphere in the bilingual group, the negativity was more right-lateralized in the monolingual group. The authors assume that in addition to the timing differences, there might also be (slightly) different underlying neural generators and/ or networks. Friederici and Wartenburger (2010) point out that a left-lateralized N400 reflects mere lexical-semantic processes, whereas a right-lateralized or global N400 reflects rather general conceptual processes. This assumption fits the results from the study by Ardal et al. (1990); monolinguals have to recruit less lexical sources and more general conceptual processes whereas bilinguals need more bare lexical-semantic processes to interpret the ongoing sentence. Importantly, the differences in the RTs and in the N400 latency did not vary as a function of age of acquisition. Comparable results come from Kutas and Kluender (1994), who report of N400 differences between a bilinguals first and second language: The peak was delayed, the N400 lasted longer, and the amplitude was decreased in the less proficient language as compared to the more proficient language.

A pioneering experiment in the field of second language processing is the study by Weber-Fox and Neville (1996). They tested 61 adult Chinese/ English bilinguals and divided them in groups according to their AoA (1-3, 4-6, 7-10, 11-13, 16+). Sentences used in the semantic condition (29) elicited N400s in all groups; there was no qualitative difference between the test and the control groups. The groups with an AoA of 11+ years showed a delay in the N400 peak latency, whereas there were no significant amplitude differences between groups. The N400 peak latency varied similarly as a function of AoA and years of experience. Further, they found a significant group effect for the accuracy of the acceptability judgments: The bilingual group performed similar than the monolingual group (though overall performance was slightly less accurate), only the 16+ group showed a significantly less accurate performance than the monolingual group. The authors conclude that the judgment accuracy level is not tied to the N400 latency, what leads to the conclusion that conscious performance (e.g. in an acceptability judgment task) might be quite independent from measureable neural responses – at least to a certain degree.

(29) The scientist criticized Max's proof/ #event of the theorem.

Hahne and Friederici (2001) investigated auditory semantic processing in Japanese/ German bilinguals with an AoA 18+ years, and an overall lower proficiency level. They found a N400 in the native German control group and in the L2 group. However, the negativity lasted about 400 ms longer in the learners' group as compared to the native group. This longer duration might reflect the additional time a learner's language processing system needs to integrate the deviant semantic input into the prior sentence context. In addition to the N400, the L2 group showed a late right anterior negativity, which might arise because the learners have to recruit additional conceptual sources to successfully interpret the sentence. However, to a certain extent this is contrary to the interpretation by Ardal et al. (1990), who argued that an L2 learner shows a more focal activation because they can *not* recruit general conceptual processes (but see also footnote 13). Further, the L2 group showed enhanced positivity in the time range from 500 – 1000 ms, and 700 – 1100 ms after the presentation of well-formed sentences. The authors suggest that this additional positivity in language learners even after processing correct sentences might indicate the recruitment of additional processes due to higher syntactic integration costs (Hahne & Friederici, 2001).

In a similar study, Hahne (2001) investigated Russian/ German bilinguals[9] by auditory presenting semantic deviances, as shown in (30). Both groups showed an N400 effect; surprisingly, the effect strength (viz. the difference between the semantically inappropriate sentence and the control sentence) was greater in the L1 group than in the L2 group, whereas the peak was about 200 ms earlier in the native group. Importantly, semantically deviant sentences yielded no significant differences in the N400 time window between the two groups. Hence, the author argues that semantic integration difficulties were similar in both groups, and, with respect to the results from Hahne and Friederici (2001), that ERP responses vary systematically as a function of L2 proficiency. After the presentation of correct sentences learners and natives showed a N400; however, the negativities' amplitude was greater, and the onset was delayed for about 100 ms in the L2 group as compared to the L1 group. Hahne explains these differences in terms of greater effort in language learners to integrate the sentence final word. Accuracy in the judgment task was higher in the L1 group as in the L2 group; response times were shorter in L1 vs. L2.

(30) #Der Ozean/ Die Tür wurde <u>geschlossen</u>.
#The ocean/ The door was being closed.

To disentangle the factors proficiency and AoA, Moreno and Kutas (2005) investigated semantic processing in Spanish/ English bilinguals. They divided their participants into groups according to their dominant language (English or Spanish), and, further, into early or late exposure to English. There was one group with early exposure to Spanish, with

[9] According to the accuracy in the acceptability judgment task, the German learners had a lower error rate than the Japanese learners of German in the study by Hahne and Friederici (2001), what might indicate a higher level of proficiency.

Spanish being the dominant language, one group with a late exposure to English, and English being the non-dominant language. Strikingly, there was one group with an early exposure to English and Spanish, and English being the more dominant language. They found clear effects of language dominance on the N400: The onset began about 10 ms earlier for the dominant language than for the non-dominant (regardless of English or Spanish, respectively). Further, the N400 peaked approximately 27 ms later in the non-dominant language than in the dominant[10]. As the authors conclude, language proficiency and age of exposure/ AoA both contribute to the processing nature of a second language, but, most strikingly, language proficiency/ dominance seems to affect the latency of the N400 in terms of shorter onset and peak delays, and thus indicating faster lexical-semantic processing, irrespective of age of exposure. The authors did not find any distributional differences of the N400. Similar results are reported by Ojima, Nakata, and Kakigi (2005): They tested Japanese speakers who learned English after childhood and reached either a low or a high proficiency level. All groups showed a N400 after processing semantic anomalies. High proficient learners showed a delayed latency compared to native speakers of English, low proficient learners' latency was increased with respect to both other groups.

Midgley, Holcomb, and Grainger (2009) propose an account of semantic L2 processing on the basis of ERP evidence from visual word processing (lexical decision task): They reversed the first and second languages English and French across experiments in order to rule out language-specific differences. The N400 in the L2 was attenuated as compared to the L1, especially over posterior sites, and they observed latency differences between L1 and L2 over anterior sites. In a follow-up experiment, they tested subjects with a similar proficiency level in their L1 and L2 (with a late AoA)[11]. The latency of the N400 over *an-*

[10] Moreover, they found differences regarding the sentence final word type: They found an effect of abstractness vs. concreteness over languages, hence assuming "[…] that bilinguals may adopt a relatively more "concrete/literal" than "abstract/metaphorical" mode of processing when encountering written or aural materials in their weaker language."(Moreno & Kutas, 2005:218). However, this remains quite speculative.

[11] AoA-influences on the N400 differences are ruled-out by a comparison of words acquired before and after the age of eight years, respectively; this comparison did not reveal differences.

terior electrode sites was decreased in the high proficient groups as compared to the low proficient, thus the authors assume that this peak-latency shift reflects language competence. However, as there was an amplitude difference of the N400 over *posterior* sites between L1 and L2, this might indicate processing differences even in high proficient L2 speakers. They argue that with the acquisition of a new language, the lexical entries of the second language are associated with the lexical entries of the first language rather than with the concepts per se. Therefore, L2 words would behave like low-frequency words of the L1 and engender (at least slightly) different N400s, for instance with timing-differences across scalp distribution. With an increase in L2 proficiency, a 'direct' computation route from L2 word forms to concepts establishes; the lexical-semantic processing becomes more and more native-like, as reflected in latency assimilation between L1 and L2 over anterior sites, and amplitude assimilation over posterior sites.

To sum up, lexical-semantic processing in a second language is quite comparable to first language processing, even when the language is learned after childhood, or even puberty. None of the studies mentioned above reports of qualitative differences, neither between a first, and a second language in one population, nor between native speakers, and learners of the same language. However, the ERP results differ with respect to the quantitative differences they found[12]:

- Decreased amplitudes in the less proficient language (Hahne, 2001; Kutas & Kluender, 1994)
- Longer N400 duration (Hahne & Friederici, 2001; Kutas & Kluender, 1994)
- Additional activations (Hahne & Friederici, 2001)

[12] Osterhout et al. (2006) note, that across-group comparisons between natives and learners are problematic as learner groups tend to show greater variability in latency and amplitude, and therefore effects might even be obscured. They argue that the absence of certain components in learners (e.g. LAN) might be due to the greater variability, what can obscure effects during the averaging process. Another problematic point is that the LAN was also reported to be absent in native speakers, what renders this component a less attractive instrument for investigating L2 processing (see footnote 2 in the paper by Osterhout et al. (2006) for investigations on this topic).

- Lateralization differences[13] (Ardal, et al., 1990; Proverbio, Cok, & Zani, 2002)
- Delayed N400 onset/peak latencies in language learners and/ or the non-dominant/ less proficient language (Ardal, et al., 1990; Hahne, 2001; Kutas & Kluender, 1994; Midgley, et al., 2009; Moreno & Kutas, 2005; Ojima, et al., 2005; Weber-Fox & Neville, 1996).

Concerning behavioral data, learners tend to show a lower accuracy rate in acceptability judgment tasks (Hahne, 2001; Hahne & Friederici, 2001; in the AoA 16+ group in Weber-Fox & Neville, 1996), and/ or longer response times (Ardal, et al., 1990; Hahne, 2001).

There is marginal consensus regarding the assumptions that greater N400 latencies reflect slower processing, that longer N400 durations reflect longer integration times, and that decreased N400 amplitudes reflect ‘weaker’ lexical-semantic processing in the non-dominant language(s); however, these are common interpretations, though – as the review above showed – there are various counter directional results. What remains even more uncertain is whether distributional differences between a dominant and a non-dominant language, and/or additional activations correlate with lower language proficiency. Especially with respect to the results reviewed above (Midgley, et al., 2009), where distributional differences in combination with latency and amplitude differences were discussed, it is unclear which factors drive these activations, and, moreover, which experimental conditions and manipulations could shed light on the fine-grained processes involved in lexical-semantic second language processing.

[13] In the study by Weber-Fox and Neville (1996) the authors do not interpret the distributional differences of the N400, because they did not survey data about handedness in the participants’ families; however, there is evidence from Kutas et al. (1988) that the familial history of handedness influences the distribution of the N400.

5.2 Syntactic Processing of a Second Language

The experiment by Weber-Fox and Neville (1996) already introduced above is a perfect example for the diverse results that most studies investigating syntactic second language processing report of. Beside the semantic condition that is outlined in section 5.1, Weber-Fox and Neville also investigated syntactic processing in a L2, using a phrase structure violation (31), a specifity constraint violation (32), and a subjacency constraint violation (33).

(31) The scientist criticized Max's <u>of</u> proof the theorem.

(32) What did the scientist criticize Max's <u>proof</u> of?

(33) What was a proof of <u>criticized</u> by the scientist?

Monolinguals showed a clear ELAN-P600 pattern after the presentation of syntactic deviances, with an additional negativity in the 300 – 500 ms time range. On the contrary, the L2 groups' ERPs were influenced by an AoA 4 < years: The ELAN and the later negativity were less lateralized than in the monolingual group, the late positivity was reduced with an AoA of 10+ years, and absent in the 16+ group[14]. With regard to the specifity constraint violation, the L2 group with an AoA of 11-13 years did not show the ERP asymmetries the monolinguals showed, the 16+ group did not show any effect. The authors interpret the decreased P600 amplitude as the learners' reduced ability to restore the sentences' meaning. The results of the acceptability judgment tasks were as follows: Subjects with an AoA of < 4 years performed similar to the control group (except for the judgments on the subjacency constraint manipulations, where the AoA 4-6 group performed better). Participants with an AoA 4+ years showed a significantly less accurate result than the control group. The error rate of the judgments increased as a function of AoA (phrase structure condition), years of experience (specificity constraint condition), or none of both (subjacency constraint condition). The authors conclude that even short delays in languages exposure alter the ERP patterns of language learners as compared to monolinguals, whereas different linguistic levels are affected distinctly.

[14] The 16+ group showed much reduced positivity beginning around 700ms post stimulus onset.

In the study by Hahne and Friederici (2001) the German control group showed a biphasic ELAN-P600 pattern after being presented with syntactic manipulations (phrase structure violations, like in (2)). The Japanese late learners of German (AoA 18+ years) did neither show an ELAN, nor a late positive *effect*. How comes that they do not even show the late effect? The reason probably lies in the processing nature of correct sentences: The L2 group showed late positivity even after processing correct sentences, thus, in the comparison between the correct and the deviant condition, no P600 effect can be observed, as there is no difference between the two conditions. Hahne and Friederici argue that this might be due to similar processing costs for both sentence types, viz. the learners' processing system has similar difficulties for the integration of correct, and erroneous sentences. Further, the authors explain the absence of the early negativity in terms of language transfer: As there are no prepositions in Japanese, the learner has no prediction about what follows prepositions in German (and what does not), hence, there was no predicted local phrase structure requiring a noun after the preposition. For this reason the violation did not elicit an ELAN.

Hahne (2001), too, reports of an ELAN-P600 pattern for phrase structure violations (2) in the control group. The Russian/ German bilinguals only showed only the late positivity with the peak latency being delayed for 150 ms. However, there were also ERP differences with respect to correct sentences: The L2 groups' N400 after the sentence final particle was delayed for approximately 100 ms, more pronounced over anterior sites, and showed a higher amplitude, as compared to native speakers. Hahne suggests that these differences for correct sentences are due to higher difficulty to integrate the sentence final word into the prior context in learners than in natives. In addition, as the ELAN seems to be highly automatic in nature, the absence of this component might be regarded as the absence of the early automatic phrase structure building processes in L2 learners. Considering the substantial qualitative differences in syntactic processing between the L1 and L2 groups, and bearing in mind that there were only timing differences in the semantic condition, the behavioral results of this study are especially interesting: The L2 learners had the highest error rate in judging correct sentences (8.1%), followed by semantic violations (7.5%), and – somewhat surprisingly – the fewest errors in the syntactic condition (7.1%). However, this raises the question whether the very restricted set of manipulations (correct, semantic, and syntactic) allows the participants to make clear predictions about the grammaticality of the sentences, because whenever a preposition was *not* followed by a noun, the sentence was ungrammatical. In contrast, correct sentences and semantic manipulations were structurally identical, thus not allowing predictions about acceptability based on word category information alone.

In a recent study by Pakulak and Neville (2010) ERPs to phrase structure violations, as in (34), were examined. They tested English native speakers and German learners of English (AoA around 11 years). Importantly, the control and the test group were matched for proficiency according to a standardized measure of English language proficiency. Thus, the two groups differed only in their age of acquisition[15]. The native group showed an early bilateral anterior negativity extending up to 1200 ms post stimulus onset; further, they showed a late positivity. In the non-native group, the violation did not elicit an early negativity; consequently, the authors conclude that persons who acquire a second language later in life have to rely on different mechanisms to achieve a proficiency level that is similar to that of some native speakers. The late positivity in the L2 group was more widespread and larger in amplitude as compared to the L1 group. Pakulak and Neville (2010:11) argue that "[…] late L2 learners rely on more general, controlled, rule-based reprocessing mechanisms in experimental conditions which place demands on second-language processing which more closely approximates those in everyday life […]"[16], as reflected by the more widespread P600.

(34) *Timmy can ride the horse at my <u>his</u> farm.

As in most studies participants had to perform semantic or grammatical judgment tasks, Guo et al. (2009) present evidence that ERP responses are modulated by different task de-

[15] Note that the groups had a similar level of proficiency because the native control group displayed a *lower* level of proficiency. However, the authors report that the proficiency and working memory span might interact in this study, as these two factors were closely tied to each other, whereas the non-native group (students and one professor) had a higher working memory span as the control group.

