

Tracking Syntactic Conflict between Languages over the Course of L2 Acquisition: A Cross-sectional Event-related Potential Study

Anne Mickan^{1,2} and Kristin Lemhöfer¹

Abstract

■ One challenge of learning a foreign language (L2) in adulthood is the mastery of syntactic structures that are implemented differently in L2 and one's native language (L1). Here, we asked how L2 speakers learn to process syntactic constructions that are in direct conflict between L1 and L2, in comparison to structures without such a conflict. To do so, we measured EEG during sentence reading in three groups of German learners of Dutch with different degrees of L2 experience (from 3 to more than 18 months of L2 immersion) as well as a control group of Dutch native speakers. They read grammatical and ungrammatical Dutch sentences that, in the conflict condition, contained a structure with opposing word orders in Dutch and German (sentence-final double infinitives) and, in the no-conflict condition, a structure for which

word order is identical in Dutch and German (subordinate clause inversion). Results showed, first, that beginning learners showed N400-like signatures instead of the expected P600 for both types of violations, suggesting that, in the very early stages of learning, different neurocognitive processes are employed compared with native speakers, regardless of L1–L2 similarity. In contrast, both advanced and intermediate learners already showed native-like P600 signatures for the no-conflict sentences. However, their P600 signatures were significantly delayed in processing the conflicting structure, even though behavioral performance was on a native level for both these groups and structures. These findings suggest that L1–L2 word order conflicts clearly remain an obstacle to native-like processing, even for advanced L2 learners. ■

INTRODUCTION

In many environments, learning to speak foreign languages not only fluently but also flawlessly has become a matter of course. This includes mastering the details of a second language's grammar. The fact that nonnative speakers who still make grammatical mistakes are perceived as generally less intelligent and even less likeable than other speakers (Queen & Boland, 2015; Ensz, 1982) demonstrates just how essential it is to attain grammatical competence in a foreign language (L2). Grammatical competence, however, is hard to achieve, especially for adult learners (Hartshorne, Tenenbaum, & Pinker, 2018). Recently, much research has been conducted to investigate how these adult learners process L2 grammar and whether it is principally possible for them to gain native-like competence in this domain (e.g., Yuan, 2017; Hopp, 2010; Birdsong, 2006; Hartsuiker, Pickering, & Veltkamp, 2004). This study investigates in how far two factors determine whether L2 learners display native-like syntactic processing in L2: the learners' degree of experience with L2 and whether the native language (L1) and L2 overlap or conflict with each other regarding the relevant syntactic feature.

We are interested not only in how learners' mastery of L2 syntax compares to that of native speakers behaviorally but also whether they employ the same neural processes during syntactic processing as native speakers do. To answer this question, many previous studies have used event-related potentials (ERPs) to syntactic violations in the EEG and compared these across L1 and L2 speakers (see Caffarra, Molinaro, Davidson, & Carreiras, 2015; Steinhauer, White, & Drury, 2009; Steinhauer, 2014; Schmid, Dusseldorp, van Hell, & Tokowicz, 2010; Kotz, 2009, for reviews). The most robustly obtained ERP effect of syntactic violations compared with grammatical sentences is the P600, a centroposterior positive deflection peaking about 600 msec after the onset of the critical word. In some cases, particularly for word category and phrase structure violations, an earlier frontal negativity is additionally found that is mostly left-lateralized (see Steinhauer & Drury, 2012). The occurrence of such ERP effects is generally taken as proof for the reader's sensitivity to the violations, thus demonstrating intact syntactic processing. A comparison of the latency, polarity, amplitude, and scalp distribution of ERP effects between L1 and L2 speakers additionally provides information on whether the processing of the manipulated syntactic features is identical on the neural level and in how far learners process L2 syntax in a native-like manner.

¹Radboud University Nijmegen, ²Max Planck Institute for Psycholinguistics, International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

One obviously important factor that is likely to affect whether an L2 learner approximates native-like ERP patterns is at which stage during the acquisition period the learner is situated. However, only few studies on L2 syntactic processing have systematically varied participants' degree of experience with L2 (e.g., see White, Genesee, & Steinhauer, 2012; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; Rossi, Gugler, Friederici, & Hahne, 2006, for exceptions). Across studies, nevertheless, certain regularities in ERP patterns of learners at different proficiency or experience levels seem to emerge, as observed by Steinhauer et al. (2009). ERP responses to syntactic violations in L2 seem to start from no response at all in beginners (when there is no sensitivity to violations yet) to N400-like signatures in intermediate learners, to more native-like patterns in highly proficient learners. These native-like patterns are characterized by the occurrence of a P600 that may, however, initially be smaller or delayed relative to natives and only reaches native-like characteristics in the most advanced speaker groups, if at all. Additionally, depending on the type of violation, the P600 may be preceded by left-lateralized anterior negativities (LANs) in the most advanced speaker groups (but see, for instance, Lemhöfer, Schriefers, & Indefrey, 2014, for results showing a LAN for L2, but not native speakers).

Interestingly, thus, rather than following a simple developmental pattern in which the native-like P600 slowly emerges from no effect at all, learners seem to go through an intermediate stage in which they do show behavioral sensitivity to syntactic violations in L2, but in which their EEG is characterized by qualitatively different ERP responses compared with native speakers, namely N400-like instead of P600 effects. It has been proposed that these qualitative changes in ERP patterns across the L2 acquisition period reflect changes in memory processes, as L2 processing becomes more efficient, more rule-based, and more dependent on procedural memory (Steinhauer et al., 2009; Ullman, 2001). However, the available evidence is too scarce and contains too few direct comparisons of learner groups differing in L2 experience to conclude that this proposed trajectory of ERP effects is a general pattern in L2 syntax acquisition across syntactic features and L1–L2 combinations. This study aims at putting the generality of the proposed trajectory to a test for two complex Dutch word order constructions, as processed by German learners with different degrees of experience with Dutch.

Besides L2 experience, other factors are likely to influence the development of L2 syntactic processing. One such factor concerns how difficult it is for a learner to acquire a given syntactic feature in the L2. The ease with which a feature of L2 grammar can be learned is likely to depend on its relation to the learner's native language. If a grammatical feature from the L2 is implemented in an identical or at least similar way as in the L1, then learners will likely be faster and ultimately more successful

in acquiring it than when L1 offers nothing to rely on or when the L1 is in conflict with the L2. As obvious as this difference may be, to our knowledge, there is very little research that directly compares the processing of similar and dissimilar structures within one general adult learner population over time, that is, either longitudinally or cross-sectionally (for an exception, see Alemán Bañón, Fiorentino, & Gabriele, 2018; Osterhout et al., 2006). Instead, existing studies typically investigate either two structures differing in similarity within one learner group (e.g., Alemán Bañón, Fiorentino, & Gabriele, 2014; Kotz, Holcomb, & Osterhout, 2008) or different learner groups in their processing of one specific syntactic structure (e.g., Bowden, Steinhauer, Sanz, & Ullman, 2013; McLaughlin et al., 2010).

What is more, the vast majority of previous L2 syntactic processing studies on the similarity between L1 and L2 have focused on grammatical structures that are unique to the L2 (e.g., grammatical gender agreement which is lacking in English). They either investigated whether L2 speakers process these unique features differently from native speakers (e.g., Meulman, Stowe, Sprenger, Bresser, & Schmid, 2014; Bowden et al., 2013; Hopp, 2010; Kotz et al., 2008; Chen, Shu, Liu, Zhao, & Li, 2007) or additionally compared them to other grammatical structures that are implemented similarly across L1 and L2 (Díaz et al., 2016; Alemán Bañón et al., 2014; Foucart & Frenck-Mestre, 2011; Zawiszewski, Gutiérrez, Fernández, & Laka, 2011; Gillon Dowens, Vergara, Barber, & Carreiras, 2010; Rossi et al., 2006). However, few studies have investigated how L2 learners deal with L2 syntactic features that are in direct conflict with the L1. Word order plays an important role here, because it has been claimed that syntactic representations are shared between the two languages of a bilingual, but possibly only if the respective constructions follow the same word order (Hartsuiker et al., 2004). Word order conflicts are frequent between languages, yet have received surprisingly little attention in studies on L2 syntactic processing. An example for word order differences is verb placement in subordinate clauses: In both German and Dutch, word order in subordinate clauses changes from SVO (also called V2) to SOV (also called verb final) order when the clause is introduced by a subordinating (as compared with a coordinating) conjunction (e.g., German: *weil*, Dutch: *omdat* "because"). These word order changes are difficult to learn for native speakers of languages without such word order variations in subordinate clauses, like English.

The few studies that examined how L2 learners deal with such conflicts have shown that word order differences in L1 and L2 remain difficult even at high proficiency levels and when the L2 was learned early on in life (Erdocia & Laka, 2018; Erdocia, Zawiszewski, & Laka, 2014). In both of these studies, learners' online processing signatures showed clear evidence of L1 Spanish influence on parsing L2 Basque sentences. Processing was additionally complicated by Basque case morphology though, so the pure effect of word order violations

remains unclear. Furthermore, these studies investigated only one group of early bilinguals and hence did not address developmental aspects of L2 acquisition as the current study intends to do.

The specific conflict in word order we investigated concerns a recursive syntactic structure, namely double infinitives in German and Dutch. In both languages, sentences containing a modal auxiliary and a content verb in present perfect, in a sentence like “she has wanted to buy the dress” are realized such that the two verbs (the modal auxiliary “want” and the content verb “buy”) are placed as infinitives at the end of the sentence (e.g., Dutch: *zij heeft de jurk willen kopen*, literally “She has the dress want_{INF} buy_{INF}”). Dutch and German differ in the order in which they place these infinitives: Whereas Dutch requires the modal auxiliary to precede the content verb (...*willen kopen*, “want buy”), German instead requires the modal auxiliary to be placed after the content verb (...*kaufen wollen*, “buy want”; see Table 3 for example sentences). Thus, German learners of Dutch will probably have little trouble learning the general principle of this particular Dutch syntactic structure but will have to learn to overcome their native tendency to place the two final infinitives in the content verb–modal auxiliary verb order and to use the reverse order instead. The influence of the L1 is likely especially strong for this language combination, given the large degree of syntactic and lexical similarity of the two languages. We compared this conflicting word order to a structure involving above-mentioned word order changes from SVO to SOV in subordinate clauses after subordinating conjunctions. Although these word order changes seem complex, they are implemented in the same way in German and Dutch. Thus, no L1–L2 conflict should arise in this condition.

In a longitudinal MEG study, Davidson and Indefrey (2009) investigated a similar verb order conflict between Dutch and German and followed German learners across their first 3 months of Dutch instruction. Participants were asked to judge whether scenes were accurately described by auditorily presented Dutch sentences containing final verb clusters which have different (preferred) orders in Dutch and German. Participants started to show a frontal LAN-like effect in the MEG to the dispreferred as compared with the preferred Dutch verb order after as early as 2 weeks of classroom instruction. This result seems to point at fast syntactic learning despite the presence of an L1–L2 conflict. However, the lack of a native speaker control group leaves open in how far the ERF signatures of learners at different stages of proficiency approached native-like levels. Furthermore, the conflicting construction was not compared with a no-conflict condition that could provide a baseline for L2 syntactic learning unhampered by L1 conflicts.

We tested three groups of German native speakers who had been immersed in a Dutch-speaking environment for different amounts of time (about 3 months, about 10 months, or at least 18 months) and assessed how they

process Dutch word orders that are consistent or inconsistent with the analogous structure in German. We expect to see the developmental trajectory as described by Steinhauer et al. (2009) for the nonconflicting structure. For the conflicting structure, however, the progression through these stages may be different, for example, slower. It is, in fact, possible that, at least in the immersion period assessed here, learners will not reach native-like processing signatures at all. On top of that, it remains to be seen whether beginning learners are sensitive to the structures tested here at all. Given that the languages are typologically very close, we assume that learners will never be entirely insensitive but might already start with qualitatively different signatures (i.e., N400). For the structure with the direct conflict between L1 and L2 (double infinitives), another intriguing possibility exists: Learners may display a “reversed P600” elicited by direct transfer from their German (L1) grammar. That is, given that the correct Dutch word order in the structure we selected is ungrammatical in German and the other way around, it is conceivable that beginning learners show a positivity to the structure that is correct in Dutch (but would be incorrect in German) rather than to the usual positivity for the ungrammatical condition. Such reversed signatures have been reported before in grammatical gender processing in L2 (Lemhöfer et al., 2014). If beginning learners indeed start off with a reversed processing mode, it will be interesting to see how soon this strong L1 bias disappears and how exactly it changes into more target-like processing.

