# The "semantic P600" in second language processing: When syntax conflicts with semantics

Xiaochen Zheng[a,b,*], Kristin Lemhöfer[a]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University, P.O. Box 9104, 6500 HE Nijmegen, the Netherlands
[b] International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ABSTRACT

In sentences like "the mouse that chased the cat was hungry", the syntactically correct interpretation (the mouse chases the cat) is contradicted by semantic and pragmatic knowledge. Previous research has shown that L1 speakers sometimes base sentence interpretation on this type of knowledge (so-called "shallow" or "good-enough" processing). We made use of both behavioural and ERP measurements to investigate whether L2 learners differ from native speakers in the extent to which they engage in "shallow" syntactic processing. German learners of Dutch as well as Dutch native speakers read sentences containing relative clauses (as in the example above) for which the plausible thematic roles were or were not reversed, and made plausibility judgments. The results show that behaviourally, L2 learners had more difficulties than native speakers to discriminate plausible from implausible sentences. In the ERPs, we replicated the previously reported finding of a "semantic P600" for semantic reversal anomalies in native speakers, probably reflecting the effort to resolve the syntax-semantics conflict. In L2 learners, though, this P600 was largely attenuated and surfaced only in those trials that were judged correctly for plausibility. These results generally point at a more prevalent, but not exclusive occurrence of shallow syntactic processing in L2 learners.

## 1. Introduction

Interpreting syntactically complex sentences, such as "the mouse that chased the cat was hungry" can be a challenge. This is because the syntactically correct interpretation (the mouse chases the cat) is contradicted by semantic and pragmatic knowledge. In sentence processing, even native speakers sometimes misinterpret this kind of sentence as the cat chasing the mouse, presumably because they apply a heuristic shortcut: They rely to a relatively large degree on semantic cues, potentially at the expense of syntactic cues, to obtain the meaning of a sentence rather than conducting a full syntactic analysis ("good enough" approach; Ferreira, 2003; Ferreira et al., 2002; Ferreira and Patson, 2007). In a second language (L2), language processing becomes slower and more effortful on all levels (e.g., Brysbaert et al., 2017; Cop et al., 2015; Meschyan and Hemandez, 2006; Morishima, 2013; Ojima et al., 2005), possibly causing a greater reliance on heuristic strategies than in native speakers (Clahsen and Felser, 2006, 2017; Guo et al., 2009; Roberts, 2012). The present study investigates how L2 learners negotiate between syntactic and semantic information during online processing, and how this differs from that in native speakers.

Online sentence processing has been commonly investigated using event-related potentials (ERPs) in the EEG. The ERP effects are elicited in response to syntactic or semantic violations during sentence reading, as compared to correct control sentences. For example, semantic violations such as "The pizza was too hot to *drink*…" give rise to an N400 effect, a centro-parietal negativity relative to control sentences, peaking around 400 ms after the onset of the anomalous word (Kutas and Federmeier, 2011; Kutas and Hillyard, 1980). On the other hand, syntactic violations such as "The spoiled child *throw* the toy…" usually elicit a parietal P600 effect, a positive-going deflection relative to control sentences, reaching its peak around 600 ms after onset of the violating word (Friederici et al., 1993; Osterhout and Holcomb, 1992). Traditionally, the semantic N400 effect and the syntactic P600 effect are considered to be qualitatively different and largely independent (Osterhout and Nicol, 1999). However, accumulating evidence suggests an interaction between syntax and semantics during online processing (e.g., Hagoort, 2003; Hahne and Friederici, 2002). Among them, a P600 effect caused by so-called semantic reversal anomalies (e.g., "The mouse that chased the cat…") has sparked a heated debate about the exact processes that are reflected in this ERP component, and in particular, whether these are exclusively syntactic in nature (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Kim and Osterhout,

---

2005; Kolk et al., 2003; Kuperberg, 2007; van Herten et al., 2006).

Semantic reversal anomalies usually involve an assignment of the two thematic roles of a verb that is opposite (reversed) to what would be expected on the grounds of semantic and pragmatic information (e.g., when the verb "chase" occurs with the nouns "cat" and "mouse", the cat is typically the agent and the mouse is the patient). Instead of the expected lexical-semantic N400 effect, semantic reversal anomalies elicit a P600 effect, the so-called "semantic P600". This effect has been reliably replicated in native speakers of Dutch (e.g., Hoeks et al., 2004; van Herten et al., 2006; van Herten et al., 2005), English (Frenzel et al., 2011; Kim and Osterhout, 2005; Paczynski and Kuperberg, 2012), and French (Shen et al., 2016). A number of hypotheses have been proposed to explain why semantic reversal anomalies evoke a P600 effect. Kim and Osterhout (2005) proposed that a semantically plausible thematic role assignment can be such a strong attractor that it dominates over the syntactically correct interpretation; consequently, the sentence is perceived as syntactically ill-formed because it contradicts this strongly preferred semantic interpretation, hence giving rise to a P600. Kuperberg (2007) assumed that the P600 results from a conflict between the output of two competing neural streams: a semantic memory-based mechanism (indicating that it should be the cat that chases the mouse), and a combinatorial mechanism which is mainly based on morphosyntactic rules (concluding that according to the syntactic structure, the mouse chases the cat in this sentence). This conflict causes a continued combinatorial analysis that is reflected in the P600. Similarly, Kolk, van Herten and colleagues (Kolk et al., 2003; van Herten et al., 2006) also proposed that semantic reversal anomalies cause a conflict between standard syntactic algorithms and a plausibility-based heuristic. This conflict in turn triggers a monitoring process based on the memory trace of the input sentence in order to resolve the conflict, which then gives rise to a P600 effect. A similar multi-stream explanation was given by Bornkessel-Schlesewsky and Schlesewsky (2008) in their extended Argument Dependency Model (eADM), where conflicting role assignments computed by two different streams (one based on mainly syntactic "prominence information" vs. one based on plausibility) fail to be integrated. However, in a later article focusing on cross-linguistic differences in ERP signatures of semantic reversal anomalies, Bornkessel-Schlesewsky et al. (2011) propose that the semantic P600 is a member of the P300b-family reflecting a late (categorical) well-formedness decision.

Theoretical debate aside, a common assumption shared by almost all the above-mentioned accounts is that the P600 is a consequence of the conflict between multiple processing streams (but see Brouwer et al., 2017; Brouwer et al., 2012), namely, between a syntactic parsing algorithm and a stream based on semantic heuristics. The relatively larger reliance on the semantic stream ("good enough" or "shallow" processing) has been shown to give rise to a high incidence of interpretation errors of complex sentences (Dąbrowska and Street, 2006; Ferreira, 2003; Lim and Christianson, 2013). Thus, while the interplay of syntax and semantics seems already complex in native speakers, we know almost nothing about this interplay where a second language is concerned. Studies on "regular" semantic or syntactic violations in L2 sentence processing have shown, first, that semantic N400 effects are usually preserved in L2 speakers, albeit sometimes with a delayed latency or reduced amplitude of the N400 (e.g., Ardal et al., 1990; Hahne, 2001; Newman et al., 2012; Ojima et al., 2005; Weber-Fox and Neville, 1996). Second, on the other hand, the P600 effect elicited by syntactic violations often only emerges gradually with increasing proficiency (for reviews, see Caffarra et al., 2015; Kotz, 2009; Steinhauer et al., 2009), even though sensitivity to syntactic violations can also occur in less than highly proficient learners, e.g. when the similarity between L1 and L2 is high with respect to the syntactic feature under investigation (e.g., Dowens et al., 2010; Foucart and Frenck-Mestre, 2011; Tolentino and Tokowicz, 2011). Thus, the available evidence suggests that L2 speakers can be less sensitive to syntactic information than native speakers are. In consequence, the interplay between syntax and

semantics during the processing of semantic reversal anomalies may take a different form in an L2. For example, as the semantic stream is weighed more and syntax is hardly used to inform sentence interpretation, the conflict between syntax and semantics may get reduced. Therefore, L2 learners may interpret the sentences as being in line with their heuristics or world knowledge (like the cat chasing the mouse, not the other way around) and not be troubled by the fact that the syntactic structure of the sentence suggests otherwise.

The current study made use of the "semantic P600" phenomenon to investigate whether L2 learners differ from native speakers in the extent to which they experience a conflict between syntactic cues and semantic heuristics. The application of "good enough" or "shallow" processing, i.e. the increased reliance on semantic compared to syntactic cues, may lead to a reduction of this conflict, as semantic reversal sentences would more often simply be interpreted as plausible. To answer this question, we compared the ERP signatures and behavioural responses of German learners of Dutch with those of native Dutch speakers while they read and made plausibility judgments to Dutch sentences with semantic reversal anomalies (e.g., "Het hondje dat tegen het meisje praatte speelde met de bal." *[The dog that talked to the girl was playing with the ball.]*).

