Applied Linguistics 2022: 43/3: 433–452
doi:10.1093/applin/amab058 Advance Access published on 4 October 2021

Is It Time to Reconsider the 'Gold Standard' for Nativelikeness in ERP Studies on Grammatical Processing in a Second Language? A Critical Assessment Based on Qualitative Individual Differences

[1,*]DOMINIK FREUNBERGER, [1,2]EMANUEL BYLUND, and [1]NICLAS ABRAHAMSSON

[1]Centre for Research on Bilingualism, Stockholm University, Sweden, [2]Department of General Linguistics, Stellenbosch University, South Africa
[*]E-mail: dominik.freunberger@biling.su.se

In most event-related potential (ERP) studies on the second language (L2) processing, the native speaker (L1) control group's grand average ERP pattern serves as the 'gold standard' that the L2 group has to reach to be labeled 'native-like'. This relies on the assumption that the grand average is representative of all or most individuals in a group. Recent research, however, has shown that there can be considerable systematic qualitative variability between individuals even in coherent L1 samples, especially in studies on morphosyntactic processing. We discuss how these qualitative individual differences can undermine previous findings from the gold standard paradigm, and critically assess the main ERP components used as markers for nativelike grammatical processing, namely the left-anterior negativity and the P600. We argue that qualitative variation reflects the dynamics characteristic of nativelike grammatical processing and propose a model for experimental designs that can capture these processing dynamics and, thereby, has the potential to provide a more fine-grained understanding of nativelike attainment in an L2.

INTRODUCTION

Can speakers of a second language (L2) become nativelike in their L2? This is one of the central questions in L2 research, and it has been enthusiastically answered in the positive (e.g. White and Genesee 1996; Birdsong 2005) about as often as it has in the negative (e.g. Bley-Vroman 1989; Gregg 1996). Typically, both positions are discussed with respect to the critical period hypothesis (CPH; Lenneberg 1967), which predicts that adult L2 learners cannot achieve nativelike attainment due to neurophysiological maturational constraints. Claims about the existence of nativelike late L2 learners (allegedly constituting counterevidence to the CPH) were often made based on behavioral measures, which may not provide the necessary granularity to make a

strong test for nativelikeness (for discussion and a candidate counterexample, see Abrahamsson and Hyltenstam 2009). Moreover, behavioral methods can at most produce indirect insights into the neurocognitive mechanisms involved in language processing. L2 speakers can, therefore, show nativelike *behavior* in some linguistic domains while employing non-nativelike *processes* (cf. Paradis 2009).

A method that contributes greatly to the understanding of neurocognitive processing in L1 and L2 is event-related potentials (ERPs). ERPs are based on the brain's electrical oscillations and provide an online insight into the mechanisms involved in language processing. In short, ERPs typically represent the average neural activation of a group of participants at individual electrodes placed on the scalp, time-locked to an event of interest (for an introduction, see Kaan 2007; Steinhauer 2014). The resulting waveform can be described in terms of timing, polarity, and scalp distribution. Importantly, for most research questions in the field of language processing, the waveform of one condition alone is not informative. Experiments, therefore, usually compare at least two conditions. One might then find, for example, a relatively more positive waveform in one condition compared with the other. Researchers would then speak of a positivity *effect* (and if this effect reached its maximum around 600 ms after the onset of the critical word, it would be labeled a P600 effect). In group designs, which are the most common in L2 research, the *grand averages* (i.e. the numerical mean of all items in one condition *and* all individuals in one group at every measured time point) of the different groups are compared with each other, relying on the assumption that a group's average neural activity pattern is representative of most or all individuals in the group. In most L2 research, the L1 group's average waveform is considered the 'gold standard' for nativelike processing against which the L2 group is compared.

In this article, we will critically assess this paradigm, with a focus on morphosyntactic L2 processing.[1] We begin by outlining the paradigm and the main ERP components associated with grammatical processing. We will then introduce some milestone studies on individual differences and discuss the consequences of their findings for the reliability and validity of the ERP result patterns most commonly used as benchmarks for nativelike grammatical processing. Finally, we will reconsider the term 'nativelikeness' in L2 ERP research before closing with suggestions and testable predictions for future studies that have the potential to advance the field and strengthen conclusions on nativelike L2 processing.

THE 'GOLD STANDARD' PARADIGM

A typical design of an ERP experiment on L2 sentence processing is the following: An L2 group as well as an L1 control group read or listen to a set of stimulus sentences that contain grammatical violations, such as agreement or structural errors, or semantic manipulations, as well as control sentences, while their electroencephalogram (EEG) is recorded. Grand average ERPs are

then computed per condition and group. L2 speakers are considered nativelike when their grand average ERPs are both qualitatively and quantitatively comparable with the grand average ERPs of native speakers. The L1 group's grand average thus serves as the gold standard of nativelikeness.

For lexical-semantic processing, this approach consistently yielded robust results: L1 speakers show a reduced N400 (i.e. a negative ERP deflection peaking around 400 ms after the onset of the critical word) when encountering words that are predictable from or congruent with their preceding context, so that lexical-semantic processing is facilitated (for an extensive overview, see Kutas and Federmeier 2011; for a recent account, see Nieuwland *et al.* 2020). In L2 speakers, minimal L2 instruction is sufficient to elicit N400 effects resembling those of native speakers (McLaughlin *et al.* 2004). L2 N400s sometimes show variation in size, onset latency, duration, or topography (e.g. Ardal *et al.* 1990; Weber-Fox and Neville 1996; Hahne 2001; Ojima *et al.* 2005), sometimes depending on different learner variables (Moreno and Kutas 2005; Newman *et al.* 2012). Crucially, there is no variation in the *quality* of the elicited ERP pattern: Participants across a wide range of populations and experimental paradigms consistently show N400 modulations as a correlate for lexical-semantic processing (e.g. Swaab *et al.* 1997; Kutas and Iragui 1998). This robust relationship between a neurocognitive process and measured ERPs seems to justify using the native speakers' grand average N400 as a benchmark for nativelike processing within this domain.

Within the domain of morphosyntax, a similar one-to-one mapping of neurocognitive processes and observed ERP patterns initially emerged: Morphosyntactic deviances embedded in sentences frequently lead to a pronounced posterior positivity with a peak around 600 ms after the onset of the critical word (P600). More importantly, agreement errors can also elicit a left anterior negativity (LAN), starting from 100 to 400 ms after the onset of the anomalous word (Osterhout and Holcomb 1992; Friederici *et al.* 1999; Knösche *et al.* 1999). Violations of local phrase structure can lead to an earlier effect with a similar distribution (hence *early* LAN or ELAN; Neville *et al.* 1991; Hahne and Friederici 2001).

The P600 has received a multitude of functional interpretations, such as syntactic repair or revision, integration or unification of different information streams, monitoring, or resolution of a conflict between different information types, thematic integration or revision, or retrieval or structure building (Hagoort *et al.* 1993; Kaan *et al.* 2000; Friederici 2002; Hagoort 2003; Kolk and Chwilla 2007; Gouvea *et al.* 2010; Brouwer *et al.* 2012). Crucially, most accounts describe the P600 as reflecting a secondary processing stage, that is, for example, after semantic information has been activated, the agreement has been established, or an initial syntactic parse computed. The LANs,[2] on the other hand, were hypothesized to index early-stage implicit, automatic, and rule-based processes of establishing local phrase structure or agreement between hierarchically dependent words (Friederici 2002; Hagoort 2003; Bornkessel and Schlesewsky 2006). The close connection of this biphasic

LAN-P600 pattern to structural and agreement processing seemed to qualify it as a marker for nativelikeness, because grammatical aspects of language have been hypothesized and shown to be more susceptible to age effects than, for instance, lexical aspects (e.g. Curtiss 1977; Patkowski 1980; Newport 1990; Mayberry 1993; for overviews of age effects in L2, see, e.g. Hyltenstam and Abrahamsson 2003; Birdsong 2005; DeKeyser and Larson-Hall 2005; Birdsong 2006; Mayberry and Kluender 2018). In particular, LANs appeared as ideal markers for nativelikeness because their functional interpretations—reflecting early, fast, automatic, etc. processes—overlap with processes that are assumed to be affected more profoundly by brain maturation (e.g. DeKeyser 2001; Ullman 2001; Paradis 2004; Clahsen and Felser 2006).