METHODS

Participants

Seventy-two German-speaking learners with various degrees of experience with Dutch and 24 native speakers of Dutch participated in this study. Two Dutch and 11 German native speakers had to be excluded either because of excessive EEG artifacts (two Dutch, one German) or because they had made too many grammaticality judgment errors (10 German participants; see Exclusion Criteria section for details). One additional German participant was excluded because of technical failure. The remaining 22 Dutch (age: $M = 23.00$, $SD = 2.78$, range = 19–29 years; 13 female) and 60 German participants (age: $M = 20.67$, $SD = 2.54$, range = 18–33 years; 51 women) were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological impairment or dyslexia. They provided informed consent and received either course credit or vouchers (10 €/hr) for their participation.

The German learners of Dutch had German as their only mother tongue and studied, or were about to study, at Radboud University in Nijmegen, the Netherlands. To be able to follow their studies in Dutch, they had all been required to pass a Dutch language exam before the start

of their first study year, for which they usually prepared by following a 6-week intensive Dutch summer course. The group of beginning learners was tested within the first months after this course and exam. The participants' knowledge of Dutch before this course was rudimentary at most. We divided our learners into three groups based on how long they had been studying and thus how long they had been immersed in Dutch: *beginning learners* ($n = 18$) had maximally 6 months of experience with Dutch, but most of them had only just finished the Dutch course. *Intermediate learners* ($n = 23$) had been immersed in the Dutch environment for 6–12 months, and *advanced learners* ($n = 19$) had been studying in the Netherlands for 18 months or longer. As assessed in a language background questionnaire, all of the learners reported knowledge of other foreign languages, including most prominently English (the first foreign language for all learners) and French, but intermediate and advanced learners listed Dutch as their currently most frequently used foreign language.¹ All other results from a language background questionnaire, including a quantification of how immersed participants were in the Dutch language (see Table 1 note for details), are summarized in Table 1 (the questionnaire itself can be inspected here: hdl.handle.net/11633/aabwx5zh).

Twelve of the 22 Dutch native speakers had learned German in high school. They rated their German very low ($M = 1.86$, $SD = 0.84$, range = 1–3; on a scale from 1 to 5), and all indicated that they hardly ever used German anymore.

Behavioral Session

Before participating in the EEG experiment, participants took part in a number of behavioral tests. For native speakers as well as intermediate and advanced learners, these tests were administered on a separate day (days before EEG session: $M = 8.3$, $SD = 3.9$, range = 2–24). For beginners, who were tested at a later point in time, the tests were administered immediately before the EEG measurements. These behavioral measures included a motivation questionnaire, a Dutch proficiency test, and the Dutch version of the LexTALE, a lexical decision-based vocabulary test (Lemhöfer & Broersma, 2012; for materials, visit this website: www.lextale.com/downloadthetest.html).

The motivation questionnaire consisted of 16 questions adapted from Gardner's attitude and motivation test battery, translated into German (Gardner, Clément, Smythe, & Smythe, 1979). The questions concerned German speakers' motivation to learn Dutch, their level of perfectionism and confidence, and their degree of perseverance and willingness to learn from mistakes. Higher scores indicate a higher degree of motivation to learn Dutch. The Dutch proficiency test was designed for the purpose of this study, that is, specifically for German learners of Dutch, and was based on two existing online proficiency tests by Transparent Language (<https://www.transparent.com/learn-dutch/proficiency-test.html>) and Klett (<https://einstufungstests.klett-sprachen.de/einstufung/index.php?questname=welkom&questclass=B1>). It

Table 1. Results from the Language Background Questionnaire Given to German Learners of Dutch

	<i>Beginners (n = 18)</i>			<i>Intermediate (n = 23)</i>			<i>Advanced (n = 19)</i>		
	<i>Mean</i>	<i>SD</i>	<i>Range</i>	<i>Mean</i>	<i>SD</i>	<i>Range</i>	<i>Mean</i>	<i>SD</i>	<i>Range</i>
Dutch experience (in months)	2.69	1.90	0–6	9.61	1.92	6–12	29.26	18.45	18–96
Immersion score ^a	0.26	0.14	0–0.33	0.30	0.20	0–0.67	0.38	0.23	0–1
Frequency of use self-ratings									
Reading	3.22	0.88	2–5	3.65	0.93	2–5	4.00	1.05	2–5
Speaking	3.44	0.92	2–5	3.57	0.99	2–5	4.21	1.03	2–5
Listening	3.89	0.90	2–5	4.39	0.66	3–5	4.63	0.60	3–5
Proficiency self-ratings									
Speaking	2.89	0.76	2–4	3.04	0.82	2–5	3.37	0.83	2–5
Listening	3.28	0.75	2–4	3.48	0.51	3–4	3.90	0.74	3–5
Writing	3.22	0.55	2–4	2.91	0.79	2–4	3.37	0.90	2–5
Reading	3.56	0.62	3–5	3.83	0.58	3–5	4.05	0.85	2–5

Self-ratings were given on a scale from 1 (*low/rarely*) to 5 (*high/very often*).

^aThe immersion score was based on whether or not a participant lived in the Netherlands (not all do, since Nijmegen is close to the German border), whether they had a job in the Netherlands, and whether they had a Dutch partner. The final immersion score was calculated as the percentage of positive (“yes”) answers to these questions.

mostly assessed knowledge of false friends and grammatical structures that are different in the two languages and thus provides a measure of how well participants are able to overcome the influence of their mother tongue. The test consisted of 48 multiple-choice questions concerning vocabulary (26) and grammar (22). Results from these tests can be inspected in Table 2.

EEG Session

Materials

During the EEG experiment, participants read two types of sentences: sentences containing a syntactic conflict in L1 and L2 with the double infinitives described above and sentences without such a conflict that were either grammatical in both languages or had verb placement errors in subordinate clauses and were thus ungrammatical in both L1 and L2 (see below for details, Table 3 for examples, and the Appendix for the full list of sentences). To ensure that participants understood the sentences content-wise, sentences only contained words that were either cognates (i.e., form-similar translations that are very common between the highly related languages Dutch and German) or high-frequency noncognates (average frequency of noncognates: 340 occurrences per million in Dutch according to the CELEX database; Baayen, Piepenbrock, & Gulikers, 1995). The proportion of cognates among the content words in our sentences was 76%, which is, given the proximity of the two languages, however, not unusually high: Although we are not aware of any official counts, our own counts of the 500 most frequent nouns and 500 most frequent verbs in Dutch in the CELEX database (Baayen et al., 1995) showed a similar proportion of cognates (namely, 75% and 68%, respectively).

Conflict Condition

Sentences in the conflict condition all followed the same structure as the example sentences in Table 3. EEG measurements were time-locked to the first infinitive in all sentences, because this is where a mistake in verb order can first be detected. This first infinitive could then either be a modal auxiliary or a content verb, depending on which order the sentence followed (grammatical vs. ungrammatical in Dutch). To avoid extensive repetition, each sentence contained a different content verb and one of five different modal auxiliary verbs: *kunnen* (“can”), *mogen* (“may”), *moeten* (“must”), *willen* (“want”), and *laten* (“let”).² With each of those five auxiliaries, 12 distinct sentences were created, resulting in a total of 60 conflict sentences. Sentences were between 7 and 12 words long ($M = 9.2$). Content verbs were between 4 and 11 letters long ($M = 7.12$, $SD = 1.68$), and modal verbs were between five and six letters long ($M = 5.6$, $SD = 0.55$). Participants saw equal numbers of grammatical (30) and ungrammatical (30) sentences. Grammatical and ungrammatical sentences

Table 2. Results of the Behavioral Tests Per Group

	Beginners			Intermediate			Advanced			Natives			Sign. Group Differences
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range	
LexTALE	61.91	5.36	47.50–68.75	66.09	7.57	55–80	70.26	10.34	42.50–83.75	93.68	3.62	85.00–98.75	b-a, b-n, i-n, a-n
Proficiency	73.22	7.37	59.09–84.27	75.43	8.17	61.36–87.76	83.13	7.88	69.06–92.31	—	—	—	b-a, i-a
Motivation	77.36	5.72	68.75–85.00	74.19	8.10	57.50–86.25	77.37	9.32	60.00–92.50	—	—	—	—

LexTALE, proficiency and motivation scores are expressed in %. Abbreviations in the last column indicate which groups differ statistically from one another (as determined in pairwise *t* tests, Games-Howell corrected): b = beginners; i = intermediate learners; a = advanced learners; n = native speakers.

Table 3. Examples for Conflict and No-conflict Sentences with German Literal and English Standard and Literal Translations

<i>Condition</i>	<i>Example Sentence</i>
Conflict–grammatical	<i>Ik heb het huiswerk voor maandag niet <u>kunnen</u> doen.</i> * <i>Ich habe die Hausaufgaben für Montag nicht können machen.</i> (I have not been able to do the homework for Monday.) [I have the homework for Monday not can do.]
Conflict–ungrammatical	* <i>Ik heb het huiswerk voor maandag niet <u>doen</u> kunnen.</i> <i>Ich habe die Hausaufgaben für Montag nicht machen können.</i> (I have not been able to do the homework for Monday.) [I have the homework for Monday not do can.]
No-conflict–grammatical	<i>Wij komen uit Berlijn maar wij <u>wonen</u> daar niet meer.</i> <i>Wir kommen aus Berlin, aber wir wohnen da nicht mehr.</i> (We're from Berlin but we no longer live there.) [We come from Berlin but we live there no more.]
No-conflict–ungrammatical	* <i>Wij komen uit Berlijn hoewel wij <u>wonen</u> daar niet meer.</i> * <i>Wir kommen aus Berlin obwohl wir wohnen da nicht mehr.</i> (We come from Berlin although we no longer live there.) [We come from Berlin although we live there no more.]
Fillers–grammatical	<i>Ik houd van de winter hoewel ik soms de sneeuw vervelend vind.</i> <i>Ich mag Winter obwohl ich manchmal den Schnee unangenehm finde.</i> (I like winter although I sometimes find the snow annoying.) [I like winter although I sometimes the snow annoying find.]
Fillers–ungrammatical	* <i>Ik houd van de winter maar ik soms de sneeuw vervelend vind.</i> * <i>Ich mag Winter aber ich manchmal den Schnee unangenehm finde.</i> (I like winter but sometimes I find the snow annoying.) [I like winter but I sometimes the snow annoying find.]

Target words are underlined. Asterisk (*) indicates ungrammatical sentences. Dutch sentences are on the first line, followed by literal German translations on the second line, English translations in round brackets, and literal translations into English in square brackets.

were always identical except for the order of the two final infinitives. Moreover, grammaticality was counterbalanced across participants, that is, participants saw each sentence only in one version (see List Construction section).

No-conflict Condition

For the no-conflict condition, we made use of coordinating and subordinating conjunctions, which differ in whether they require main clause or inverted verb order in the sentence they introduce. Both in German and in Dutch, subordinating conjunctions like *omdat*_{DUT}/*weil*_{GER} (“because”) require the verb to be placed in final position, whereas coordinating conjunctions like *want*_{DUT}/*denn*_{GER} (“thus”) require the normal (SVO) main clause word order. Grammatical sentences all contained coordinating conjunctions in combination with the correct SVO order. Ungram-

matical sentences were identical to these sentences, but contained a subordinating conjunction instead, which rendered the SVO order in the subordinate sentence ungrammatical (see Table 3). ERPs were measured on the finite verb, where the grammatical violation became apparent. Conveniently, both German and Dutch have pairs of (directly translatable) coordinating and subordinating conjunctions with identical or very similar meaning. We chose three such conjunction pairs and created an even number of sentences with each of them: *want-omdat* [“because”; in German: *denn-weil*] (20), *maar-hoewel* [“but/although”; in German: *aber-obwohl*] (20), and *dus-waardoor* [“thus/whereby”; in German: *also-wodurch*] (10).³ In total, participants saw 50 such sentences, half of which were grammatical and the other half ungrammatical. Sentences were between 7 and 12 words ($M = 10.0$) long. Target words (the finite verbs after the conjunction) were between two and eight letters long

($M = 4.84$, $SD = 1.60$) and fell between the sixth and the ninth word of the entire sentence ($M = 7.4$). As for the sentences in the conflict condition, grammaticality and sentence presentation order were counterbalanced across participants.

Filler Sentences

Given that type of conjunction was confounded with grammaticality in the no-conflict condition sentences (those with coordinating conjunctions were grammatical, those with subordinating conjunctions ungrammatical), we included filler sentences with the opposite grammaticality mapping: Ungrammatical filler sentences contained coordinating conjunctions followed by an inverted verb order, whereas grammatical filler sentences contained subordinating conjunctions (see Table 3). Participants saw 40 filler sentences (*maar-boewel* [16], *want-omdat* [16], *dus-waardoor* [8]), half of them grammatical and half of them ungrammatical. Sentence length was between 8 and 12 words ($M = 10.2$). EEG measurements for filler sentences were not analyzed.

