Second language processing: electrophysiological studies and data analysis methods

Meulman, Nienke

DOI:
10.33612/diss.651426336

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
chapter 1

general introduction
Chapter 1
This dissertation explores second language (L2) processing through electrophysiological studies, employing sophisticated data analysis methods. In this first chapter the concepts that underlie this type of research are introduced.

A major question in L2 acquisition research is to what extent and under which circumstances L2 speakers can become native-like users of their L2 (i.e., reach the same level of proficiency and rely on the same neurocognitive mechanisms as native speakers). This dissertation uses Event-Related Potential (ERP) measures of native-likeness to investigate certain aspects of this question. ERPs are known to be highly sensitive to the immediate, unconscious, online cognitive processing mechanisms underlying language comprehension. As such they provide an excellent tool to determine how native-like an L2 speaker truly is, as they allow insight, not only into which kinds of structures the speaker treats as (un)grammatical, but also the cognitive processes underlying the responses.

Two important factors that are thought to influence whether L2 speakers can attain native-like processing are the age at which they started L2 acquisition and the similarity between their first language (L1) and their L2. These are the central themes in this dissertation. Both factors are closely related, but the first is often used to investigate maturationally determined constraints on language acquisition (i.e., nature), while the second emphasizes an individual's learning experience (i.e., nurture) as a constraining factor. In Section 1.1, the age effect in L2 acquisition is introduced — that is, the fact that L2 speakers that start at a younger age generally reach higher proficiency levels than those that start at an older age — and arguments for and against one of the central theories that have been put forward in the literature to account for this effect (i.e., the ‘critical period hypothesis’) are discussed. In Section 1.2, the influence of the L1 on L2 acquisition is discussed. Here, the focus is on how the presence of certain linguistic structures in the L1 affects the acquisition of new ones in the L2, and especially whether similarity between old and new structures is a hindrance or an advantage.

The other central themes of this dissertation are methodological in nature, and concern experimental and data analysis methods of ERP studies on L2 acquisition. Age effects have traditionally been investigated using group comparisons of the grand mean waveforms (in a pre-specified time window) of L2 speakers versus native speakers, or early- versus late-onset L2 speakers. The problem with this approach, as this dissertation will argue, is that it does not take into account the full range of variability in both age of acquisition as well as in the timing of the ERP effects. One of the studies in this dissertation will aim to demonstrate that the traditional approach may lead to erroneous conclusions, and will investigate whether an alternative approach (using a fine-grained generalized additive modeling analysis; GAM; Wood, 2017) is superior in uncovering the true shape of the age of acquisition effect. Section 1.3 will introduce the ERP method (a thorough description, situated within the literature, can be found in Chapter 2), as well as the methodological issues that arise in the current line of research.
Another disadvantage of the traditional analysis methods is that they offer limited tools to investigate individual differences in the ERP signal. The final aim of this dissertation, therefore, is to investigate how (systematic) individual variability in ERPs may be more accurately modelled. A range of different approaches to tackle this issue will be applied throughout this dissertation (ranging from visual inspection through linear regression to GAMs), culminating in the introduction of a new set of GAM-based individual difference measures for L2 ERP research that will be introduced. In Section 1.4, a brief introduction on the relevance of the study of individual differences in ERPs is provided.

Finally, Section 1.5 gives an overview of the structure of this dissertation.

1.1 Age effects in L2 acquisition: a critical analysis

Second language speakers that start at a younger age appear to have an advantage over those that start when they are older, which is evident in the fact that they systematically reach higher levels of proficiency (e.g., DeKeyser 2012). Age of onset of acquisition (in the following: age of acquisition)\(^1\), which in this dissertation is used to refer to the age at which the L2 speakers are immersed in the L2 context as immigrants, has been reliably shown to be the strongest predictor of the ‘final’ state of L2 proficiency (Birdsong, 2006). The explanation for this age effect has been the topic of a large volume of research for many decades and until this day continues to inspire debates among linguists (see Birdsong, 2018; Singleton & Leśniewska, 2021).