To avoid that syntactic differences between L1 and L2 would distort the effects, we chose a pair of languages that are very similar in terms of the relevant syntactic feature, i.e. German (L1) and Dutch (L2). Particularly, the experimental sentences used in the current study were identical in terms of word order between Dutch and the German translation ("Das Hündchen, das mit dem Mädchen sprach, spielte mit dem Ball."). Thus, the question was whether despite the structural overlap between L1 and L2, the mere fact that L2 learners are processing their weaker non-native language would lead them to adopt a different negotiation strategy between syntactic and semantic information than L1 speakers.

One possibility for such a different strategy would be that L2 learners put less weight on the syntactic information and are therefore more likely to accept semantic reversal anomalies as plausible. This might also mean that they experience less conflict between syntax and semantics as reflected in the "semantic P600", leading to a reduced or even absent P600 in this group. It is also possible that the combination of the two streams works in a qualitatively different way from that in native speakers, for example, causing an N400 instead of a P600 (cf. Guo et al., 2009). A third possibility is that L2 learners are no different from native speakers in processing these sentences, resulting in a native-like P600 effect and similar behavioural response patterns.

Among the three possibilities, we are more inclined to a gradual difference between the L1 and L2 group due to the challenge of L2 processing. The challenges of L2 sentence processing might result from, on the one hand, a lack of knowledge of L2 syntax and semantics, and on the other hand, more general disadvantages such as fewer available cognitive resources or a lack of automaticity during online processing. To further explore the nature and determinants of L2 sensitivity to the syntax-semantics conflict, we also included a few individual difference measurements. According to a meta-analysis by Caffarra and colleagues, the most influential factor for the size of the P600 effect as response to a syntactic violation is L2 proficiency (Caffarra et al., 2015). Similarly, Tanner and colleagues (Tanner et al., 2014, 2013) have shown that the occurrence of a P600 in standard syntactic violation processing may depend on individual differences, most notably proficiency. Furthermore, working memory has been found to be a predictor for ERP correlates in L1 sentence processing, in particular the P600 and N400 (Kim et al., 2018; Nakano et al., 2010). Therefore, we included a few measurements of L2 proficiency and working memory capacity, and investigated in an exploratory way whether there was strong evidence for the modulation of the processing of semantic reversal anomalies in L2 by these variables. However, note that the current study was not designed as a "full-blown" individual differences study and also does not possess the necessary sample size to function as one.

**Table 1**
L2 learners' language background and Dutch proficiency.

| | *Mean* | *SD* | *Range* |
|---|---|---|---|
| Months of exposure to the Dutch language | 14.2 | 5.7 | 6–24 |
| Months of living in the Netherlands | 10.4 | 6.8 | 0–23 |
| Self-rated frequency of using Dutch[a] | | | |
| – reading | 3.8 | 0.9 | 1–5 |
| – speaking | 3.8 | 1.0 | 2–5 |
| – listening | 4.4 | 0.8 | 2–5 |
| Self-rated proficiency in Dutch[a] | | | |
| – reading | 3.9 | 0.6 | 3–5 |
| – speaking | 3.2 | 0.8 | 2–5 |
| – listening | 3.7 | 0.7 | 2–5 |
| – writing | 3.1 | 0.9 | 1–5 |
| Dutch vocabulary size | | | |
| – LexTALE test score | 68.2 | 7.4 | 55–86[b] |

Note. SD = Standard Deviation.

[a] Self-ratings were given on a scale from 1 = *very rare/bad* to 5 = *very often/good*.

[b] The score is a weighted % correct score, i.e. 50 is chance level, 100 is the maximum score.

## 2. Method

### 2.1. Participants

A total of 24 native Dutch speakers (L1 group, $M_{age}$ = 22.8, nine males) and 64 German learners of Dutch (L2 group, $M_{age}$ = 21.0, ten males), mostly students at Radboud University, participated for course credit or vouchers. More L2 learners than native speakers were tested because we also looked at individual differences in the L2 group. All the L2 learners were raised with German as the only mother tongue and had maximally two years of experience of learning Dutch in the context of their Dutch-language study program. All participants were right-handed, had normal or corrected-to-normal vision, and no reading disabilities. We excluded three L2 learners in the ERP analysis due to excessive artefacts and measurement errors. For the sake of consistency, we also left out these three participants in the behavioural analysis, leaving the final set with 61 L2 learners ($M_{age}$ = 21.0, nine males). Table 1 shows the L2 learners' language background and their Dutch vocabulary size measured with the Dutch version of the LexTALE test (Lemhöfer and Broersma, 2012).

We also measured all participants' working memory capacity and L2 learners' proficiency in Dutch. We used the working memory tests developed by Klaus and Schriefers (2016), which consisted of an operation span test and a reading span test, and were performed in participants' native language (Dutch or German). The Dutch proficiency test was specifically tailored to German learners of Dutch and was adapted from the Dutch Proficiency Test (Transparent Language, Inc, 2015) and the Placement Test in the Dutch Language (Ernst Klett Sprachen, GmbH, 2015). The proficiency test consisted of 48 multiple-choice questions with four alternatives, measuring both grammar and vocabulary. More information about the tests can be found in Appendix B.

### 2.2. Materials

The materials included 60 sentence pairs (i.e., plausible vs. implausible) in the experimental condition, 50 sentence pairs in the control condition, and 10 filler sentence pairs (see Appendix A). Each participant saw either the plausible or the implausible variant of each sentence pair once (i.e., each participant read in total 120 different sentences, 60 of which were plausible and 60 implausible). All the sentence pairs in the experimental condition had semantic reversal anomalies in their implausible versions and used a subject-clause structure in the simple past tense (see Table 2 for an overview of conditions and sub-conditions). We also had a control condition with

**Table 2**
Example sentences in the current study (four sub-conditions).

| Experimental Condition | |
|---|---|
| Plausible (PE) | Het meisje dat tegen het hondje <u>praatte</u> speelde met de bal. |
| | *The girl that talked to the dog was playing with the ball.* |
| | *[Literal: The girl that to the dog talked played with the ball.]* |
| Implausible (IE) | *Het hondje dat tegen het meisje <u>praatte</u> speelde met de bal. |
| | *The dog that talked to the girl was playing with the ball.* |
| | *[Literal: The dog that to the girl talked played with the ball.]* |

| Control Condition | |
|---|---|
| Plausible (PC) | De gangsters sprongen in de <u>wagen</u> en verlieten de stad. |
| | *The gangsters jumped into the car and left the city.* |
| Implausible (IC) | *De gangsters sprongen in de <u>vraag</u> en verlieten de stad. |
| | *The gangsters jumped into the question and left the city.* |

Note. The critical words on which ERPs were measured are underlined.

standard semantic anomalies as a check of general semantic sensitivity, especially in the L2 learners. To increase the variation of sentence types, we constructed the control condition with 10 pairs of subject-relative clauses, 20 pairs of objective-relative clauses, and 20 pairs of conjunctions.

All the sentences were adapted from van Herten et al. (2006) using words and phrases that were comprehensible for L2 learners: All the words in a sentence were cognates between Dutch and German, or high-frequency Dutch words generally known by L2 learners, except for the critical verb in the experimental condition (i.e., the verb of the subject clause, underlined in Table 2), to avoid that the sentence was processed purely in "L1 mode". The critical verbs were non-cognates between Dutch and German, or non-identical cognates with different simple past forms in Dutch and German, but were of high frequency in Dutch[1]; all the relative clauses used an indirect object, to avoid the ambiguous subject-object or object-subject-verb structure in Dutch.

A majority of the implausible sentences used in the experiment have an inanimate subject (e.g., "De ladder die op de acrobaat klom viel plotseling naar beneden." *[The ladder that climbed on the acrobat suddenly fell down.]*). To minimize the association between inanimate subjects and implausible sentences, we also included ten filler sentences which were subject-relative clauses with an inanimate subject in the plausible version, and semantic reversal anomalies in the implausible version.

We constructed four pseudo-randomized lists and assigned each participant randomly to one of the four lists. Each list consisted of 120 sentences, with half of the sentences being plausible and the other half implausible. Plausibility and item order was counterbalanced across lists. We restricted the randomization such that (1) there were no more than three contiguous sentences from the same condition with the same plausibility; (2) no more than three contiguous implausible or plausible sentences; (3) no more than three contiguous sentences in the experimental condition.