List Construction

For counterbalancing, we constructed four experimental lists, each of which contained 150 sentences (60 conflict, 50 no-conflict, 40 filler sentences). Each sentence was read in its grammatical form by half of the participants, whereas the other half saw it in its ungrammatical form. Furthermore, to control for sentence order effects, trial orders in the two lists above were reversed, resulting in a third and fourth list. The order of sentences was pseudo-randomized (but fixed for a given list) with the restriction that no more than three grammatical (or ungrammatical) sentences and no more than two sentences with the same conjunction pair (in no-conflict sentences) or the same modal auxiliary verb (in conflict sentences) occurred in immediate succession.

Procedure

Participants were tested individually in a quiet room and were seated comfortably in a chair at a viewing distance of about 50 cm from a computer screen. Their task was to read each sentence carefully and to judge its grammaticality by pressing one of two buttons (left = ungrammatical, right = grammatical). Sentences were presented word by word in black letters (36 point Arial font) in the center of a gray screen using the experimental software Presentation (Version 18.1, Neurobehavioral Systems). Each new sentence was introduced by a 500-msec fixation cross. The words were subsequently presented for 450 msec each and were always preceded by a blank screen for 250 msec. Experimental trials were preceded by six practice trials (two example sentences from each condition). Participants did not get feedback during these practice trials; the purpose of the practice was merely to familiarize

participants with the task, the presentation pace, and the button assignment. The main experiment consisted of three blocks, each containing 50 sentences. After the experiment, German participants were additionally given a list of all sentences on paper and were asked to circle words that were unknown to them; the sentences containing unknown words in critical position or before were later excluded from the analyses for the respective participant (resulting mean percentages of excluded trials for each group for conflict and no-conflict conditions together: advanced: $M = 0\%$, $SD = 1$, range = 0–3%; intermediate: $M = 1\%$, $SD = 1$, range = 0–3%; beginners: $M = 2\%$, $SD = 3$, range = 0–8%). The complete EEG session took 2–2.5 hr for the German and 1.5–2 hr for the Dutch participants.

EEG Recording

Continuous EEG was recorded from 27 active Ag/AgCl electrodes embedded in an elastic cap (ActiCAP 32ch Standard-2, Brain Products) as well as from electrodes placed on the left and right mastoids and the forehead. The electrode on the forehead served as the ground. EEG was referenced online to the left mastoid and re-referenced offline to the averaged activity over both mastoids. Eye movements were recorded from a bipolar montage consisting of electrodes placed above and below the right eye and electrodes on the left and right temples. The EEG was amplified with a BrainAmp DC/MR plus amplifier, digitized with a 500-Hz sampling rate and filtered online with a high cutoff at 125 Hz and a low cutoff at 0.016 Hz. Impedances for EEG electrodes were kept below 10 k Ω and, for EOG electrodes, below 15 k Ω .

EEG Preprocessing

All offline EEG data preprocessing was done using the EEGLAB toolbox (Delorme & Makeig, 2004) in MATLAB (MathWorks). EEG signals were segmented into epochs from 200 msec before until 1000 msec after onset of each target word. The data were low-pass filtered at 30 Hz and high-pass filtered at 0.02 Hz. After segmentation, baseline correction was carried out based on the average EEG activity in the 200-msec interval before target onset. Subsequently, ocular correction was performed using the Gratton and Coles algorithm (Gratton, Coles, & Donchin, 1983) as implemented in EEGLAB. Finally, trials with amplitudes below $-100 \mu\text{V}$ or above $100 \mu\text{V}$, trials containing peak-to-peak activity greater than $150 \mu\text{V}$ (assessed in 200 msec windows every 10 msec), as well as trials with a voltage change of more than $75 \mu\text{V}$ between adjacent sample points or step-like activity greater than $30 \mu\text{V}$ within segments of 400 msec (assessed every 10 msec) were removed automatically (9.46% of all trials).

EEG Analysis

EEG data were analyzed using nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) as

implemented in Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011; calculations run with 2000 permutations). As mentioned in the Introduction, previous literature has shown that syntactic ERP effects in L2 learners vary a lot not only in terms of latency but also in polarity (e.g., N400 vs. P600). Conventional ERP analyses based on a predefined time window thus seemed inappropriate for our L2 data. Cluster-based permutation tests require fewer a priori choices and allow for a more data-driven analysis while still correcting for multiple comparisons (Maris & Oostenveld, 2007). Using such cluster-based permutation tests, we first tested for interaction effects of group (four levels: native speakers, beginners, intermediate and advanced learners) by grammaticality (two levels: ungrammatical and grammatical) for each of the two conditions (conflict and no-conflict) separately (critical alpha was set at $p = .05$). Significant interactions were followed up with pairwise comparisons, the alpha levels of which were corrected for multiple comparisons using the Games-Howell procedure (recommended for data with unequal sample sizes; see Ruxton & Beauchamp, 2008). To find out how exactly the groups' ERP signatures differed from each other, we furthermore conducted separate permutation tests per group (for details, see Results section). The latter were run as two-tailed tests to test for both positive and negative ERP components, for which the critical alpha level had to be adjusted to $p = .025$ (cf. Maris & Oostenveld, 2007).

RESULTS

Exclusion Criteria

As mentioned earlier, we excluded 2 Dutch and 11 German native speakers from analysis: Participants for whom artifact rejection resulted in cell sizes of less than 15 trials (two Dutch, two German participants), as well as participants whose d' score for the grammaticality judgments of either the conflict or the no-conflict sentences was below 1, indicating performance at or below chance level (nine German participants; Macmillan & Creelman, 2005). We reasoned that participants who were unable to discriminate grammatical from ungrammatical sentences would likely engage different processes during sentence reading compared with the other participants; however, the number of participants for whom this was true did not allow for a separate analysis of this group. In total, the above criteria resulted in the rejection of two native speakers, four beginning, five intermediate, and two advanced learners. Average d' scores for the final set of participants can be inspected in Figure 1.

Behavioral Results—Grammaticality Judgments during EEG

Mean error rates for grammaticality judgments of both the conflict and the no-conflict sentences can be inspected in

Figure 1. Mean d' scores for grammaticality judgments per group and condition with error bars. Error bars reflect the standard error of the condition means. Significant ($p = .05$) differences between groups (as determined via Games-Howell adjusted post hoc pairwise t tests) are marked with an asterisk.

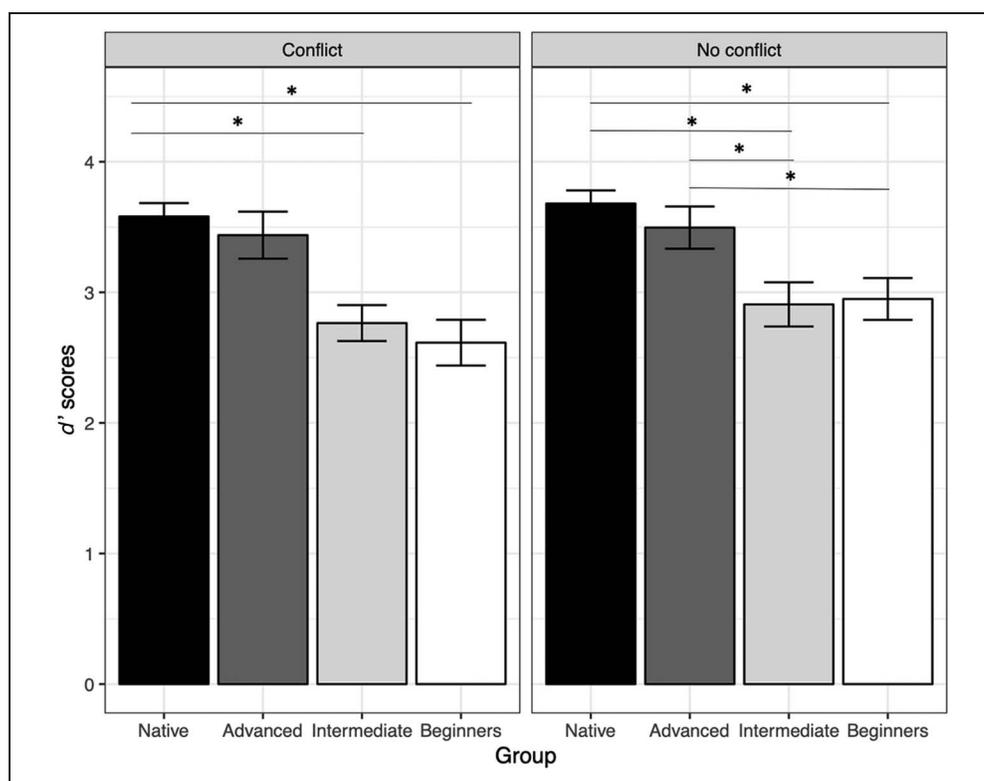


Table 4. Mean Grammaticality Judgment Error Rates (in %) Per Group, Condition, and Grammaticality Condition

	<i>Conflict Condition</i>		
	<i>Grammatical Sentences Mean (SE)</i>	<i>Ungrammatical Sentences Mean (SE)</i>	<i>Average (Over Grammaticality) Mean (SE)</i>
Natives	4.70 (1.09)	2.12 (0.52)	3.41 (0.64)
Advanced	1.40 (0.53)	10.53 (3.53)	5.97 (1.78)
Intermediate	4.06 (0.82)	17.25 (3.83)	10.65 (2.05)
Beginners	4.26 (1.14)	15.56 (3.75)	9.91 (1.90)
	<i>No Conflict Condition</i>		
	<i>Grammatical Sentences Mean (SD)</i>	<i>Ungrammatical Sentences Mean (SD)</i>	<i>Average (Over Grammaticality) Mean (SD)</i>
Natives	2.00 (0.73)	5.09 (1.06)	3.55 (0.72)
Advanced	4.84 (1.92)	6.74 (2.95)	5.79 (1.79)
Intermediate	10.78 (1.73)	10.61 (2.57)	10.70 (1.38)
Beginners	8.44 (1.96)	17.34 (4.71)	12.89 (2.29)

Table 4; the resulting d' scores per group and condition are shown in Figure 1. An ANOVA on d' scores with Group as a between-subject factor for each of the two conditions separately revealed significant main effects of Group in both conditions (conflict condition: $F(3, 78) = 6.89, p < .001, \eta_p^2 = .209$; no-conflict condition: $F(3, 78) = 10.45, p < .001, \eta_p^2 = .287$). Games-Howell adjusted pairwise comparisons revealed that, for both the conflict and the no-conflict sentences, the d' scores of Dutch native speakers were significantly higher compared with those of beginners (conflict: $t(29.41) = 3.86, p = .003$; no-conflict: $t(28.00) = 4.75, p < .001$) and intermediate learners (conflict: $t(35.71) = 3.92, p = .002$; no-conflict: $t(40.27) = 4.75, p < .001$), but not compared with d' scores of advanced learners ($ps > .7$).

The three learner groups did not significantly differ from each other in the conflict condition (beginners vs. intermediate: $p = .999$; though there was a trend for advanced learners to show higher d' scores than beginners, $p = .096$, and intermediate learners, $p = .073$). In the no-conflict condition, however, the advanced group had significantly higher d' scores than the beginners, $t(35.00) = 3.27, p = .012$, and intermediate learners, $t(35.32) = 2.97, p = .026$. The intermediate learners and beginners again did not differ from each other, $t(34.32) = 0.67, p = .907$. These results paint a clear developmental pattern: while advanced learners are already on a native-like level in judging sentences for grammaticality, beginning learners as well as intermediate learners are still struggling, equally so for conflict and no-conflict sentences.

EEG Results

EEG analyses and grand averages are based on all trials, independent of whether or not participants made errors in their grammaticality judgments, as advocated by VanRullen (2011).