[16] "Everyday life" refers to communicative and interactive needs, as opposed to focusing on bare syntactic cues.

mands: In their study, participants had to read sentences for comprehension rather than for semantic or syntactic deviances, therefore maybe focusing on the kind of information (semantic or syntactic) that is better available to them. The native control group showed a late positivity for verb sub-categorization violations, as in (35). The Chinese/ English bilinguals, however, showed a N400 rather than a P600. The authors point out that Natives use syntactic information to process verb sub-categorization violations, as the P600 is correlated to structural/ syntactic processes. Contrary to that, the language learners recruit lexical-semantic information for these processes (see also the results of the longitudinal studies outlined in section 5.3). Nevertheless, the only differences that are directly addressed to the different task demands are distributional differences of the reported components that are not further discussed here.

(35) Joe's father didn't *show/ let him drive the car.

Yet, the results by Guo and colleagues (2009) can also be explained in terms of bare word (category) serialization rules: Language learners might employ a strategy that relies on the knowledge of possible (and impossible) word sequences; that is to say, that after *Joe's father didn't show him* ___ learners would expect a noun rather than any other category. Therefore, the violation could also be regarded as a word category violation, and moreover as a violation of lexical properties of the expected continuation of the sentence (at least in L2 learners); this is further supported by findings suggesting that language learners process words as holistic lexemes rather than as word stems plus functional endings, therefore learners are more likely to increasingly use lexical-semantic mechanisms rather than syntactic ones, even for the processing of syntactic information (see section 5.3 for more on this issue); of course this remains speculative.

Similar results are reported in a study investigating the processing of morphosyntactic features of a second language (Weber & Lavric, 2008): They tested highly proficient German/ English participants in their L1 (German) and in their L2 (English), as well as a native English control group. The morphosyntactic violation elicited a LAN/P600 pattern in the English group and in the German groups' L1; in the L2 condition, the syntactic violation elicited a N400/P600 pattern. Because the manipulations were at sentence-final positions, the authors explain the N400 in the L2 group as a wrap-up effect due to slower/ weaker morphosyntactic processing, hence "[…] exacerbat[ing] the pressures on the wrap-up semantic integration mechanism, leading to an N400 in the syntactic condition in L2." (Weber & Lavric, 2008:925).

Regarding the results mentioned above, the study by Friederici, Steinhauer, and Pfeifer (2002) is especially interesting: They investigated the processing of an artificial miniature language (BROCANTO) in a trained group with high BROCANTO proficiency[17] and a control group that was untrained. The syntactic violations clearly elicited an early anterior negativity, followed by an additional negativity, and a late posterior positivity. In the control group, no effect was found. On the behavioral level, trained participants showed a high performance in the grammatical judgment task (93%), and in the probe identification task (89%). Furthermore, they ruled out the possibility that the learned BROCANTO syntax was simply transferred from the participants' L1 German, by using violations that do not exist in German. They conclude that the level of proficiency, rather than AoA, is crucial for establishing early automatic processes, as reflected in the early anterior negativity. Of course the authors are well aware of the fact, that they investigated non-natural language processing, since BROCANTO consists only of a very restricted set of rules and vocabulary, therefore enabling high predictability about the incoming linguistic material. Nevertheless, these results can be regarded as evidence that early processes can be established even in adulthood and are not entirely inhibited by maturational constraints, but at least interact with contextual (or experimental) properties.

Are such native-like patterns in late language learners restricted to artificial languages? Rossi and colleagues (2006) present evidence that this is not the case. They tested 84 German/Italian and Italian/German bilinguals with high and low proficiency, and with a late AoA. Stimulus material contained word category violations, morphosyntactic violations,

[17] The trained group had to reach an accuracy level of 95% in both, production and perception. The training was bound to a chess-like board game (on a computer) where participants had to communicate their moves while playing against each other; overall training sessions lasted up to 25 hours distributed across several days.

and a combined violation containing both of the former. Strikingly, they found a native-like (E)LAN/P600 pattern for all three experimental conditions in both high proficient groups. Further, the low proficient groups showed qualitatively the same pattern for the word category violation and the combined violation (with longer latency times and slightly reduced amplitudes). Only in the agreement violation lower proficient learners displayed a P600 without a LAN. Of course this is evidence that even late language learners can – under certain circumstances – show native-like processing patterns, even in natural languages. But why did even the low proficient learners display a native-like pattern? The authors, as well as others (e.g. Pakulak & Neville, 2010), argue, that these astonishing findings are due to the very restricted set of manipulation, therefore enabling participants to concentrate on just a few relevant syntactic rules. This leads to high predictability in the experimental context. Rossi and colleagues suggest that one difference with respect to former studies (especially Hahne, 2001; Hahne & Friederici, 2001) is the use of simple, active sentences in the 2006 study, and additionally a small set of used words and rules. Ojima et al. (2005) report of similar results: They also found a native-like pattern in late L2-learners. Importantly, they did not have filler sentences and used a slow presentation rate. Hence, the results can – once more – be attributed to the experimental conditions rather than to language proficiency *per se*. Regarding the results from artificial language processing (Friederici, et al., 2002), one might argue that experimental conditions crucially interact with language proficiency and/or AoA, and therefore have an essential impact on the ERP results.

In order to rule out problems related to stimulus material, Steinhauer et al. (2006; cited from Steinhauer, et al., 2009) counterbalanced critical items and their context, respectively, as shown in (36). They tested Chinese/English and French/English bilinguals with low or high language proficiency and a late AoA. In the low proficient groups, they report of a P600. The high proficient group exhibited a native-like LAN/P600 pattern, despite the high AoA, and independent of the first languages' orthographical properties (French and Chinese).

(36) a. The man hoped to <u>enjoy</u> the meal with friends.

b. *The man hoped to <u>meal</u> the enjoy with friends.

c. The man made the meal to enjoy with friends.

d. *The man made the enjoy to meal with friends.

These results suggest that a high proficiency level can lead to native-like processing patterns, even in natural languages, with more diverse manipulations as in former studies. But even though the stimulus material was constructed in order to avoid unwanted material effects, word order was highly predictable in this context, leading to similar conclusions as in other studies facing similar circumstances (Friederici, et al., 2002; Ojima, et al., 2005; Rossi, et al., 2006).

Remarkable results in regard to material-dependent proficiency differences come from a study by Hahne et al. (2006), who investigated the processing of noun plurals and past participles in late Russian/German bilinguals. Participants were more proficient with respect to the past participle manipulation as compared to noun plurals, as revealed by behavioral data[18]. However, the manipulation-specific proficiency difference was also reflected in the ERP data: For the past participle manipulation, bilinguals displayed a biphasic LAN/P600 pattern; for the noun plurals the anterior negativity was absent. The native control group showed a LAN/P600 pattern as a response to both manipulations. Strikingly, the ERP patterns differed with respect to irregular vs. regular morphology, whereas violating applications of regular inflection led to an anterior negativity, the violations of irregular inflectional rules elicited a N400 rather than an LAN. The authors argue that ill-inflected irregular words are processed like pseudowords (in terms of holistic word form storage), therefore leading to a N400. Nevertheless, this can be regarded as evidence for the interaction of proficiency and stimulus material, as even in the offline tasks participants exhibited proficiency differences relative to stimulus material.

An interesting manipulation was used in a recent study examining number and gender agreement violations within and across phrases (Dowens, Vergara, Barber, & Carreiras, 2009). English/Spanish bilinguals with a late AoA and a high proficiency level were tested.

[18] It has to be noted that two different behavioral tasks were employed for the two manipulations: A nonce-word production task for the past participles, and an acceptability judgment task for the noun plurals.

Regarding the violations within phrases, learners showed the same LAN/P600 pattern as the native control group did. In the across phrase condition, the learners displayed a P600 but no anterior negativity, whereas the control group did. However, in all conditions the learners showed latency and amplitude differences as compared to the control group. Beside factors like AoA, proficiency with respect to different linguistic features seems to be crucial for the observed neural activity.

Morgan-Short et al. (2007; cited from Steinhauer, et al., 2009) present material that supports the idea that beside speaker-internal properties (e.g. AoA, proficiency), and material-bound effects, also the type of language learning affects ERP patterns. They used the miniature language BROCANTO2 and tested one group with implicit training, and another with explicit, classroom-like instruction, whereas it is important to note that both groups reached a similar high level of proficiency. A LAN/P600 pattern for word category and agreement violations was only elicited in the implicitly trained group, hence leading to the assumption that brain responses are also shaped by the type of instruction, such as learning through interaction vs. formal classroom instruction.

As most studies investigated syntactic processing using violation paradigms, Kotz, Holcomb, and Osterhout (2008) investigated the processing of local ambiguities in highly proficient Spanish/English bilinguals (AoA ~5 years). They used sentences like in the study by Osterhout and Holcomb (1992), as displayed in (21). Both groups displayed a P600 to local ambiguity (and to phrase structure violations); the authors argue that this indicates similar processing of grammatical but locally ambiguous sentences. In the 'offline' acceptability judgment task, where participants had unlimited time to respond, acceptability judgments were equal in natives and non-natives. In the 'online' acceptability judgment task (as part of the ERP experiment), native speakers responded more accurately as compared to learners. The accuracy of the learners in the phrase structure condition was at chance level, though they showed similar ERP patterns; hence the findings assume higher automaticity in native speakers despite the early AoA and the high proficiency of the L2 group.

Taken together, the studies reviewed above clearly demonstrate that syntactic L2 processing is much more amenable to modifications of AoA, years of experience, ultimate proficiency, stimulus material, and tasks demands than semantic processing is. There are numerable reported ERP differences between native speakers of a language and language learners:

- No (E)LAN (Dowens, et al., 2009; Hahne, 2001; Hahne, et al., 2006; Pakulak & Neville, 2010; Rossi, et al., 2006; Weber-Fox & Neville, 1996)
- Less lateralized frontal negativity (Weber-Fox & Neville, 1996)
- Increased amplitudes (Weber-Fox & Neville, 1996)
- Reduced P600 (Dowens, et al., 2009; Rossi, et al., 2006; Weber-Fox & Neville, 1996)
- Delayed P600 (Hahne, 2001)
- Distributional differences of the P600 (Guo, et al., 2009; Pakulak & Neville, 2010)
- No P600 (Weber-Fox & Neville, 1996)
- 'Syntactic' N400s (Guo, et al., 2009; Weber & Lavric, 2008)
- No effects (Hahne & Friederici, 2001; Weber-Fox & Neville, 1996).

Remarkably, some studies were able to elicit native-like patterns in high proficient learners (Dowens, et al., 2009; Friederici, et al., 2002; Kotz, et al., 2008; Ojima, et al., 2005; Rossi, et al., 2006; Steinhauer et al., 2006; cited from Steinhauer, et al., 2009), whereas the respective experimental settings, such as presentation rate and stimulus material remain critical points. Regarding behavioral data, language learners tend to give less accurate responses[19] as compared to natives (e.g. Guo, et al., 2009; Hahne, 2001; Hahne & Friederici, 2001; Hahne, et al., 2006; Rossi, et al., 2006; Weber-Fox & Neville, 1996). In addition, data from one study (Kotz, et al., 2008) suggests that there can be a discrepancy between explicit responses, as reflected in behavioral data, and ERP patterns, whereas the assimilation between natives and non-natives in behavioral data might precede the assimilation in ERP data. In the next section, studies that employed longitudinal designs are reviewed in order

[19] Note that in the study by Rossi et al. (2006) there was a main effect of group. However, the high proficient learners had an accuracy rate of 93.9% (German) and 96.6% (Italian), the low proficient learners an accuracy rate of 90% (German) and 86.6% (Italian). The high accuracy rates are also addressed to the restricted set of manipulations.

to shed light on changes in the brain, and in explicit responses related to increasing language proficiency.

5.3 Longitudinal Designs

Novice learners of French (L1 English) were tested in a longitudinal study by Osterhout et al. (2006; 2008). Participants had to decide whether the presented word pairs were related, unrelated, or pseudowords of French. ERPs were measured on the target word (second word) of each pair, respectively. The first session was after ~14 hours of instruction, the second session after ~60 hours of instruction, and the last session after ~140 hours of classroom instruction. ERP plots of this study are depicted in Figure 6. Strikingly, they found a N400 effect between word and pseudoword targets in the learners group after only ~14 hours of instruction. With increasing language exposure, learners showed also a N400 to unrelated target words. In the last experimental session, learner showed the typical graded N400 amplitudes, with the greatest negativity for pseudoword targets, followed by unrelated targets, and finally related targets (see also the plots from the study by Roehm et al. (2007); Figure 4). Regarding the behavioral data, participants' performance in the lexical decision task was at chance level in the first session.

(37)	a. Related:	chien – chat	("dog – cat")
	b. Unrelated:	maison – soif	("house – thirst")
	c. Pseudoword:	mot – noisier	("word – *pseudoword*")

In addition to the lexical-semantic task, Osterhout and colleagues investigated the processing of morphosyntactic features during the above-mentioned three experimental sessions in the same groups of participants. The question was whether features of a L2 that are absent in a person's L1 can be acquired. They used semantic violations (38a), verbal (38b) and number agreement (38c) violations; verbal agreement is very restricted in English but not in French, and is phonologically realized in French; number agreement is not restricted to French, but is phonologically unrealized in French (at least in the used material). In the violation in (38c) the noun disagrees in number with the determiner.