An important question concerns the possible existence of a biologically based ‘critical period’ for L2 acquisition.\(^2\) According to this hypothesis, there is an ideal, maturationally determined window of brain development during which L2 acquisition needs to take place (Lenneberg, 1967). The critical period theory predicts that an age of acquisition after puberty (defined in various studies between 12 and 18 years of age) would prevent a L2 speaker from achieving nativelike competence, because by that point the brain is no longer sufficiently sensitive to language input or sufficiently ‘plastic’ (i.e., flexible) to fully represent some new grammatical structures in the same way as native speakers. By contrast, other theories that argue against a critical period for L2 acquisition assume that late-onset L2 speakers can establish nativelike representations, but this becomes increasingly difficult with increasing age, due to general cognitive aging, increased competition from the L1 and other language-external factors (Bialystok, 2001).

Immersion in the L2 environment is one of the conditions that benefits language learning tremendously. It is why immigrants generally outperform classroom L2 speakers who live in an L1 setting. It is also why this population of immersed L2 speakers is the focus

---

1 Note that in the literature some studies use the terms ‘age of exposure’, ‘age of immersion’ or ‘age of arrival’.

2 The original critical period hypothesis was formulated for first-language development (Penfield & Roberts, 1959), and later extended to a potential critical period for second language acquisition.
of our current investigation, as we are interested in questions concerning the performance of the most highly proficient L2 speakers. Indeed, the behavioral measures used in the studies discussed in this dissertation show that our L2 speakers perform at or close to native levels. All of the L2 speakers that took part in these studies had been living in the L2 country for many years. Their work, studies and social activities required them to use the L2 throughout their daily lives. Concentrating on some challenging grammatical properties of the L2, our research focuses on the areas that even these highly proficient speakers may find difficult. Our investigation concentrates on understanding how their age of acquisition modulates L2 speakers' processing of certain (notoriously difficult) grammatical aspects of the L2, and whether these age effects support or contradict either of the theoretical positions on the mechanisms behind second language acquisition.

Underlying the critical period hypothesis are assumptions about biological factors that constrain language development after a certain age. Lenneberg (1967) linked L2 learning difficulties in adulthood with hemispheric functional specialization, as during childhood the left hemisphere becomes progressively more dominant for language functions. Pulvermüller and Schumann (1994) indicated that as the brain matures, until around puberty, myelination of the axons of the neurons reduces plasticity. After puberty and for the remainder of the L2 speaker's lifespan, plasticity then remains low. With respect to hemispheric specialization, it is worth noting that more recent neurological research suggests that different time frames exist for the lateralization of different language functions (for discussion see, for example, Singleton 2005). This would support the idea of multiple critical periods, with the phonetic/phonological critical period coming to an end (much) earlier in childhood than the one for grammar, for example. In this dissertation the focus is on grammar, which is what traditionally has been the topic of most of the research into the critical period in L2.

Although there are many diverse and competing versions of the critical period hypothesis (for an overview, see Birdsong, 2017), crucially, proponents of the critical period hypothesis claim that these biological factors cause the ‘susceptibility’ or ‘sensitivity’ to language input to vary as a function of age, with late-onset L2 speakers being less able to implement grammatical properties based on input alone than early-onset L2 speakers. In order to qualify as a critical period as this is understood in the biological sciences, the optimal window for language acquisition would have to end in a relatively abrupt manner: the relation between learning and age would have to be non-linear (i.e., different inside than outside the critical period). Furthermore, this decline in the ability to successfully build the target-language grammar is domain-specific (that is, it applies to the learning of language).

An alternative to the critical period hypothesis is that other factors, not specific to language, are responsible for the age effect in L2 acquisition. There are a number of cognitive processes that decline across the lifespan in normal cognitive aging, such as working memory capacity, cognitive processing speed, and attention. These factors obviously impact the mechanisms necessary for learning a new language (cf. Hakuta et al,
Cognitive resource limitations may also be responsible for the fact that L2 speakers perform better in offline experiments than in online ones (e.g., Hopp, 2010; López et al., 2014).