No-conflict Condition

The grand-averaged waveforms for the no-conflict condition for all four groups can be found in Figure 2. A cluster-based permutation test on the entire data set revealed significant main effects of Grammaticality (first marginally significant negative cluster: 266–432 msec over the whole scalp, $p = .041$; second marginally significant negative cluster: 738–1000 msec over frontal electrodes only, $p = .026$; positive cluster: 494–1000 msec over centroposterior electrodes, $p = .001$) and Group (positive cluster: 154–356 msec over entire scalp, $p = .016$), as well as, most importantly, a marginally significant interaction between the two (positive cluster: 846–956 msec, over centroposterior electrodes, $p = .070^4$). To further explore the nature of this (albeit only marginally significant) interaction, we calculated difference scores per participant, based on the average activity of grammatical minus ungrammatical sentences for all channel–time pairs involved in the two clusters from the interaction analysis. Subsequent Games-Howell corrected pairwise t tests on these difference scores revealed that intermediate and advanced learners' signatures did not differ from those of native speakers or from each other in the interaction window (all $ps = 1.00$), whereas beginners' ERPs did, both

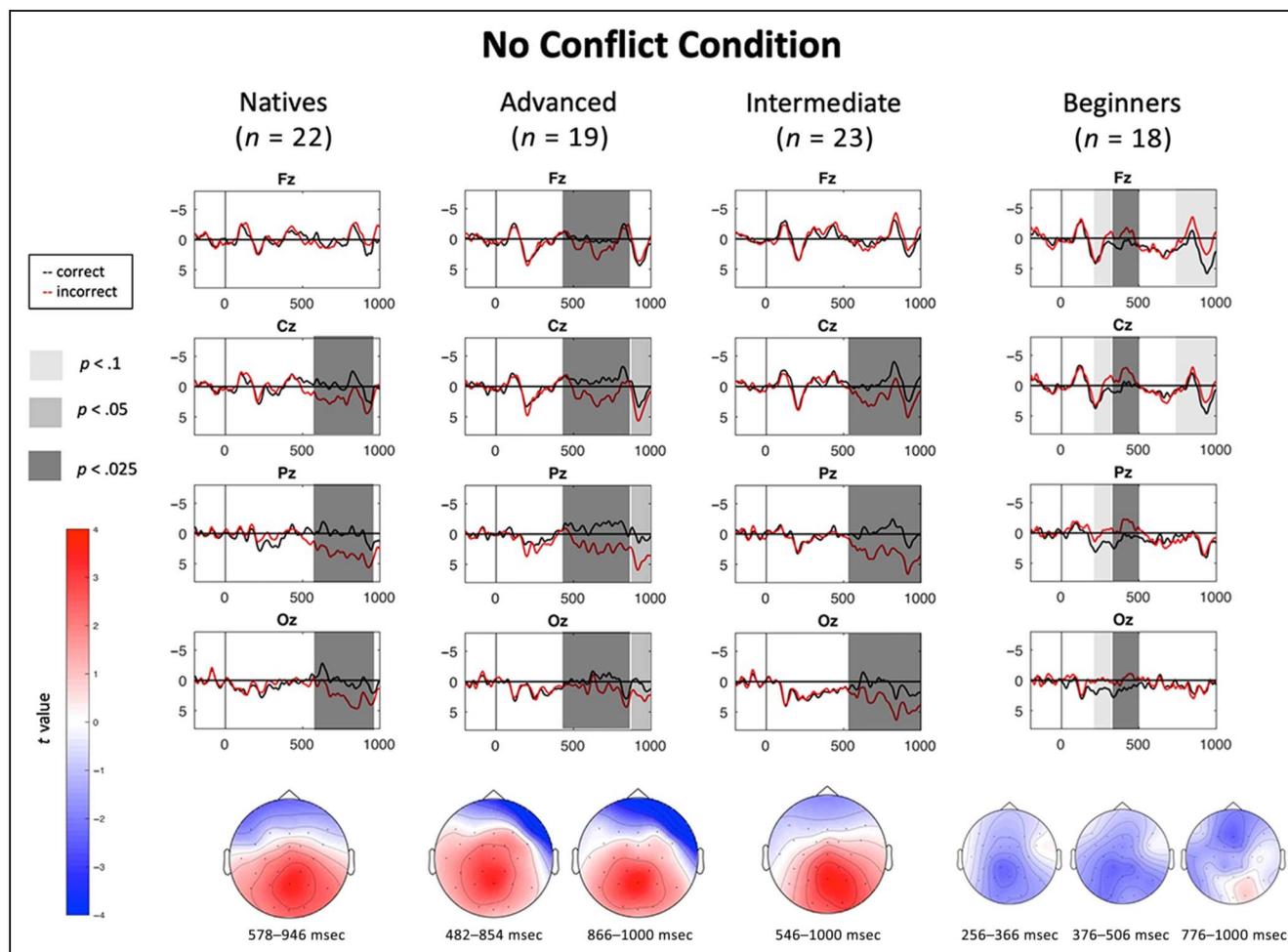


Figure 2. Grand-averaged ERP waveforms for the no-conflict condition for all groups. Clusters revealed by the permutation tests are marked in shades of gray, depending on the significance level of the cluster statistic. For each cluster, a topographic plot is included at the bottom, showing the respective scalp topography.

from native speakers, $t(35.59) = 3.25$, $p = .013$, as well as from intermediate, $t(33.25) = 3.46$, $p = .008$, and advanced learners, $t(35.00) = 3.06$, $p = .021$.

Based on these tests, however, we cannot make any claims as to the polarity and latency of individual group signatures. To get a better grasp of how the individual groups react to grammatical violations in the no-conflict sentences, we subsequently conducted separate permutation tests per group looking for effects of grammaticality over the entire time window (0–1000 msec).⁵

These additional permutation tests revealed a large positive cluster for incorrect compared with correct verb placement in subordinate clauses for Dutch native speakers from 578 to 946 msec over central and posterior electrode sites ($p = .002$). In terms of latency and scalp distribution, this corresponds to a classical P600 effect to grammatical violations (Hagoort, Brown, & Groothusen, 1993). Similarly, advanced (positive cluster: 482–854 msec, $p = .002$; marginally significant positive cluster: 866–998 msec, $p = .030$) and intermediate learners (544–1000 msec, $p = .001$) also showed large P600 effects in response to violations of verb placement rules

in subordinate clauses. The effect was most prominent over centroposterior electrode sites in both these learner groups. The ERP signal of beginning L2 learners, instead, was dominated by a widely distributed negativity: Permutation tests revealed a marginally significant negative cluster over the whole scalp from 256 to 366 msec ($p = .068$), followed by a significant negative cluster from 376 to 506 msec ($p = .020$, over the whole scalp), as well as another marginally significant negative cluster from 776 to 1000 msec ($p = .059$, over centroposterior electrodes).⁶

Overall, violations of the inversion rule—a structure implemented similarly in German and Dutch—elicited native-like online processing signatures in the two more advanced learner groups: Both advanced and intermediate learners showed robust P600 effects that were comparable to those in native speakers in terms of latency, size, and scalp distribution. Beginners instead were still struggling with the no-conflict sentences, as manifested in a prolonged negativity, a qualitatively very different ERP signature compared with natives and more advanced learners.

Conflict Condition

Grand averages for all groups for the conflict condition are displayed in Figure 3. Permutation tests on the whole data set again revealed significant main effects of Grammaticality (negative cluster: 250–532 msec over the entire scalp, $p = .013$; positive cluster: 394–1000 msec over the whole scalp, $p = .001$) and Group (first positive cluster: 136–486 msec over the entire scalp, $p = .006$; second positive cluster: 524–1000 msec over the entire scalp, $p = .001$), as well as a significant interaction of Group and Grammaticality (positive cluster: 422–868 msec, $p = .001$, distributed over the entire scalp), which was followed up in the same way as for the no-conflict sentences. Games-Howell corrected pairwise t tests revealed that, in contrast to the no-conflict sentences, all learner groups differed in that time window from the native speakers in their ERP signatures to violations of infinitive verb order (advanced vs. native speakers: $t(32.04) = 3.59, p = .006$; intermediate vs. native speakers: $t(42.68) = 5.02, p < .001$; beginners vs. native speakers: $t(32.51) = 5.45, p < .001$). Learner

groups themselves, instead, did not differ from each other in the interaction window (all $ps > .50$).

Separate permutations per group over the whole time window revealed that Dutch native speakers showed a robust P600 (cluster from 504 to 976 msec, $p < .001$), spread over the whole scalp. Advanced and intermediate learners also showed P600s, though much later, less robust, and much more restricted in terms of topography (advanced learners: marginally significant positive cluster from 780 to 1000 msec over posterior and right hemisphere electrodes only, $p = .036$; intermediate learners: positive cluster from 822 to 1000 msec over centroposterior electrodes only, $p = .018$). Intermediate learners additionally showed a trend toward an earlier negative cluster (300–462 msec, $p = .098$, over left-central posterior electrodes), though this effect is statistically very weak and should not be overinterpreted. For the beginners, similar to the no-conflict condition, permutation tests revealed two negative clusters: a marginally significant cluster from 312 to 450 msec (whole scalp, $p = .045$; remember that $\alpha = .025$) and another marginally significant cluster from 486 to 620 msec over the whole scalp

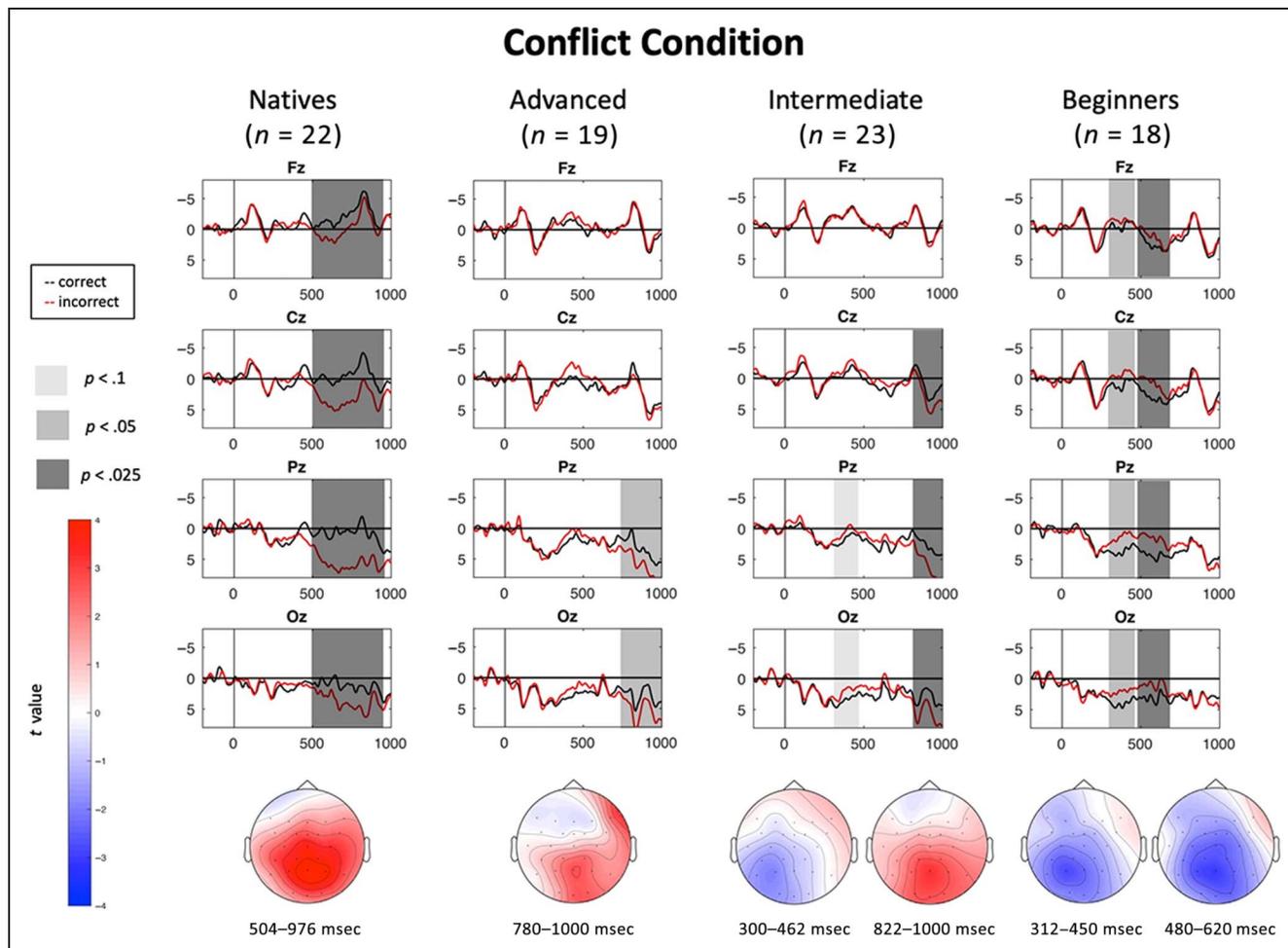


Figure 3. Grand-averaged ERP waveforms for the conflict condition for all groups. Clusters revealed by the permutation tests are marked in shades of gray, depending on the significance level of the cluster statistic. For each cluster, a topographic plot is included at the bottom, showing the respective scalp topography.