A number of studies that applied this logic reported L2 ERPs that were different from the biphasic L1 benchmark for grammatical processing (e.g. Weber-Fox and Neville 1996; Hahne 2001; Hahne and Friederici 2001; Weber and Lavric 2008; Pakulak and Neville 2011). The nonnative groups that exhibited greater differences were typically late learners, and, therefore, such results are in line with theories that predict maturational constraints. Others found that with higher proficiency even adult L2 learners can show nativelike ERP responses (Hahne *et al.* 2006; Rossi *et al.* 2006; Kotz *et al.* 2008; Bowden *et al.* 2013), in contrast to the CPH.

There are, however, certain challenges associated with the LAN and the P600, which center mainly around their observed individual variability. In recent years, some labs have taken a systematic look at such individual differences and have tried to identify possible sources of variation.

QUALITATIVE INDIVIDUAL DIFFERENCES IN L1 AND L2 SENTENCE PROCESSING

One approach to investigating individual differences is to look for sources of variability between subjects on a unidimensional variable, such as reaction times.[3] A classic example is the processing of relative clauses (RCs): Object RCs typically take longer to read than subject RCs (e.g. Ford 1983). King and Just (1991), for example, found that subjects with a lower verbal working memory (VWM) are slower than subjects with a higher VWM. One can thus conclude that individual differences in VWM explain some part of the variability in the reading times of syntactically taxing structures. Reading times can vary only along the temporal dimension. This unidimensionality of the dependent variable does not allow drawing inferences about *qualitative* differences. The data show that people with lower VWM take longer to process object RCs. We can, however, not know whether they rely on the *same processes* as people with higher VWM, and that these processes are simply slowed down, or if they have to employ *different or additional processes* and that this causes a delay.

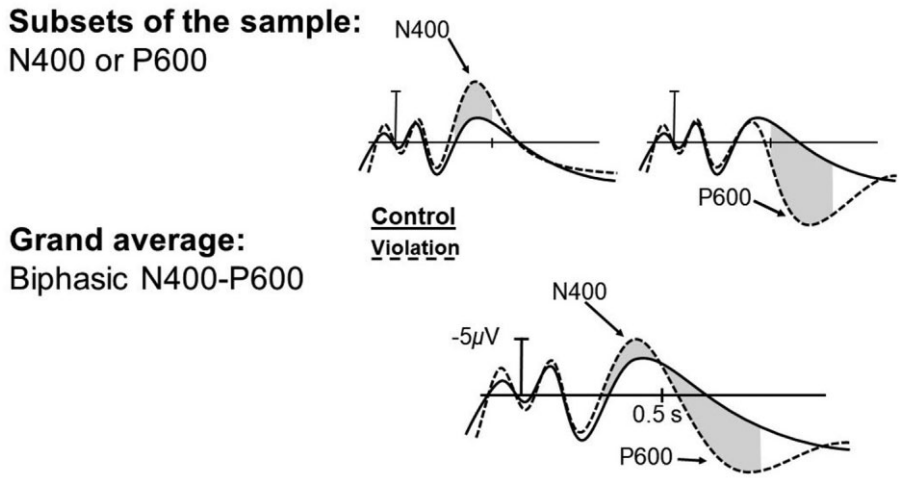

Figure 1: Schematic illustration of how the averaging procedure can produce a pattern that is not representative of individuals. When subsets of the sample show an N400 and another subset a P600, averaging over the two can create a biphasic pattern that is not representative of any individual (cf. Osterhout 1997). Shaded areas indicate typical time windows for statistical analyses (300–500 ms for the N400/LAN, 500–800 ms for the P600).

An EEG signal, on the other hand, can vary in timing, distribution, and polarity, and this multidimensionality can offer a window into the underlying neural mechanisms. Variability between subjects can thus also point to *qualitative* differences, that is, individuals might rely on different neurocognitive mechanisms (for an overview of individual differences in EEG/ERP research, see Boudewyn 2015). For example, in Experiment 2, Osterhout (1997) observed a grand mean biphasic N400–P600 response in garden-path sentences. Strikingly, 'no individual subject showed a clear biphasic response' (p. 509); instead, some participants showed the expected P600, while others showed an unexpected N400, a component typically associated with semantic processing. The biphasic pattern in the grand average was not representative of how individuals actually responded when they were led down the garden path (see Figure 1). Any conclusions based on the grand mean alone would have been unwarranted—despite a supporting statistical result.

Similarly, Tanner and Van Hell (2014) found that individuals' ERP profiles in two morphosyntactic manipulations (subject–verb agreement and verb tense) varied between being either negativity-dominant, biphasic, or positivity-dominant, while the grand averages displayed a LAN-P600 pattern. Importantly, the negativity effect in the negativity-dominant individuals was neither clearly left-lateralized nor predominantly anterior. Instead, the distribution closely resembled an N400. The positivity in the positivity-dominant group had the typical right-posterior distribution of the P600. The LAN in the

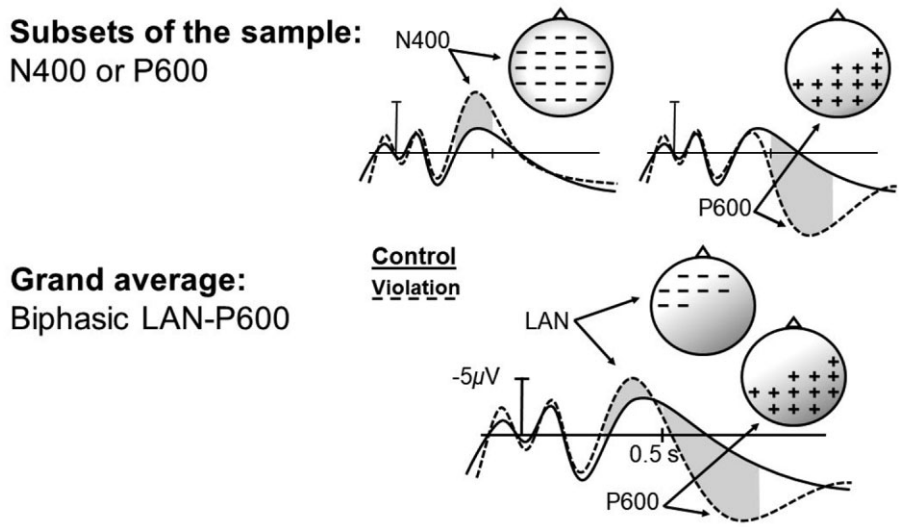

Figure 2: Schematic illustration of how the averaging procedure can produce an artificial effect. When subsets of the sample show a broadly distributed N400 and another subset a posterior right P600, the posterior right part of the N400 can be canceled out by the P600, resulting in a LAN followed by a P600 in the grand average (cf. Tanner and Van Hell 2014). The shaded areas in the ERP plots indicate the typical time windows for statistical analyses (300–500 ms for the N400/LAN and 500–800 ms for the P600). Topographic plots illustrate the distribution of the effects (i.e. violation minus control) with darker areas and '+' indicating positive voltages and lighter areas and '−' indicating negative voltages.

grand average was the product of the P600 canceling out the right-posterior parts of the N400, leaving a left-anterior negativity effect (see also Osterhout *et al.* 2004; Tanner 2019). Figure 2 illustrates how two distinct effects in subgroups of the sample can create an artefactual effect in the grand average. Whether one observes an N400, a P600, or a biphasic LAN-P600 pattern in the grand average can, therefore, depend on the ratio of N400 to P600 responses in the sample. In turn, Tanner and Van Hell argue that the proportion of N400s in the sample depends on the depth of lexical, semantic, or discourse processing necessary to resolve the morphosyntactic anomalies. Strikingly, the individual variation in response type can also lead to a null-effect on the group level (i.e. positivity and negativity ultimately cancelling each other out) in the grand average (Van Hell and Abdollahi 2017).