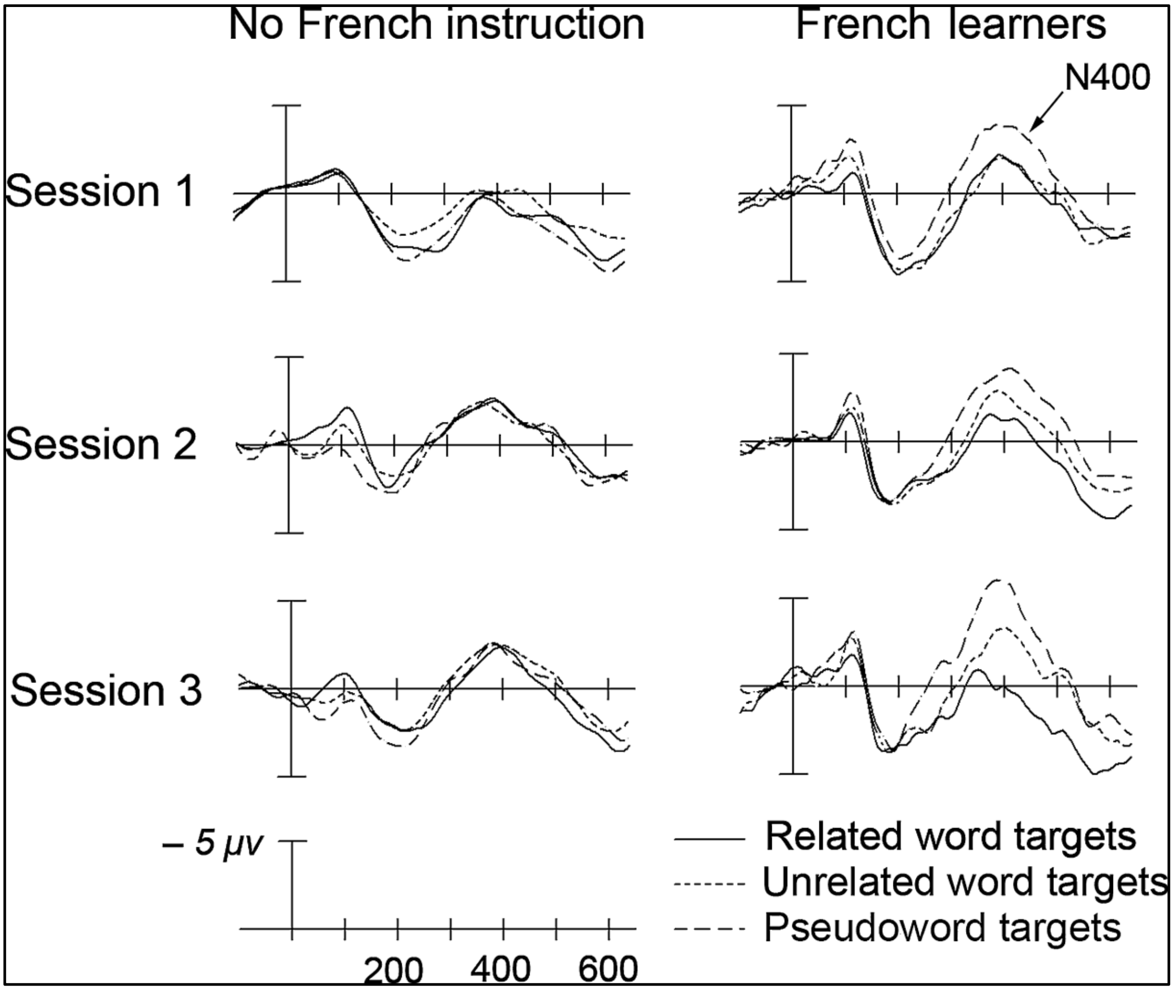

Figure 6 shows the grand average waveforms of the untrained control group (left panel) and the French learners (right panel) after ~14 hours (session 1), ~60 hours (session 2), and after ~140 hours (session 3) of language instruction. ERP measurements are plotted for target words that were semantically related to the prime (solid line), unrelated to the prime (small dashes), or pseudowords (large dashes). (adapted from Osterhout, et al., 2006:210)

(38) a. Sept plus cinq/#livre font douze.

"Seven plus five/#book equals twelve."

b. Tu adores/*adorez le francais.

"You love/*loves French."

c. Tu manges des hamburgers/*hamburger pour dîner.

„You eat hamburgers/*hamburger for dinner."

For the noun phrase agreement conditions, there were no significant differences between test and control sentences. But, as can be seen in the ERP waveforms (Figure 7), the learners showed a N400 rather than a P600 as a response to verbal person violations in the first session. After ~60 hours of instruction (session 2) the L2 group showed late positivity (P600); in the last session the P600 was more pronounced. These findings go hand in hand with other studies that report of N400s after processing syntactic deviances in lower proficient language learners (e.g. Guo, et al., 2009). Osterhout et al. (2006) argue that the change from a N400 to P600 in syntactic processing marks the point, where 'grammaticalization' takes place. Grammaticalization means that certain rules are internalized in a native-like way. Before this point, words are processed as holistic items, therefore requiring lexical-semantic processes. An ill-inflected verb therefore is recognized as an unfamiliar or unexpected combination with a pronoun (e.g. **tu adorez*). After grammaticalization words are recognized as lexical items plus functional properties (e.g. in terms of word stem plus ending), hence being processed via syntactic routes, as reflected in the P600. In other words, after grammaticalization learners do not have to rely on lexical information alone in order to successfully understand a second language.

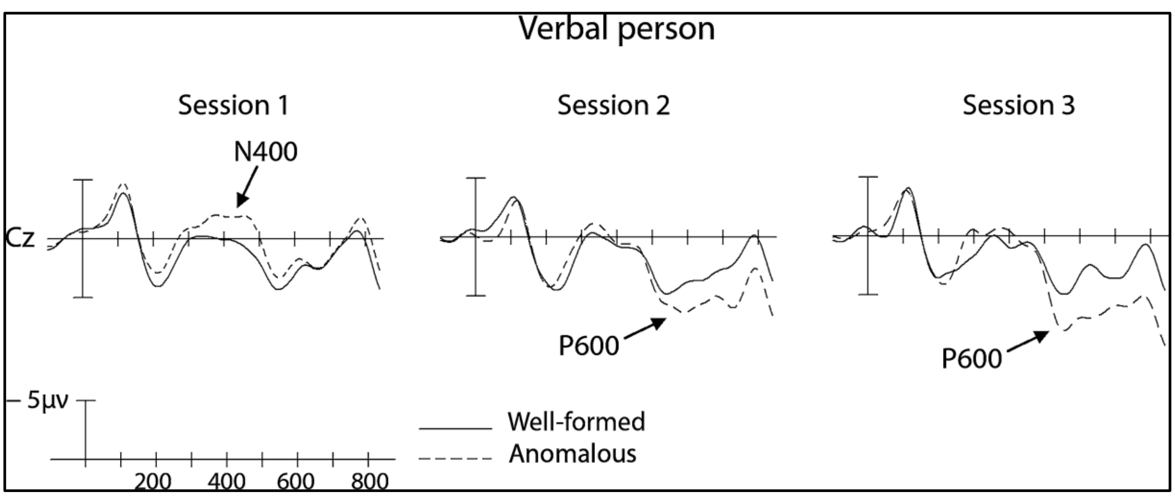

Figure 7. ERP plots of the French learners for the three testing sessions to verbal person agreement violations (dashed line) and to the well-formed control (solid line). (Adapted from Osterhout et al. 2006:219)

With respect of the discrepancy between behavioral responses and ERP signatures in the first session, the authors assume, that the participants' "[…] brains could discriminate between actual words and pseudowords, even if the learners themselves could not do so." (Osterhout, et al., 2006:211). Hence, the results can be viewed as evidence that even little language exposure suffices to establish brain signatures comparable to those of natives, and that with increasing proficiency neural responses assimilate to those of L1 speakers. Similar findings of brain responses preceding conscious responses are reported by McLaughlin, Osterhout, and Kim (2004), whereas they conclude that the most elemental linguistic aspects of a language (e.g. word forms) are acquired even after minimal instruction; therefore the brain did detect word-pseudoword differences, the learners themselves did not. Another interesting insight comes from a study that examined the processing of a miniature version of Japanese (Mini-Nihongo), where German participants were trained to a high level of proficiency (Mueller, Hirotani, & Friederici, 2007). Case violations elicited late positivity in both groups, however, the Japanese natives showed an additional N400; the German learners showed an anterior negativity to one violation type. Whatsoever, whereas the ERP patterns of the two groups clearly differed, the behavioral responses were nearly identical across groups. Davidson and Indefrey (2009) also present evidence that after only one day of instruction of German agreement rules, Dutch natives showed a P600 after the processing of declension violations. Their participants received further training and accuracy in the acceptability judgment task increased in all conditions (they reached the same accuracy level as natives did), whereas ERPs revealed only a P600 to one type of violation. Of course, these results suggest that offline measures, like judgment tasks, are not able to detect early progressions in language learners. Further, behavioral data from natives and non-natives can be identical although brain responses differ significantly; thus, behavioral data alone can not reveal fine-grained processes and processing differences in a way ERPs do. But nevertheless, acceptability judgments provide a fruitful insight in learners' conscious awareness of deviances in a newly learned language.

The longitudinal studies outlined above clearly demonstrate that even after a short period of instruction, specific brain signatures occur in response to certain violations. With increasing language proficiency, the N400's amplitude to semantic deviances increases, and with respect to syntactic processing, a P600 replaces a N400 after grammaticalization, with an increasing P600 amplitude as a higher proficiency level is reached (McLaughlin, et al., 2004; Osterhout, et al., 2006; 2008). Moreover, the distinctive natures of behavioral and ERP data were revealed, since ERP responses might precede conscious, behavioral responses (Mueller, et al., 2007; Osterhout, et al., 2006), or vice versa (Davidson & Indefrey, 2009).

6 The Present Study

Aim of the Study. The initial point of the experiment was to determine neuro-cognitive correlates of second language proficiency, using ERP technique in a violation paradigm. Therefore we designed material containing deviances that were either detectable by beginners, by intermediates, or by advanced learners of English. We tested and compared two groups of participants: One group were native speakers of German with English as a foreign language (L2 group), and the control group were native speakers of English; participants in the L2 group were at least intermediate learners of English. Since former studies that investigated second language proficiency used either a very restricted set of materials, and/ or only very simple constructions, we wanted to create a more natural linguistic environment (in terms of more diverse constructions). Moreover, we wanted to investigate how learners of English with a different proficiency level process structures containing deviances that are detectable by highly proficient speakers only; in addition, we used deviances that are legal in German, but ill-formed in English, in order to examine possible transfer effects from the participants' first language German.

As could be seen from the review of second language processing experiments in the chapters above, one major question in the investigation of second language processing is whether it is possible for non-natives to unfold the same (implicit) predictive mechanisms as native speakers of a language do. This was predominantly done in the syntactic domain by investigating the automaticity of structure-building processes. So, one interesting question is if such predictions are also computed in the semantic domain (or at the interface to more general cognitive domains), in terms of an implicit prediction of the properties of incoming lexical elements. The results that are presented in this thesis exclusively deal with this question; hence, only results from one condition, namely a semantic manipulation, are presented here.

Shortcomings of the Experiment. Initially, the L2 group should have been split up according to the proficiency level of the participants. Even though half of them were students of English and half of them came from a different field, splitting was not possible for the following reasons: We evaluated their English skills only with a language history questionnaire; neither the self-ratings of their language skills, nor factors like AoA, time spent in an English speaking country, and daily exposure to English provided a base for dividing the group; the same applies to the results of the acceptability judgment task, as well as to the collected response times. Another shortcoming of the experiment is of course the low number of participants, as well as the dissimilar number of participants in each group. Further, the task in our experiment was to judge sentences according to their *acceptability* (rather than *grammaticality*); for this reason, it was not possible to disentangle 'real' grammatical knowledge from individual preferences, since we cannot say, for instance, whether a sentence was judged as 'not-acceptable' due to grammatical deviances that were detected, or due to personal matters (like 'words that are unusual').

6.1 Participants

In the L2 group (L1 = German; L2 = English), 14 healthy students of the University of Salzburg (12 female; mean age 23.8 years; age range 22 – 27 years) participated in the experiment; seven participants study English (two graduates), seven studied Linguistics (one graduate). We had 7 participants in the control group, with their native language being English (3 female; mean age 24.7; age range 18 – 43 years). In the native control group, 2 participants acquired a second language early in childhood: One participant was exposed to Filipino from the age of 2 years, and one participant learned German at the age of 6. All participants were right-handed (according to an adapted version of the Edinburgh

Handedness Inventory; Oldfield, 1971), had normal or corrected-to-normal vision, and knew of no neurological or psychological disorders at the time of the experiment.

6.2 Materials

We designed 360 English sentences using nine different conditions, whereas there were 20 pairs of sentences in every condition (well-formed and deviant, respectively). The conditions were grouped into three different stimulus categories according to the degree of difficulty to detect the deviances: Deviances detectable by beginners (A) contained a semantic violation (A1), a morphosyntactic condition violating do-support (A2), a morphosyntactic manipulation violating verbal congruency rules (A3), and, finally, a 'word order' manipulation violating the default SVO-order in English (A4). Deviances detectable by intermediate learners of English (B) contained a *wh*-in-situ violation (B1), a violation of modal infinitival construction rules (B2), and a violation of the compactness of head-initial VPs (B3). The third stimulus category, with deviances detectable by advanced learners only, contained two violations of an exceptional complement selection (C1 and C2). Example stimuli are shown in Table 2. A full list of all stimulus sentences is provided in the appendix A. In this thesis, results from condition A1 will be reported; the results from the other conditions will be presented elsewhere.

Condition	Type of violation	Example stimuli
A1	Lexical-semantic	#The crown was too precious to bark.
		The crown was too precious to wear.
A2	Do-support	*They not watch television.
		They don't watch television.
A3	Congruency	*He play guitar pretty lousy.
		He plays guitar pretty lousy.
A4	Word order (OS)	*The safe took he from the house.
		The thief took it from the house.
B1	*wh*-in-situ: Version 1	*Who crashed the car how?
		Who crashed the car where?
	Version 2	*What did who buy yesterday?
		Who bought what yesterday?

B2	Modal infinitival construction	*The game was <u>not</u> to win.
		The game was <u>hard</u> to win.
B3	Compactness of head-initial VPs	*She intended to help <u>sincerely</u> her parents.
		She intended to care <u>sincerely</u> for her parents.
C1/ C2	Exceptional complement selection	*Amanda believed <u>to</u> meet the singer.
		Amanda hoped <u>to</u> meet the singer.
		*Amanda hoped <u>herself</u> to meet the singer.
		Amanda believed <u>herself</u> to meet the singer.

Table 2: Example sentences for each of the conditions. The test sentence is the upper one, the control sentence underneath, respectively, whereas the critical element is underlined.

The semantic violation in A1 is due to the sentence-final verb, which either does not fit the prior context, and/ or is a mismatch between the thematic properties of the subject and the verb (e.g. the verb ‘to bark’ requires an animate subject, except for special contexts). A very important remark is that the sentential context allows for a very restricted – or even unique – prediction about the sentence-final verb. In A2, the do-support-properties of the negation are violated; that is, negation in these sentences requires do-support, what is not complied in the deviant sentences. In this condition, two positions are of interest: First, the negation itself (**They <u>not</u> drink coke*), as well as the position of the subsequent verb (**They not <u>drink</u> coke*), because, incrementally, the *not* following the subject is still legal, as in constructions like ‘*They not only talk about it…’*. Therefore the actual *violation* is at the position of the verb. A3 is a congruency violation, because the third person subjects require s-marking on the verbs that are used here. A4 violates the obligatory SO-order in English, what is manifested in an incongruous inflected pronoun (Nominative instead of Accusative). It should be noted that condition A4 is more than a bare congruency violation: If the pronoun would be inflected correctly, the sentence would contain a semantic deviation, as, for instance, the verb ‘to take’ requires (or at least prefers) an animate subject. Incrementally, this lexical-semantic mismatch would occur at the position of the verb. In German, this restriction in word order does not apply. Hence, as ‘the’ in English is case-ambiguous, a German learner of English could parse the first NP as the Accusative-object and the Nominative pronoun as the subject, what is perfectly legal in German. In this case, the sentence could be properly interpreted and no processing difficulties would occur.