### 2.3. Procedure

We tested the participants individually in a sound-proof lab and ran the experiment using the software package Presentation (Version 17.0, Neurobehavioural System Inc, Berkeley, U.S.). The computer screen (Benq XL2420Z, screen size 24 in.) was set to grey, with a resolution of 1920 × 1080 pixels, at a refresh rate of 120 Hz. All the sentences were presented word-by-word in the centre of the screen in black 36 point Arial letters on a light grey background, with a viewing distance of approximately 60 cm. Before the beginning of each sentence, a fixation

---

[1] We piloted all the critical verbs with a separate group of L2 learners (N = 10) and replaced all insufficiently familiar words prior to the EEG experiment.

cross appeared in the centre of the screen for 500 ms, followed by a blank screen for 500 ms. Each word appeared on the screen for 450 ms and was followed by a blank screen for 250 ms. The final word of each sentence was presented together with a period. After that, the question "plausible?" showed up on the screen. Participants then made a judgment about whether the sentence was "plausible" or not by pressing one of two buttons (right for "yes", left for "no") on the button box. The next trial started once participants responded or after 10 s if participants did not make a response. Participants were instructed not to blink during sentence presentation. The experiment started with four practice sentences. Participants' EEG was recorded during all the sentences. After the EEG measurement, we gave all the L2 learners a list with the sentences they had read in the experimental condition and asked them to circle all the words of which they did not know the meaning. Individual sentences in which the participants did not know the critical verbs were excluded for later analyses (more details are given in 2.4). The EEG session took about 1.5 h.

Roughly one week before the EEG session, participants completed the language background questionnaire, the working memory tests, the proficiency test (only for L2 learners) and the LexTALE vocabulary test in a separate behavioural session.

### 2.4. EEG recording and data analysis

We recorded the EEG using an elastic cap containing 27 active Ag-AgCl electrodes (ActiCAP 32Ch Standard-2, Brain Products). Three additional electrodes were placed on both mastoids and the forehead. The EEG signals were amplified with a BrainAmp DC/MR plus amplifier, digitized at a sampling rate of 500 Hz and online-filtered with the low cutoff of 0.016 Hz and the high cutoff of 125 Hz, and were online referenced to the left mastoid electrode. The forehead electrode served as the ground. Impedances for EEG electrodes were kept below 20 kΩ. EOG was measured with four electrodes placed above and below the right eye (to monitor for vertical eye movements) and on the outer side of both eyes (to monitor for horizontal eye movements).

We performed all the ERP analyses using the Fieldtrip open source Matlab toolbox (Oostenveld et al., 2011) and custom analysis scripts using Matlab v.8.6.0 (R2015b, The Math Works, Inc). We re-referenced the continuous EEG to the average of the right and left mastoid electrodes and then band-pass filtered the data at 0.05 to 30 Hz. The continuous EEG was segmented into stimulus-locked epochs from 200 ms before to 1000 ms after critical word onset in each sentence. We rejected trials with atypical artefacts (e.g., jumps and drifts) by visual inspection and then removed EOG artefacts (eye blinks and saccades) using independent component analysis (ICA). Afterwards, a baseline correction was applied based on the average EEG activity in the 200 ms interval before critical word onset. Then, trials with remaining artefacts (e.g., muscle artefacts) were removed by visual inspection. Individual EEG channels with bad signals were disabled before ICA and interpolated by a weighted average of the data from neighbouring channels of the same participant.

For the experimental condition, sentences for which individual L2 learners had reported the critical verb as unknown by circling them in the sheet after the experiment were excluded from further behavioural and ERP analyses (on average, 15.5% of the data were excluded, *SD* = 7.3%, range = 3.0–35.0%).[2] For the control condition, we excluded one sentence which contained a syntactic error. We also excluded participants from the ERP analysis with less than 15 remaining trials after all exclusion procedures in any sub-condition (see below).

## 3. Results

### 3.1. Behavioural results

Fig. 1 shows the percentages of correct plausibility judgments in each sub-condition (i.e., PE, IE, PC, IC) for the L1 and the L2 groups. We computed a *d'* score ($d' = Z_{Hit} − Z_{False\ Alarm}$) for each participant in both the experimental and the control conditions.[3] A mixed ANOVA of Condition × Group of these *d'* scores showed: first, L2 learners were worse at detecting anomalies than the native speakers ($Mean_{L1}$ = 1.94, $SE_{L1}$ = 0.10, $Mean_{L2}$ = 1.58, $SE_{L2}$ = 0.06, $F(1, 83)$ = 9.02, $p$ = .004, $\eta_p^2$ = .10); second, anomalous sentences in the experimental condition were more difficult to detect than those in the control condition ($Mean_{exp}$ = 1.56, $SE_{exp}$ = 0.07, $Mean_{ctr}$ = 1.96, $SE_{ctr}$ = 0.06, $F(1, 83)$ = 47.15, $p$ < .001, $\eta_p^2$ = .36); and most importantly, an interaction effect of Condition × Group ($F(1, 83)$ = 5.76, $p$ = .019, $\eta_p^2$ = .07) suggested that the L2 learners were especially worse than the native speakers in terms of detecting semantic reversal anomalies ($Mean_{L1}$ = 1.81, $SD_{L1}$ = 0.60, $Mean_{L2}$ = 1.32, $SD_{L2}$ = 0.54, $t(83)$ = 3.68, $p$ < .001, Cohen's $d$ = 0.81), but not in terms of detecting the standard semantic anomalies in the control condition ($Mean_{L1}$ = 2.07, $SD_{L1}$ = 0.55, $Mean_{L2}$ = 1.85, $SE_{L2}$ = 0.53, $t(83)$ = 1.68, $p$ = .098, Cohen's $d$ = 0.37). In addition, the fact that sentences with semantic reversal anomalies were more difficult to detect than those with standard semantic anomalies held especially for the L2 group ($Mean_{exp}$ = 1.31, $SD_{exp}$ = 0.54, $Mean_{ctr}$ = 1.85, $SE_{ctr}$ = 0.53, $t(60)$ = -8.87, $p$ < .001, Cohen's $d$ = -1.00) but also for the L1 group ($Mean_{exp}$ = 1.81, $SD_{exp}$ = 0.60, $Mean_{ctr}$ = 2.07, $SE_{ctr}$ = 0.55, $t(23)$ = -2.53, $p$ = .019, Cohen's $d$ = -0.66).
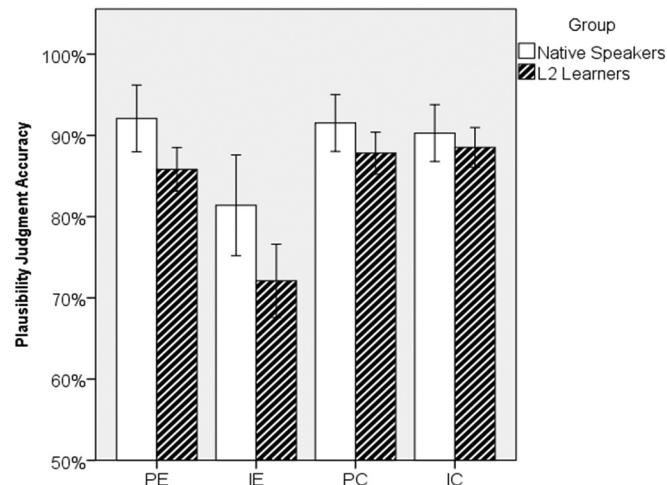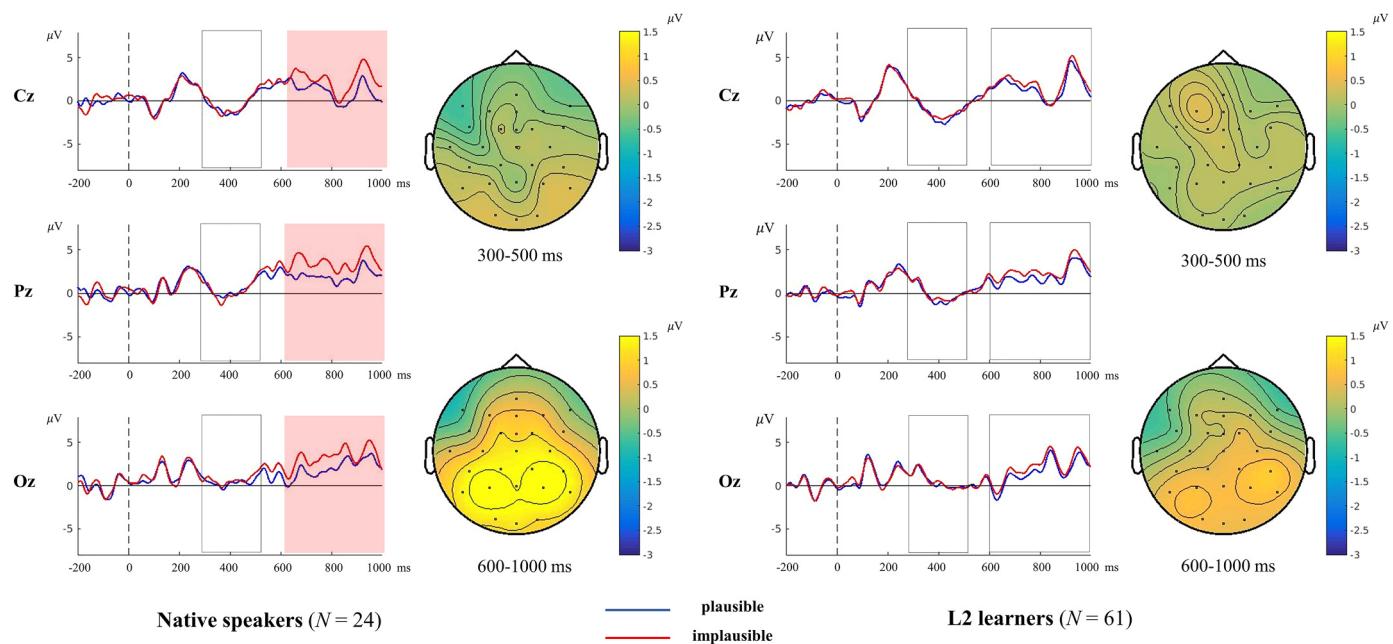
**Fig. 1.** Plausibility judgment accuracy in each sub-condition ($N_{L1}$ = 24; $N_{L2}$ = 61). PE = *plausible sentences in the experimental condition*; IE = *implausible sentences in the experimental condition*; PC = *plausible sentences in the control condition*; IC = *implausible sentences in the control condition*. Error bars indicate 95% CI.