However, the view that the LAN is but an artefact from the averaging procedure is controversial and vividly debated (Molinaro *et al.* 2011; Molinaro *et al.* 2015; Tanner 2015; see also Nichols and Joanisse 2019, for a recent detection of a LAN without a subsequent P600). Caffarra *et al.* (2019), among

others, argue that the LAN is in fact distinct from the N400 and the most common component found as a response to morphosyntactic violations—at least when elicited by local agreement violations in a morphologically rich language (such as Spanish; for contrasting results see Alemán Bañón *et al.* 2012, who used local noun-adjective gender agreement, Alemán Bañón *et al.* 2014, who used local possessive-noun gender and number agreement, as well as Foucart and Frenck-Mestre (2011, 2012), who tested, respectively, local determiner–noun and noun–adjective gender agreement in French). Caffarra *et al.* (2019) reported a biphasic pattern in about two-thirds of the participants, as opposed to roughly one third in Tanner and Van Hell's (2014) report. Moreover, they detected a 'true' LAN (i.e. a LAN that is not the result of averaging) in about half of the participants, trials, and items, but also N400s and P600s in about one-quarter, respectively.

For L2 speakers, studies report a similar distribution of individual responses: About half of the L1 Spanish L2 English speakers in Tanner *et al.* (2014) showed an N400 effect, the other half a P600 effect to noun–verb agreement errors. Again, the biphasic N400–P600 pattern—despite supporting results from the ANOVA—was not fully representative of the individual results (although some individuals did show a biphasic response). In Tanner *et al.* (2013), morphosyntactic violations elicited a P600 in third-year L1 English learners as well as native speakers of German. First-year learners, in contrast, showed either an N400 or a P600, while, again, the grand mean result for this group was a biphasic response.

The variability in these studies suggests that even individuals with comparable linguistic backgrounds and cognitive traits can display different (and apparently inconsistent) ERP patterns in both their L1 and L2. The group averages typically conceal these individual differences and are thus not always representative of the individual responses. In particular, the processing of agreement errors can lead to considerable variability, and the view that agreement errors systematically engender LANs seems no longer tenable. A grand average biphasic LAN-P600 pattern, for instance, can be the result of individuals displaying either a 'true' LAN and/or a P600, or an N400 and/or a P600. On the one hand, such variation between individuals highlights the dynamics in language processing: People of comparable age, education, gender, and cognitive characteristics can employ different neurocognitive mechanisms to achieve the same goal.[4] On the other hand, qualitative variability undermines the key assumption of the gold standard paradigm, namely that comparable speakers show fairly uniform activation patterns, and that grand average ERPs reflect this. Furthermore, to our knowledge, no current model of L1 sentence processing predicts an N400 as a regular response to a range of grammatical violations. Therefore, such a view requires a revision of neurocognitive models on both L1 grammatical processing and L2 acquisition. Nevertheless, if indeed only some LANs stem from N400s, then even initial-stage and low-proficient L2 learners display nativelike grammatical processing (cf. Osterhout *et al.* 2006, 2008)—at least with respect to the processing stage preceding the

P600. In the following, we address how these insights can affect the conclusions drawn from two components often used as markers for nativelike morphosyntactic processing, the LAN and the P600.

The LAN as a marker of nativelikeness?

The LAN does not appear to be an ideal candidate as a marker of nativelike grammatical processing. Many studies did not find LANs as a response to agreement violations, even in homogeneous L1 groups (e.g. Hagoort and Brown 1999; Allen *et al.* 2003; Nevins *et al.* 2007; Alemán Bañón *et al.* 2012). Those who did find a LAN often report considerable variability in timing and distribution (e.g. Osterhout and Nicol 1999; Hagoort *et al.* 2003; Kaan and Swaab 2003; Silva-Pereyra and Carreiras 2007; Mancini *et al.* 2011; Dillon *et al.* 2012). Even more importantly, the estimation that—given certain linguistic limitations—a true LAN is found in half of a nativespeaker sample (Caffarra *et al.* 2019) seems to cast principal doubts on the LAN as a reliable benchmark, especially when the other half of the sample displays qualitatively different components. Furthermore, a LAN in the grand average is (at least) complex to interpret, since there are multiple possibilities why a LAN emerges, some of which are but the result of the averaging procedure (see Tanner *et al.* 2018, for a review).

Certainly, these issues can impose substantial limitations on findings from studies employing the gold standard paradigm. For example, in Bowden *et al.* (2013), the higher proficiency L2 group showed a LAN-P600 response to word-order violations, which was statistically indistinguishable from the L1 group, while a lower proficiency L2 group showed an early positivity and a smaller and more global P600. The authors interpret their results in line with models that argue that '(. . .) L2 learners initially rely on largely different substrates than (similarly aged) L1 individuals, but, with increasing experience or proficiency, can gradually come to depend on L1 neurocognitive mechanisms' (p. 2509; cf. Steinhauer *et al.* 2009; Ullman 2001, 2016; but see also Clahsen and Felser 2006). While Bowden and colleagues consider the possibility that the averaging procedure obscured a LAN in some individuals in the lower proficiency group, they ignore the possibility that the averaging procedure also obscured *absent LANs* in individuals from the other two groups. Thus, it might very well be that some native and advanced L2 speakers showed the same pattern (i.e. only a P600) as some low proficiency L2 speakers. In line with the conclusion quoted above, this would also mean that some L1 speakers do not rely on L1 neurocognitive mechanisms, making the gold standard much less golden.

Several influential neurocognitive models of L2 acquisition hinge on grammatical processes that are 'automatic', 'fast', 'early', 'rapid', 'implicit', 'procedural', or 'unconscious' (Ullman 2001; Paradis 2004; Clahsen and Felser 2006; Paradis 2009; Steinhauer *et al.* 2009; Ullman 2016). Only near-native or nativelike L2 speakers can rely on these processes, and a disruption—for

example, by means of a morphosyntactic violation—is typically predicted to lead to a LAN. The LAN, as we have seen, is inconsistently and unreliably elicited, and in some cases a mere artefact from data processing. This challenges the view that it reflects the final stage of achievement in L2 grammatical development.

The P600 as a marker of nativelikeness?