Furthermore, with increasing age, competition and transfer from the L1 also become stronger, affecting learning as well as processing of the L2 (MacWhinney, 2018). These competition and transfer effects of the L1 are further discussed in Section 1.2 of the current chapter. In contrast to the critical period hypothesis, none of these changes (effects of cognitive aging and increased competition from L1) predict a discontinuity of the age of acquisition curve. Instead they predict a gradual decline across the life span.

Part of this dissertation is focused on this fundamental question: What is causing the age effect in L2 acquisition? Is it the critical period, or other factors related to cognitive aging and increased competition from L1? In the study presented in Chapter 4, we apply a fine-grained generalized additive modeling (GAM; Wood, 2017) analysis to the ERP data, to investigate whether there is any evidence of a discontinuity in the age curve, which would support a critical period in L2 acquisition. This study, specifically, also illustrates the methodological theme of this dissertation, as we investigate whether a GAM approach is more suitable to identify the age of acquisition effect than the more traditional analysis of variance (anova) approach.

1.2 How the L1 influences L2 acquisition: the case of grammatical gender

Any L2 that is not learned simultaneously with the L1 is by definition learned under different circumstances: the L1 is already in place. Research has shown that the L1 cannot simply be “switched off”. Even highly advanced L2 speakers who have target knowledge of the L2 grammar show evidence of L1 effects during L2 sentence comprehension (e.g., Bauke, 2020; Rankin, 2014). The question is therefore how the presence of certain linguistic structures affects the acquisition of new ones, and especially whether similarity between old and new structures is a hindrance or an advantage.

Much of the evidence of L1 effects in L2 processing comes from studies on grammatical gender, and this is also the feature that is used for the investigations in this dissertation. Grammatical gender is a classification system for nouns (Corbett, 1991) which allows speakers to establish grammatical cohesion between the elements in a phrase through agreement. Learning grammatical gender involves acquiring knowledge of a word’s gender class (i.e., gender assignment; e.g., in German, “Schule” [school] → feminine) and of how gender is expressed syntactically in related elements (i.e., gender agreement; e.g., “eine_{FEM} schöne_{FEM} Schule...” [a beautiful school]).

Languages can differ in the characteristics of their gender systems. The differences that are most relevant to the current dissertation are presented in Table 1.1, and will be discussed in detail later in this section. Attaining gender processing that is comparable
General introduction

to that of native speakers has been shown to pose a major challenge to L2 speakers, and even those with the highest proficiency levels do not always show target-like L2 gender processing (e.g., Foucart & Frenck-Mestre, 2012; Sabourin & Stowe, 2008).

There is some evidence that having a gender system in the L1 is beneficial when it comes to acquiring an L2 gender system (e.g., Hopp, 2013; Grüter et al., 2012). This finding supports models proposing that the L1 restricts L2 acquisition, such that late-onset L2 speakers will not have the ability to represent abstract gender features if gender is not instantiated in a L2 speaker’s native language (Clahsen & Felser, 2006; Hawkins & Chan, 1997). At the same time, there is also evidence of L2 speakers without gender systems in their L1 who do appear to show full acquisition of grammatical gender (e.g., Alemán Bañón et al., 2014), arguing against such a restriction (Schwartz & Sprouse, 1996; Prévost & White, 2000).

Moreover, other studies on L2 gender processing suggest that L1 effects may be more varied than previously thought. When an L1 gender system is already in place, it may cause interference when the L2 gender system is dissimilar. Several examples of such negative transfer effects have been found. Sabourin & Stowe (2008) found that German L1 speakers of L2 Dutch were able to show native-like processing, whereas Romance L2 speakers of Dutch were not, which was attributed to the fact that the German and Dutch gender systems are typologically more similar. Similarly, Foucart (2008) found that German L2 speakers of French were not able to show native-like brain responses to structures which are gender-marked in the L2 but not in the L1 (i.e., plural noun phrases). Another example, this time for negative transfer at the lexical level, comes from a study by Lemhöfer et al. (2008), who found that items which are gender-congruent between L1 and L2 were facilitated, while incongruent items remained problematic.