Condition B1 violates the restriction, that the English adverbial interrogatives ‘how’ and ‘why’ are grammatical in fronted clause positions only. This is not the case for ‘when’ and ‘where’. A grammatical version of the sentence shown above would be ‘Who crashed the car *and* how’. In German, this rule does not apply. Again, the sentence could be parsed – analogous to German – as well-formed. B2 violates the restriction, that infinitive clauses like the ones we used can be modulated with an adverb, but not with a negation. There could be a negative transfer effect again, as the literal translation of the example sentence shown above (‘Das Spiel war nicht zu gewinnen’) is well-formed in German, but not in English. Further, in this condition there are two interesting positions, because the negation following the verb *per se* is – incrementally spoken – no violation, as in constructions like ‘*The game was not too easy to win*.’ Therefore, the following ‘to’ is of interest as well. B3 violates the principle of compactness of head-initial VPs in English; that means that the verb and the object must be adjacent and intervening adverbs render the sentence ungrammatical. For PP Objects this does not apply. In German, intervening adverbs are legal and typical. Condition C1 and C2 use the exceptional status of English verbs, like ‘believe’, which do not select sentential infinitival clauses, as most other verbs of this class do. For a grammatical version, an Accusative subject in the infinitival clause is required (control sentence for C2); this does not apply to verbs like ‘hope’. In German, however, the ‘believe’-constructions in C1 are well-formed and common.

6.3 Procedure

Behavioral Data. All participants filled out a language history questionnaire to provide information about their language background (Age of Acquisition; years of learning; further language skills; time spent in L2-speaking country; L2 language usage per day in percent)[20]. Additionally, participants were asked to rate their speaking, writing, reading, and listening skills in their L2(s) on a seven-point-scale (1 = very poor; 7 = native-like). The questionnaire is attached in the appendix B. Furthermore, all participants filled out an

[20] The Native control group filled out only the questions regarding language proficiency in other languages than their L1 English, the age of acquisition of other languages, and time spent in countries with other languages than their L1, and daily exposure rate to other languages.

adapted version of the Edinburgh Handedness Inventory (Oldfield, 1971), as attached in the appendix C.

EEG Recording. Participants were given written instructions (appendix D) and performed a short training session in order to get familiar with the procedure. Further, they were asked to give acceptability judgments after each sentence, as well as to identify a probe item, whereas half of the sentences contained the probe item, half of them did not. Participants were seated in a chair approximately 100 cm in front of a CRT. Stimulus sentences were presented in the middle of the computer screen as yellow letters on a blue screen. They were asked to avoid eye movements and blinking and press the buttons only when the response signals were visible on the screen. To fixate participants' eyes at the center of the screen and to indicate the beginning of a sentence, each trial began with the presentation of a fixation cross (1000 ms). NPs were presented for 450 ms, single words for 350 ms with an inter-stimulus-interval (ISI) of 100 ms. Every sentence was followed by a blank screen for 550 ms, before a green response cue appeared in the center of the screen for 3000 ms, indicating participants to press a key in order to judge the sentences' acceptability. This was followed by a blank screen for 200 ms before the probe item appeared on the CRT for another 3000 ms, where participants had to judge whether the word was in the sentence or not (see Figure 8). Left and right keys for 'acceptable'/'probe in the sentence' and 'unacceptable'/'probe not in the sentence', respectively, were counter-balanced across participants in order to avoid handedness effects on the responses. Sentences were presented in eight blocks, with each block containing 45 sentences. Additionally, the rest-EEG was recorded from every participant (except from two participants from the native group) for 4 minutes before, and 4 minutes after the experimental session (2 minutes with eyes open, 2 minutes with eyes shut, respectively). The overall experimental session took about 2 ½ hours, and participants were paid for this duration.

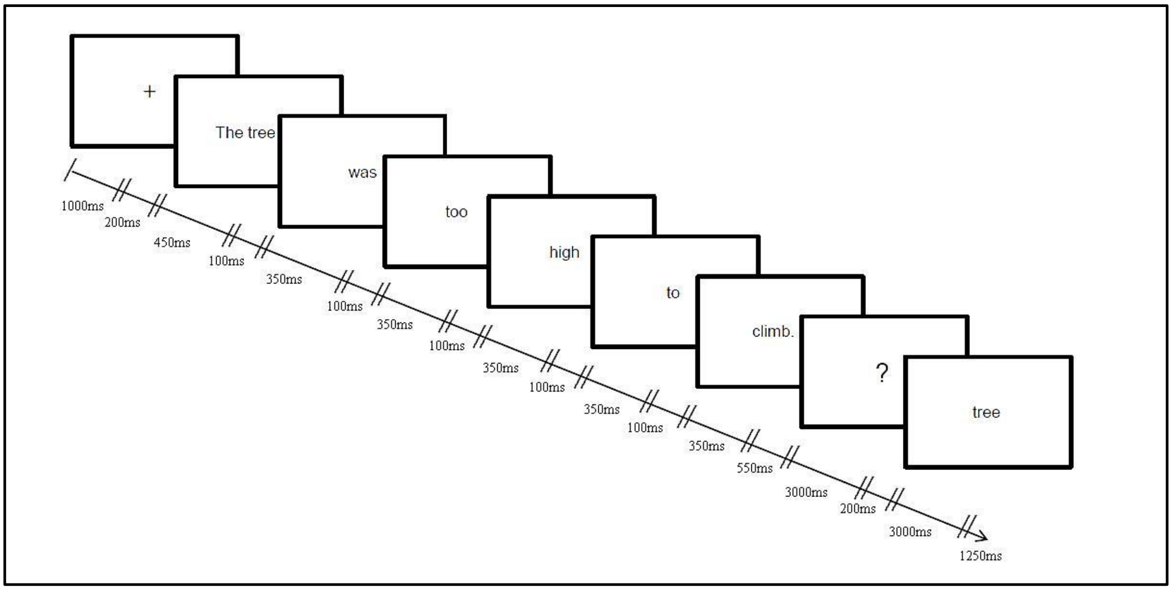

Figure 8. Sequence of a single trial: Fixation cross for 1000 ms, followed by a blank screen for 200 ms; presentation of the sentence (450 ms for NPs, 350 ms for single words), with an ISI of 100 ms; blank screen for 550 ms, followed by the response cue (3000 ms); blank screen for 200 ms; probe item for 3000 ms; blank screen for 1250 ms. Responses were only valid within a 50 – 3000 ms range while the probe item and the cue appeared on the screen.

The EEG was recorded with 24 Ag/ AgCl electrodes that were mounted in an elastic cap according to the 10/20 system (Jasper, 1958). The scalp electrode locations were Fz, FCz, Cz, CPz, Pz, POz, F7/8, F3/4, FC5/6, FC1/2, CP5/6, CP1/2, P7/8, P3/4, O1/2. The horizontal electrooculogramm (HEOG) was recorded from positions at the outer canthus of each eye, the vertical electrooculogramm (VEOG) was recorded from electrodes placed below and above the left eye. Recordings were online referenced against the left mastoid. The EEG signal was collected at a sampling rate of 500Hz with a high pass filter of 250Hz and a notch filter (50Hz). Scalp impedances were kept below 5kΩ (except for two participants in the native control group, where 3 and 5 electrodes, respectively, had impedances between 5 and 10kΩ).

6.4 EEG Evaluation

EEG recordings were manually and semi-automatically screened for artifacts (muscle, noise, eye, and other artifacts) [gradient: maximal 50 μV/ms; difference: maximal 200 μV/200 ms; amplitude: maximal 200 μV, minimal -200 μV; low activity: minimal allowed activity 0.5 μV/ 100ms] and filtered offline (Butterworth Zero Phase Filters) with a low-pass filter of 0.3Hz (48 dB/Oct) and a high-pass-filter of 20Hz (12 dB/Oct). All electrodes were re-referenced offline to the averaged A2 (right mastoid), and the sampling rate was reduced to 250 Hz. Only trials with correct probe responses and trials without artifacts were included in the evaluation. This led to the removal of 9.25% of all trials (L2-group: 4.23%; Native group: 18.57%). Further, EEG data was segmented into 1400 ms epochs consisting of 300 ms before the onset of the critical item, and 1100 ms following the critical item. Due to too many artifacts data of one participant from the L2 group had to be excluded from analysis. Mean averages for the following time windows were computed for every participant (on the base of visual inspection of the waveforms): 220 – 350 ms, 270 – 470 ms, and 510 – 780 ms. We defined the following ROIs: anterior-left (F7, F3, FC1), anterior-right (F8, F4, FC2), central-left (FC5, CP5, CP1), central-right (FC6, CP6, CP2), posterior-left (P7, P3, O1), and posterior-right (P8, P4, O2). Midline electrodes for analysis were Fz, Cz, and Pz. Repeated measure analyses of variance (ANOVAs) were computed, including the following factors: GROUP, WELL-FORMEDNESS, TIME WINDOW, ROI, and MIDLINE as well as separate ANOVAs for each group, if an interaction with group reached significance.

7 Results

7.1 Behavioral Data

Questionnaire Data. A summary of the data for the L2 group obtained from the language history questionnaire is given in Table 3. ‘Time spent in English speaking country’ was excluded because only 1 participant spent more than 3 months in an English speaking country. All of the participants in the L2 group (n=14) started learning English exclusively in school. 12 participants learned English mainly in a formal classroom setting, 1 participant mostly in formal classroom setting and occasionally through interaction, and 1 participant mostly through interaction and occasionally in a formal classroom setting.

Participant Variable	***Component***	***Score***
Self-rated skills	Reading	6
	Writing	6
	Speaking	5
	Listening	6
Age of acquisition/ exposure	Age (mean, range in years)	23.8, 5
	Age of acquisition (mean, range in years)	9.9, 4
	L2 learning duration (mean, range in years)	12.7, 11
	Exposure rate (percent per day, frequency)	0-25, 6
		25, 5
		50, 2
		75, 1
Other languages	Number of spoken languages (mean, range)	2.6, 1-5
	Reading skills in other languages	2
	Writing skills in other languages	1.5
	Speaking skills in other languages	1
	Listening skills in other languages	1

Table 3. Characteristics of the L2 group. ‘Self rated skills’ and skills in ‘Other languages’ are the median ratings on a 7-point scale, with 7 as highest level of proficiency (“native-like”). ‘Number of languages spoken’ is the mean number of spoken languages other than the native language (*including* English); skills in other languages refers to other languages *excluding* English.

Acceptability Judgments. The well-formed sentences in condition A1 were judged as acceptable in 91.2 % of all cases [L2 group: 89.9 %; Native group: 93.6 %], semantically-deviant sentences were judged as acceptable in 5.8 % of all cases [L2 group: 5.5 %; Native group: 6.4 %]. Note that only responses that were in the legal response time window are included in this analysis; responses with a response time < 50 ms and > 3000 ms were excluded [0.77% of all trials]. There was a main effect of condition on the acceptability judgments [$F(1,41) = 816.42$; $p < .001$]. There was no interaction with group [$F(1,40) > 1$].

Response Times. Mean response time (RT) for this condition was 552.78 ms (SD = 185.97) [L2 group: 540.35 ms, SD = 205.13; Native group: 577.64 ms, SD = 143.84]. The mean RT for the well-formed sentences was 558.67 ms (SD = 173.23) [L2 group: 561.02, SD = 196.75; Native group: 553.97, SD = 126.96]; the mean RT for semantically deviant sentences was 546.89 ms (SD = 202.04) [L2 group: 519.98 ms, SD = 218.54; Native group: 601.31 ms, SD = 165.54]. There was no effect of well-formedness on the RTs [$F(1,41) > 1$], as well as no interaction with group [$F(1,40) > 1$].

7.2 ERP Data

As can be seen in Figure 9 (Native group), and Figure 10 (L2 group), responses to well-formed sentences differed from the semantically incongruous sentences in several aspects: The semantically deviant sentences engendered a strong broadly distributed negativity, starting approximately 220 ms post stimulus onset, with a peak around 400 ms, and being most pronounced at central to parietal electrode sites; this negativity effect was followed by a positivity that also was strongest over central to parietal electrode sites. Regarding the waveforms to well-formed sentences, the Native group showed an early positive deflection peaking around 300 ms post stimulus onset, most pronounced at frontal/central-left sites, whereas in the L2 group, well-formed sentences elicited a more negative going waveform as compared to the Native group.

In the first time window (220 – 350 ms), there was a main effect of ROI [$F(1,19) = 18.68$; $p < .001$], and well-formedness [$F(1,19) = 22.31$; $p < .001$], as well as a significant interaction of ROI x well-formedness [$F(1,19) = 3.96$; $p < .005$], whereas the effects' maximum was at central-right electrode sites [$F(1,19) = 22.97$; $p < .001$]. Even though the interaction well-formedness x group did not reach significance [$F(1,18) = 3.02$; $p = .099$], analysis revealed that the effect of well-formedness was stronger in the Native group [$F(1,6) = 18.49$; $p < .01$], as compared to the L2 group [$F(1,12) = 7.55$; $p < .05$]. Regarding midline electrode sites, there was a main effect of group [$F(1,18) = 4.66$; $p < .05$], midline [$F(1,19) = 47.26$; $p < .001$], and well-formedness [$F(1,19) = 25.35$; $p < .001$].

In the 270 – 470 ms time-window, there was a main effect of group [$F(1,18) = 7.88$; $p < .05$], ROI [$F(1,18) = 7.82$; $p < .001$], and well-formedness [$F(1,18) = 105$; $p < .001$], as well a an interaction of ROI x group [$F(1,19) = 2.23$; $p = .058$], and ROI x well-formedness [$F(1,19) = 14.05$; $p < .001$]. Further analysis revealed that the effect of well-formedness was strongest at central-right electrode sites [$F(1,19) = 113.69$; $p < .001$]; the effect of group was strongest at posterior-right sites [$F(1,18) = 11.42$; $p < .005$]. Separate analysis for the two groups showed that the effect of well-formedness was most pronounced at posterior-right sites in the Native group [$F(1,6) = 39.18$; $p < .001$], and central-right sites in the L2 group [$F(1,12) = 87.83$; $p < .001$]. Analysis of the midline electrodes revealed a main effect of group [$F(1,18) = 8.98$; $p < .01$], midline [$F(1,19) = 23.35$; $p < .001$], and well-formedness [$F(1,19) = 102.66$; $p < .001$]. There was a significant interaction of midline x well-formedness [$F(1.19) = 16.18$; $p < .001$]. The effect of well-formedness was strongest at the electrode Pz [$F(1,19) = 112.81$; $p < .001$].