A large number of studies in fact adopted the P600 as the benchmark for nativelike grammatical processing (e.g. Osterhout *et al.* 2006, 2008; McLaughlin *et al.* 2010; Tanner *et al.* 2013; Alemán Bañón *et al.* 2014, 2018), probably because it is most reliably elicited by agreement and word-order violations (for overviews, see Molinaro *et al.* 2011; Bowden *et al.* 2013; Caffarra *et al.* 2015). Crucially, none of the P600 accounts (e.g. Hagoort *et al.* 1993; Kaan *et al.* 2000; Friederici 2002; Hagoort 2003; Kolk and Chwilla 2007; Gouvea *et al.* 2010; Brouwer *et al.* 2012) describes the underlying processes as fast, early, or rapid. Its reported relationship to conscious processing (e.g. White *et al.* 2012; Tanner *et al.* 2013; Sassenhagen *et al.* 2014; Van Gaal *et al.* 2014; Rohaut and Naccache 2017) makes it furthermore questionable if it represents implicit, unconscious, or automatic processes—attributes that represent the core characteristics of nativelike grammatical processing in some theories, and that are predicted to be most susceptible to maturational constraints (e.g. Ullman 2001; Paradis 2004; Clahsen and Felser 2006; Paradis 2009; Steinhauer *et al.* 2009; Ullman 2016). In addition, the P600 is neither exclusive to the morphosyntactic (Kuperberg *et al.* 2003; Kuperberg 2007; Rigoulot *et al.* 2020) nor even to the linguistic domain (Kok 1986; Coulson *et al.* 1998; Sassenhagen *et al.* 2014; Sassenhagen and Bornkessel-Schlesewsky 2015).

What does it mean then when L2 speakers show a grand average P600 that is indistinguishable from that of the L1 group? The minimal assumption certainly holds: L2 speakers show nativelike ERPs, thus assuming that the underlying neurocognitive processing was also nativelike. However, the nativelikeness reflected in the P600 might be unrelated to morphosyntactic or linguistic processes, and might only reflect a downstream processing stage *after* linguistic analyses have taken place. Note that this does not entirely spoil the game: As Sassenhagen *et al.* (2014: 37) point out, 'it may become necessary to give up on the idea of a specific EEG index of structural or combinatorial processing, [but] a reliable measure for these [domain general] mechanisms and how they contribute to language processing may be won instead'. Nonetheless, if the P600 is to serve as *the* benchmark for nativelike grammatical processing, theories of what characterizes nativelikeness have to be fundamentally reformulated, or we have to be satisfied with assuming upstream processes that we cannot (yet) measure with ERPs.

Although the P600 has been identified as the most common ERP component as a response to grammatical disruptions, the individual difference

studies by Tanner and colleagues show an N400-P600 tradeoff, which indicates that some individuals show only a N400 *without* a subsequent P600, and vice versa (but see Caffarra *et al.* 2019, where almost all subjects and items show some positivity effect). Thus, without knowing the individual results pattern, a grand average P600 might be misleading and not accurately representing the entire sample. If the P600 alone serves to set the bar for nativelike processing, it should first be demonstrated that most or all L1 individuals show this effect.

THE DYNAMICS OF NATIVELIKENESS

In most behavioral nativelikeness studies, the benchmark for nativelikeness has not been the grand average of the L1 group, but either one or two standard deviations from the L1 average (e.g. Flege *et al.* 1999) or the absolute range of the L1 results (e.g. Abrahamsson and Hyltenstam 2008, 2009; Abrahamsson 2012). In line with Gullberg *et al.* (2008), who suggest that 'the range of variation defines what is "nativelike" and allows for an equal range of possible behaviors for L2 learners that would still qualify as "nativelike"' (p. 158–9), we also believe that the full L1 range 'opens for a more gradient and sophisticated view of L2 performance' (Gullberg *et al.* 2008). If we transfer this to the multidimensional ERP signal, the range of possible ERP patterns in relatively coherent L1 samples—ERP responses that should logically or objectively qualify as *nativelike*—in studies on morphosyntactic processing includes a LAN-P600, an N400, an N400-P600, and a P600.[5] Does this leave us with a situation in which almost any result can be viewed as reflecting nativelike processing? Under the (superficial) view that any L2 result that is statistically indistinguishable from any L1 result reflects nativelike processing, then yes. It seems, however, that the answer is more complex.

For example, some L2 acquisition models (Ullman 2001; Steinhauer *et al.* 2009; Ullman 2016) strongly hinge on a transition from N400 responses at early stages of L2 acquisition to P600s at later stages. In simple terms, such models state that the initial N400 reflects the learners' reliance on lexical-semantic processing. Only at higher proficiency levels can learners rely on the same mechanisms as native speakers and show a P600 (i.e. after 'grammaticalization'; Osterhout *et al.* 2006, 2008). Then again, some L1 speakers also show an N400 to morphosyntactic manipulations (e.g. Tanner and Van Hell 2014). Does this mean that some L1 speakers do not pass the stage of grammaticalization and have to rely on lexical-semantic mechanisms, or that some early-stage learners who show an N400 are nativelike? The answer might be a matter of necessity: While L1 speakers *can* employ lexical-semantic processes to resolve morphosyntactic anomalies, early-stage L2 learners *need to* rely on it more often. This is supported by the fact that P600s are not systematically elicited at initial and low proficiency L2 stages (but see Gabriele et al., In press). In addition, children seem to rely on the same structure-based parsing mechanisms as adults (Clahsen and Felser 2006), which indicates that this is the

default morphosyntactic parsing mode (whether this is reflected in a true LAN and/or a downstream P600 is a matter of experimental inquiry, but it should *not* be reflected in an N400). It is possible that only experienced speakers with relatively more lexical knowledge and/or cognitive resources *can* resort to a rather meaning-based parsing, as reflected in an N400.

The available evidence to pin down factors that influence individual processing strategies is, at most, inconclusive. In the studies of Tanner *et al.* (2013, 2014), age of arrival in an English-speaking country and motivation to speak like a native speaker as well as online behavioral performance have been associated with response type (i.e. N400 or P600), while—contrary to what the longitudinal studies of Osterhout *et al.* (2006, 2008) would suggest—proficiency influenced neither response type nor effect magnitude. Others found that in a syntactic anomaly condition, the N400 effect was influenced by performance, while the P600 effect was mainly influenced by daily usage (Fromont *et al.* 2020). Still, others found no relationship between any predictor variables and response type or effect magnitude (Alemán Bañón *et al.* 2018). Furthermore, in their metanalysis, Caffarra *et al.* (2015) report that longer L2 immersion duration increases the proportion of elicited LANs, whereas the proportion of N400 and P600 responses is increased by an earlier age of L2 acquisition and higher proficiency, respectively.

Importantly, these studies show that processing differs *between* individuals. It is, however, mostly unclear if individual strategies change dynamically, for instance, depending on language experience or experimental variables. Tanner (2019) found that individual response types (N400 or P600) were comparable across two different instantiations of subject-verb agreement (lexical and morphosyntactic). Importantly, all experimental sentences were short and simple, and, therefore, this robustness in response type might be limited to such sentences. A lexical-semantic processing strategy to resolve agreement errors might only be accessible in simpler sentences, thus resulting in some individuals showing an N400.

In fact, most studies on grammatical processing use sentences that are short and simple, and the critical manipulation and its position can sometimes be predictable within an experiment (e.g. Rossi *et al.* 2006; Bowden *et al.* 2013; Tanner *et al.* 2013; for discussion and counterexample, see Gouvea *et al.* 2010). While this is often inevitable to control a set of stimuli, results from these studies might not provide a complete view of what characterizes native grammatical processing. In L2 experiments, relatively easy-to-process sentences, sometimes together with other processing-friendly experimental parameters (Hahne *et al.* 2006; such as a slower presentation rate, e.g. Alemán Bañón *et al.* 2012, 2018; Tanner *et al.* 2013), might not present a strong test for native-like grammatical processing, and converging L1 and L2 ERP patterns might be limited to relatively simple-to-process input. This can lead to an overestimation of the prevalence of individuals who would qualify as nativelike and, thus, individuals who present evidence against the CPH (cf. Long 2005; Abrahamsson and Hyltenstam 2009).