From these studies it can be concluded that even if both the L1 and the L2 have grammatical gender, differences in the amount of overlap in how they realize grammatical gender at the lexical and at the morphosyntactic level may cause difficulties for L2 speakers. These findings therefore argue against an abstract gender feature, and in favor of theories that put emphasis on cross-linguistic competition (i.e., Competition Model; Bates & MacWhinney, 1987) and transfer (i.e., Full Transfer / Full Access Model; Schwartz & Sprouse, 1996). The characteristics of the target language itself have also been linked to failure to show native-like processing. In two studies on gender processing in L2 Dutch (Loerts, 2012; Sabourin & Stowe, 2008) it was argued that the relatively opaque and non-salient Dutch gender system may have contributed to the results, in particular if the L2 speaker’s L1 is more transparent. In Loerts’ (2012) study, for example, only the most proficient Polish-Dutch L2 speakers are able to show native-like effects, which are additionally somewhat smaller in amplitude compared to natives. As shown in Table 1.1, Romance languages (such as Spanish) and Slavic languages (such as Polish) have phonologically highly transparent gender markers for the majority of nouns (e.g., “-a” for feminine in Spanish). By contrast, gender marking in Dutch and German is relatively
opaque. This is particularly true for Dutch, as German does have some morphological and phonological cues (Köpcke & Zubin, 2009), but these are less prevalent and less reliable compared to the more transparent gender systems mentioned.

Dutch, furthermore, differs from German, as can be seen in Table 1.1, in the number and distribution of gender categories and the saliency of the gender system (see Koster & Loerts, 2020, for a comparison of the Dutch and German gender systems). In terms of saliency, the interaction with the case system makes German gender markings far more complex, but also more prominent and relevant for (the meaning of) a sentence, compared to Dutch. In Dutch, the saliency of grammatical gender is very low (Cornips & Hulk, 2008). The only clear evidence can be found in the singular definite and demonstrative determiners, but even there the saliency is lowered by a large frequency difference between common and neuter nouns (i.e., 80% of Dutch root nouns are common gender; Van Berkum, 1996). This difference in saliency may be why, in monolinguals, German children acquire their gender system earlier than Dutch children (Bewer, 2004; Van der Velde, 2004). If such characteristics of the L1 gender system indeed (also) affect L2 processing, it would provide another argument against accounts that involve an abstract gender feature to explain failure of L2 speakers to achieve nativelike gender processing.

In this dissertation the effects of cross-linguistic similarity and of the characteristics of the target gender system on L2 processing are further investigated. Two ERP studies on gender processing are presented, involving 1) Romance L2 speakers of Dutch and 2) Slavic L2 speakers of German. Note that in order to investigate the effects of gender transparency and saliency from L1 to L2 comprehensively, other language combinations would need to be included in the comparison. For example, we did not directly compare our Romance-Dutch speakers to Slavic-Dutch speakers, nor did we include L2 speakers whose L1 does not instantiate grammatical gender. However, as discussed above, these populations were investigated in previous studies with similar research paradigms (e.g., the Loerts’ 2012 study on Polish L2 speakers of Dutch). Together with the studies presented in this dissertation we may therefore come to a better understanding of the influence of particular characteristics of gender systems (e.g., transparency and saliency) on L2 gender acquisition. Since L2 gender processing has less frequently been studied in populations having a transparent L1 and learning an opaque (and less salient) L2 gender system, this will be the focus of this dissertation. This population is particularly interesting because the lack of morpho-phonological and semantic cues for gender in the L2 may make acquisition of the gender system particularly difficult. It therefore provides a true test which will reveal to what extent L2 speakers may become nativelike with respect to syntax processing.

Both language combinations present typologically distant languages. Therefore, effects of cross-linguistic similarity at the lexical level are expected to be minimal, and will not be investigated in this dissertation. Rather, we focus on cross-linguistic similarity effects at the morpho-syntactic level.
Table 1.1 Overview of the main characteristics of the gender systems that are relevant to the current dissertation.