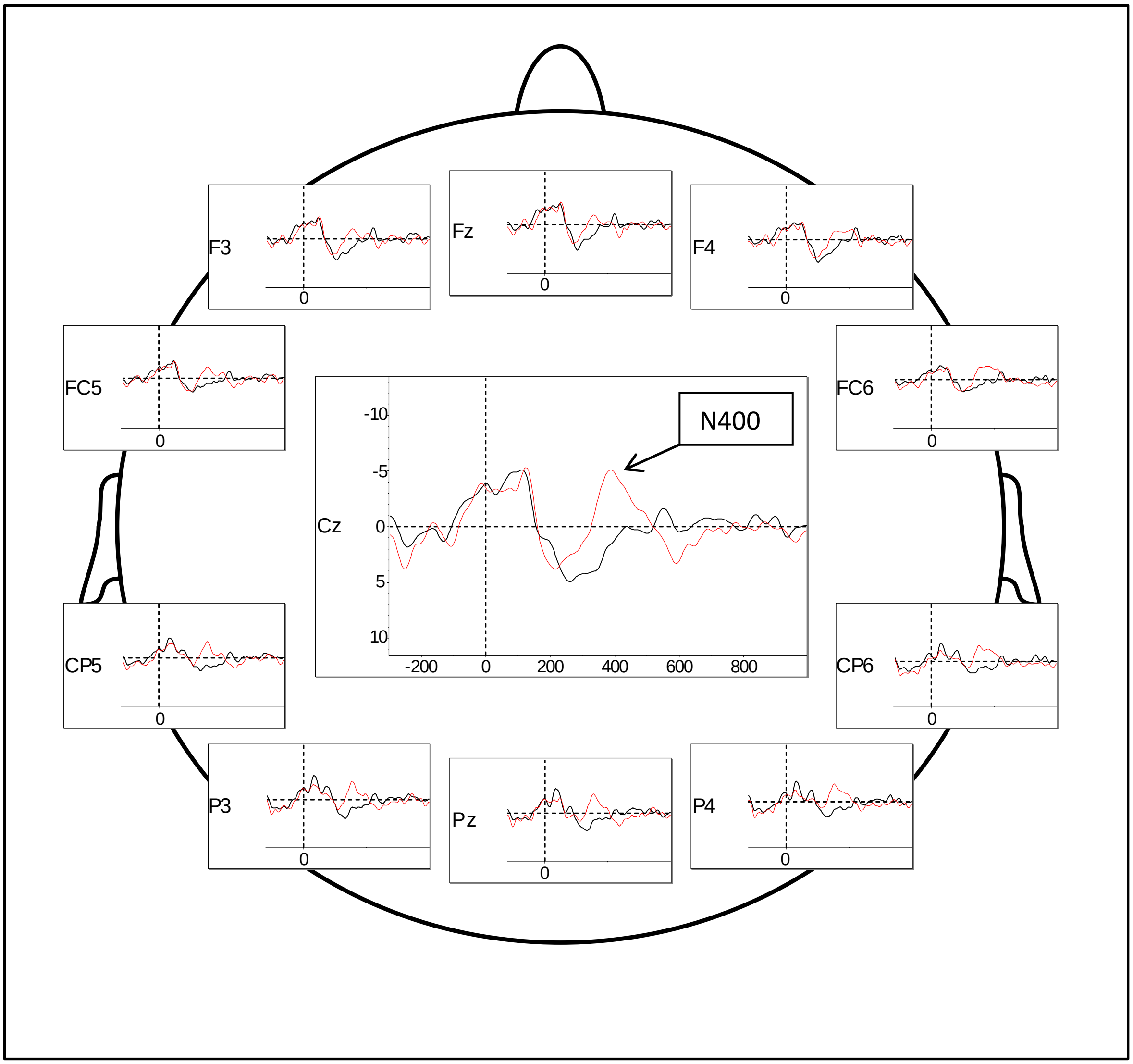

Figure 9. Grand average waveforms for the semantic violation (red line), and control sentence (black line) of the Native group. Negativity is plotted upwards. The vertical line at zero marks the presentation onset of the critical stimulus.

In the last time-window (510 – 780 ms) there was a main effect of ROI [$F(1,19) = 5.65$; $p < .01$], as well as a significant interaction of ROI x well-formedness [$F(1,19) = 4.71$; $p < .05$], whereas the effect of well-formedness was most pronounced at central-left electrode sites [$F(1,19) = 4.17$; $p = .056$]. At midline electrode sites, there was a main effect of mid-line [$F(1,19) = 5.52$; $p < .01$], and an interaction of midline x well-formedness [$F(1,19) = 3.25$; $p = .05$], with the effect of well-formedness being strongest at Pz [$F(1,19) = 6$, $p < .05$].

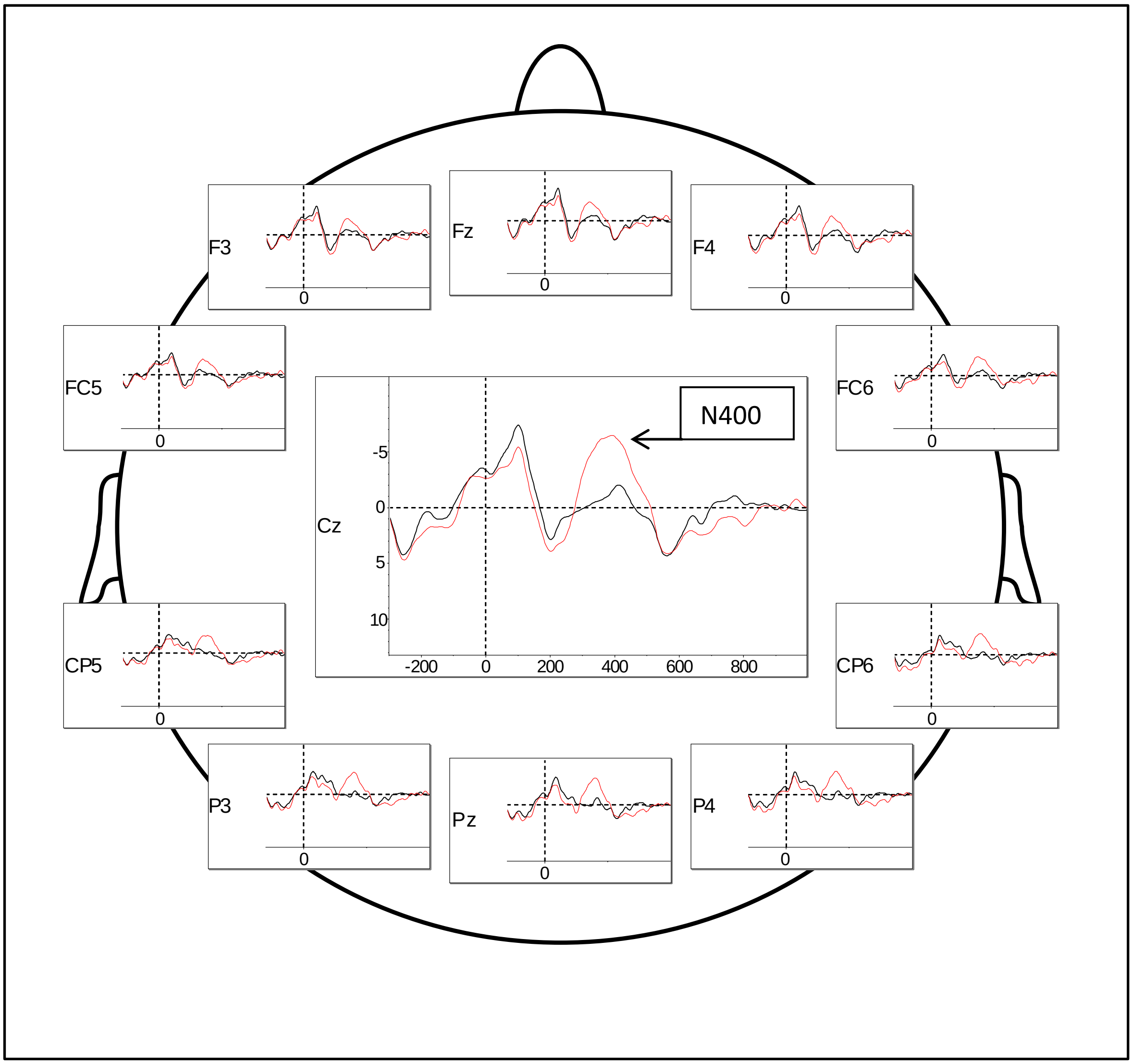

Figure 10. Grand average waveforms of the semantic violation (red line), and control sentence (black line) of the L2 group. Negativity is plotted upwards. The vertical line at zero marks the presentation onset of the critical stimulus.

8 Discussion

In this experiment we compared the processing of semantically deviant to well-formed English sentences in two different groups: Participants in one group were native speakers of English (Native control group), participants in the other group were German native speakers who started learning English around the age of 10, with a mean learning duration of 12.7 years (L2 group). Even though half of the participants in the L2 group study English at university and half of the participants came from a different field, there were no differences in the self-ratings of their English skills; moreover, the median rating was 'very good' for reading, writing, and listening, and 'good' for speaking (on a seven point scale). Taken together, data from the language history questionnaire and the performance in the acceptability judgment task – where both groups performed equally well – participants from the L2 group can be classified as at least intermediate proficient. The semantic violation in this experiment was due to an incongruity of the sentence final verb and the preceding sentential context. As expected, both groups showed a N400 effect for the semantically deviant as compared to the control sentence. In addition to the N400 effect, we observed a positive deflection as a response to the well-formed control sentences in the Native group but not in the L2 group.

Processing Semantically Incongruous Words in a Second Language. The negative deflection for the deviant sentences did not differ significantly across the two groups, what might indicate similar semantic processing in natives and language learners. This is in one line with other studies that report of (qualitatively) similar electrophysiological manifestations of the processing of semantically deviant structures in language learners and native speakers (e.g. Ardal, et al., 1990; Hahne, 2001; Kutas & Kluender, 1994; Midgley, et al., 2009; Moreno & Kutas, 2005; Ojima, et al., 2005; Weber-Fox & Neville, 1996). These studies conclude that even after a short period of language learning semantic processing becomes alike to that of native speakers. Our participants from the L2 group had an relatively early AoA and long learning duration as compared to the participants from other studies, which also report of similar semantic processing in native speakers and language learners (e.g. Hahne, 2001; Hahne & Friederici, 2001; Weber-Fox & Neville, 1996). Therefore we assume that the similar ERP responses to semantically deviant sentences we observed for natives and language learners were in fact due to similar underlying processing mechanisms.

With respect to the distribution of the N400, we observed the negativity being most pronounced over the right hemisphere in both groups. Friederici and Wartenburger (2010) call attention to the lateralization of the N400: According to the authors, a right-lateralized N400 indicates rather general conceptual than bare lexical-semantic processes; a more left-lateralized negativity reflects the recruitment of more bare lexical-semantic processes. This in turn can be regarded as an indicator for language proficiency, as with increasing language proficiency less lexical-semantic sources have to be recruited in order to successfully interpret a sentence. Of course this assumption faces several findings from the opposite direction, since various dissimilar distributional differences between native speakers and language learners have been found.

Task Demands, Parsing Strategies, and Component Overlap. Another observation is the positive deflection with a peak around 300 ms post stimulus onset for semantically congruous sentence endings in the Native group, but not in the L2 group. The statistical analysis did not yield significant differences in the time window of this positivity (220 – 350 ms). Both groups showed a significant difference between deviant and well-formed sentences, whereas the effect was stronger in the Native group as compared to the L2 group. The fact that both groups showed a significant effect in this time window might be due to the general problem of overlapping components, what is especially problematic for statistical evaluation. With the methods used here, it is impossible to tell whether the differences concerning the processing of well-formed structures reflect a decrease of the N400, or an increase of the P300 in the Native group relative to the L2 group. However, visual inspection of the waveforms strongly suggests a difference between the two groups: Semantically congruent sentence endings engendered a strong positivity – starting at approximately 250 ms post stimulus onset – in the Native group relative to the L2 group (Figure 11).

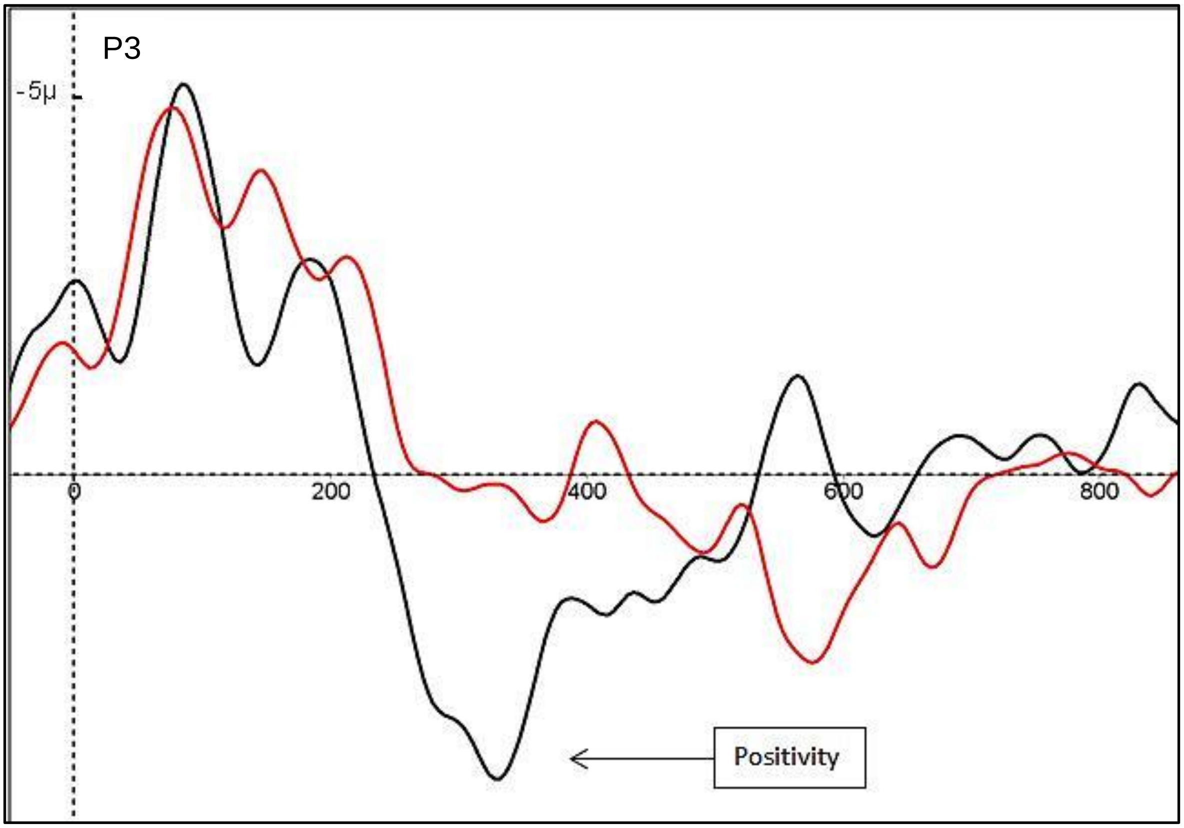

Figure 11. Grand Average waveforms for the semantically congruent sentence ending for the Native group (black line), and L2 group (red line) at electrode P3. Negativity is plotted upwards. The vertical line at zero marks the presentation onset of the critical stimulus. It can be seen that the waveforms of the two groups clearly differ from one another; the Native group shows a strong positivity whereas the L2 group does not.

The voltage maps[21] in Figure 12 (L2 group), and Figure 13 (Native group) highlight the difference even more: In the first time window (200 – 228 ms), both groups show a similar voltage distribution. In subsequent windows (232 – 380 ms), the voltages of the L2 group hardly change, whereas the Native group shows a broadly distributed positivity in all following time windows.

[21] Note that the maps here represent *absolute* voltages rather than difference maps between two conditions; differences maps would entail the same problems that were responsible for the lack of significance between the two groups in the statistical analysis.