Sentences that are cognitively more demanding, on the other hand, might force individuals to rely on the default structural processing mode to resolve grammatical difficulties—given that this mode is available to them. Such increased processing demands could be achieved with increased sentence length and syntactic complexity (MacDonald *et al.* 1992; McElree *et al.* 2003), more distance between dependent words (Clifton and Frazier 1989; Phillips *et al.* 2005; Alemán Bañón *et al.* 2012), the use of less frequent words to increase lexical processing load (Ferreira *et al.* 1996; Baayen *et al.* 2016), speaker variability in auditory experiments (Weatherholtz and Jaeger 2016), reduced intelligibility through masking (Rabbitt 1968; Wendt *et al.* 2016) or faster presentation rates (Just *et al.* 1982; Griffiths 1990 ; see also Experiments 3 and 4 in Hopp 2010), or interference tasks (Nicol *et al.* 2006; Durlik *et al.* 2016).

The ability to dynamically adapt the parsing strategy to the processing demands might eventually be limited to L1 speakers. Figure 3 provides an overview of the hypothesized availability and sufficiency of the two processing modes as a function of proficiency and input processing demands. At the very lowest proficiency stages, only a lexical-semantic/meaning-based processing strategy might be available, reflected in N400s for low-demand input (cf. Osterhout *et al.* 2006, 2008; sextant number (1)). At slightly higher but still low proficiency, a structure-based mode might also become available, yet limited to low processing demands, reflected in LANs and/or P600s as well as N400s (cf. Rossi *et al.* 2006; Bowden *et al.* 2013; Tanner *et al.* 2013; sextant number (2)). L1 and proficient L2 speakers can show the same N400 and LAN/P600 variation at low processing demands, indicating that they can employ both modes, and that both modes are sufficient to process the input (cf. Bowden *et al.* 2013; Tanner *et al.* 2013, 2014; Tanner and Van Hell 2014; Tanner 2019; sextant number (3)). With high processing demands, the meaning-based mode, which is the only one available to low proficient L2 speakers, might not suffice for nativelike online processing and comprehension might be delayed, slowed down, or even fail with this parsing strategy. This could be reflected in delayed N400 effects, null effects, or other unsystematic ERP modulations (sextant numbers (4) and (5)). High input demands might force L1 and possibly nativelike L2 speakers to resort to structure-based parsing, reflected in LANs and/or P600s but no (or fewer) N400s (sextant number (6)).

Note that once a mode is available it is not replaced by the emergence of another mode. Rather, we hypothesize that some linguistic input can be processed with either mode. Mode selection is likely subject to situational factors, input materials, and speaker experience, whereas the precise variables governing this selection will have to be determined empirically. Likewise, there are no studies that have systematically looked into qualitative individual differences when processing demands are high, and, therefore, these predictions also require empirical testing.

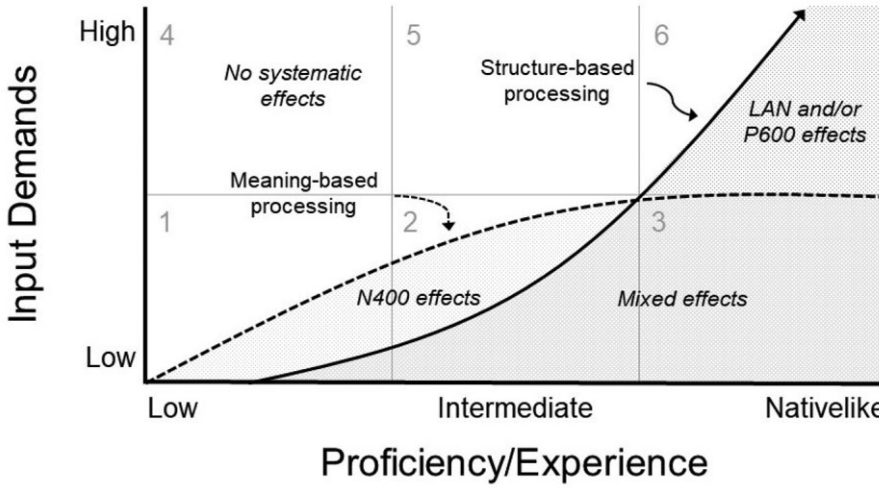

Figure 3: Schematic illustration of the availability and sufficiency of the two processing modes as a function of input processing demands and proficiency/experience. At the very lowest proficiency stages, only a lexical-semantic/meaning-based processing strategy might be available, reflected in N400s for low-demand input. At slightly higher but still low proficiency, a structure-based mode might also become available, yet limited to low processing demands, reflected in LANs and/or P600s as well as N400s. L1 and proficient L2 speakers can show the same N400 and LAN/P600 variation at low processing demands, indicating that they can *employ both modes, and that both modes are sufficient to process the input. With high processing demands, the meaning-based mode, which is the only one available to low proficient L2 speakers, might not suffice for nativelike online processing, possibly reflected in unsystematic ERP modulations. High input demands might force L1 and possibly nativelike L2 speakers to resort to structure-based parsing, reflected in LANs and/or P600s but no (or fewer) N400s. The numbered sextants are included for ease of reference to certain input demand/proficiency combinations in the main text.*

CONCLUSIONS

Even in coherent L1 samples, individuals employ different parsing strategies to process morphosyntactic deviations, as reflected in a set of distinct ERP responses that are not always represented in the grand average. Methodological concerns surrounding the LAN might limit its applicability as a benchmark for nativelike processing, and domain-general functional views of the P600 could necessitate a rethinking of L2 acquisition models if this component is to serve as a benchmark. Naturally, this can limit conclusions from

earlier studies using the gold standard paradigm (cf. Tanner 2019). Nativelike grammatical processing should no longer be imperatively defined by a single grand mean ERP pattern but rather by the full range of systematic patterns that L1 speakers show. It is, therefore, essential that future studies on L2 grammatical processing show that the grand averages are representative of most or all participants in the sample, or report the range of individual response patterns and discuss the L2 results with respect to this range.

The qualitative variation observed in the outlined individual difference studies clearly demonstrates the dynamics in language comprehension. To systematically pinpoint these dynamics, manipulating the processing demands posed by the input could prove a useful tool and complement the picture that emerged from previous studies with lower processing demands. Through this, systematic variation observed in language comprehension can be integrated into the concept of nativelikeness. After all, processing dynamics might provide a stronger and more fine-grained test of age effects and enable a deeper understanding of nativelike attainment in an L2.

ENDNOTES

1 Please note that individual variation is just as big a challenge for many L1 research designs as it is for L2 research designs. L2 research presents a model example for challenges associated with individual variation, because in L2 research, the gold standard paradigm prevails without much criticism, whereas for L1, this topic has been addressed and reviewed to a greater extent (e.g. Tanner *et al.* 2018).

2 The validity and reliability of the ELAN have been critically challenged because it might be a spurious effect resulting from baseline correction in EEG data processing in designs that are unbalanced with respect to the word type preceding the critical words across conditions. This has been discussed in depth (Steinhauer and Drury 2012), and attempts to find ELAN effects in balanced designs have failed (Bowden *et al.* 2013; Fromont *et al.* 2016, 2020). We will, therefore, not specifically discuss this component any further.

3 Further examples are reading times, the score in a grammaticality or acceptability judgment task, the number of recalled elements in a memory task, the number of correct answers in a cloze-test, or the ERP amplitude difference between two conditions in a specific time window.

4 Certainly, there is some relativity and vagueness of 'comparable linguistic and educational backgrounds', since studies run in different countries have participants with different curricula, different class times allocated for L2 learning, different goals, and methods of learning, etc.