<table>
<thead>
<tr>
<th>Language (family)</th>
<th>Studied here as L1 or L2?</th>
<th>Example</th>
<th>Gender categories</th>
<th>Transparency (how visible and systematic?)</th>
<th>Saliency of gender marking in NP (how prominent and important for meaning?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romance: e.g., French, Spanish</td>
<td>L1</td>
<td>French: <em>un bon lit</em> – ‘a good bed’ <em>une bonne table</em> – ‘a good table’</td>
<td>masculine, feminine</td>
<td>transparent: many reliable morpho-phonological and semantic cues (particularly true in Spanish, somewhat less in French)</td>
<td>salient: gender marking on determiners and adjectives</td>
</tr>
<tr>
<td>Slavic: e.g., Polish, Russian</td>
<td>L1</td>
<td>Polish: <em>dobry stół</em> – ‘a good table’ <em>dobra lampa</em> – ‘a good lamp’ <em>dobro łóżko</em> – ‘a good bed’</td>
<td>masculine, feminine, neuter (and animacy distinction)</td>
<td>transparent: many reliable morpho-phonological and semantic cues</td>
<td>salient: gender marking on nouns and adjectives (no determiners), plus interaction with case system</td>
</tr>
<tr>
<td>Germanic: German</td>
<td>L2</td>
<td>German: <em>ein guter Tisch</em> – ‘a good table’ <em>eine gute Lampe</em> – ‘a good lamp’ <em>ein gutes Bett</em> – ‘a good bed’</td>
<td>masculine, feminine, neuter</td>
<td>opaque: only some, less reliable, morpho-phonological and semantic cues</td>
<td>salient: gender marking on determiners and adjectives, plus interaction with case system</td>
</tr>
<tr>
<td>Germanic: Dutch</td>
<td>L2</td>
<td>Dutch: <em>een goede tafel</em> – ‘a good table’ <em>een goed bed</em> – ‘a good bed’</td>
<td>common, neuter (80% common)</td>
<td>opaque: very few morpho-phonological cues</td>
<td>non-salient: marking on determiners and adjectives, but in practice clear gender marking evidence is scarce</td>
</tr>
</tbody>
</table>

*Note.* The two ERP studies presented in this dissertation investigate 1) Romance L2 speakers of Dutch and 2) Slavic L2 speakers of German.
Chapter 1

The first experiment of this dissertation (Chapter 3) is a replication of the Sabourin and Stowe (2008) study, but with a more nuanced analysis of proficiency of the L2 speakers and with an added comparison of stimulus modality (see Section 1.3 below). Specifically, we assess whether the problems Romance L2 speakers were found to have with the opaque and non-salient Dutch gender system are robust. The second experiment on the Slavic-German group (Chapter 4) aims to assess whether L2 speakers can show native-like gender processing in an opaque but highly salient L2 gender system, when their L1 gender system is transparent, and to what extent this is dependent on their age of acquisition.

Both studies that will be presented investigate to what extent the L2 speakers may exhibit native-like brain responses, as measured by the ERP signal, using a violation paradigm (i.e., mismatches in gender agreement). The benefits of the ERP approach are discussed in Section 1.3 below. We hypothesize that it is not just the presence of the grammatical gender property in the L1 that affects L2 speakers’ ERP responses, but rather that the degree of L1-L2 similarity in terms of transparency and saliency of the gender systems impacts L2 acquisition, particularly when acquisition has started later in life.

1.3 Methodological challenges in L2 ERP research

In this dissertation, the focus is on highly advanced L2 speakers. Based on behavioral measures, they perform close to nativelike. However, event-related brain potentials (ERPs) offer information about online processing and can therefore reveal processing differences at the moment they arise, even if there are no behavioral differences. I.e., they reflect actual implicit grammatical or linguistic processing (see Paradis, 2008), rather than possibly reflecting participants’ knowledge of explicitly learned rules. This allows us to investigate whether or not L2 speakers differ fundamentally with respect to the way in which they process grammatical gender.