Roehm et al. (2007) present evidence that in a highly predictable context, a P300 effect can be observed when the prediction is met, given the task enables or requires such a prediction (see also Figure 4). The sentences in the correct condition of our experiment enabled a solid prediction about the upcoming input (see appendix A)[22]; this might be a possible explanation for the positive deflection in the Native group. Hence, one might conclude that the mechanisms that enable this prediction are different in a first and a second language – even if an intermediate L2-proficiency level is met. Former studies that investigated changes in ERP responses to semantic deviances with increasing language proficiency (e.g. McLaughlin, et al., 2004; Osterhout, et al., 2006; Osterhout, et al., 2008) used either word pairs with different semantic relatedness, or a sentence context that did not facilitate expectancy in a sufficient way in order to make a unique prediction.

In the study by Roehm et al. (2007), a P300 could only be observed when the context was narrow enough to raise the likelihood of a certain incoming lexical element (even to the degree of unambiguous expectation); this was the case in sentential contexts (*The opposite of black is <u>white</u>*), but not in word pair contexts (*black* – *<u>white</u>*) with a *lexical decision task*. Additionally, Roehm and colleagues found ERP differences that can be attributed to different parsing strategies; participants who employed a prediction-based parsing strategy in the word-pair context – with an *antonym judgment task* – did show a P300, participants who employed a kind of ‘wait-and-see’ strategy did not show this positivity. Of course this assumption remains speculative and is only supported by reaction time differences between the two strategy groups, although these differences did not reach significance in the statistical analysis. The group that showed the positivity gave faster responses to semantically related sentences than the group that did not show the positivity.

Regarding the results from our study, similar assumptions can be drawn on the base of reaction time differences (although they did not reach significance between the two groups): The RTs of the Native group to well-formed sentences were shorter as compared to the L2 group. This is in one line with the findings by Roehm et al., who also report of shorter RTs of the ‘positivity-group’. Whereas the differences in the study by Roehm et al. are bound to

[22] Of course this assumption requires further verification, for instance with a cloze probability experiment.

different task demands *and* parsing strategies, respectively, the differences we observed in this study can only be attributed to different parsing strategies, since the task demands were the same in both groups. The parsing strategies that are *accessible* to native speakers and non-native speakers must be responsible for the difference in the ERP-pattern. This in turn means that the mechanisms that enabled the prediction in the Native group were not accessible to the L2 group. If the prediction was met, a positive deflection could be observed, if there was no prediction (what was the case in the L2 group), no positivity was engendered.

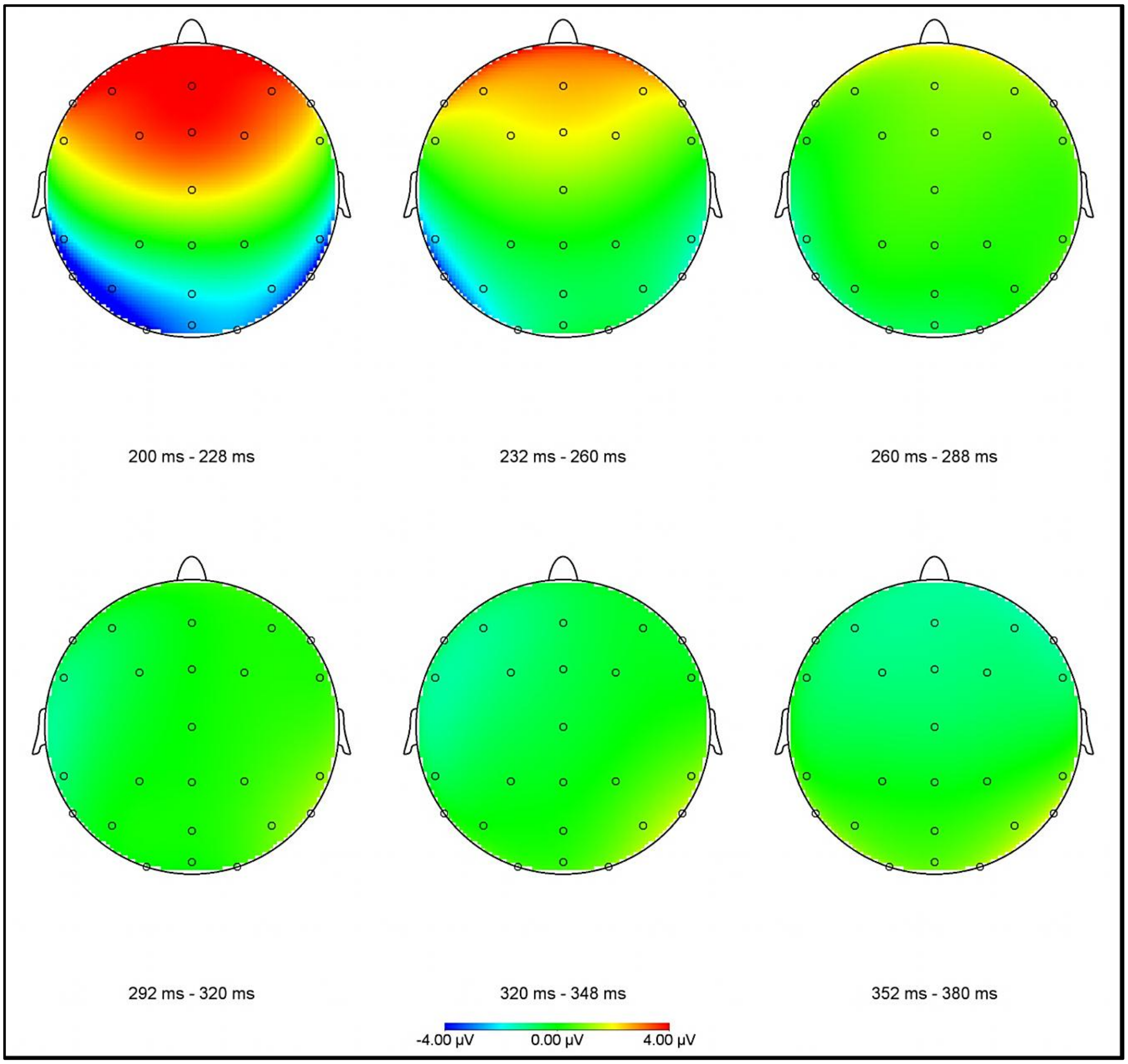

Figure 12. Voltage maps of the L2 group for the semantically congruent sentence endings from 200 – 380 ms post onset of the critical stimulus. As can be seen, the slight positivity in the first time window changes to an absolute voltage around 0.00 µV that remains stable in all subsequent time windows.

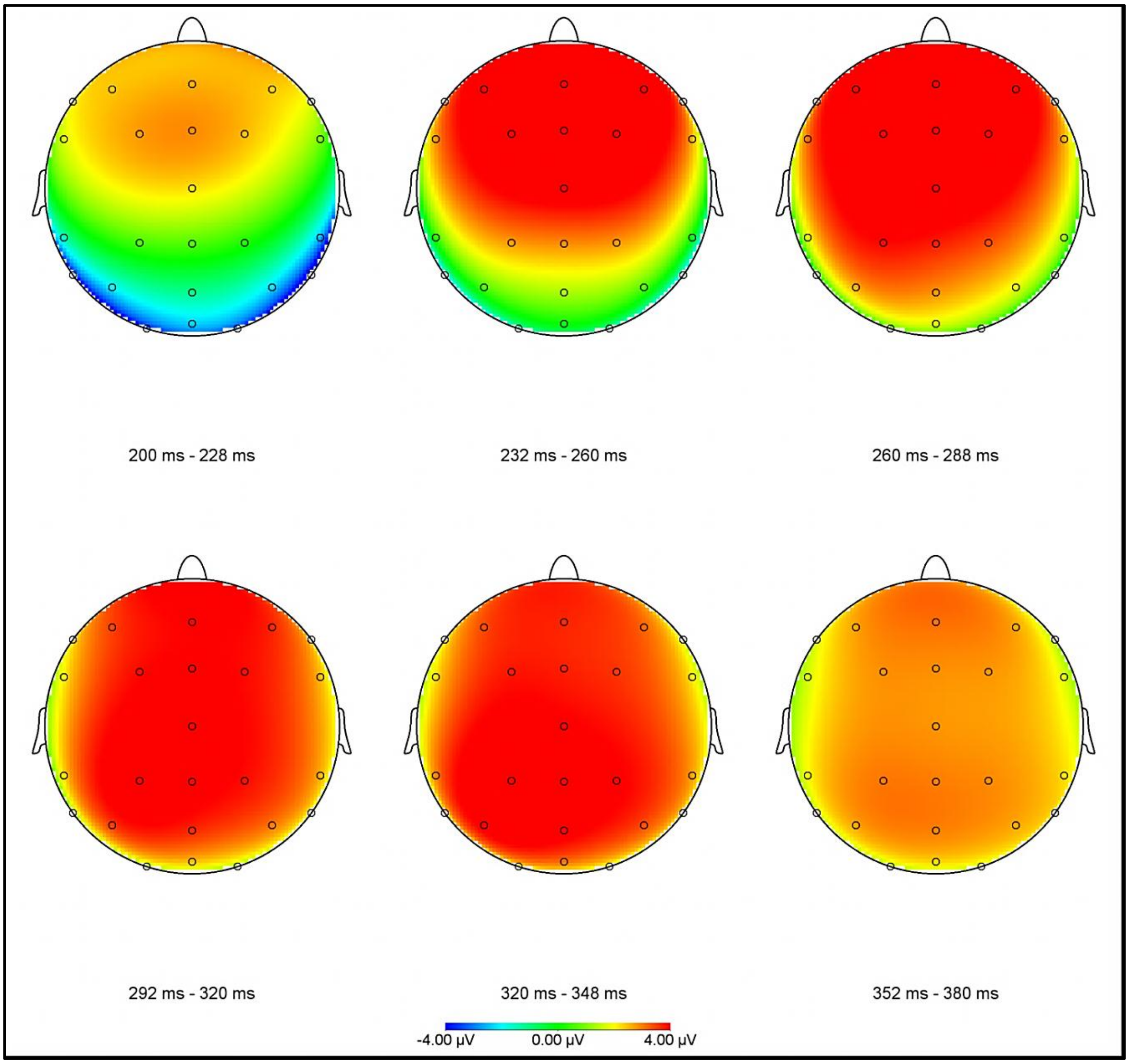

Figure 13. Voltage maps of the Native group for the semantically congruent sentence endings from 200 – 380 ms post onset of the critical stimulus. The slight positivity in the first time window changes to a pronounced broadly distributed positivity over all following time windows.

Consequences for Correlates of Second Language Proficiency. Coming back to the initial aim of the study, namely to spot neuro-cognitive correlates of language proficiency, the ERP differences that we found in lexical-semantic processing between native speakers and language learners seem to be a promising new perspective on the investigation of second language processing. Moreover, if these findings are supported by further experiments, prediction-based proficiency correlates can make an essential contribution to L2 processing models. At this point, a short (modified) outline of Steinhauer, White, and Drury's (2009)[23] classification of ERP results according to different proficiency levels should be drawn. They introduce a six stage model of the temporal dynamics of (late) second language acquisition:

- In phase 1, language learners can not discriminate grammatical from ungrammatical structures; the ERP responses to grammatical and ungrammatical structures also do not differ from one another (e.g. the control group from Friederici, et al., 2002; Osterhout, et al., 2006; Osterhout, et al., 2008).
- In phase 2, semantic and syntactic manipulations elicit N400s, because words are processed rather holistically than being analyzed as word stems and functional endings (Guo, et al., 2009; Osterhout, et al., 2006; 2008).
- Phase 3 is characterized by the *grammaticalization* of certain rules (see also section 5.3). Syntactic deviances elicit a P600, rather than a N400, though this late positivity often is qualitatively different from that of natives (Guo, et al., 2009; Hahne, et al., 2006; Osterhout, et al., 2006; 2008; Pakulak & Neville, 2010; Weber-Fox & Neville, 1996; Weber & Lavric, 2008).
- In phase 4, reanalysis processes assimilate to those of native speakers, as reflected in a qualitative assimilation of the late positivity (Hahne, 2001; Hahne & Friederici, 2001; Weber-Fox & Neville, 1996).
- In phase 5, a biphasic (E)LAN/ P600-pattern as a response to (morpho-) syntactic deviances can be observed, though there might be differences in amplitude and distribution when compared to native speakers (Friederici, et al., 2002; Ojima, et al., 2005; Rossi, et al., 2006).
- In the last stage of second language learning, and therefore with the highest possible language proficiency, ERP responses from language learners and native speakers do not differ from one another (Hahne, et al., 2006).

[23] This model is a compendium that adequately reflects the conclusions about second language proficiency of the majority of the studies that are cited here.

All these observations rely on error-related ERP responses. In Steinhauer, White, and Drury's (2009) classification, proficiency is measured in terms of (implicit neural) error detection; that is, the higher proficient a learner is in a second language, the more automatic are the neural responses to deviances, what is reflected in the occurrence of either an early negativity ([E]LAN), and shorter onset and/ or peak latencies of a LAN, and P600. In the case of semantic processing, shorter onset and/ or peak latencies of the N400 reflect greater language proficiency. This assumption appears plausible and is supported by the results of many experiments. Nonetheless, the foundation of this definition of language proficiency is a facilitation (and acceleration) of error detection rather than a facilitation of interpretational processes[24]. Hence, a future direction for the examination of second language processing (or at least a valuable contribution) might be the investigation of the development of mechanisms that enable predictions based on prior linguistic context.

[24] This means, that with increasing proficiency, more resources are available for interpretational, and/ or predictive processes, and less resources are required in order to simply process a sentence.

9 Acknowledgements

First of all I would like to thank my brother Roman, his wife Bibiane, and their gorgeous daughter Lilia Claire for being either fruitful help, inspiration, recreation (in terms of a cool beer, and/or babbling like babies for hours), or simple but necessary distraction from this thesis.

It is impossible to express in mere words my gratitude to Susanne, not just for coordinating the proofreading procedure, ingestion, and dish-washing, but also for the million little things that made my days easier during the last months.

A substantial contribution came from my parents, who gave me the possibility to study whatever I want – so I ended up in Linguistics.

A massive 'Thank You!' goes to Colleen Caffrey, Kathleen Omerza, and Jessica Replogle for proofreading this thesis and other important papers I had to write in English!

Last but not least, my gratitude goes to Dietmar Röhm and Hubert Haider, who gave me the chance to see science (with all its up- and downsides) beyond characters on paper, all the colleagues who helped in data acquisition, and of course all the patient and uncomplaining participants, who gave their blood, sweat, tears, and brain signatures.

"I am glad, it is finished!" (Freunberger, 2005:51)

10 References

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language, 39*(2), 187-205.

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*(2), 201-212.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *European Archives of Psychiatry and Clinical Neuroscience, 87*(1), 527-570.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). *Processing syntax and morphology: a neurocognitive perspective*: Oxford University Press.

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language, 51*(4), 495-522.

Bornkessel, I., & Schlesewsky, M. (2006). The Extended Argument Dependency Model: A Neurocognitive Approach to Sentence Comprehension Across Languages. *Psychological Review, 113*(4), 787-821.

Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002). Grammar overrides frequency: evidence from the online processing of flexible word order. *Cognition, 85*(2), B21-B30.

Bornkessel, I., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: an individual differences approach. *Cognitive Brain Research, 21*(1), 11-21.

Brown, C. M., & Hagoort, P. (1999). *The Neurocognition of Language*. New York: Oxford University PressBrown, C.M.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*(2), 159-168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, *18*(17), 1851-1854.