5 We would like to clarify that we propose that every *systematic* native-speaker response pattern should qualify as a benchmark for nativelike processing. The EEG signal can be noisy, and a single individual showing odd patterns is not uncommon. Clearly, such outlier responses should not be included in the result patterns that qualify as nativelike.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers and José Alemán Bañón, Guillermo Montero-Melis, and Gunnar Norrman for helpful comments on earlier versions of the manuscript. Thanks also go to Lamont Antieau for editing our English writing. Any remaining errors are entirely our own.

FUNDING

This research was funded by Vetenskapsrådet grant 2016-01630 (to N.A. and E.B.).

CONFLICT OF INTEREST

The authors report no conflict of interest.

REFERENCES

Abrahamsson, N. 2012. 'Age of onset and native-like L2 ultimate attainment of morphosyntactic and phonetic intuition,' *Studies in Second Language Acquisition* 34: 187–214.

Abrahamsson, N. and **K. Hyltenstam**. 2008. 'The robustness of aptitude effects in near-native second language acquisition,' *Studies in Second Language Acquisition* 30: 481–509.

Abrahamsson, N. and **K. Hyltenstam**. 2009. 'Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny,' *Language Learning* 59: 249–306.

Alemán Bañón, J., **R. Fiorentino**, and **A. Gabriele**. 2012. 'The processing of number and gender agreement in Spanish: An event-related potential investigation of the effects of structural distance,' *Brain Research* 1456: 49–63.

Alemán Bañón, J., **R. Fiorentino**, and **A. Gabriele**. 2014. 'Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance,' *Second Language Research* 30: 275–306.

Alemán Bañón, J., **R. Fiorentino**, and **A. Gabriele**. 2018. 'Using event-related potentials to track morphosyntactic development in second language learners: The processing of number and gender agreement in Spanish,' *PLoS One* 13: e0200791.

Allen, M., **E. Badecker**, and **L. Osterhout**. 2003. 'Morphological analysis in sentence processing: An ERP study,' *Language and Cognitive Processes* 18: 405–30.

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

Baayen, R. H., **P. Milin**, and **M. Ramscar**. 2016. 'Frequency in lexical processing,' *Aphasiology* 30: 1174–220.

Birdsong, D. 2005. 'Nativelikeness and non-nativelikeness in L2A research,' *International Review of Applied Linguistics in Language Teaching* 43: 319–28.

Birdsong, D. 2006. 'Age and second language acquisition and processing: A selective overview,' *Language Learning* 56: 9–49.

Bley-Vroman, R. 1989. 'What is the logical problem of foreign language learning?' in S. Gass and J. Schachter (eds): *Linguistic Perspectives on Second Language Acquisition*. Cambridge University Press, 1–68.

Bornkessel, I. and **M. Schlesewsky**. 2006. 'The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages,' *Psychological Review* 113: 787–821.

Boudewyn, M. A. 2015. 'Individual differences in language processing: Electrophysiological approaches,' *Language and Linguistics Compass* 9: 406–19.

Bowden, H. W., **K. Steinhauer**, **C. Sanz**, and **M. T. Ullman**. 2013. 'Native-like brain processing

of syntax can be attained by university foreign language learners,' *Neuropsychologia* 51: 2492–511.

Brouwer, H., **H. Fitz**, and **J. Hoeks**. 2012. 'Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension,' *Brain Research* 1446: 127–43.

Caffarra, S., **M. Mendoza**, and **D. Davidson**. 2019. 'Is the LAN effect in morphosyntactic processing an ERP artifact?,' *Brain and Language* 191: 9–16.

Caffarra, S., **N. Molinaro**, **D. Davidson**, and **M. Carreiras**. 2015. 'Second language syntactic processing revealed through event-related potentials: An empirical review,' *Neuroscience and Biobehavioral Reviews* 51: 31–47.

Clahsen, H. and **C. Felser**. 2006. 'Grammatical processing in language learners,' *Applied Psycholinguistics* 27: 3–42.

Clifton, C. and **L. Frazier**. 1989. 'Comprehending sentences with long-distance dependencies' in G. N. Carlson, and M. K. Tanenhaus (eds): *Linguistic Structure in Language Processing*. Springer, 273–317.

Coulson, S., **J. W. King**, and **M. Kutas**. 1998. 'Expect the unexpected: Event-related brain response to morphosyntactic violations,' *Language and Cognitive Processes* 13: 21–58.

Curtiss, S. 1977. *Genie: A Psycholinguistic Study of a Modern Day "Wild Child"*. Academic Press.

DeKeyser, R. M. 2001. 'Automaticity and automatization' in P. Robinson (ed): *Cognition and Second Language Acquisition*. Cambridge University Press, 125–151.

DeKeyser, R. and **J. Larson-Hall**. 2005. 'What does the Critical Period really mean?' in J. F. Kroll and A. M. B. DeGroot (eds.): *Handbook of Bilingualism: Psycholinguistic Perspectives*. Oxford University Press, 88–108.

Dillon, B., **A. Nevins**, **A. C. Austin**, and **C. Phillips**. 2012. 'Syntactic and semantic predictors of tense in Hindi: An ERP investigation,' *Language and Cognitive Processes* 27: 313–44.

Durlik, J., **J. Szewczyk**, **M. Muszyński**, and **Z. Wodniecka**. 2016. 'Interference and inhibition in bilingual language comprehension: Evidence from Polish-English interlingual homographs,' *PLoS One* 11: e0151430.

Ferreira, F., **J. M. Henderson**, **M. D. Anes**, **P. A. Weeks**, and **D. K. McFarlane**. 1996. 'Effects of lexical frequency and syntactic complexity in spoken-language comprehension: Evidence from the auditory moving-window technique,' *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22/2: 324–35.

Flege, J. E., **G. H. Yeni-Komshian**, and **S. Liu**. 1999. 'Age constraints on second-language acquisition,' *Journal of Memory and Language* 41: 78–104.

Ford, M. 1983. 'A method of obtaining measures of local parsing complexity throughout sentences,' *Journal of Verbal Learning and Verbal Behavior* 22: 203–18.

Foucart, A. and **C. Frenck-Mestre**. 2011. 'Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity,' *Bilingualism: Language and Cognition* 14: 379–99.

Foucart, A. and **C. Frenck-Mestre**. 2012. 'Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking,' *Journal of Memory and Language* 66: 226–48.

Friederici, A. D. 2002. 'Towards a neural basis of auditory sentence processing,' *Trends in Cognitive Sciences* 6: 78–84.

Friederici, A. D., **D. Y. Von Cramon**, and **S. A. Kotz**. 1999. 'Language related brain potentials in patients with cortical and subcortical left hemisphere lesions,' *Brain: A Journal of Neurology* 122 (Pt 6): 1033–47.

Fromont, L. A., **P. Royle**, **I. Perlitch**, and **K. Steinhauer**. 2016. 'Re-evaluating the dynamics of phrase-structure processing using Event Related Potentials: The case of syntactic categories in French,' *International Journal of Psychophysiology* 108: 86.

Fromont, L. A., **P. Royle**, and **K. Steinhauer**. 2020. 'Growing Random Forests reveals that exposure and proficiency best account for individual variability in L2 (and L1) brain potentials for syntax and semantics,' *Brain and Language* 204: 104770.

Gabriele, A., **J. Alemán Bañón**, **L. Hoffman**, **L. Covey**, **A. Rossomondo**, and **R. Fiorentino**. In press. 'Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials,' *Journal of Experimental Psychology: Learning, Memory, and Cognition*

Gouvea, A. C., **C. Phillips**, **N. Kazanina**, and **D. Poeppel**. 2010. 'The linguistic processes underlying the P600,' *Language and Cognitive Processes* 25: 149–88.

Gregg, K. R. 1996. 'The logical and developmental problems of second language acquisition' in W.