ERPs are derived from the EEG (electroencephalogram), and they represent changes in the electrical potential of neural cells measured on the scalp and time-locked to the stimulus. In this dissertation, the stimulus is a word in a sentence that a participant either reads or hears. In our studies we use ERPs to investigate whether and when L2 speakers show native-like processing. Specifically, we compare their brain responses to those of native speakers. Any deviations from the native norm may reveal when L2 speakers are having difficulties, and/or are using different processing strategies (Steinhauer, 2009). Combined with their behavioral data, we can make further inferences about these differences. As different ERP components (i.e., certain well-known response patterns) are linked to particular processes in the brain, modulations of these components may inform us about differences in processing strategies between L2 speakers and natives (e.g., Foucart & French-Mestre, 2012; Tanner et al., 2013).

While in this sense the ERP technique is powerful, it also comes with a great deal of methodological challenges (see Luck, 2014, for a comprehensive guide to the practicalities of
conducting ERP research). Chapter 2 describes various practical and methodological issues regarding ERP studies on L2 processing, and offers an in-depth description and discussion of the research design for the studies presented in this dissertation.

One of the issues of measuring ERPs is that acquisition of the data needs to take place in a carefully controlled environment. Sentence materials are traditionally presented visually, with one word at a time being displayed on a computer screen. This is quite different from what L2 speakers will typically experience in daily life and may lead to artefacts that reflect the artificial nature of the task. Auditory sentence presentation, by contrast, does not have that disadvantage. Therefore, the present dissertation includes an experiment comparing auditory vs. visual sentence presentation, to investigate whether a more natural listening paradigm benefits L2 processing (Chapter 3).

Another methodological challenge of the ERP technique is that the signal is relatively noisy. As a consequence, a considerable amount of data is needed to identify any significant effects (Luck, 2014). To achieve adequate signal-to-noise ratio, the traditional approach entails averaging the ERP signal over participants and over a pre-specified time window. However, this method is problematic in various ways, amongst others sensitivity to outliers and potential obscuring of interesting time-course effects. Averaged ERPs are furthermore not well-suited to investigating individual differences (for further discussion on this issue see Section 1.4 below). In this dissertation (Chapter 4), the disadvantages of the traditional approach are discussed, and a more powerful technique to model the data (by means of so-called generalized additive modeling: GAM; Wood, 2017) is presented, which does not require averaging across a time window.

1.4 Individual differences in L2 processing as measured by ERPs: a set of new measures

Second language learning is subject to substantial individual variation. However, traditional ERP studies on L2 processing are often restricted to group comparisons of the grand mean waveforms of L2 speakers versus natives, essentially regarding individual differences as noise. When group differences are observed, the underlying causes could be the influence of the L2 speakers’ L1 and/or difference in the use of processing strategies between L2 speakers and native speakers. However, variability in the L2 speakers’ proficiency levels, or working memory capacity, for example, perhaps in combination with other factors, could also account for such results (Roberts, 2012). More recently, (relatively few) studies have therefore opted to investigate individual differences, using within-group correlation and regression-based statistics assessing the impact of these factors on the ERP response of an L2 speaker or native speaker (e.g., Alemán Bañón et al., 2017; Beatty-Martínez et al., 2021; Bice & Kroll, 2021; Tanner et al., 2014).

In their overview paper, Tanner and colleagues (2018) present several examples of cases in which grand mean average waveforms and accompanying omnibus statistical analyses
Chapter 1

of mean amplitudes indicate one picture, but closer inspection of between-subject variation another. Their paper highlights the importance of taking individual variation into account (particularly in heterogeneous L2 populations), but also exposes the fact that we have limited tools to do so. Tanner and colleagues (2018) challenge future research to develop more powerful ways to capture variation across all of the dimensions that ERPs have to offer, to help us gain a better understanding of the underlying mechanisms and to be able to incorporate these findings in current models of language processing.