Burkhardt, P., & Roehm, D. (2007). Differential effects of saliency: An event-related brain potential study. *Neuroscience Letters, 413*(2), 115-120.

Caton, R. (1875). Electrical Currents of the Brain. *The Journal of Nervous and Mental Disease, 2*(4), 610.

Coulson, S., King, J. W., & Kutas, M. (1998). Expect the Unexpected: Event-related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes, 13*(1), 21 - 58.

Curtiss, S. (1977). *Genie: a psycholinguistic study of a modern-day "wild child"*. New York: Academic Press.

Davidson, D. J., & Indefrey, P. (2009). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience, 21*(3), 433-446.

Donchin, E., & Ritter, W.,& McCallum, W.C. (1978). Cognitive Psychophysiology: The endogenous Components of the ERP. In E. Callaway, Tueting, P., and Koslow, S. (Ed.), *Event-related brain potentials in man.* (pp. 349-411). New York: Academic Press.

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2009). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870-1887.

Federmeier, K. D., & Kutas, M. (1999). A Rose by Any Other Name: Long-Term Memory Structure and Sentence Processing. *Journal of Memory and Language, 41*(4), 469-495.

Filik, R., Sanford, A. J., & Leuthold, H. (2008). Processing Pronouns without Antecedents: Evidence from Event-related Brain Potentials. *Journal of Cognitive Neuroscience, 20*(7), 1315-1326.

Freunberger, R. (2005). *Source Specific Analysis of Evoked and Induced Gamma Activity in Picture Perception.* Master Thesis. Paris-Lodron University of Salzburg, Salzburg.

Friederici, A. D. (1995). The Time Course of Syntactic Activation During Language Processing: A Model Based on Neuropsychological and Neurophysiological Data. *Brain and Language, 50*(3), 259-281.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct Neurophysiological Patterns Reflecting Aspects of Syntactic Complexity and Syntactic Repair. *Journal of Psycholinguistic Research, 31*(1), 45-63.

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*(3), 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 of the United States of America, 99*(1), 529-534.

Friederici, A. D., & Wartenburger, I. (2010). Language and brain. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 150-159.

Frisch, S. (2000). *Verb-Argument-Struktur, Kasus und thematische Interpretation beim Sprachverstehen*: Dissertation. Max Planck Institute of Cognitive Neuroscience.

Frisch, S., & Schlesewsky, M. (2001). The N400 reflects problems of thematic hierarchizing. *Neuroreport, 12*(15), 3391-3394.

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*(4), 556-568.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics, 22*(2), 123-134.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *NeuroImage, 20*(Supplement 1), S18-S29.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439 - 483.

Hagoort, P., Brown, C.M., Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown, Hagoort, P. (Ed.), *Neurocognition of language* (pp. 273-316). Oxford: Oxford University Press.

Hahne, A. (2001). What's Different in Second-Language Processing? Evidence from Event-Related Brain Potentials. *Journal of Psycholinguistic Research, 30*(3), 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*(7), 1302-1318.

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*(02), 123-141.

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*(2), 194-205.

Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological Processing in a Second Language: Behavioral and Event-related Brain Potential Evidence for Storage and Decomposition. *Journal of Cognitive Neuroscience, 18*(1), 121-134.

Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology, 74*(3), 337-346.

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*(10), 371-375.

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass, 1*(6), 571-591.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159-201.

Kaan, E., & Swaab, T. Y. (2003). Repair, Revision, and Complexity in Syntactic Analysis: An Electrophysiological Differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98-110.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225.

Kluender, R., & Kutas, M. (1993). Bridging the Gap: Evidence from ERPs on the Processing of Unbounded Dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language, 100*(3), 257-261.

Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica, 128*(3), 514-527.

Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117-129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology, 11*(2), 99-116.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., & Kluender, R. (1994). What is who violating? A reconsideration of linguistic violations in light of event- related brain potentials. In H. J. Heinze, Münte, T.F. & Mangun, G.R. (Ed.), *Cognitive electrophysiology: Basic and clinical applications.* (pp. 183-210). Boston: Birkhäuser.

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics Electrified II (1994-2005). In J. T. Matthew & A. G. Morton (Eds.), *Handbook of Psycholinguistics (Second Edition)* (pp. 659-724). London: Academic Press.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Luck, S. J. (2005a). *An Introduction to the event related potential technique*. Cambridge: MIT Press.

Luck, S. J. (2005b). Ten simple Rules for designing ERP Experiments. In T. C. Handy (Ed.), *Event Related Potentials: A Methods Handbook*. Massachusetts: MIT Press.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience, 7*(7), 703-704.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory and Cognition, 23*(4), 477-494.

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics, 22*(3), 281-300.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research, 22*(2), 205-220.

Mueller, J., Hirotani, M., & Friederici, A. D. (2007). ERP evidence for different strategies in the processing of case markers in native speakers and non-native learners. *BMC Neuroscience, 8*(1), 18.

Munte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395*(6697), 71-73.

Neville, H., 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*(2), 151-165.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP Study of Second Language Learning after Childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience, 17*(8), 1212-1228.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*(6), 785-806.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice Learners, Longitudinal Designs, and Event-Related Potentials: A Means for Exploring the Neurocognition of Second Language Processing. *Language Learning, 56*, 199-230.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21*(6), 509-521.

Otten, L. J., and Rugg, M.D. (2005). Interpreting Event- Related Brain Potentials. In T. C. Handy (Ed.), *Event Related Potentials: A Methods Handbook* (pp. 3-16). Massachusetts: MIT Press.

Pakulak, E., & Neville, H. J. (2009). Proficiency Differences in Syntactic Processing of Monolingual Native Speakers Indexed by Event-related Potentials. *Journal of Cognitive Neuroscience, 22*(12), 2728-2744.

Pakulak, E., & Neville, H. J. (2010). Maturational Constraints on the Recruitment of Early Processes for Syntactic Processing. *Journal of Cognitive Neuroscience*, 1-14.

Phillips, C., Mattout, J., Rugg, M. D., Maquet, P., & Friston, K. J. (2005). An empirical Bayesian solution to the source reconstruction problem in EEG. *NeuroImage, 24*(4), 997-1011.

Proverbio, A., Cok, B., & Zani, A. (2002). Electrophysiological measures of language processing in bilinguals. *Journal of Cognitive Neuroscience, 14*(7), 994 - 1017.

Roehm, D., Bornkessel-Schlesewsky, I., Rösler, F., & Schlesewsky, M. (2007). To predict or not to predict: influences of task and strategy on the processing of semantic relations. *Journal of Cognitive Neuroscience, 19*(8), 1259-1274.

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The Impact of Proficiency on Syntactic Second-language Processing of German and Italian: Evidence from Event-related Potentials. *Journal of Cognitive Neuroscience, 18*(12), 2030-2048.

Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters, 384*(3), 228-233.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition, 18*(4), 367-379.

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research, 25*(1), 13-41.

Swaab, T. Y., Camblin, C. C., & Gordon, P. C. (2004). Electrophysiological Evidence for Reversed Lexical Repetition Effects in Language Processing. *Journal of Cognitive Neuroscience, 16*(5), 715-726.

van Berkum, J. J., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: evidence from the N400. *Journal of Cognitive Neuroscience, 11*(6), 657-671.

van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early Referential Context Effects in Sentence Processing: Evidence from Event-Related Brain Potentials. *Journal of Memory and Language, 41*(2), 147-182.

van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition, 19*(1), 95-112.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational Constraints on Functional Specializations for Language Processing: ERP and Behavioral Evidence in Bilingual Speakers. *Journal of Cognitive Neuroscience, 8*(3), 231-256.

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology, 45*(6), 920-925.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology, 36*(5), 559-570.

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Research, 1130*, 167-180.

Appendices (in the electronic version only)

A. Full list of all stimuli

B. Language History Questionnaire

C. Adapted Version of the Edinburgh Handedness Inventory

D. Written Instruction (left version only)

Appendix A: Full list of stimuli, whereas (a) is the control and (b) the test sentence, except for condition C1/C2, where (a) is the control for the test sentence (b), and (d) is the control for (c).

A1	1	a	The tree was too high to climb.
A1	1	b	The tree was too high to laugh.
A1	2	a	The bread was too old to eat.
A1	2	b	The bread was too old to paint.
A1	3	a	The pants were too small to fit.
A1	3	b	The pants were too small to smile.
A1	4	a	His voice was too soft to hear.
A1	4	b	His voice was too soft to swim.
A1	5	a	The box was too heavy to move.
A1	5	b	The box was too heavy to chew.
A1	6	a	The bone was too damaged to repair.
A1	6	b	The bone was too damaged to intend.
A1	7	a	The rent was too high to pay.
A1	7	b	The rent was too high to dive.
A1	8	a	The concert was too crowded to enjoy.
A1	8	b	The concert was too crowded to ring.
A1	9	a	The writing was too hard to read.
A1	9	b	The writing was too hard to glow.
A1	10	a	The task was too difficult to understand.
A1	10	b	The task was too difficult to beg.
A1	11	a	The crown was too precious to wear.
A1	11	b	The crown was too precious to bark.
A1	12	a	The movie was too scary to watch.
A1	12	b	The movie was too scary to bake.
A1	13	a	The fruit was too sour to consume.
A1	13	b	The fruit was too sour to train.
A1	14	a	The air was too polluted to breathe.
A1	14	b	The air was too polluted to smear.
A1	15	a	The river was too wide to cross.
A1	15	b	The river was too wide to record.
A1	16	a	The mother was too busy to phone.
A1	16	b	The mother was too busy to fold.
A1	17	a	The dog was too fast to catch.
A1	17	b	The dog was too fast to print.
A1	18	a	The rabbit was too cute to kill.
A1	18	b	The rabbit was too cute to fuel.
A1	19	a	The birds were too injured to cure.
A1	19	b	The birds were too injured to edit.
A1	20	a	The scorpions were too tiny to find.
A1	20	b	The scorpions were too tiny to plan.
A2	1	a	They don’t drink coke.
A2	1	b	They not drink coke.
A2	2	a	They don’t eat pasta.
A2	2	b	They not eat pasta.
A2	3	a	They don’t watch television.
A2	3	b	They not watch television.
A2	4	a	They don’t do sports.
A2	4	b	They not do sports.
A2	5	a	They don’t write poems.
A2	5	b	They not write poems.
A2	6	a	They don’t wash socks.
A2	6	b	They not wash socks.
A2	7	a	They don't talk rubbish.
A2	7	b	They not talk rubbish.
A2	8	a	They don't steal food.
A2	8	b	They not steal food.
A2	9	a	They don't start fights.
A2	9	b	They not start fights.
A2	10	a	They don't remember jokes.
A2	10	b	They not remember jokes.
A2	11	a	He doesn’t like music.
A2	11	b	He not likes music.

A2	12	a	He doesn't climb mountains.
A2	12	b	He not climbs mountains.
A2	13	a	He doesn't touch animals.
A2	13	b	He not touches animals.
A2	14	a	He doesn't water plants.
A2	14	b	He not waters plants.
A2	15	a	He doesn't answer questions.
A2	15	b	He not answers questions.
A2	16	a	She doesn't need meat.
A2	16	b	She not needs meat.
A2	17	a	She doesn't count money.
A2	17	b	She not counts money.
A2	18	a	She doesn't finish work.
A2	18	b	She not finishes work.
A2	19	a	She doesn't drive cars.
A2	19	b	She not drives cars.
A2	20	a	She doesn't receive emails.
A2	20	b	She not receives emails.
A3	1	a	He knows movies pretty well.
A3	1	b	He know movies pretty well.
A3	2	a	He describes events quite appropriately.
A3	2	b	He describe events quite appropriately.
A3	3	a	He recognizes faces very fast.
A3	3	b	He recognize faces very fast.
A3	4	a	He takes walks very often.
A3	4	b	He take walks very often.
A3	5	a	He plays guitar pretty lousy.
A3	5	b	He play guitar pretty lousy.
A3	6	a	He hates cooking very much.
A3	6	b	He hate cooking very much.
A3	7	a	He speaks English fairly fluently.
A3	7	b	He speak English fairly fluently.
A3	8	a	He cleans windows very seldom.
A3	8	b	He clean windows very seldom.
A3	9	a	He likes hats rather small.
A3	9	b	He like hats rather small.
A3	10	a	He drinks beer every day.
A3	10	b	He drink beer every day.
A3	11	a	She enjoys massages pretty much.
A3	11	b	She enjoy massages pretty much.
A3	12	a	She smokes cigarettes fairly often.
A3	12	b	She smoke cigarettes fairly often.
A3	13	a	She draws faces very well.
A3	13	b	She draw faces very well.
A3	14	a	She sings lullabies every day.
A3	14	b	She sing lullabies every day.
A3	15	a	She buys jewellery rather often.
A3	15	b	She buy jewellery rather often.
A3	16	a	She sells realties pretty well.
A3	16	b	She sell realties pretty well.
A3	17	a	She wears glasses quite often.
A3	17	b	She wear glasses quite often.
A3	18	a	She lends movies very frequently.
A3	18	b	She lend movies very frequently.
A3	19	a	She makes debts pretty seldom.
A3	19	b	She make debts pretty seldom.
A3	20	a	She sheds tears hardly ever.
A3	20	b	She shed tears hardly ever.
A4	1	a	The thief took it from the house.
A4	1	b	The safe took he from the house.
A4	2	a	The postman put it in the mailbox.
A4	2	b	The letter put he in the mailbox.
A4	3	a	The player shot it in the goal.
A4	3	b	The ball shot he in the goal.
A4	4	a	The driver took it to the factory.
A4	4	b	The commodity took he to the factory.
A4	5	a	The fire-fighter extinguished it in the building.
A4	5	b	The fire extinguished he in the building.
A4	6	a	The vocalist sang it in the opera.
A4	6	b	The song sang he in the opera.
A4	7	a	The nurse prepared it in the hospital.