### 3.2. ERP Results

#### 3.2.1. Experimental condition

*3.2.1.1. Analysis of all trials.* Fig. 2 shows the ERPs of the experimental (semantic reversal) condition in both groups. Based on the literature

---

[2] To validate that the L2 learners actually knew the meaning of the words they claimed to know (i.e., they did not circle in the post-reading questionnaire), we asked a subset of the participants ($N$ = 20) to translate the critical verbs. The result showed that when participants did not circle a given verb, they indeed translated it correctly in 90% of the cases.

[3] To calculate the *d'* score, extreme values in the raw data such as 1 in hit rate and 0 in false alarm rate were replaced with $(n − .5)/n$ and $.5/n$, respectively, where n was the number of trials per sub-condition (Macmillan and Kaplan, 1985).

**Fig. 2.** Grand-averaged ERPs and topographies for the critical verb in the experimental condition (all trials) for the native speakers (left, $N = 24$) and the L2 learners (right, $N = 61$), from three midline electrodes (Cz, Pz, Oz). The time windows of analysis (N400: 300–500 ms; P600: 600–1000 ms) are marked in light red (significant) or as an empty frame (not significant). Significances are calculated based on the lateral analysis, not the actual electrodes that are shown. Scalp maps show the topographies of the difference between plausible and implausible sentences.

and visual inspection of the waveforms, we selected the time from 300 to 500 ms as the time window for the N400 component, and 600 to 1000 ms for the P600 component. We calculated the mean amplitude of the N400 and the P600 time windows for each participant in each condition and performed repeated measures MANOVAs on the two groups separately. We included six clusters of lateral electrodes (Left anterior: F3, FC5, FC1; left medial: C3, CP5, CP1; left posterior: P7, P3, O1; right anterior: F4, FC2, FC6; right medial: C4, CP2, CP6; right posterior: P4, P8, O2), with the variables Hemisphere (left vs. right), Region (anterior vs. medial vs. posterior) and Plausibility (plausible vs. implausible). Since the analysis of the midline electrodes (Fz, FCz, Cz, Pz, Oz) did not provide additional informative evidence, we only report

the results of the lateral analysis. In Table 3, we report all the F-statistics for the main effects and interactions with the variable Plausibility; in the main text, we follow up significant interactions (i.e., $p < .05$) by simple effect tests.

*3.2.1.1.1. Native speakers.* No significant effects involving plausibility were found in the N400 time window. In the P600 time window, the results showed a marginally significant effect of plausibility and a significant interaction between plausibility and region. Follow-up analyses of the anterior, medial, and posterior regions showed a significant effect of plausibility at posterior sites ($F(1, 23) = 10.88$, $p = .003$, $\eta_p^2 = .32$) and a marginally significant effect at medial sites ($F(1, 23) = 4.16$, $p = .053$, $\eta_p^2 = .15$), but no effect at

**Table 3**
F-statistics of the $2 \times 2 \times 3$ lateral analysis performed on the N400 and the P600 time windows in the ERP analysis for both groups (experimental condition).

| Group/Condition | | Experimental Condition | | Experimental Condition | |
|---|---|---|---|---|---|
| | | (all trials) | | (correct trials only) | |
| | Factors*/Time | N400 | P600 | N400 | P600 |
| Native speakers | P | $F < 1$ | $F = 3.95$, $p = .059$, $\eta_p^2 = .15$ | $F < 1$ | $F = 4.11$, $p = .056$, $\eta_p^2 = .17$ |
| | P×H | $F < 1$ | $F < 1$ | $F < 1$ | $F = 1.27$, $p = .273$, $\eta_p^2 = .06$ |
| | P×R | $F < 1$ | **$F = 3.78$, $p = .039$, $\eta_p^2 = .26$** | $F = 1.10$, $p = .352$, $\eta_p^2 = .10$ | **$F = 5.95$, $p = .010$, $\eta_p^2 = .39$** |
| | P×H×R | $F < 1$ | $F < 1$ | $F = 1.10$, $p = .352$, $\eta_p^2 = .10$ | $F < 1$ |
| L2 learners | P | $F < 1$ | $F = 1.69$, $p = .198$, $\eta_p^2 = .03$ | $F < 1$ | **$F = 4.19$, $p = .047$, $\eta_p^2 = .09$** |
| | P×H | $F < 1$ | $F = 1.94$, $p = .169$, $\eta_p^2 = .03$ | $F = 1.06$, $p = .310$, $\eta_p^2 = .02$ | $F = 3.77$, $p = .059$, $\eta_p^2 = .08$ |
| | P×R | $F < 1$ | $F = 2.43$, $p = .096$, $\eta_p^2 = .08$ | $F < 1$ | **$F = 4.28$, $p = .020$, $\eta_p^2 = .17$** |
| | P×H×R | $F = 2.10$, $p = .131$, $\eta_p^2 = .07$ | $F = 1.60$, $p = .210$, $\eta_p^2 = .05$ | $F = 1.52$, $p = .231$, $\eta_p^2 = .07$ | **$F = 3.35$, $p = .045$, $\eta_p^2 = .14$** |
| Between-group comparison | G×P | $F < 1$ | $F = 3.71$, $p = .058$, $\eta_p^2 = .04$ | $F < 1$ | $F < 1$ |
| | G×P×H | $F < 1$ | $F < 1$ | $F = 1.81$, $p = .183$, $\eta_p^2 = .03$ | $F < 1$ |
| | G×P×R | $F < 1$ | **$F = 3.17$, $p = .047$, $\eta_p^2 = .07$** | $F < 1$ | **$F = 3.36$, $p = .041$, $\eta_p^2 = .10$;** |
| | G×P×H×R | $F < 1$ | $F < 1$ | $F < 1$ | $F = 1.44$, $p = .245$, $\eta_p^2 = .04$ |

*P = Plausibility; H = Hemisphere; R = Region; G = Group.
Significant effects are printed in bold.

anterior sites ($p = .339$).

*3.2.1.1.2. L2 learners.* No significant effects were found in the N400 time window, or in the P600 time window.

*3.2.1.1.3. Between-group comparison.* To compare the ERP effects between the L1 and the L2 groups, we introduced a new variable Group (L1 vs. L2 group) as a between-subject factor. No difference between groups was found in the N400 time window. In the P600 time window, results showed a marginally significant interaction between plausibility and group as well as a significant three-way interaction between plausibility, region, and group. Follow-up analyses showed an interaction of plausibility and group, reflecting a larger effect of plausibility in the native speakers than in the L2 learners, in the posterior region ($F(1, 83) = 8.35$, $p = .005$, $\eta_p^2 = .09$) and the medial region ($F(1, 83) = 4.12$, $p = .046$, $\eta_p^2 = .05$; for the anterior region $p = .495$).