Direct Comparison of Conflict and No Conflict Condition

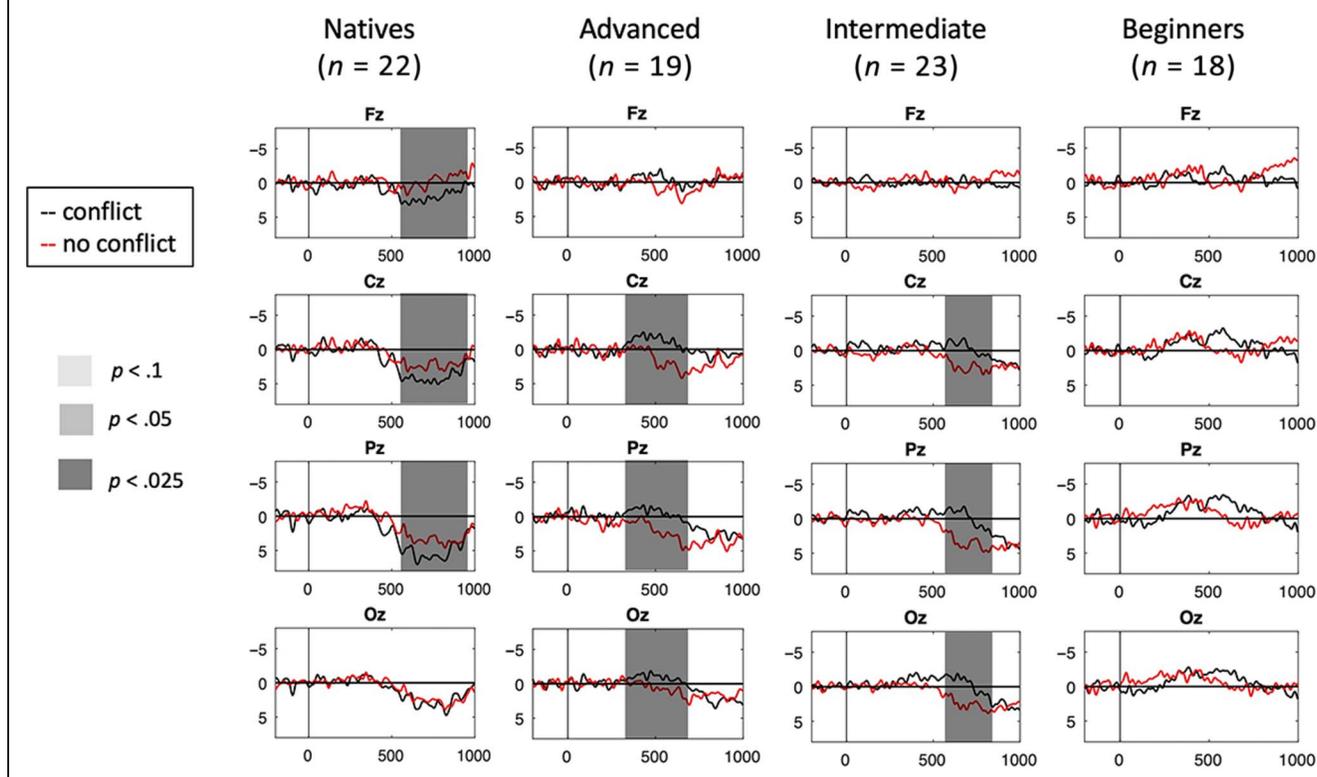


Figure 4. Direct comparison of conflict and no-conflict conditions for native speakers and all learner groups. The black lines represent grand-averaged differences waves (ungrammatical–grammatical) for the conflict sentences, and the red lines reflect grand-averaged difference waves for the no-conflict condition.

($p = .030$). Based on visual inspection, we interpret these two clusters as one, and just as for the no-conflict sentences, we interpret these negativities to ungrammatical conflict-containing sentences as N400-like effects.

To sum up, learners at all stages of L2 acquisition showed nonnative ERP signatures to violations of infinitive verb order in Dutch, as revealed by the pairwise t tests reported above. Sentences containing conflicting syntactic rules in Dutch and German thus pose difficulties even to advanced learners, despite native-like behavioral performance for that group. Over the interaction time window, learner groups did not differ statistically from each other. Separate permutation tests, though, point to clear differences between learner groups' signatures: Whereas the advanced and intermediate learners already show an (albeit delayed and for the advanced learners only marginally significant) P600, this P600 is entirely absent in beginners, who only show a prolonged N400.

Comparing Conflict and No-conflict Conditions

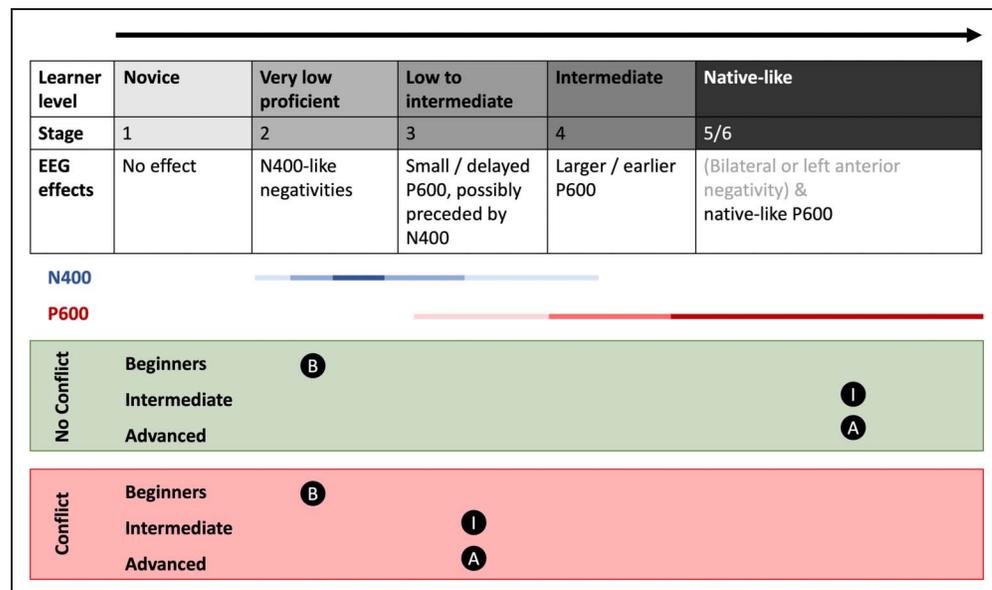
Because of the properties of Dutch and German, it was impossible to find two exactly parallel structures that differ only in whether they are implemented in the same or a conflicting manner in the two languages. Our conflict-

ing structure was a double infinitive construction, whereas the non-conflicting structure concerned the placement of a single finite verb in a subordinate clause. It is thus possible that violations of those structures are processed differently and thus give rise to different ERP signatures already in native speakers. To investigate this issue, we analyzed whether the two structures give rise to similarly sized P600 effects in native speakers. We calculated difference waves (ungrammatical – grammatical sentences) for the two structures for the native speakers, as well as for comparison and completeness, for each of the learner groups. The results are shown in Figure 4. Native speakers indeed differed in how they processed the two structures: The P600 effect in the conflict (double infinitive) condition was larger than in the (subordinate clause) no-conflict condition. Importantly though, this effect goes in the opposite direction to what we see for (intermediate and advanced) learners, who showed smaller P600s for the conflict than for the no-conflict condition. Thus, the differences between sentence types for learners cannot be explained by baseline difference regarding how the two structures are processed by native speakers, something which was already implicated in the comparison of learner groups with native speakers in the main analysis above.

DISCUSSION

This study examined the role of L1–L2 similarity in the development of syntactic processing ability. More specifically, we investigated how German learners of Dutch deal with a directly conflicting syntactic rule in their first and second language, as compared with a rule that is implemented similarly in both languages. Furthermore, we asked how online processing signatures for these two different instances of foreign grammar develop with increasing L2 proficiency. To that end, we tested three groups of German native speakers with differing amounts of experience with Dutch: Beginners had been immersed in Dutch for an average of 3 months, intermediate learners for an average of 10 months, and advanced learners for at least 18 months. Learners' ERP signatures to sentences containing violations of infinitive verb order (conflicting structure) were compared with those for violations of verb placement in subordinate clauses (not conflicting structure). For both structures, learners' signatures were also compared with those of Dutch native speakers. In line with Steinhauer et al.'s (2009) developmental model of L2 syntactic processing, we expected learners at different levels of L2 proficiency or experience to differ in their EEG signatures to syntactic violations: Beginning learners should (after acquiring sensitivity to the syntactic feature at all) start out with qualitatively different ERP effects in the form of N400-like negativities, whereas advanced learners should show beginnings of a P600 that should become more native-like with rising degrees of L2 experience and possibly be preceded by a LAN. For the sentences with the direct conflict, we assumed that this process would be altered, that is, slowed down and delayed as compared with the no-conflict sentences, and that possibly beginning learners would start out with reversed processing signatures driven by L1 parsing strategies.

Figure 5. Stages of L2 syntactic development (adapted from Steinhauer et al., 2009) and schematic representation of this study's results.



Changes in ERPs as a Function of L2 Experience and L1–L2 Similarity

As we had expected and as illustrated in Figure 4, the processing of the non-conflicting structure largely developed in the stages proposed by Steinhauer et al. (2009): Beginning learners showed a prolonged N400-like negativity, whereas already from 6 months of L2 experience onward, learners showed robust P600 effects that are comparable in scalp topography and magnitude to the pattern observed for Dutch native speakers. In contrast, for the conflicting sentences with double infinitives, all three learner groups showed clear signs of an ongoing struggle with the conflict between German and Dutch word order. Beginners again showed a prolonged negativity, whereas intermediate and advanced learners both displayed delayed, smaller, and thus less native-like P600 signatures than in the no-conflict condition, preceded by a trend toward an N400. Behaviorally, only the advanced learners were on a native-like level in judging sentences from both conditions for grammaticality.

Our results paint a clear developmental pattern. In terms of Steinhauer et al.'s (2009) model for L2 syntactic processing, our beginning learners can be placed at Stage 2 for both conflict and no-conflict sentences (see Figure 5). This stage is characterized by N400-like effects, the absence of a P600 and non-native behavioral performance. Both intermediate and advanced learners are located at Stage 3 for the conflicting sentences as evidenced by a delayed P600 preceded by small N400 effects. However, these same two learner groups have already advanced to Stage 5/6 for the no-conflict sentences, for which they show native-like P600s (see Figure 5). These EEG signatures go together with native-like behavioral performance for the advanced learners, but higher and thus not native-like error rates in intermediate learners. We did not observe LANs in our study, neither in the learners nor in the natives.

LANs are—if at all—usually found for morphosyntactic violations, which are of a very different type than the word order violations in our study (note that even for morphosyntactic violations LANs are not always reported though; cf. Caffarra et al., 2015; Steinhauer & Drury, 2012). Given this lack of LAN effects in the native reference group, we had no reason to distinguish between Stages 5 and 6.

Steinhauer et al. (2009) describe the developmental stages as a function of what they call “general proficiency” in the foreign language. In our case, following the tradition of (cross-sectional) developmental studies, we divided the German learners into groups based on how long they have had experience with Dutch. Our results thus show a developmental pattern across these different levels of experience with Dutch and not necessarily general Dutch proficiency. This is important to note, because developmental studies on syntactic processing that compare learners at different learning stages often define these stages in quite different ways: according to amount of experience with the L2 (e.g., Bowden et al., 2013; Osterhout et al., 2006), according to performance on a standardized proficiency test (e.g., Alemán Bañón et al., 2018), or even based on a mix of self-reported proficiency and test performance (e.g., Rossi et al., 2006). These differences in definitions are often overlooked, even though they are important for the interpretation of the results. Because we chose to sample Dutch learners at different time points during their acquisition period, including the presumably earliest one where fluent L2 sentence reading can be expected (beginners, tested after their introductory L2 intensive course), we opted for experience level as a grouping criterion, as is often done in (cross-sectional or longitudinal) developmental learning studies (e.g., Bowden et al., 2013; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Osterhout et al., 2006; Flege & Liu, 2001). Also, it should be noted that the proficiency tests we took (a self-made, not validated proficiency test covering different linguistic domains that are typically difficult for German learners, and a vocabulary size test for advanced L2 speakers) were collected as secondary measures to provide basic information on the participants’ proficiency but do not seem appropriate as grouping criteria. Accordingly, in the critical conflict condition in our study, the length of experience with Dutch (in months) was a better predictor of d' scores ($r_s = .32, p = .012$) than was the score on either of the two tests (proficiency test: $r_s = .20, p = .129$; LexTALE: $r_s = .21, p = .117$). Nevertheless, our groups defined by level of experience of course also differed in average proficiency (see Table 3). Future research seems necessary for developmental L2 theories and models such as that by Steinhauer et al. (2009) to specify whether syntactic development is primarily driven by increases in proficiency or by increases in amount of experience (cf. Alemán Bañón et al., 2018).