A4	7	b	The medicine prepared she in the hospital.
A4	8	a	The student lost it in the library.
A4	8	b	The script lost he in the library.
A4	9	a	The secretary sent it to the businessman.
A4	9	b	The e-mail sent she to the businessman.
A4	10	a	The teacher gave it to the pupils.
A4	10	b	The test gave he to the pupils.
A4	11	a	The policeman placed it under the bridge.
A4	11	b	The radar placed he under the bridge.
A4	12	a	The daughter ate it from the basket.
A4	12	b	The apple ate she from the basket.
A4	13	a	The owner forgot it at the university.
A4	13	b	The bike forgot he at the university.
A4	14	a	The girl threw it on the bed.
A4	14	b	The skirt threw she on the bed.
A4	15	a	The father attached it to the wall.
A4	15	b	The picture attached he to the wall.
A4	16	a	The mason dropped it from the balcony.
A4	16	b	The brick dropped he from the balcony.
A4	17	a	The attorney revealed it before the judge.
A4	17	b	The evidence revealed he before the judge.
A4	18	a	The mechanic repaired it in the garage.
A4	18	b	The car repaired he in the garage.
A4	19	a	The waiter spilled it on the table.
A4	19	b	The soup spilled he on the table.
A4	20	a	The king read it to the audience.
A4	20	b	The statute read he to the audience.
B1	1	a	Who crashed the car where?
B1	1	b	Who crashed the car how?
B1	2	a	Who joined the dinner when?
B1	2	b	Who joined the dinner why?
B1	3	a	Who entered the meeting when?
B1	3	b	Who entered the meeting why?
B1	4	a	Who called the boss where?
B1	4	b	Who called the boss how?
B1	5	a	Who bought the ticket when?
B1	5	b	Who bought the ticket why?
B1	6	a	Who slapped the girl where?
B1	6	b	Who slapped the girl how?
B1	7	a	Who visited the patient where?
B1	7	b	Who visited the patient why?
B1	8	a	Who used the dictionary when?
B1	8	b	Who used the dictionary how?
B1	9	a	Who took the pill when?
B1	9	b	Who took the pill why?
B1	10	a	Who killed the bird where?
B1	10	b	Who killed the bird how?
B1	11	a	Who bought what yesterday?
B1	11	b	What did who buy yesterday?
B1	12	a	Who stole what recently?
B1	12	b	What did who steal recently?
B1	13	a	Who recorded what lately?
B1	13	b	What did who record lately?
B1	14	a	Who broke what upstairs?
B1	14	b	What did who brake upstairs?
B1	15	a	Who drank what abroad?
B1	15	b	What did who drink abroad?
B1	16	a	Who announced what annually?
B1	16	b	What did who announce annually?
B1	17	a	Who indicated what yesterday?
B1	17	b	What did who indicate yesterday?
B1	18	a	Who published what accidentally?
B1	18	b	What did who publish accidentally?
B1	19	a	Who burned what nearby?
B1	19	b	What did who burn nearby?
B1	20	a	Who ordered what generously?
B1	20	b	What did who order generously?
B2	1	a	The book was hard to read.
B2	1	b	The book was not to read.
B2	2	a	The grass was hard to cut.
B2	2	b	The grass was not to cut.

B2	3	a	The bread was hard to chew.
B2	3	b	The bread was not to chew.
B2	4	a	The fruits were easy to harvest.
B2	4	b	The fruits were not to harvest.
B2	5	a	The play was nice to watch.
B2	5	b	The play was not to watch.
B2	6	a	The meal was excellent to eat.
B2	6	b	The meal was not to eat.
B2	7	a	The game was hard to win.
B2	7	b	The game was not to win.
B2	8	a	The offer was hard to reject.
B2	8	b	The offer was not to reject.
B2	9	a	The rules are easy to obey.
B2	9	b	The rules are not to obey.
B2	10	a	The helmet was perfect to wear.
B2	10	b	The helmet was not to wear.
B2	11	a	The candle was easy to light.
B2	11	b	The candle was not to light.
B2	12	a	The machines were easy to handle.
B2	12	b	The machines were not to handle.
B2	13	a	The bottles are easy to empty.
B2	13	b	The bottles are not to empty.
B2	14	a	The profession was easy to learn.
B2	14	b	The profession was not to learn.
B2	15	a	The knife was difficult to sharpen.
B2	15	b	The knife was not to sharpen.
B2	16	a	The path was hard to follow.
B2	16	b	The path was not to follow.
B2	17	a	The journey was easy to plan.
B2	17	b	The journey was not to plan.
B2	18	a	The roses were easy to grow.
B2	18	b	The roses were not to grow.
B2	19	a	A bank is hard to rob.
B2	19	b	A bank is not to rob.
B2	20	a	A license is easy to get.
B2	20	b	A license is not to get.
B3	1	a	They tried to believe firmly in the mission.
B3	1	b	They tried to fulfil firmly the mission.
B3	2	a	She wanted to stare constantly at the moon.
B3	2	b	She wanted to admire constantly the moon.
B3	3	a	They decided to wait patiently for the rain.
B3	3	b	They decided to watch patiently the rain.
B3	4	a	They intended to ask politely for their drinks.
B3	4	b	They intended to finish politely their drinks.
B3	5	a	Nobody wants to pay constantly for that service.
B3	5	b	Nobody wants to use constantly that service.
B3	6	a	He hated to fall madly for that girl.
B3	6	b	He hated to love madly that girl.
B3	7	a	He promised to drive carefully to the party.
B3	7	b	He promised to join carefully the party.
B3	8	a	She struggled to smile lovely at the boy.
B3	8	b	She struggled to kiss lovely the boy.
B3	9	a	They longed to look intensively at the painting.
B3	9	b	They longed to absorb intensively the painting.
B3	10	a	He decided to look carefully after the house.
B3	10	b	He decided to clean carefully the house.
B3	11	a	Nobody wants to speak harshly to the kids.
B3	11	b	Nobody wants to punish harshly the kids.
B3	12	a	They tried to pass slowly around the baby.
B3	12	b	They tried to calm slowly the baby.
B3	13	a	Nobody wants to cheat steadily on the partner.
B3	13	b	Nobody wants to hurt steadily the partner.
B3	14	a	She intended to care sincerely for her parents.
B3	14	b	She intended to help sincerely her parents.
B3	15	a	They decided to close completely down the company.
B3	15	b	They decided to destroy completely the company.
B3	16	a	He managed to climb entirely up the ladder.
B3	16	b	He managed to build entirely the ladder.
B3	17	a	She promised to confess completely to the police.
B3	17	b	She promised to trust completely the police.
B3	18	a	They intended to fight fiercely against the troops.

B3	18	b	They intended to support fiercely the troops.
B3	19	a	She loved to dress beautifully up her doll.
B3	19	b	She loved to paint beautifully her doll.
B3	20	a	Nobody likes to go anxiously to the doctor.
B3	20	b	Nobody likes to see anxiously the doctor.
C1/C2	1	a	Amanda hoped to meet the singer.
C1/C2	1	b	Amanda believed to meet the singer.
C1/C2	1	c	Amanda hoped herself to meet the singer.
C1/C2	1	d	Amanda believed herself to meet the singer.
C1/C2	2	a	Sarah managed to catch the train.
C1/C2	2	b	Sarah believed to catch the train.
C1/C2	2	c	Sarah managed herself to catch the train.
C1/C2	2	d	Sarah believed herself to catch the train.
C1/C2	3	a	Suzie failed to pass the exam.
C1/C2	3	b	Suzie believed to pass the exam.
C1/C2	3	c	Suzie failed herself to pass the exam.
C1/C2	3	d	Suzie believed herself to pass the exam.
C1/C2	4	a	Alex promised to finish the treatment.
C1/C2	4	b	Alex believed to finish the treatment.
C1/C2	4	c	Alex promised himself to finish the treatment.
C1/C2	4	d	Alex believed himself to finish the treatment.
C1/C2	5	a	Tony failed to leave the maze.
C1/C2	5	b	Tony believed to leave the maze.
C1/C2	5	c	Tony failed himself to leave the maze.
C1/C2	5	d	Tony believed himself to leave the maze.
C1/C2	6	a	Erica hoped to get another chance.
C1/C2	6	b	Erica believed to get another chance.
C1/C2	6	c	Erica hoped herself to get another chance.
C1/C2	6	d	Erica believed herself to get another chance.
C1/C2	7	a	Mary managed to find a hostel.
C1/C2	7	b	Mary believed to find a hostel.
C1/C2	7	c	Mary managed herself to find a hostel.
C1/C2	7	d	Mary believed herself to find a hostel.
C1/C2	8	a	Matt failed to see the whales.
C1/C2	8	b	Matt expected to see the whales.
C1/C2	8	c	Matt failed himself to see the whales.
C1/C2	8	d	Matt expected himself to see the whales.
C1/C2	9	a	Gwen hoped to get the job.
C1/C2	9	b	Gwen expected to get the job.
C1/C2	9	c	Gwen hoped herself to get the job.
C1/C2	9	d	Gwen expected herself to get the job.
C1/C2	10	a	John managed to solve the riddle.
C1/C2	10	b	John expected to solve the riddle.
C1/C2	10	c	John managed himself to solve the riddle.
C1/C2	10	d	John expected himself to solve the riddle.
C1/C2	11	a	Melissa failed to get political asylum.
C1/C2	11	b	Melissa expected to get political asylum.
C1/C2	11	c	Melissa failed herself to get political asylum.
C1/C2	11	d	Melissa expected herself to get political asylum.
C1/C2	12	a	Colleen promised to bring her notebook.
C1/C2	12	b	Colleen expected to bring her notebook.
C1/C2	12	c	Colleen promised herself to bring her notebook.
C1/C2	12	d	Colleen expected herself to bring her notebook.
C1/C2	13	a	Thom hoped to overcome this problem.
C1/C2	13	b	Thom expected to overcome this problem.
C1/C2	13	c	Thom hoped himself to overcome this problem.
C1/C2	13	d	Thom expected himself to overcome this problem.
C1/C2	14	a	Deborah managed to take the lead.
C1/C2	14	b	Deborah expected to take the lead.
C1/C2	14	c	Deborah managed herself to take the lead.
C1/C2	14	d	Deborah expected herself to take the lead.
C1/C2	15	a	Kyle managed to win the race.
C1/C2	15	b	Kyle imagined to win the race.
C1/C2	15	c	Kyle managed himself to win the race.
C1/C2	15	d	Kyle imagined himself to win the race.
C1/C2	16	a	Randy promised to publish that story.
C1/C2	16	b	Randy imagined to publish that story.
C1/C2	16	c	Randy promised himself to publish that story.
C1/C2	16	d	Randy imagined himself to publish that story.
C1/C2	17	a	Paul hoped to lead the league.
C1/C2	17	b	Paul imagined to lead the league.

C1/C2	17	c	Paul hoped himself to lead the league.
C1/C2	17	d	Paul imagined himself to lead the league.
C1/C2	18	a	Robert failed to achieve the award.
C1/C2	18	b	Robert imagined to achieve the award.
C1/C2	18	c	Robert failed himself to achieve the award.
C1/C2	18	d	Robert imagined himself to achieve the award.
C1/C2	19	a	Jim hoped to cross the desert.
C1/C2	19	b	Jim imagined to cross the desert.
C1/C2	19	c	Jim hoped himself to cross the desert.
C1/C2	19	d	Jim imagined himself to cross the desert.
C1/C2	20	a	June managed to give a speech.
C1/C2	20	b	June imagined to give a speech.
C1/C2	20	c	June managed herself to give a speech.
C1/C2	20	d	June imagined herself to give a speech.

Appendix B. Language History Questionnaire.

L2 Language History Questionnaire

Contact Information	Name:		Email:	
	Telephone:		Today's Date:	

Please answer the following questions to the best of your knowledge!

1. Age (in years):	
2. Sex:	☐ Female ☐ Male
3a. Education (degree obtained or school level attended):	
3b. Fields of study (Major):	
4a. Country of origin:	
4b. Country of Residence:	
5a. How long have you lived in a foreign country where your second language is spoken? (in years)	
5b. How long have you been in the country of your current residence? (in years)	
6. What is your native language? (If you grew up with more than one language, please specify)	
7. Do you speak a second language? Which?	☐ YES: ☐ NO

8. Please specify the age at which you started to learn your second language in the following situations (write <u>age</u> next to any situation that applies).	At home:	
	In school:	
	After arriving in the second language speaking country:	

9. How did you learn your second language up to this point? (check all that apply)	Through formal classroom instruction	Mainly ☐	Mostly ☐	Occasionally ☐
	Through interacting with people	Mainly ☐	Mostly ☐	Occasionally ☐
	A mixture of both, but…	More class-room ☐	More interac-tion ☐	Equally both ☐

10. L2- Skills (1 = very poor, 2 = poor, 3 = fair, 4 = functional, 5 = good, 6 = very good, 7 = native like

Language	Reading	Writing	Speaking	Listening

11. Provide the age at which you were first exposed to each foreign language in terms of speaking, reading, and writing, and the number of years you have spent on learning each language.

Language	Age first exposed to the language			Years of Learning
	Speaking	Reading	Writing	

12. Estimate, in terms of percentages, how often you use your native language and other languages per day (in all daily activities combined, circle one that applied):

Native Language	< 25%	25%	50%	75%	100%
Second	< 25%	25%	50%	75%	100%
Other (which?)	< 25%	25%	50%	75%	100%

13. If you have lived or travelled in other countries for more than three months, please indicate the name(s) of the country or countries, your length of stay, and the language(s) you learned or tried to learn.

14. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

15. Do you want some information about the results of the study? ☐ **YES** ☐ **NO**

THANK YOU FOR YOUR PARTICIPATION!!!

Appendix C. Adapted Version of the Edinburgh Handedness Inventory.

Handedness Questionnaire

For each oft the activities below, please indicate:

Which hand you prefer for that activity?
Do you ever use the other hand for the activity?

	left	**No prefe-rence**	**right**	**Do you ever use the other hand?**	
Writing:	☐	☐	☐	Yes	No
Drawing:	☐	☐	☐	Yes	No
Throwing:	☐	☐	☐	Yes	No
Using scissors:	☐	☐	☐	Yes	No
Using a toothbrush:	☐	☐	☐	Yes	No
Using a knife (without a fork)	☐	☐	☐	Yes	No
Using a spoon	☐	☐	☐	Yes	No
Using a broom (upper hand)	☐	☐	☐	Yes	No
Striking a match	☐	☐	☐	Yes	No
Opening a box (holding the lid)	☐	☐	☐	Yes	No

Appendix D. Instruction (left version only).

In this experiment you will see sentences presented word by word on a computer screen.

Your first task is to read each sentence and at the end of each sentence you have to judge whether the sentence was acceptable or unacceptable for you. Unacceptability might refer to sentences, which are grammatically wrong (Example: *He are a young man*.), or which have a nonsensical meaning (Example: *The honey was killed yesterday*.).

After that, you will see a single word presented on the screen. Now you have to decide whether the word was in the sentence or not.

The sequence is as follows:

First, you see a fixation cross in the middle oft the screen: +

As soon as you see the fixation cross, you should turn your attention to the following sentence. The sentence is presented word by word in the middle of the screen:

The honey was eaten yesterday.

After the end of each sentence, you will see a questions mark: ?

When you see the question mark, you should press a key on the keyboard, depending on whether the sentence was acceptable for you, or not. Please press the left shift-key, when the sentence was acceptable, and the right shift-key, when it was unacceptable.

After the question mark disappears, you will see a single word:

honey

Please press the left shift-key, when the word was in the sentence, and the right shift-key, when the word was not in the sentence.

After that, the screen turns blue, and the next sentence starts.

Once again:

LEFT SHIFT-Key means **YES** (= sentence was acceptable, word was in the sentence)

RIGHT SHIFT-Key means **NO** (= sentence was unacceptable, word was not in the sentence)

Further, please note:

Please don’t press the key until the question mark is on the screen. When you press to soon, there are motor artifacts in the EEG and we can not evaluate the data. But also do not press to late. Turn your attention to the next sentence, as soon as you can see the fixation cross on the screen.

During the whole experiment, please try to avoid eye movements. This is most important during the presentation of the sentence. Until the question mark appears, please try to avoid blinking as well, because also blinking leads to artifacts in the EEG. As soon as you see the questions mark, and as long as the screen is blank, you may blink.

All in all there will be 8 blocks of sentences, after each one there’s a little break. Then you can move, or cough, or drink…

If you have any questions now, please ask…