C. Ritchie and T. K. Bhatia (eds): *Handbook of Second Language Acquisition*. Academic Press, 49–81.

Griffiths, R. 1990. 'Speech rate and nonnative speaker comprehension: A preliminary study in the time-benefit analysis,' *Language Learning* 40: 311–36.

Gullberg, M., **K. De Bot**, and **V. Volterra**. 2008. 'Gestures and some key issues in the study of language development,' *Gesture* 8/2: 149–79.

Hagoort, P. 2003. 'How the brain solves the binding problem for language: A neurocomputational model of syntactic processing,' *Neuroimage* 20 Suppl 1: S18–29.

Hagoort, P. and **C. M. Brown**. 1999. 'Gender electrified: ERP evidence on the syntactic nature of gender processing,' *Journal of Psycholinguistic Research* 28/6: 715–28.

Hagoort, P., **C. Brown**, and **J. Groothusen**. 1993. 'The syntactic positive shift (SPS) as an ERP measure of syntactic processing,' *Language and Cognitive Processes* 8: 439–83.

Hagoort, P., **M. Wassenaar**, and **C. M. Brown**. 2003. 'Syntax-related ERP-effects in Dutch,' *Brain Research. Cognitive Brain Research* 16: 38–50.

Hahne, A. 2001. 'What's different in second-language processing? Evidence from event-related brain potentials,' *Journal of Psycholinguistic Research* 30/3: 251–66.

Hahne, A. and **A. D. Friederici**. 2001. 'Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials,' *Bilingualism: Language and Cognition* 4: 123–41.

Hahne, A., **J. L. Mueller**, and **H. Clahsen**. 2006. 'Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition,' *Journal of Cognitive Neuroscience* 18: 121–34.

Hopp, H. 2010. 'Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers,' *Lingua* 120: 901–31.

Hyltenstam, K. and **N. Abrahamsson**. 2003. 'Maturational constraints in SLA' in C. J. Doughty and M. H. Long (eds): *The Handbook of Second Language Acquisition*. Blackwell Publishing, 539–588.

Just, M. A., **P. A. Carpenter**, and **J. D. Woolley**. 1982. 'Paradigms and processes in reading comprehension,' *Journal of Experimental Psychology: General* 111: 228–47.

Kaan, E. 2007. 'Event-related potentials and language processing: A brief overview,' *Language and Linguistics Compass* 1: 571–91.

Kaan, E., **A. Harris**, **E. Gibson**, and **P. Holcomb**. 2000. 'The P600 as an index of syntactic integration difficulty,' *Language and Cognitive Processes* 15: 159–201.

Kaan, E. and **T. Y. Swaab**. 2003. 'Electrophysiological evidence for serial sentence processing: A comparison between non-preferred and ungrammatical continuations,' *Brain Research. Cognitive Brain Research* 17: 621–35.

King, J. and **M. A. Just**. 1991. 'Individual differences in syntactic processing: The role of working memory,' *Journal of Memory and Language* 30: 580–602.

Knösche, T. R., **B. Maess**, and **A. D. Friederici**. 1999. 'Processing of syntactic information monitored by brain surface current density mapping based on MEG,' *Brain Topography* 12/2: 75–87.

Kolk, H. and **D. Chwilla**. 2007. 'Late positivities in unusual situations,' *Brain and Language* 100: 257–61.

Kok, A. 1986. 'Effects of degradation of visual stimuli on components of the event-related potential (ERP) in go/nogo reaction tasks,' *Biological Psychology* 23: 21–38.

Kotz, S. A., **P. J. Holcomb**, and **L. Osterhout**. 2008. 'ERPs reveal comparable syntactic sentence processing in native and non-native readers of English,' *Acta Psychologica* 128: 514–27.

Kuperberg, G. R. 2007. 'Neural mechanisms of language comprehension: Challenges to syntax,' *Brain Research* 1146: 23–49.

Kuperberg, G. R., **P. J. Holcomb**, **T. Sitnikova**, **D. Greve**, **A. M. Dale**, and **D. Caplan**. 2003. 'Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies,' *Journal of Cognitive Neuroscience* 15: 272–93.

Kutas, M. and **K. D. Federmeier**. 2011. 'Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP),' *Annual Review of Psychology* 62: 621–47.

Kutas, M. and **V. Iragui**. 1998. 'The N400 in a semantic categorization task across 6 decades,' *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 108: 456–71.

Lenneberg, E. 1967. *Biological Foundations of Language*. Wiley.

Long, M. H. 2005. 'Problems with supposed counter-evidence to the critical period hypothesis,' *International Review of Applied Linguistics in Language Teaching* 43: 287–317.

MacDonald, M. C., **M. A. Just**, and **P. A. Carpenter**. 1992. 'Working memory constraints on the processing of syntactic ambiguity,' *Cognitive Psychology* 24: 56–98.

Mancini, S., **N. Molinaro**, **L. Rizzi**, and **M. Carreiras**. 2011. 'A person is not a number: Discourse involvement in subject–verb agreement computation,' *Brain Research* 1410: 64–76.

Mayberry, R. I. 1993. 'First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language,' *Journal of Speech, Language, and Hearing Research* 36/6: 1258–70.

Mayberry, R. I. and **R. Kluender**. 2018. 'Rethinking the critical period for language: New insights into an old question from American Sign Language,' *Bilingualism: Language and Cognition* 21: 886–905.

McElree, B., **S. Foraker**, and **L. Dyer**. 2003. 'Memory structures that subserve sentence comprehension,' *Journal of Memory and Language* 48: 67–91.

McLaughlin, J., **L. Osterhout**, and **A. Kim**. 2004. 'Neural correlates of second-language word learning: Minimal instruction produces rapid change,' *Nature Neuroscience* 7: 703–4.

McLaughlin, J., **D. Tanner**, **I. Pitkänen**, **C. Frenck-Mestre**, **K. Inoue**, **G. Valentine**, and **L. Osterhout**. 2010. 'Brain potentials reveal discrete stages of L2 grammatical learning,' *Language Learning* 60: 123–50.

Molinaro, N., **H. A. Barber**, and **M. Carreiras**. 2011. 'Grammatical agreement processing in reading: ERP findings and future directions,' *Cortex* 47: 908–30.

Molinaro, N., **H. A. Barber**, **S. Caffarra**, and **M. Carreiras**. 2015. 'On the left anterior negativity (LAN): The case of morphosyntactic agreement,' *Cortex* 66: 156–9.

Moreno, E. M. and **M. Kutas**. 2005. 'Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals,' *Brain Research. Cognitive Brain Research* 22: 205–20.

Neville, H., **J. L. Nicol**, **A. Barss**, **K. I. Forster**, and **M. F. Garrett**. 1991. 'Syntactically based sentence processing classes: Evidence from event-related brain potentials,' *Journal of Cognitive Neuroscience* 3: 151–65.

Nevins, A., **B. Dillon**, **S. Malhotra**, and **C. Phillips**. 2007. 'The role of feature-number and feature-type in processing Hindi verb agreement violations,' *Brain Research* 1164: 81–94.

Newman, A. J., **A. Tremblay**, **E. S. Nichols**, **H. J. Neville**, and **M. T. Ullman**. 2012. 'The influence of language proficiency on lexical semantic processing in native and late learners of English,' *Journal of Cognitive Neuroscience* 24: 1205–23.

Newport, E. L. 1990. 'Maturational constraints on language learning,' *Cognitive Science* 14: 11–28.

Nichols, E. S. and **M. F. Joanisse**. 2019. 'Individual differences predict ERP signatures of second language learning of novel grammatical rules,' *Bilingualism: Language and Cognition* 22: 78–92.

Nicol, J., **D. Swinney**, **T. Love**, and **L. Hald**. 2006. 'The on-line study of sentence comprehension: An examination of dual task paradigms,' *Journal of Psycholinguistic Research* 35: 215–31.