The present pattern of results, next to providing empirical evidence for the developmental stages proposed

by Steinhauer et al. (2009), clearly shows that progression through these different stages is not solely dependent on experience (or proficiency) but is crucially modulated by L1–L2 similarity. To our knowledge, we are among the first to do so within one single study and one general (and hence comparable) adult learner population (though see Alemán Bañón et al., 2018; Osterhout et al., 2006) and, importantly, involving a structure that is not only different in L1 and L2 but in fact in direct conflict in the two languages. Although native-like processing is possible for structures that are implemented similarly in L1 and L2 after less than a year of L2 immersion, the very same learners are still struggling with processing structures where L1 rules are in direct conflict with L2 rules. The progression from qualitatively different signatures to native-like P600s for syntactic violations is thus slower for cases where L1 and L2 syntax are at odds with each other. Beginning learners, in contrast, seem to employ different neurocognitive mechanisms in L2 syntactic processing regardless of L1–L2 structural similarity.

The developmental stages we observe in our study and the online processing signatures associated with them correspond and map well onto previous research findings from, on the one hand, studies comparing structures of different L1–L2 similarity within learners at one specific stage of L2 proficiency (Alemán Bañón et al., 2014; Lemhöfer et al., 2014; Foucart & Frenck-Mestre, 2011; Zawiszewski et al., 2011; Kotz et al., 2008) and, on the other hand, from studies investigating one specific type(s) of structure(s) over time, either cross-sectionally (Bowden et al., 2013; Rossi et al., 2006; Ojima, Nakata, & Kakigi, 2005) or (though rare) longitudinally (White et al., 2012; McLaughlin, Osterhout, & Kim, 2004). We will subsequently go through each of the developmental stages in turn and discuss how the signatures we observed fit with previous studies and in how far they add to them.

N400-like Signatures as Markers of Low L2 Proficiency

We see N400 effects in both conditions for the beginning learners in our study. Peaking between 300 and 500 msec after stimulus onset with a centroposterior scalp distribution, these signatures correspond to canonical N400 effects, as they are typically observed in native speakers in response to semantic violations (Kutas & Federmeier, 2011). In the no-conflict condition, the effect starts very early on mostly over posterior sites but later develops into a typical N400. Although such an early beginning is unusual for native speakers, it is not uncommon in studies on L2 syntactic processing (McLaughlin et al., 2010; Osterhout et al., 2006). In general, N400 responses to syntactic violations are a common finding in the L2 literature (Tanner, Inoue, & Osterhout, 2014; McLaughlin et al., 2010; Chen et al., 2007). Given the typical finding of an

N400 to semantic violations in natives, these N400 effects have been interpreted to indicate greater involvement of the lexico-semantic system in the early stages of L2 acquisition (Steinhauer et al., 2009; Ullman, 2001). Rather than decomposing word sequences morphosyntactically as native speakers and more advanced learners would do, beginners are thought to memorize individual word sequences. Violations to, for example, word order or subject–verb agreement then result in surprising, “novel” word sequences, which consequently trigger N400s rather than signatures of syntactic reanalysis and repair, such as the P600 (Steinhauer et al., 2009).

The observed N400s in our beginning learners are clear evidence for the fact that these participants were sensitive to the grammatical violations in both conditions (albeit in a different way than native speakers and more advanced learners). We thus do not find evidence for “Stage 1” in Steinhauer et al.’s (2009) model, characterized by complete insensitivity to grammar at the very beginning of L2 learning. It is possible that the high typological proximity of Dutch and German facilitates the early stages of learning and gives German learners of Dutch (and vice versa) a head start, allowing them to skip this initial phase. However, given that our beginning learners had already been in intensive contact with Dutch for, on average, about 3 months and passed a language exam, it is questionable to what extent they really represent the very early stages of learning. We might have found evidence for Stage 1 had we tested participants at the beginning of their Dutch course. However, our aim to test the different groups with identical materials required the ability to routinely read Dutch sentences of average difficulty, that is, a certain minimal level of word and syntactic knowledge, a prerequisite that would not have been met at such an early stage in learning.

For the conflict condition, we alternatively hypothesized that beginners may start out with opposite parsing strategies, relying on L1 syntactic knowledge, and thus showing a P600 to grammatical instead of ungrammatical Dutch sentences or, put differently, a reversed P600 to violations of Dutch infinitive verb order. Direct conflicts for which the correct L2 word order matches the incorrect L1 word order and vice versa have received surprisingly little attention in the L2 processing literature, despite the fact that they may be especially difficult for learners to acquire. Learners do not only have to learn the L2 rule but need to simultaneously inhibit their L1, which may fail in the early stages of L2 learning, thus leading to reversed processing signatures. Although such reversed processing signatures have been reported before for cross-linguistically incompatible grammatical gender (Bultena, Danielmeier, Bekkering, & Lemhöfer, 2017; Lemhöfer et al., 2014) and word order (Erdocia & Laka, 2018), we did not find evidence for such a reversed P600 effect. The L1 bias was thus either not strong enough or had already been overcome. The latter seems most

likely, especially given the low behavioral error rates in the current study (less than 10% in the conflict condition).

What does remain puzzling is the late anterior negativity (starting at about 770 msec after target word onset) that we observed in beginners, but only in the no-conflict condition. Such signatures are not common in syntactic processing studies but have occasionally been reported for both native speakers (Sabourin & Stowe, 2008) and advanced L2 learners (Gillon Dowens, Guo, Guo, Barber, & Carreiras, 2011; Chen et al., 2007; Hahne, 2001). Post hoc interpretations for these unexpected negativities usually attribute their presence to increased working memory demands. Whatever the functional significance of this late component may be, it lends further support to the fact that beginners were, despite their low proficiency, sensitive to the syntactic violations, both in the conflict and no-conflict conditions, just in a different way than native speakers and more advanced learners. Overall, it seems that beginning learners employ generally different neurocognitive mechanisms in L2 syntactic processing than native speakers or advanced learners do, regardless of L1–L2 structural similarity.

From Delayed P600 Effects to Native-like Processing

Intermediate and advanced learners clearly distinguished themselves from the beginners. In both these learner groups, we already saw fully fledged native-like P600 effects to word order violations in the no-conflict condition and delayed P600s in the conflict condition. Our results are in line with earlier studies, which showed that native-like processing is possible for structures that are implemented similarly in L1 and L2 (Foucart & Frenck-Mestre, 2011; Hopp, 2010), in our case after on average only 10 months of L2 immersion. Likewise, delayed P600 responses have been reported before in structures that are unique to the L2 or at least implemented differently (Hahne, 2001). The delay in the P600 effect is usually taken as evidence that it takes learners longer to realize the mistake in word order and to initiate syntactic reanalysis/repair processes. It remains unclear from our study whether what we see in advanced (and already in intermediate) learners reflects the end stage of acquisition, that is, the ultimate attainment that is possible for conflicting structures. It is possible that direct syntactic L1–L2 conflicts always remain problematic, regardless of how long people have been immersed in the L2, especially when the languages are otherwise so similar and, in most other instances, lend themselves well to transfer. The delayed P600 in the conflict condition in the intermediate learner group was furthermore preceded by a small N400 signature. Given that it was not statistically reliable though, one should be cautious with interpreting this effect. Truly (i.e., statistically robust) biphasic N400–P600 patterns have been reported for L2 syntactic processing before (for subject–verb agreement violations: Tanner et al., 2014; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013).

Tanner et al. (2013, 2014), however, showed that these group averages often misrepresent participants' individual signatures, which in their studies tended toward either an N400 or a P600, rather than both. The authors linked differences in N400–P600 dominance to proficiency and motivation to learn the L2. It is possible that this is also true for our study: Considerable individual variation may underlie the lack of statistical power to reliably detect the N400 effect on the grand average level. In fact, a look at individual N400 and P600 effect magnitudes for our intermediate learners in the conflict condition indeed shows a significant negative correlation between the two components ($r = -.42, p = .046$).⁸ This negative correlation suggests that, on average, one signature tended to dominate over the other, that is, that a strong N400 was associated to a weak P600 and vice versa and that the biphasic pattern we observed was indeed mainly a result of averaging across these two dominance patterns. The rather small group sample size unfortunately does not allow to relate each individual's N400/P600 dominance to individual differences beyond level of L2 experience, but this would certainly be a promising direction for future research.

Behavioral versus Neural Sensitivity

Finally, we would like to briefly discuss similarities and differences between the behavioral and the electrophysiological patterns of results. In error rates, at least descriptively, intermediate learners patterned with beginning learners, both making significantly more errors than native speakers. Advanced learners stood out behaviorally as they were statistically indistinguishable from native speakers. On the neural level though, this pattern was different: Here, advanced and intermediate learners patterned together (delayed P600 in conflict condition, native-like P600 in no-conflict condition) and clearly distinguished themselves from the beginners (N400 in both conditions). In other words, although behaviorally the

change to native-like processing only seems to occur after about 18 months of immersion, a clear qualitative shift (toward native-like or at least P600-dominated signatures, depending on L1-conflict) appears to be happening after only 10 months of immersion on the neural level. Possibly, the neural shift is necessary to enable the behavioral performance improvement. It is not unheard of for neural sensitivity to precede behavioral sensitivity during the course of L2 acquisition (McLaughlin et al., 2004); after all, the EEG is a much more sensitive measure than behavioral grammaticality judgments, especially if the latter are delayed as in the current study. Note though that this is at odds with the much more common finding that neural processing lags behind in L2 learners despite native-like behavioral performance (e.g., Chen et al., 2007; Ojima et al., 2005; see also the results for our advanced learners in the conflict condition: nonnative ERPs despite native-like behavioral accuracy).

Conclusions

The current study is among the first to adopt a developmental approach toward the processing of syntactic structures that are not only implemented differently but for which L2 syntactic rules are in direct conflict with L1. Our results show that such difficult instances of foreign grammar remain difficult to process and are an obstacle to native-like processing even in advanced L2 learners, when behavioral performance is already native-like. Importantly, the very same learners were already at native-like levels in processing a structure that is compatible in L1 and L2. The developmental trajectory of online processing signatures, traditionally believed to be determined by L2 proficiency, is thus significantly influenced and modulated by such direct L1–L2 conflicts. Finally, beginning learners, in contrast, seem to employ different neurocognitive mechanisms in L2 syntactic processing regardless of L1–L2 structural similarity.

Appendix. Experimental Sentences Split by Condition

Conflict Condition

Ik heb mijn vriendin vanwege de harde muziek niet kunnen verstaan/*verstaan kunnen.

*I have my friend because of the hard music not can understand/*understand can.*

Ik heb mijn nieuwe rok vandaag voor het eerst kunnen dragen/*dragen kunnen.

*I have my new skirt today for the first time can wear/*wear can.*

De politie heeft gisteren twee inbrekers kunnen pakken/*pakken kunnen.

*The police has yesterday two burglars can catch/*catch can.*

De vrouw heeft haar zoontje met zijn huiswerk kunnen helpen/*helpen kunnen.

*The woman has her son with his homework can help/*help can.*

Mijn ouders hebben in het weekend hun probleem kunnen oplossen/*oplossen kunnen.

*My parents have in the weekend their problems can solve/*solve can.*

Onze kinderen hebben een goede opleiding kunnen genieten/*genieten kunnen.

*Our children have a good education can enjoy/*enjoy can.*

Ik heb geen taart voor zijn verjaardag kunnen bakken/*bakken kunnen.

*I have no cake for his birthday can bake/*bake can.*

Ik heb het huiswerk voor maandag niet kunnen doen/*doen kunnen.

*I have the homework for Monday not can do/*do can.*

Hij heeft de maaltijd wegens zijn allergie niet kunnen eten/*eten kunnen.

*He has the meal because of his allergy not can eat/*eat can.*

Appendix. (continued)

Conflict Condition

Elena heeft haar geheim niet kunnen verbergen/*verbergen kunnen.

*Elena has her secret not can hide/*hide can.*

De kinderen hebben het gesprek van de ouders niet kunnen volgen/*volgen kunnen.

*The children have the discussion of the parents not can follow/*follow can.*

Wij hebben de sleutels van het huis niet kunnen vinden/*vinden kunnen.

*We have the keys to the house not can find/*find can.*

Ik heb mijn jas vandaag op het werk laten liggen/*liggen laten.

*I have my jacket today at work let lie (=to lose)/*lie let.*

Ik heb mijn mooiste vaas per ongeluk laten vallen/*vallen laten.

*I have my prettiest vase accidentally let fall/*fall let.*

Karljin heeft haar zoontje de nieuwe laptop laten gebruiken/*gebruiken laten.

*Karljin has her son the new laptop let use/*use let.*

Mijn baas heeft me dit jaar meer vakantie laten nemen/*nemen laten.

*My boss has me this year more vacation let take/*take let.*

Maartens ouders hebben hem aan de excursie laten deelnemen/*deelnemen laten.

*Maarten's parents have him in the excursion let participate/*participate let.*

De kinderen hebben de hond in het bos vrij laten lopen/*lopen laten.

*The children have the dog in the forest free let walk/*walk let.*

Ik heb de natte kleren niet lang genoeg laten drogen/*drogen laten.