*3.2.1.2. Analysis of correctly responded trials.* The above results show that the native speakers displayed a P600 effect when processing semantic reversal sentences, but the L2 learners did not. However, this analysis included trials with incorrect responses, which occurred more often in L2 learners than in native speakers. It is possible that in incorrectly judged sentences, readers did not even experience a conflict between syntactic and heuristic-semantic information, because they largely ignored the first and settled for the latter. To investigate L2 learner's processing in trials where they must have used the syntactic information (indicated by a correct response), we carried out a second analysis in which we included only those trials where the L2 learners made a correct plausibility judgment. To make the between-group results more comparable, we applied the same criteria to the L1 group as well. In this analysis, we excluded three native speakers and 17 additional L2 learners because of their high error rates in the judgment task. Fig. 3 shows the ERPs of both groups.
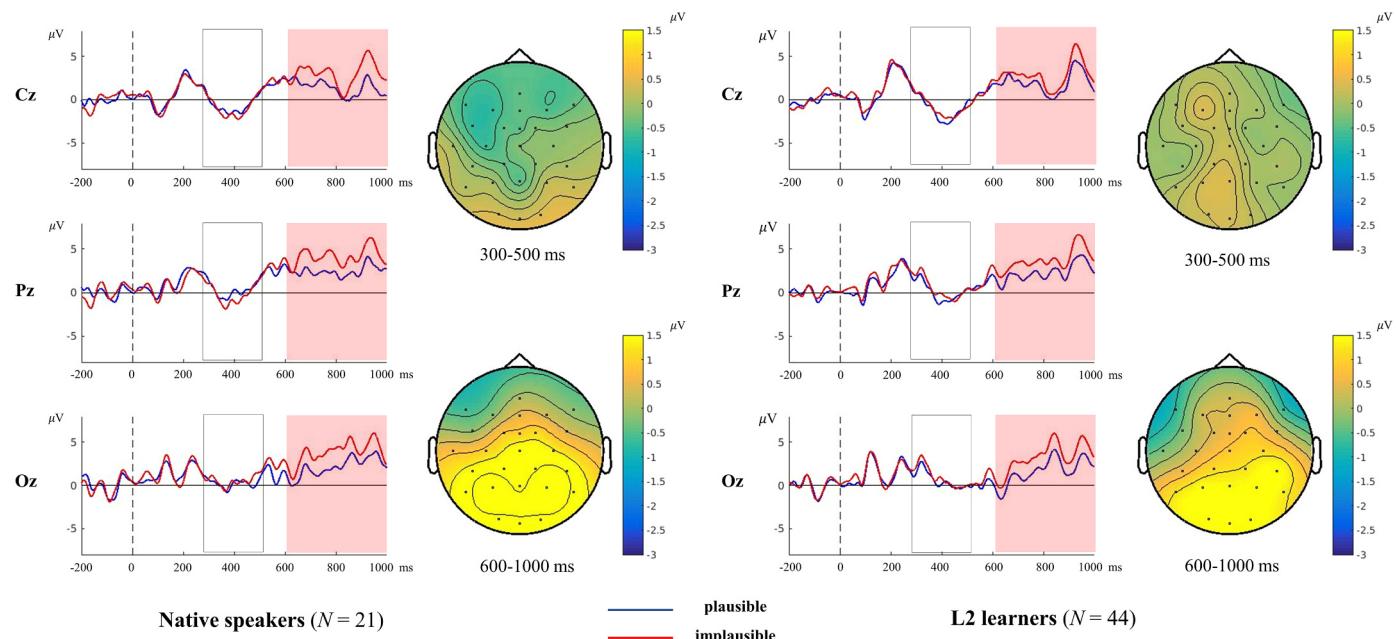
*3.2.1.2.1. Native speakers.* No differences between plausibility conditions were revealed in the N400 time window (see Table 3). In the P600 time window, there was a marginally significant effect of plausibility and a significant interaction between region and

plausibility. Follow-up analyses showed that the effect of plausibility was significant at posterior sites ($F(1, 20) = 15.40$, $p = .001$, $\eta_p^2 = .44$), marginally significant at medial sites ($F(1, 20) = 4.24$, $p = .053$, $\eta_p^2 = .18$), and n.s. at anterior sites ($p = .601$).
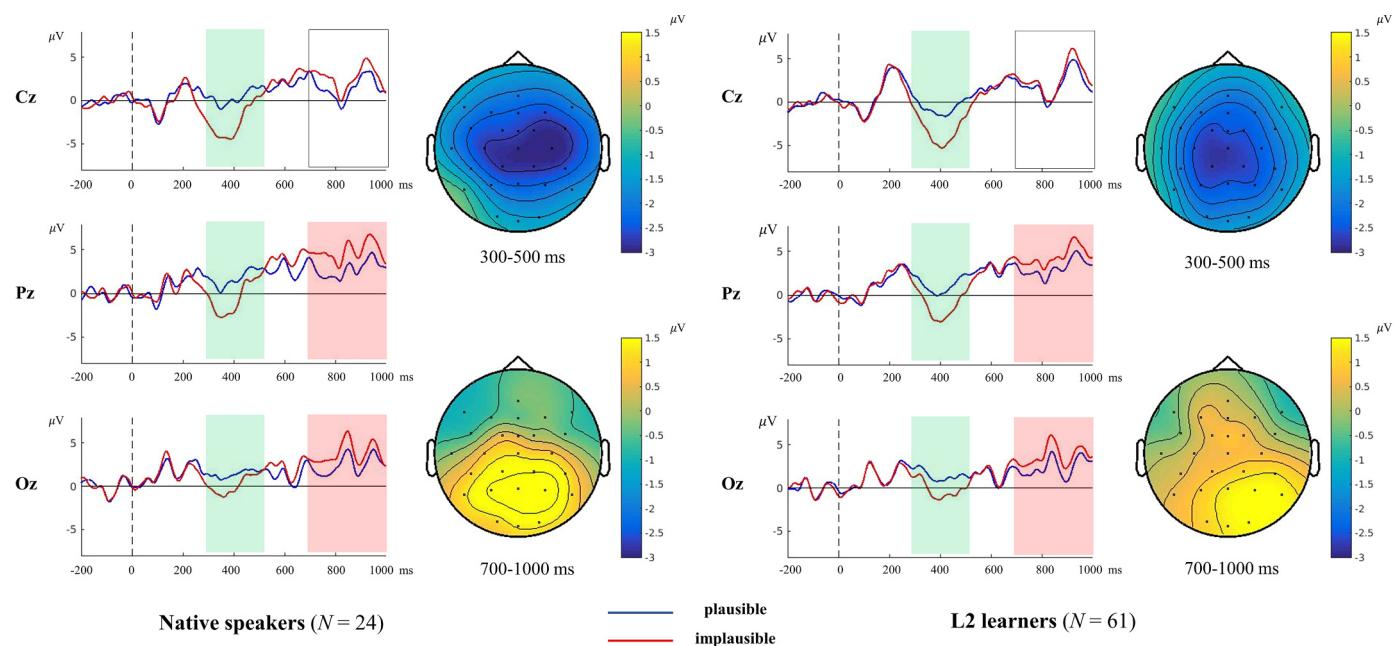
*3.2.1.2.2. L2 learners.* In the N400 time window, no significant effects were found. In the P600 time window, there was a significant effect of plausibility, an interaction between plausibility and region, a three-way interaction between plausibility, region and hemisphere, and a marginally significant interaction between plausibility and hemisphere. Follow-up analyses for each region showed that: at posterior sites, there was a significant effect of plausibility ($F(1, 43) = 12.48$, $p = .001$, $\eta_p^2 = .23$), but no interaction between plausibility and hemisphere ($p = .709$); at medial sites, there was a marginally significant effect of plausibility ($F(1, 43) = 3.77$, $p = .059$, $\eta_p^2 = .08$), as well as an interaction between plausibility and hemisphere ($F(1, 43) = 6.15$, $p = .017$, $\eta_p^2 = .13$), with a significant effect of plausibility at the right hemisphere ($F(1, 43) = 5.19$, $p = .028$, $\eta_p^2 = .11$) but not at the left hemisphere ($F(1, 43) = 2.06$, $p = .158$, $\eta_p^2 = .05$). There were no significant effects at anterior sites (both $ps > .107$).

*3.2.1.2.3. Between-group comparison.* No difference between groups was found in the N400 time window. In the P600 time window, there was a significant three-way interaction between plausibility, region, and group. However, follow-up analyses revealed no significant interaction between plausibility and group in any of the three regions (all $ps > .436$).

*3.2.1.3. Preliminary summary.* When including both correct and incorrect responses, we found a P600 effect for implausible relative to plausible sentences in the native speakers, but not in the L2 learners. However, when only the correctly responded trials were taken into account, the L2 learners showed a P600 effect as the native speakers did. The P600 effect in the L2 learners was slightly lateralized to the right hemisphere compared to the native speakers. Nevertheless, as the lateralization did not occur as an interaction between groups, the overall patterns remain identical in both groups.



**Fig. 3.** Grand-averaged ERPs and topographies for the critical verb in the experimental condition (correctly responded trials only), for the native speakers (left, $N = 21$) and the L2 learners (right, $N = 44$), from three midline electrodes (Cz, Pz, Oz). The time windows of analysis (N400: 300–500 ms; P600: 600–1000 ms) are marked in light red (significant) or as an empty frame (not significant). Significances are calculated based on the lateral analysis, not the actual electrodes that are shown. Scalp maps show the topographies of the difference between plausible and implausible sentences.