This dissertation aims to take a step in this direction, by introducing and evaluating new approaches that are able to capture (systematic) variability between individuals, while, importantly, still including the complete time course of the ERP signal (i.e., without having to average over an arbitrarily selected time window). In Chapter 4, we aim to demonstrate that using GAMs allows us to investigate the full range of variability in both age of acquisition of the L2 speakers and the complete time course of their ERP responses to grammatical violations.

In Chapter 5, we introduce a method to extract a set of GAM-based measures from the ERP waveforms, which can be used to study individual differences in L2 processing. Potential problems with the traditional approach of extracting measures such as the ‘average response magnitude’ as a measure of nativelike processing are outlined, and the benefits of the new approach are illustrated with data from the study presented in Chapter 4. Our method is based on participants’ complete waveforms for a given time series (modeled using GAMs), and aimed at detecting the relevant modulations of this waveform due to the grammatical manipulations. In our empirical example we investigate which of the different measures (which are based on peak, area and latency effects in the modeled waveform) are optimal for characterizing differences between native and L2 speakers, and capturing proficiency differences between L2 speakers.

In doing so, we also aim to pose a more fundamental question: what does the size of an ERP reflect in L2 processing? Traditionally, the size of the ERP response (i.e., amplitude) of an L2 speaker is compared to that of a native speaker to assess ‘nativelikeness’, but perhaps a measure of ‘robustness’ of an L2 speaker’s ERP response (i.e., how reliably they are able to show a similar response as a native speaker) may be more appropriate. Chapter 5 investigates this question and concludes by discussing implications of our results for future studies on L2 processing.

1.5 Structure of the dissertation

In Chapter 2, we will first elaborate on ERPs in detail, as a general introduction to this technique. A thorough examination of the methodological challenges involved in designing, executing and analyzing data from (large scale) ERP experiments on second language processing is also provided. In addition, we will describe the experimental setup of our own experiments in detail.
Chapter 3 covers the first experimental study of this dissertation. This ERP study investigates late-onset L2 speakers of Dutch with a Romance language background (e.g., French, Spanish), and their processing of L2 gender violations. The study aims to identify whether late-onset L2 speakers can show nativelike processing for a syntactic structure that is present in both L1 and L2, but with some cross-linguistic differences, assessing the potential impact of L1 transfer. In addition, the study compares auditory with visual presentation of the sentence materials, to investigate potential effects of presentation modality.

Chapter 4 discusses the second ERP study. This experiment focuses on L2 speakers of German with a Slavic language background (e.g., Russian, Polish), and with varying ages of onset of acquisition. Again, we look at how these L2 speakers process L2 violations of gender. The aim of the study is twofold: First, it focuses on the question whether and how age of onset of acquisition impacts the processing of language properties with a different surface form in L2 vs. L1. Second, it addresses methodological concerns. Particularly, it demonstrates the problems that arise from dichotomizing the age variable (early- vs. late-onset of acquisition) as has often been done, rather than treating it as a continuous variable. Instead, it investigates whether generalized additive modeling (GAM; Wood, 2017) provides a better framework for analyzing the effect of a continuous variable on ERP waveforms, taking into account the full variability in timing of the ERP effects. Both the theoretical as well as the methodological research questions of this study add to the debate on the existence of a critical period for L2 acquisition.

Chapter 5 is purely methodological in nature. Here we introduce a new set of tools to study individual differences in the ERP signal, in the form of a (series of) individual ERP outcome measures, which are based on GAMs. Using the data from the study presented in Chapter 4, we investigate which of the measures is best able to distinguish native and L2 speakers, and more proficient from less proficient L2 speakers. We also assess whether or not these measures outperform the traditional (averaged) measure. Furthermore, we hypothesize that it is not the size (i.e., amplitude) of an ERP response that is most relevant in studying ‘nativelike’ effects in L2 speakers, but rather the robustness of a L2 speaker’s ERP response.

The final chapter (Chapter 6) offers general conclusions, based on the work presented in the previous chapters. It addresses the theoretical implications of our findings on age effects and L1 influence on L2 processing. In addition, it discusses methodological implications about how the ERP technique can best be used to study L2 acquisition.