*I have the wet clothes not long enough let dry/*dry let.*

Appendix. (continued)

Conflict Condition

Ik heb de kinderen niet alleen in de zee laten zwemmen/*zwemmen laten.

*I have the children not alone in the sea let swim/*swim let.*

De bankier heeft de dief niet laten ontkomen/*ontkomen laten.

*The banker has the thief not let escape/*escape let.*

Jordy heeft zijn zoontje het bier natuurlijk niet laten proeven/*proeven laten.

*Jordy has his son the beer of course not let try/*try let.*

De andere kinderen hebben de kleine Luuk niet laten meespelen/*meespelen laten.

*The other children have small Luuk not let join in playing/*join let in playing.*

De ouders hebben hun dochter gisteren niet laten uitgaan/*uitgaan laten.

*The parents have their daughter yesterday not let go out/*go out let.*

Ik heb in de vakantie een bijbaantje moeten zoeken/*zoeken moeten.

*I have during the vacation a minijob must search/*search must.*

Ik heb mijn hele vakantie met rugpijn moeten doorbrengen/*doorbrengen moeten.

*I have my entire vacation with back pain must spend/*spend must.*

De docent heeft de studenten altijd moeten corrigeren/*corrigeren moeten.

*The lecturer has the students always must correct/*correct must.*

Marlijn heeft na de operatie veel pillen moeten slikken/*slikken moeten.

*Marlijn has after the operation a lot of pills must swallow/*swallow must.*

Wij hebben in de hete zomer veel water moeten drinken/*drinken moeten.

*We have in the hot summer lots of water must drink/*drink must.*

Appendix. (continued)

Conflict Condition

De kinderen hebben hun speelgoed met elkaar moeten delen/*delen moeten.

*The children have their toys with each other must share/*share must.*

Ik heb mijn fiets vandaag voor het eerst moeten repareren/*repareren moeten.

*I have my bike today for the first time must repair/*repair must.*

Ik heb mijn scriptie in het Nederlands moeten schrijven/*schrijven moeten.

*I have my thesis in Dutch must write/*write must.*

Marijke heeft vandaag de was moeten doen/*doen moeten.

*Marijke has today the laundry must do/*do must.*

Guido heeft het tentamen helaas moeten overdoen/*overdoen moeten.

*Guido has the exam unfortunately must redo/*redo must.*

De bewoners hebben deze maand een hogere huur moeten betalen/*betalen moeten.

*The residents have this month a higher rent must pay/*pay must.*

Wij hebben de conferentie dit jaar weer moeten plannen/*plannen moeten.

*We have the conference this year again must plan/*plan must.*

Ik heb de presentatie met mijn beste vriend mogen maken/*maken mogen.

*I have the presentation with my best friend may make/*make may.*

Ik heb een nieuwe keuken voor ons huis mogen uitzoeken/*uitzoeken mogen.

*I have a new kitchen for our house may choose/*choose may.*

De firma heeft veel goede reacties mogen ontvangen/*ontvangen mogen.

*The company has many good reactions may receive/*receive may.*

Appendix. (continued)

Conflict Condition

Bart heeft al voor zijn zestiende verjaardag bier mogen drinken/*drinken mogen.

*Bart has already before his sixteenth birthday beer may drink/*drink may.*

De kinderen hebben de hele dag televisie mogen kijken/*kijken mogen.

*The children have the whole day television may watch/*watch may.*

De mensen hebben gisteren een nieuwe regering mogen kiezen/*kiezen mogen.

*The people have yesterday a new government may elect/*elect may.*

Ik heb mijn cadeaus niet voor mijn verjaardag mogen openen/*openen mogen.

*I have my presents not for my birthday may open/*open may.*

Ik heb de fiets van mijn moeder niet mogen gebruiken/*gebruiken mogen.

*I have the bike from my mother not may use/*use may.*

Marlies heeft tijdens haar dieet geen chocolade mogen eten/*eten mogen.

*Marlies has during her diet no chocolate may eat/*eat may.*

Bram heeft niet van het feest mogen weten/*weten mogen.

*Bram has not of the party may know/*know may.*

Wij hebben ons huis in dit gebied niet mogen bouwen/*bouwen mogen.

*We have our house in this area not may build/*build may.*

De studenten hebben het examen niet zonder de leraar mogen beginnen/*beginnen mogen.

*The students have the exam not without the teacher may start/*start may.*

Ik heb al lang meer over het brein willen weten/*weten willen.

*I have a long time more about the brain want know/*know want.*

Appendix. (continued)

Conflict Condition

Ik heb mijn oude fiets al lang willen verkopen/*verkopen willen.

*I have my old bike a long time want sell/*sell want.*

Lisa heeft haar oma al lange tijd willen schrijven/*schrijven willen.

*Lisa has her grandma a long time want write/*write want.*

Vincent heeft Maaïke voor haar verjaardag willen feliciteren/*feliciteren willen.

*Vincent has Maaïke for her birthday want congratulate/*congratulate want.*

De vriendinnen hebben Chris in Amsterdam willen bezoeken/*bezoeken willen.

*The friends have Chris in Amsterdam want visit/*visit want.*

De vrienden hebben vanavond een film willen kijken/*kijken willen.

*The friends have tonight a movie want watch/*watch want.*

Ik heb de soep van mijn grootmoeder niet willen eten/*eten willen.

*I have the soup from my grandmother not want eat/*eat want.*

Ik heb mijn huiswerk voor maandag niet willen doen/*doen willen.

*I have my homework for Monday not want do/*do want.*

Martijn heeft het Engelse boek niet willen lezen/*lezen willen.

*Martijn has the English book not want read/*read want.*

Aniek heeft haar vriend natuurlijk niet willen bedriegen/*bedriegen willen.

*Aniek has her friend of course not want threaten/*threaten want.*

Johanna en Tom hebben het feest niet willen verlaten/*verlaten willen.

*Johanna and Tom have the party not want leave/*leave want.*

De mensen hebben de waarheid niet willen geloven/*geloven willen.

*The people have the truth not want believe/*believe want.*

Appendix. (continued)

No-conflict Condition

Ik wil op vakantie maar/*hoewel ik heb geen tijd.

*I want on vacation but/*although I have no money.*

Ik wil naar de stad gaan maar/*hoewel ik ben ziek.

*I want to the city go but/*although I am sick.*

Ik kan niet slapen maar/*hoewel ik ben wel heel moe.

*I cannot sleep but/*although I am very tired.*

Ik houd van actiefilms maar/*hoewel ik vind James Bond niet leuk.

*I love action movies but/*although I find James Bond not cool.*

Ik verdien weinig geld maar/*hoewel ik ben altijd aan het werk.

*I earn little money but/*although I am always at work.*

Ik houd van dansen maar/*hoewel ik ben er niet goed in.

*I love dancing but/*although I am there not good at.*

Mijn broer haalt lage cijfers maar/*hoewel hij is eigenlijk heel intelligent.

*My brother gets low grades but/*although he is actually very intelligent.*

De vrouw werkt door maar/*hoewel ze voelt zich ziek.

*The woman works through (continues working) but/*although she feels sick.*

Het is warm maar/*hoewel het is niet zo zonnig.

*It is warm but/*although it is not so sunny.*

Sara gaat naar de kerk maar/*hoewel ze is niet religieus.

*Sara goes to church but/*although she is not religious.*

Hij speelt poker maar/*hoewel hij begrijpt de regels niet.

*He plays poker but/*although he understands the rules not.*

Elisa koopt graag nieuwe kleren maar/*hoewel zij heeft geen geld.

*Elisa likes to buy clothes but/*although she has no money.*

Appendix. (continued)

No-conflict Condition

Elsa lijkt rustig maar/*hoewel ze is echt heel nerveus.

*Elsa appears calm but/*although she is really very nervous.*

Maria en Tom hebben ruzie maar/*hoewel zij houden wel van elkaar.

*Maria and Tom have a fight but/*although they do love each other.*

We willen een hond maar/*hoewel we hebben er echt geen tijd voor.

*We want a dog but/*although we have there really no time for.*

Ron en Denise zijn gelukkig samen maar/*hoewel ze zijn heel verschillend.

*Ron and Denise are happy together but/*although they are very different.*

De jongens roken maar/*hoewel ze vinden het eigenlijk niet lekker.

*The boys smoke but/*although they find it actually not tasty.*

Mijn collega's helpen me graag maar/*hoewel zij hebben het zelf ook druk.

*My colleagues like to help me but/*although they are also busy themselves.*

Zij studeren in China maar/*hoewel zij spreken geen Chinees.

*They study in China but/*although they speak no Chinese.*

Wij komen uit Berlijn maar/*hoewel wij wonen daar niet meer.

*We are from Berlin but/*although we don't live there anymore.*

Ik heb heel veel gegeten dus/*waardoor ik heb geen honger meer.

*I have very much eaten thus/*whereby I have no hunger anymore.*

Ik eet geen vlees dus/*waardoor ik heb extra proteïnen nodig.

*I eat no meat thus/*whereby I need extra proteins.*

Ik ga vaak op reis dus/*waardoor ik ken veel verre landen.

*I go often on trips thus/*whereby I know many far away countries.*

Appendix. (continued)

No-conflict Condition

Hij werkt zestig uur per week dus/*waardoor hij verdient veel geld.

*He works sixty hours a week thus/*whereby he earns a lot of money.*

De oude vrouw loopt veel dus/*waardoor ze is nog vitaal.

*The old woman walks a lot thus/*whereby she is still agile.*

Peter leest altijd de krant dus/*waardoor hij weet veel over politiek.

*Peter reads always the newspaper thus/*whereby he knows a lot over politics.*

Janneke woont ver weg dus/*waardoor ze ziet haar familie zelden.

*Janneke lives far away thus/*whereby she sees her family seldomly.*

De kleren zijn op maat gemaakt dus/*waardoor ze zijn duur.

*The clothes are tailored to fit thus/*whereby they are expensive.*

De mannen spelen voetbal dus/*waardoor ze hebben een goede conditie.

*The men play football thus/*whereby they have (are in) a good condition.*

De Hollanders fietsen veel dus/*waardoor ze zijn fit.

*The Dutch bike a lot thus/*whereby they are fit.*

Ik ga met de fiets want/*omdat het is mooi weer.

*I go by bike because/*since it is nice weather.*

Ik leef gezond want/*omdat ik kook elke dag zelf.

*I live healthy because/*since I cook every day myself.*

Ik geef een feest want/*omdat ik ben jarig.

*I give a party because/*since it's my birthday.*

Ik slaap lekker want/*omdat ik heb een heel fijn bed.

*I sleep well because/*since I have a very nice bed.*

Ik doe een jas aan want/*omdat ik heb het koud.

*I do the laundry because/*since I feel cold.*

Appendix. (continued)

No-conflict Condition

Ik kan vandaag niet werken want/*omdat ik heb de griep.

*I can today not work because/*since I have the flu.*

Jeroen geeft veel concerten want/*omdat hij is een beroemde zanger.

*Jeroen gives a lot of concerts because/*since he is a famous singer.*

Hij leert Spaans want/*omdat hij gaat volgend jaar naar Spanje.

*He learns Spanish because/*since he goes next year to Spain.*

Zij draagt een blauwe jurk want/*omdat zij houdt van blauw.

*She wears a blue dress because/*since she loves blue.*

Ton slaapt vandaag uit want/*omdat hij heeft vrij.

*Ton sleeps in today because/*since he is free.*

Jan komt niet naar het feest want/*omdat hij heeft geen tijd.

*Jan does not come to the party because/*since he has no time.*

Hij is steeds moe want/*omdat hij kan 's nachts niet slapen.

*He is still tired because/*since he can at night not sleep.*

Bas is niet populair want/*omdat hij denkt alleen aan zichzelf.

*Bas is not popular because/*since he thinks only about himself.*

De kinderen zijn blij want/*omdat ze zijn op vakantie.

*The children are happy because/*since they are on vacation.*

De hardlopers zijn nerveus want/*omdat ze hebben een grote wedstrijd.

*The runners are nervous because/*since they have a big competition.*

Eva en Simon kopen een huis want/*omdat zij krijgen een kind.

*Eva and Simon buy a house because/*since they are having a baby.*

De meisjes spelen samen want/*omdat ze vinden elkaar aardig.

*The girls play together because/*since they like each other.*

Appendix. (continued)

No-conflict Condition

Mijn nichtjes krijgen een kat want/*omdat ze willen een huisdier.

*My cousins get a cat because/*since they want a pet.*

De honden springen want/*omdat ze zijn blij.

*The dogs jump because/*since they are happy.*

Wij bezoeken onze neef want/*omdat wij zien hem anders nooit.

*We visit our cousin because/*since we see him otherwise never.*

Filler Sentences

Ik houd van de winter hoewel/*maar ik soms de sneeuw vervelend vind.