**Fig. 4.** Grand-averaged ERPs and topographies for the critical verb in the control condition (all trials) for the native speakers (left, *N* = 24) and the L2 learners (right, *N* = 61), from three midline electrodes (Cz, Pz, Oz). The time windows of analysis (N400: 300–500 ms; P600: 700–1000 ms) with significant effects are marked in light green and light red (significant), respectively, or as an empty frame (not significant). Significances are calculated based on the lateral analysis, not the actual electrodes that are shown. Scalp maps show the topographies of the difference between plausible and implausible sentences.

### 3.2.2. Control condition

Fig. 4 shows the ERPs of the control (standard semantic anomalous) condition in both groups, including both correct and incorrect responses. Inspection of the data showed that a P600 effect occurred in both groups, but was shifted to a later time window compared to the experimental condition, probably due to the prior occurrence of the N400. Therefore, we used a slightly later time window (700 to 1000 ms) for the analysis of the P600 effect. In Table 4, we report all the *F*-statistics for the main effects and interactions with the variable Plausibility; in the main text we follow up significant interactions (i.e., *p* < .05) by simple effect tests.

### 3.2.2.1. Native speakers.
In the N400 time window, the analysis revealed a significant effect of plausibility and a significant interaction between plausibility and region. Follow-up analyses revealed a larger N400 effect in the anterior (*F*(1, 23) = 23.15, *p* < .001, $\eta_p^2$ = .50) and medial regions (*F*(1, 23) = 21.41, *p* < .001, $\eta_p^2$ = .48) than that in the posterior region (*F*(1, 23) = 12.76, *p* = .002, $\eta_p^2$ = .36). In the P600 time window, the analysis showed a significant interaction between plausibility and region. Follow-up analyses showed that the effect of plausibility was restricted to posterior sites (*F*(1, 23) = 5.58, *p* = .027, $\eta_p^2$ = .19; both other two regions: *ps* > .104).

**Table 4**
F-statistics of the 2 × 2 × 3 lateral analysis performed on the N400 and the P600 time windows in the ERP analysis for both groups (control condition).

| Group/Condition | | Control Condition | |
| --- | --- | --- | --- |
| | | (all trials) | |
| | Factors*/Time | N400 | P600 |
| Native speakers | P | ***F* = 20.75, *p* < .001, $\eta_p^2$ = .47** | *F* = 1.78, *p* = .186, $\eta_p^2$ = .07 |
| | P × H | *F* = 2.40, *p* = .135, $\eta_p^2$ = .09 | *F* < 1 |
| | P × R | ***F* = 11.02, *p* < .001, $\eta_p^2$ = .50** | ***F* = 7.29, *p* = .004, $\eta_p^2$ = .40** |
| | P × H × R | *F* < 1 | *F* < 1 |
| L2 learners | P | ***F* = 108.64, *p* < .001, $\eta_p^2$ = .64** | ***F* = 5.98, *p* = .017, $\eta_p^2$ = .09** |
| | P × H | *F* < 1 | *F* = 3.61, *p* = .062, $\eta_p^2$ = .06 |
| | P × R | ***F* = 9.15, *p* < .001, $\eta_p^2$ = .24;** | ***F* = 3.72, *p* = .030, $\eta_p^2$ = .11** |
| | P × H × R | *F* < 1 | ***F* = 5.01, *p* = .009, $\eta_p^2$ = .15** |
| Between-group comparison | G × P | ***F* = 27.46, *p* < .001, $\eta_p^2$ = .25** | *F* < 1 |
| | G × P × H | ***F* = 5.02, *p* = .028, $\eta_p^2$ = .06** | *F* < 1 |
| | G × P × R | ***F* = 5.61, *p* = .005, $\eta_p^2$ = .12** | ***F* = 4.70, *p* = .012, $\eta_p^2$ = .10** |
| | G × P × H × R | *F* = 2.53, *p* = .086, $\eta_p^2$ = .06 | *F* < 1 |

*P = Plausibility; H = Hemisphere; R = Region; G = Group.
Significant effects are printed in bold.

*3.2.2.2. L2 learners.* In the N400 time window, the analysis showed a significant effect of plausibility and a significant interaction between plausibility and region. The effect of plausibility was widely distributed across the scalp, with slightly larger effects in the medial ($F$(1, 60) = 97.13, $p$ < .001, $\eta_p^2$ = .62) and posterior region ($F$(1, 60) = 93.70, $p$ < .001, $\eta_p^2$ = .61) than in the anterior region, though it was still highly significant there ($F$(1, 60) = 71.28, $p$ < .001, $\eta_p^2$ = .54). In the P600 time window, the analysis showed a significant effect of plausibility, an interaction between plausibility and region, and a three-way interaction between plausibility, region and hemisphere. Follow-up analyses of the three regions showed: At posterior sites, there was a significant effect of plausibility ($F$(1, 60) = 17.27, $p$ < .001, $\eta_p^2$ = .22) and a significant interaction between plausibility and hemisphere ($F$(1, 60) = 9.93, $p$ = .003, $\eta_p^2$ = .14), with the effect of plausibility being significant in both hemispheres, but larger in the right ($F$(1, 60) = 25.36, $p$ < .001, $\eta_p^2$ = .30) than in the left hemisphere ($F$(1, 60) = 7.79, $p$ = .007, $\eta_p^2$ = .12). At medial sites, there was a significant effect of plausibility ($F$(1, 60) = 5.27, $p$ = .025, $\eta_p^2$ = .08) and a marginally significant interaction between plausibility and hemisphere ($F$(1, 60) = 3.97, $p$ = .051, $\eta_p^2$ = .06), with the effect being restricted to the right hemisphere ($F$(1, 60) = 7.48, $p$ = .008, $\eta_p^2$ = .11) rather than in the left hemisphere ($F$(1, 60) = 2.58, $p$ = .114, $\eta_p^2$ = .04). No significant effects were observed at anterior sites (both $ps$ > .425).

*3.2.2.3. Between-group comparison.* The analysis of the N400 time window showed a significant interaction between plausibility and groups, with the main difference between implausible and plausible sub-conditions being larger in the L1 than the L2 group (L1: *Mean_diff*$_{Cz}$ = 1.16, $SD_{Cz}$ = 4.40; L2: *Mean_diff*$_{Cz}$ = 0.64, $SD_{Cz}$ = 3.12). The interaction between plausibility and region and the interaction between plausibility and hemisphere were both smaller for the L2 group than the L1 group (see Table 4). The analysis of the P600 time window showed a significant interaction between plausibility, region, and group. However, subsequent analyses for the three regions revealed no significant interaction between plausibility and group in any region (all $ps$ > .455).[4]

*3.2.2.4. Preliminary summary.* Both the L1 and the L2 groups showed a robust N400 effect in the control condition. In addition, both groups showed a P600 effect. The N400 effect in the native speakers had a more centro-frontal distribution compared to the L2 learners, whereas the P600 effect in the L2 learners was more lateralized to the right hemisphere compared to the native speakers.

*3.3. Relationship with individual differences*

From the results above, we learned that when the L2 learners made a correct judgement as the native speakers did on the semantic reversal sentences, they showed a native-like P600 effect. However, when including all trials, also those where they incorrectly responded to a semantic reversal sentence, the P600 effect was attenuated. It is possible that the extent of this attenuation differs greatly between individuals. We explored in a multiple linear regression analysis how L2 learners' behavioural sensitivity to semantic reversal anomalies (reflected in the *d'* score) was related to their P600 amplitude in this condition, and to what extent both of these could be potentially predicted by individual

---

[4] To better compare the results of the control condition to those of the experimental condition, we applied the same analysis to the subset of trials where the participants made a correct plausibility judgement. The L2 learners made fewer mistakes in the control condition and no one had to be excluded based on the error rates. Anyhow, for the sake of consistency, we used the same subset of participants as in the experimental condition. The pattern of results did not differ from the ones of the analysis of all trials.

**Table 5**
Means, standard deviation and Pearson correlation matrix for *d'*, P600 amplitude, and individual difference measures in L2 learners (N = 61).

|  | M | SD | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| 1. *d'* | 1.32 | 0.54 |  |  |  |  |  |  |
| 2. P600 amplitude | 0.89 | 3.47 | .28* |  |  |  |  |  |
| 3. OST | 0.80 | 0.15 | .35** | .01 |  |  |  |  |
| 4. RST | 0.74 | 0.13 | .24 | .07 | .53** |  |  |  |
| 5. vocabulary proficiency | 83.46 | 12.48 | .54** | .13 | .18 | .29* |  |  |
| 6. grammar proficiency | 72.19 | 11.42 | .49** | .14 | .29* | .28* | .64** |  |
| 7. LOE | 14.16 | 5.70 | .40** | .09 | .15 | .23 | .53** | .47** |

Note. OST = *operation span test*; RST = *reading span test*; LOE = *length of exposure*, SD = *Standard Deviation*.
\* $p$ < .05.
\*\* $p$ < .01.