Nieuwland, M. S., **D. J. Barr**, **F. Bartolozzi**, **S. Busch-Moreno**, **E. Darley**, **D. I. Donaldson**, **H. J. Ferguson**, **X. Fu**, **E. Heyselaar**, **F. Huettig**, **E. Matthew Husband**, **A. Ito**, **N. Kazanina**, **V. Kogan**, **Z. Kohút**, **E. Kulakova**, **D. Mézière**, **S. Politzer-Ahles**, **G. Rousselet**, **S.-A. Rueschemeyer**, **K. Segaert**, **J. Tuomainen**, and **S. Von Grebmer Zu Wolfsthurn**. 2020. 'Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials,' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 375: 20180522.

Ojima, S., **H. Nakata**, and **R. Kakigi**. 2005. 'An ERP study of second language learning after childhood: Effects of proficiency,' *Journal of Cognitive Neuroscience* 17: 1212–28.

Osterhout, L. 1997. 'On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences,' *Brain and Language* 59: 494–522.

Osterhout, L., **J. McLaughlin**, **A. Kim**, and **K. Inoue**. 2004. 'Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning' in M. Carreiras and C. Clifton Jr. (eds): *The on-Line Study of Sentence Comprehension:*

Eyetracking, ERP, and Beyond. Psychology Press, 271–308.

Osterhout, L., **J. McLaughlin**, **I. Pitkänen**, **C. Frenck-Mestre**, and **N. Molinaro**. 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., **A. Poliakov**, **K. Inoue**, **J. McLaughlin**, **G. Valentine**, **I. Pitkanen**, **C. Frenck-Mestre**, and **J. Hirschensohn**. 2008. 'Second-language learning and changes in the brain,' *Journal of Neurolinguistics* 21: 509–21.

Osterhout, L. and **P. J. Holcomb**. 1992. 'Event-related brain potentials elicited by syntactic anomaly,' *Journal of Memory and Language* 31: 785–806.

Osterhout, L. and **J. Nicol**. 1999. 'On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies,' *Language and Cognitive Processes* 14: 283–317.

Pakulak, E. and **H. J. Neville**. 2011. 'Maturational constraints on the recruitment of early processes for syntactic processing,' *Journal of Cognitive Neuroscience* 23: 2752–65.

Paradis, M. 2004. *A Neurolinguistic Theory of Bilingualism*, vol. 18. John Benjamins Publishing.

Paradis, M. 2009. *Declarative and Procedural Determinants of Second Languages*, vol. 40. John Benjamins Publishing.

Patkowski, M. S. 1980. 'The sensitive period for the acquisition of syntax in a second language,' *Language Learning* 30: 449–68.

Phillips, C., **N. Kazanina**, and **S. H. Abada**. 2005. 'ERP effects of the processing of syntactic long-distance dependencies,' *Brain Research. Cognitive Brain Research* 22: 407–28.

Rabbitt, P. M. 1968. 'Channel-capacity, intelligibility and immediate memory,' *The Quarterly Journal of Experimental Psychology* 20: 241–8.

Rigoulot, S., **X. Jiang**, **N. Vergis**, and **M. D. Pell**. 2020. 'Neurophysiological correlates of sexually evocative speech,' *Biological Psychology* 154: 107909.

Rohaut, B. and **L. Naccache**. 2017. 'Disentangling conscious from unconscious cognitive processing with event-related EEG potentials,' *Revue Neurologique* 173: 521–8.

Rossi, S., **M. F. Gugler**, **A. D. Friederici**, and **A. D. Hahne**. 2006. 'The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials,' *Journal of Cognitive Neuroscience* 18: 2030–48.

Sassenhagen, J. and **I. Bornkessel-Schlesewsky**. 2015. 'The P600 as a correlate of ventral attention network reorientation,' *Cortex* 66: A3–20.

Sassenhagen, J., **M. Schlesewsky**, and **I. Bornkessel-Schlesewsky**. 2014. 'The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned,' *Brain and Language* 137: 29–39.

Silva-Pereyra, J. F. and **M. Carreiras**. 2007. 'An ERP study of agreement features in Spanish,' *Brain Research* 1185: 201–11.

Steinhauer, K. 2014. 'Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2,' *Applied Linguistics* 35: 393–417.

Steinhauer, K. and **J. E. Drury**. 2012. 'On the early left-anterior negativity (ELAN) in syntax studies,' *Brain and Language* 120: 135–62.

Steinhauer, K., **E. J. White**, and **J. E. Drury**. 2009. 'Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials,' *Second Language Research* 25: 13–41.

Swaab, T., **C. Brown**, and **P. Hagoort**. 1997. 'Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit,' *Journal of Cognitive Neuroscience* 9: 39–66.

Tanner, D. 2015. 'On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A commentary on" grammatical agreement processing in reading: ERP findings and future directions" by Molinaro *et al.*, 2014,' *Cortex* 66: 149–55.

Tanner, D. 2019. 'Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach,' *Cortex* 111: 210–37.

Tanner, D., **M. Goldshtein**, and **B. Weissman**. 2018. 'Individual differences in the real-time neural dynamics of language comprehension' in K. D. Federmeier and D. G. Watson (eds): *Psychology of Learning and Motivation*. Academic Press, 299–335.

Tanner, D., **K. Inoue**, and **L. Osterhout**. 2014. 'Brain-based individual differences in online L2 grammatical comprehension,' *Bilingualism: Language and Cognition* 17: 277–93.

Tanner, D., **J. McLaughlin**, **J. Herschensohn**, and **L. Osterhout**. 2013. 'Individual differences reveal stages of L2 grammatical acquisition: ERP evidence,' *Bilingualism: Language and Cognition* 16: 367–82.

Tanner, D. and **J. G. Van Hell**. 2014. 'ERPs reveal individual differences in morphosyntactic processing,' *Neuropsychologia* 56: 289–301.

Ullman, M. T. 2001. 'The neural basis of lexicon and grammar in first and second language: The declarative/procedural model,' *Bilingualism: Language and Cognition* 4: 105–22.

Ullman, M. T. 2016. 'The declarative/procedural model: A neurobiological model of language learning, knowledge, and use' in G. Hickok and S. L. Small (eds): *Neurobiology of Language*. Academic Press, 953–968.

Van Gaal, S., **L. Naccache**, **J. D. Meuwese**, **A. M. Van Loon**, **A. H. Leighton**, **L. Cohen**, and **S. Dehaene**. 2014. 'Can the meaning of multiple words be integrated unconsciously?,' *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369: 20130212.

Van Hell, J. G., and **F. Abdollahi**. 2017. Individual variation in syntactic processing in the second language: Electrophysiological approaches. Developmental Perspectives in Written Language and Literacy, (February 2018), 257–273.

Weatherholtz, K. and **T. F. Jaeger**. 2016. 'Speech perception and generalization across talkers and accents,' *Oxford Research Encyclopedia of Linguistics.*

Weber, K. and **A. Lavric**. 2008. 'Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first,' *Psychophysiology* 45: 920–5.

Weber-Fox, C. M. and **H. J. Neville**. 1996. 'Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers,' *Journal of Cognitive Neuroscience* 8: 231–56.

Wendt, D., **T. Dau**, and **J. Hjortkjær**. 2016. 'Impact of background noise and sentence complexity on processing demands during sentence comprehension,' *Frontiers in Psychology* 7: 345.

White, L. and **F. Genesee**. 1996. 'How native is near-native? The issue of ultimate attainment in adult second language acquisition,' *Second Language Research* 12: 233–65.

White, E. J., **F. Genesee**, and **K. Steinhauer**. 2012. 'Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners,' *PLoS One* 7: e52318.