*I love winter although/*but I sometimes the snow annoying find.*

Ik praat zes verschillende talen hoewel/*maar ik pas twaalf jaar ben.

*I speak six different languages although/*but I only twelve am.*

Ik houd van paarden hoewel/*maar ik soms bang voor ze ben.

*I love horses although/*but I sometimes afraid of them am.*

Ik heb een slecht gebit hoewel/*maar ik altijd mijn tanden poets.

*I have bad teeth although/*but I always my teeth brush.*

Ik verf mijn haar blond hoewel/*maar ik nu bruin haar mooier vind.

*I die my hair blond although/*but I now brown hair prettier find.*

Marta's mobieltje is kapot hoewel/*maar ze altijd goed oppast.

*Marta's mobile is broken although/*but she always attention pays.*

Iris eet graag paprika hoewel/*maar ze meestal broccoli kookt.

*Iris likes to eat paprika although/*but she usually broccoli cooks.*

Appendix. (continued)

No-conflict Condition

Mijn vader doet de boodschappen hoewel/*maar hij nooit zelf kookt.

*My dad does the grocery shopping although/*but he never cooks himself.*

Kees houdt van boeken hoewel/*maar hij langzaam leest.

*Kees loves books although/*but he slowly reads.*

Anne heeft mooi haar hoewel/*maar ze zelden naar de kapper gaat.

*Anne has pretty hair although/*but she seldomly to the hairdresser goes.*

Koen studeert rechten hoewel/*maar hij eigenlijk van biologie houdt.

*Koen studies law although/*but he actually biology loves.*

Wij gaan naar Thailand hoewel/*maar wij meestal nooit ver weg gaan.

*We go to Thailand although/*but we usually never far away go.*

De broers lijken veel op elkaar hoewel/*maar ze geen tweeling zijn.

*The brothers look very similar although/*but they not twins are.*

De vrienden gaan uit hoewel/*maar ze eigenlijk te jong zijn.

*The friends go out although/*but they actually too young are.*

De acteurs treden op hoewel/*maar zij slechts met drie man zijn.

*The actors perform although/*but they only with three people are.*

Wij verhuizen morgen hoewel/*maar we eigenlijk niet weg willen.

*We move tomorrow although/*but we actually not leave want.*

Ik moet ver fietsen waardoor/*dus ik soms te laat kom.

*I need to bike far whereby/*thus I sometimes late am.*

Appendix. (continued)

No-conflict Condition

Ik luister veel muziek waardoor/*dus ik veel liedjes ken.
*I listen to a lot of music whereby/*thus I a lot of songs know.*

Het huis is weer schoon waardoor/*dus het nu lekker ruikt.
*The house is again clean whereby/*thus it now good smells.*

Sophie is zwanger waardoor/*dus ze nu een dikke buik heeft.
*Sophie is pregnant whereby/*thus she now a big belly has.*

Thijs is soldaat waardoor/*dus hij vaak weg van huis is.
*Thijs is a soldier whereby/*thus he often away from home is.*

Wij hebben geen sleutel waardoor/*dus we niet naar binnen kunnen.
*We have no keys whereby/*thus we not inside can.*

De artiesten accepteren geen kritiek waardoor/*dus ze erg arrogant lijken.
*The artists accept no criticism whereby/*thus they very arrogant seem.*

Mijn ouders hebben veel zorgen waardoor/*dus ze's nachts slecht slapen.
*My parents have a lot of worries whereby/*thus they at night badly sleep.*

Ik koop geen cd's omdat/*want ik nooit naar muziek luister.
*I buy no cds since/*because I never to music listen.*

Ik drink geen koffie omdat/*want ik niet van de smaak houd.
*I drink no coffee since/*because I not the taste like.*

Ik vind wiskunde vervelend omdat/*want ik nooit hoge cijfers krijg.
*I find math annoying since/*because I never high grades get.*

Ik zie mijn huisgenoten zelden omdat/*want ik nooit thuis ben.
*I see my housemates seldomly since/*because I never home am.*

Ik heb een vieze kamer omdat/*want ik nooit schoonmaak.
*I have a dirty room since/*because I never clean.*

Appendix. (continued)

No-conflict Condition

Pim sport niet meer omdat/*want hij vaak last van zijn benen heeft.
*Pim no longer does sports since/*because often leg problems has.*

De ijsbeer vangt vis omdat/*want hij erg hongerig is.
*The ice bear catches fish since/*because he very hungry is.*

Lea komt met de trein omdat/*want ze momenteel geen auto heeft.
*Lea comes by train since/*because she currently no car has.*

Rens heeft geen vriendin omdat/*want hij altijd thuis zit.
*Rens doesn't have a girlfriend since/*because he always home stays.*

Tina heeft hulp nodig omdat/*want zij nog nieuw in de stad is.
*Tina needs help since/*because she still new to the city is.*

De studenten zijn relaxed omdat/*want ze nu geen tentamens hebben.
*The students are relaxed since/*because they now no exams have.*

Mijn ooms worden steeds dikker omdat/*want zij veel bier drinken.
*My uncles keep getting fatter since/*because they a lot of beer drink.*

De gasten blijven niet lang omdat/*want ze vandaag haast hebben.
*The guests don't stay long since/*because they today in a rush are.*

Wij gebruiken veel peper omdat/*want wij graag pittig eten.
*We use a lot of pepper since/*because we like spicy to eat.*

De muizen vluchten omdat/*want ze heel bang voor de kat zijn.
*The mice flee since/*because they very afraid of the cat are.*

Uilen jagen 's nachts omdat/*want ze dan beter zien.
*Owls hunt at night since/*because they then better see.*

Asterisk (*) marks ungrammatical sentence continuation. Original Dutch sentences are in normal print; English literal translations are in italics.

Acknowledgments

We would like to thank Iris Verpaalen and Julia Egger for help with data acquisition, as well as Stephen Politzer-Ahles for useful discussions on data analysis. This research was supported by a VIDI grant to K. L. (grant 016.134.322). The raw data, as well as analysis scripts for the current study, can be accessed via the following URL: <http://hdl.handle.net/11633/aabwx5zh>.

Reprint requests should be sent to Anne Mickan, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Montessorilaan 3, Nijmegen, The Netherlands, 6525 HR, or via e-mail: a.mickan@donders.ru.nl.

Notes

1. We will refer to Dutch as a second language (L2) throughout the paper, even though it was technically L3, or even L4, for most participants.
2. Please note that the repetition of modal verbs, though not ideal, is inherent to the sentences under investigation: The number of modal auxiliaries that are cognates with German and Dutch is limited, hence calling for repetition of the few available ones. Repeating content verbs the same amount of times to match this repetition, in turn, would have rendered the sentences extremely monotonous, which seemed undesirable. A comparison of the first and second half of the experiment for each of the groups thankfully revealed, however, that this repetition difference did not affect the ERPs in our study: Repetition effects should show in reduced N400 effects in the second compared to the first half of the experiment in the condition where the modal auxiliary verbs were the target (grammatical sentences in the conflict condition), which was not the case.
3. In a small set of pilot data, it turned out that “dus” sentences were difficult for some beginning German learners of Dutch. However, for the sake of variability, we did not want to drop it completely; thus, we decided to include fewer sentences (10 dus/waardoor) containing this conjunction than sentences containing the other conjunctions (20 maar/hoewel; 20 want/omdat).
4. Remember that, for the interaction analysis, a one-tailed F test was used, and hence, the critical p value here is $p = .05$ rather than $p = .025$, as for all other (two-tailed) permutation tests.
5. We chose to do the analysis over the whole time window for two reasons: First, searching for an effect of grammaticality in a window based on an analysis that already yielded an effect involving the factor grammaticality would inflate Type I error rate. Second, visual inspection suggests that some groups show effects outside the interaction time window. These effects would be neglected entirely if we were to limit the follow-up permutation tests to the interaction time window.
6. Two-tailed permutation tests are thresholded at an alpha level of .025 (instead of .05). Because permutation tests are of a rather conservative nature though, it is justified and common to consider marginal effects as well (cf. Politzer-Ahles, Schluter, Wu, & Almeida, 2016).
7. As explained earlier, the critical alpha level for a two-tailed permutation test is $p = .025$ instead of $p = .05$; hence, the positive cluster in the advanced learners is only marginally significant. Please be reminded of the conservative nature of the cluster-based permutation tests though, especially when it comes to detecting more narrowly distributed effects (Groppe, Urbach, & Kutas, 2011), such as the P600 in the advanced group.
8. Following Tanner et al. (2013, 2014), N400 effect magnitudes were calculated as the mean activity in the grammatical conflict condition sentences minus the ungrammatical conflict condition sentences in a time window from 300 to 460 msec (spanning the N400 cluster from the intermediate learners) over

a centroposterior ROI (C3, Cz, C4, P3, Pz, P4). P600 effect magnitudes were calculated as the mean amplitude in the ungrammatical minus the grammatical conflict condition sentences in a time window from 820 to 1000 msec over the same centroposterior ROI, again based on the latency and topography of the P600 clusters in the intermediate learner group.

REFERENCES

- Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (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., Fiorentino, R., & Gabriele, A. (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.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database. Retrieved from celex.mpi.nl/.
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning, 56*, 9–49.
- Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia, 51*, 2492–2511.
- Bultena, S., Danielmeier, C., Bekkering, H., & Lemhöfer, K. (2017). Electrophysiological correlates of error monitoring and feedback processing in second language learning. *Frontiers in Human Neuroscience, 11*, 29.
- Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Reviews, 51*, 31–47.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition, 10*, 161–174.
- Davidson, D. J., & Indefrey, P. (2009). Plasticity of grammatical recursion in German learners of Dutch. *Language and Cognitive Processes, 24*, 1335–1369.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.
- Díaz, B., Erdocia, K., de Menezes, R. F., Mueller, J. L., Sebastián-Gallés, N., & Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Frontiers in Psychology, 7*, 133.
- Ensz, K. Y. (1982). French attitudes toward typical speech errors of American speakers of French. *Modern Language Journal, 66*, 133–139.
- Erdocia, K., & Laka, I. (2018). Negative transfer effects on L2 word order processing. *Frontiers in Psychology, 9*, 337.
- Erdocia, K., Zawiszewski, A., & Laka, I. (2014). Word order processing in a second language: From VO to OV. *Journal of Psycholinguistic Research, 43*, 815–837.
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition, 23*, 527–552.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition, 14*, 379–399.
- Gardner, R. C., Clément, R., Smythe, P. C., & Smythe, C. L. (1979). *Attitudes and motivation test battery: Revised manual*

- (Research Bulletin, no. 15). London, Ontario, Canada: Department of Psychology, University of Western Ontario.
- Gillon Dowens, M., Guo, T., Guo, J., Barber, H. A., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish—Evidence from event related potentials. *Neuropsychologia*, *49*, 1651–1659.
- Gillon Dowens, M., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*, 1870–1887.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, *48*, 1711–1725.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–483.
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, *30*, 251–266.
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277.
- Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, *15*, 409–414.
- Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.
- Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*, 68–74.
- Kotz, S. A., Holcomb, P. J., & Osterhout, L. (2008). ERPs reveal comparable syntactic sentence processing in native and non-native readers of English. *Acta Psychologica*, *128*, 514–527.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343.
- Lemhöfer, K., Schriefers, H., & Indefrey, P. (2014). Idiosyncratic grammars: Syntactic processing in second language comprehension uses subjective feature representations. *Journal of Cognitive Neuroscience*, *26*, 1428–1444.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., et al. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*, 123–150.
- Meulman, N., Stowe, L. A., Sprenger, S. A., Bresser, M., & Schmid, M. S. (2014). An ERP study on L2 syntax processing: When do learners fail? *Frontiers in Psychology*, *5*, 1072.
- Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*, 933–947.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869.
- Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199–230.
- Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1547–1570.
- Queen, R., & Boland, J. E. (2015). I think your going to like me: Exploring the role of errors in email messages on assessments of potential housemates. *Linguistics Vanguard*, *1*, 283–293.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*, 2030–2048.
- Ruxton, G. D., & Beauchamp, G. (2008). Time for some a priori thinking about post hoc testing. *Behavioral Ecology*, *19*, 690–693.
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*, 397–430.
- Schmid, M. S., Dusseldorp, E., van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*, 43–74.
- 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., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*, 135–162.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*, 13–41.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*, 277–293.
- Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*, 367–382.
- Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*, 717–726.
- VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 365.
- White, E. J., Genesee, F., & Steinhauer, K. (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.
- Yuan, B. (2017). Can L2 sentence processing strategies be native-like? Evidence from English speakers' L2 processing of Chinese base-generated-topic sentences. *Lingua*, *191–192*, 42–64.
- Zawiszewski, A., Gutiérrez, E., Fernández, B., & Laka, I. (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition*, *14*, 400–411.