**Table 6**
Beta-coefficients, *t*-test statistics and *p* values for the regression model (final step) with *d'* as dependent variable.

|  | B | SE B | β | t | p |
|---|---|---|---|---|---|
| vocabulary score | 0.02 | 0.01 | 0.35 | 2.35 | .023 |
| grammar score | 0.01 | 0.01 | 0.15 | 1.03 | .305 |
| LOE | 0.01 | 0.01 | 0.13 | 0.99 | .328 |
| OST | 0.95 | 0.47 | 0.26 | 2.02 | .048 |
| RST | − 0.27 | 0.54 | − 0.06 | − 0.49 | .627 |

Note. OST = *operation span test*; RST = *reading span test*; LOE = *length of exposure*, B = *Unstandardized Coefficients*, SE = *Standard Error*, β = *Standardized Coefficients*.

differences. The P600 amplitude for each participant was calculated by averaging the difference between the implausible and plausible sub-conditions in all trials across all posterior sites (P7, P8, Pz, P4, P3, O1, Oz, O2) in the 600 to 1000 ms time window. Table 5 shows the means and standard deviations for these predictor and dependent variables as well as the intercorrelations between them.

We used five individual difference measures: L2 vocabulary proficiency, L2 grammar proficiency, the length of exposure to the Dutch language, and operation span as well as reading span (as two measures of working memory). First, there was a significant positive correlation between the amplitude of the P600 effect and the *d'* scores. Furthermore, the majority of individual difference measures were significantly and positively correlated with *d'*, except for reading span. In contrast, none of the individual measures were correlated with the P600 amplitude.

We then performed two separate hierarchical linear regressions using the five individual differences to predict the *d'* score and the P600 amplitude, respectively. Predictors were introduced in a stepwise fashion in three clusters: L2 proficiency, length of exposure, and working memory. Our reasoning was that proficiency differences would probably be most fundamental in explaining L2 processing variance, followed by length of exposure, which also contributes to proficiency, and then working memory. We found that including the first block of predictors, L2 proficiency (Model 1), explained a significant 32% of the variance of *d'* scores ($F_{change}$ (2, 58) = 13.83, $p$ < .001), with vocabulary scores being a significant predictor (β = .38, $t$ = 2.72, $p$ = .008): The larger an L2 learner's vocabulary size, the higher his or her *d'* was. The grammar score was not significant as a predictor ($p$ = .090). Adding length of exposure as a second block of predictor (Model 2) did not significantly improve the prediction of *d'* scores ($\Delta R^2$ = .01, $F_{change}$ (1, 57) = 0.95, $p$ = .333), nor did adding the two measures of working memory (Model 3; $\Delta R^2$ = .05, $F_{change}$ (2, 55) = 2.22, $p$ = .119). The latter result occurred despite the fact that the operation span score reached significance as a predictor of *d'* scores. Beta-coefficients for the final model are reported in Table 6.

On the contrary, we found no significant explanation of the variance

in P600 amplitudes (all *p*s > .510), with no predictor being significant (all *p*s > .529).

To reduce the potential influence of task-specific variance and to examine general contributions of possible latent variables, we also did an exploratory factor analysis and then used the latent factor scores as predictors in the regression. The factor analysis revealed two factors. All predictors loaded on the first factor about equally, which explained the most variance of *d'* but not the P600 amplitude (neither did the second factor), and therefore does not answer the question which (latent) variable is the best predictor. We provide more details on this factor analysis in Appendix C.

### 3.3.1. Preliminary summary

The linear regression analyses showed that individuals' vocabulary size and, to some extent, their working memory capacity, in particular, operation span, can positively predict how well they can detect semantic reversal anomalies behaviourally (even though adding operation span to the regression model did not significantly increase the amount of explained variance). In contrast, we did not find any significant predictors of individual's P600 amplitudes.

## 4. Discussion

In the current study, we investigated how the processing of syntax and semantics interact in L2 learners during sentence comprehension. To this end, we made use of the "semantic P600" effect which has previously been reported only for L1 speakers, and which has generally been interpreted as an indicator of the syntax-semantics conflict (e.g., Kim and Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007). We compared behavioural performance and neural signatures of L2 learners and native speakers when processing sentences with semantic reversal anomalies which usually give rise to the semantic P600. By this, we examined whether "shallow" or "good enough" processing (i.e., relying more on semantics and less on syntax during (complex) sentence interpretation) occurs to a greater extent in L2 learners than in native speakers (Clahsen and Felser, 2006, 2017; Guo et al., 2009; Roberts, 2012).

First of all, our behavioural results show that German learners of Dutch were as able to detect standard semantic violations in the control condition as were native speakers, indicating equivalent degrees of general semantic sensitivity in both groups. Compared to standard semantic anomalies, detecting semantic reversal anomalies was already harder for native speakers (i.e., smaller *d'*), confirming the observation that native speakers often use a "good enough" approach when processing syntactically complex, non-canonical sentences (e.g., Ferreira, 2003; Ferreira and Patson, 2007; Traxler, 2014). The processing problem became even more salient for German learners of Dutch (i.e., even smaller *d'*): L2 learners more often judged the sentences with reversed roles ("The dog that talked to the girl…") as "plausible", supporting the idea that L2 learners misinterpreted the sentences as being in line with their world knowledge. Thus, they seem to have engaged more often in shallower, less complete, and even less syntactically accurate processing compared to native speakers. Note, however, that L2 speakers also judged the non-reversed sentences ("The girl that talked to the dog…") less often as "plausible" than the native speakers did, pointing at an increased level of general uncertainty in interpreting the relative-clause sentences in the experimental condition. Recall that the mere order of nouns in the sentence was not a valid cue to role assignment, as some of our filler sentences contained the object of the relative clause in the first position (e.g., "The water from which the bull drank…"). Thus, the "NVN heuristic" (the first noun is the agent; Ferreira, 2003) did not work in this experimental context. Rather, each sentence required a complete syntactic analysis for its correct meaning to be derived, something that L2 speakers apparently struggled with to a greater extent than native speakers did.

In terms of ERPs, we found a P600 effect elicited by semantic

reversal anomalies in native Dutch speakers; in contrast, the German learners of Dutch failed to show a P600 effect when both correct and incorrect responses were included in the analysis. Together with the behavioural results, this reflects a generally reduced sensitivity to the semantic reversal manipulation in L2 learners. In native speakers, the conflict between the semantically plausible and the syntactically correct interpretation in the semantic reversal anomaly condition was associated with a P600, while – when looking at all trials regardless of the plausibility judgments – this was not the case for the L2 group (see VanRullen, 2011, for arguments in favour of including all trials, not only correctly responded ones, in EEG analyses). The fact that L2 learners were indistinguishable from L1 speakers in terms of behaviour and ERP effects in the control condition rules out the possibility that a general difference in semantic sensitivity caused the observed non-nativelike effects for L2 learners for semantic reversal anomalies. Note, though, that we did not probe sentence interpretation directly, as in, for instance, a picture matching task. Therefore, we cannot be certain about the exact roles which participants assigned to the sentence constituents. However, most errors that L2 speakers made were those where they judged semantic reversal sentences as "plausible", suggesting that they often assumed the most plausible role assignment despite it being in contrast with the sentence's syntax.

Nevertheless, as shown by the second ERP analysis of correctly responded trials only, when L2 learners successfully detected the conflict caused by semantic reversal anomalies (i.e., made an "implausible" response), they also showed a native-like semantic P600 effect. This suggests that native-like processing and resolution of the syntax-semantics conflict was not absent in L2 speakers, but occurred in fewer trials, perhaps only in those sentences that they found relatively easy and that taxed attentional resources to a lesser degree. In contrast, at least in a substantial number of trials with incorrect responses, this conflict was either not detected (e.g., because the participants immediately settled for the semantically plausible, but incorrect sentence interpretation), or not attempted to be resolved as it would be the case in native speakers. Apparently, even with extremely high lexical (all words except the critical verb were cognates) and syntactic overlap (the sentences can be translated word-by-word) between German and Dutch, the knowledge of L2 syntax can be hard to access in real-time comprehension (Clahsen and Felser, 2006). Thus, the elimination of the syntax-semantics conflict due to increased reliance on semantic heuristics seems to occur only for a certain number of mostly incorrectly responded sentences, which is however large enough to reduce the P600 effect when averaging within conditions.

Our study replicated the semantic P600 effect established in L1 speakers (e.g., Frenzel et al., 2011; Hoeks et al., 2004; Kim and Osterhout, 2005; Paczynski and Kuperberg, 2012; van Herten et al., 2006, 2005) and extended it to the population of L2 learners. Unlike previous studies (e.g., van Herten et al., 2006), where participants read the sentences merely for comprehension, we used a plausibility judgment task, with the following motivation: First, L2 learners are different from native speakers in that they are more likely to misinterpret a sentence, especially when the sentence is structurally complex. Therefore, a judgment task can help provide at least some approximate information about processing differences between sentences, which has also been proven critical in the current analyses. Second, it is very common in the sentence processing ERP literature to use such a judgment task. Although a judgment task may give rise to a more pronounced P600 effect (Gunter and Friederici, 1999; Kolk et al., 2003), our results in native speakers did not seem to differ much from those in van Herten et al. (2006) in latencies or amplitudes.

To extend our investigation to an L2 while avoiding the possible confounding from the syntactic differences between L1 and L2, we chose the similar language pair of Dutch and German. However, we have to consider an alternative explanation for the observed differences between German learners of Dutch and Dutch native speakers. Compared to the other languages (e.g., English and Dutch) investigated

in previous L1 studies, German, our L2 learners' native language, is more flexible in word order because of the use of case markers. As a consequence, German-Dutch speakers may rely less on word order as a cue to interpret sentences also in their L2 Dutch and consequently experience less conflict between syntax (here: word order) and semantics in the case of semantic reversal anomalies (Bornkessel-Schlesewsky et al., 2011). Nevertheless, this alternative explanation can be also questioned: by using an identical sentence structure between Dutch and German, the experimental sentences in the current study (e.g., "Het hondje dat tegen het meisje praatte speelde met de bal." *[The dog that talked to the girl was playing with the ball.]*) are as unambiguous in their German translation ("Das Hündchen das mit dem Mädchen sprach spielte mit dem Ball.") as they are in Dutch (i.e., the dog has to be the subject because of the preposition before the indirect object). The only study so far on how German native speakers process semantic reversal anomalies (Schlesewsky and Bornkessel-Schlesewsky, 2009) used materials which were not comparable to our current study (i.e., case marking played a major role to resolve role ambiguity). In that study, a biphasic N400-P600 pattern was found. Therefore, the fact that we observed a null effect overall and a native-like P600 pattern in the correctly responded trials makes it unlikely that a transfer from L1 to L2 is responsible for the null effect in our ERP results.

Apparently, L2 learners display sensitivity to conflicting syntactic and semantic information only occasionally, while native speakers show such sensitivity consistently. While even in native speakers, "the language comprehension system uses a mixture of heuristics and syntactic algorithms" (Ferreira, 2003, p. 197), L2 speakers may more often resort to the heuristic route, possibly due to attentional and cognitive limitations. This is also supported by the finding of a significant positive correlation between the amplitude of the P600 effect and the *d'* for semantic reversal anomalies: individuals with better behavioural discrimination ability between anomalous and canonical sentences tended to show larger P600 amplitudes. Although individuals can also differ in their L1 performance, the influence of individual differences seems to be far greater during L2 performance in general (e.g., Bialystok and Hakuta, 1994). Therefore, in additional analyses of our dataset that necessarily had to be exploratory in nature given the limited sample size, we investigated whether aspects of a learner's background modulated their overall sensitivity to our manipulation. We found that L2 proficiency, and in particular, L2 vocabulary, predicted L2 learners' ability to detect semantic reversal anomalies: the more proficient an individual was in the L2, the better he or she was able to detect the conflict between syntactic parsing and heuristics in the L2 (as reflected by the *d'* score). This suggests that high-proficient L2 learners may assign more weight to syntactic information than low-proficient learners, and thus be more similar to native speakers in that respect. Interestingly, it was vocabulary size rather than grammar skill that played the larger role here. A possible explanation is that individuals with higher vocabulary skills may possess more stable and well-defined lexical representations. Low-quality L2 lexical representation may delay and attenuate effects of syntactic structure in L2 sentence processing (Hopp, 2018). Note, however, that the proficiency tests we used in the current study have not been standardized and therefore the results need to be interpreted with caution. Future research should explore the role of vocabulary knowledge in L2 (syntactic) sensitivity in depth. On the other hand, we found a positive relationship between operation span, as

a measurement of working memory capacity, and sensitivity to semantic reversal anomalies (*d'*) (but note that working memory capacity, when added as a final predictor cluster after proficiency, did not significantly increase the quality of the prediction any more). It is likely that L2 processing alone is more cognitively demanding than L1 processing due to higher cognitive effort at all processing levels, leaving less room for additional complex computations (e.g., McDonald, 2006; Hopp, 2014). During the interplay between syntax and semantics, individual working memory capacity determines how well one can maintain not yet integrated syntactic information in working memory during on-going sentence processing. When one's working memory capacity is low, the capacity-demanding syntactic route becomes weaker and is more likely to be overruled by the less demanding heuristic routine. Surprisingly, all of the above-mentioned individual difference measurements failed to predict individual P600 amplitudes in L2 speakers. This may seem contradictory to the findings by Tanner and colleagues (Tanner et al., 2014, 2013), who succeeded in identifying aspects of learners background (e.g., proficiency, age of arrival) that can predict the relative brain response type (N400 or P600) during L2 morphosyntactic processing. However, the situation in the present study on semantic reversal anomalies was different, as we did not observe a subgroup of L2 speakers who predominantly showed an N400 instead of a P600.

### 4.1. Summary

The present study showed that during online sentence processing, L2 learners are less sensitive than native speakers to the conflict between syntactically- and semantically- (heuristics-) based processing. While even native speakers sometimes rely on a "good enough" approach in processing syntactically complex sentences (Ferreira et al., 2002; Ferreira and Patson, 2007), syntactic processing in L2 learners appears to be even shallower and less detailed than that in native speakers (Clahsen and Felser, 2006, 2017). This difference was observable even in an L1-L2 combination with extremely high lexical and syntactic overlap, German and Dutch. However, in the situation where L2 learners do successfully detect the conflict as reflected in correct "implausible" judgments for semantic reversal anomalies, they did show native-like neural responses, i.e. a P600. This is the first time the semantic P600 effect has been investigated in an L2 population, providing new insights on the interaction between syntax and semantics during online L2 processing.

### Acknowledgements

### Data availability

Data are available from the Donders Institute for Brain, Cognition and Behaviour repository at http://hdl.handle.net/11633/di.dcc.DSC_2016.00216_591.

### Appendix A. Experimental materials

*A.1 Experimental condition*

(60 sentences, **IC**: incorrect version with semantic reversal anomalies, the critical words are underlined)

1. Het meisje dat tegen het hondje <u>praatte</u> speelde met de bal. *The girl that <u>talked</u> to the dog was playing with the ball.*
   **IC:** Het hondje dat tegen het meisje <u>praatte</u> speelde met de bal.
2. De pinguïn die onder de ijsschots <u>dook</u> bevond zich op de Zuidpool. *The penguin that <u>dived</u> under the ice shelf was located at the South Pole.*

**IC:** De ijsschots die onder de pinguïn <u>dook</u> bevond zich op de Zuidpool.

3. De acrobaat die op de ladder <u>klom</u> viel plotseling naar beneden. *The acrobat that <u>climbed</u> on the ladder suddenly fell down.*
   **IC:** De ladder die op de acrobaat <u>klom</u> viel plotseling naar beneden.

4. De apen die naar de bananen <u>graaiden</u> hingen in de bomen. *De monkeys that <u>grasped</u> the bananas were hanging in the trees.*
   **IC:** De bananen die naar de apen <u>graaiden</u> hingen in de bomen.

5. De jongeman die tegen de bal <u>schopte</u> brak de vaas. *The young man who <u>kicked</u> the ball broke the vase.*
   **IC:** De bal die tegen de jongeman <u>schopte</u> brak de vaas.

6. De student die van de fiets <u>afstapte</u> was in het centrum van de stad. *The student who <u>got off</u> the bike was in the city centre.*
   **IC:** De fiets die van de student <u>afstapte</u> was in het centrum van de stad.

7. De Italiaan die van de koffie <u>proefde</u> had een goede smaak. *The Italian who <u>tasted</u> of the coffee had a good taste.*
   **IC:** De koffie die van de Italiaan <u>proefde</u> had een goede smaak.

8. De man die over de bus <u>mopperde</u> stond te wachten bij de bushalte. *The man who <u>grumbled</u> about the bus was waiting at the bus stop.*
   **IC:** De bus die over de man <u>mopperde</u> stond te wachten bij de bushalte.

9. De kat die naar de vogel <u>uithaalde</u> werd niet gezien door de oude vrouw. *The cat that <u>took a swing</u> at the bird was not seen by the old woman.*
   **IC:** De vogel die naar de kat <u>uithaalde</u> werd niet gezien door de oude vrouw.

10. De poedels die naar de kinderen <u>blaften</u> waren van de directeur. *The puddles that <u>barked</u> to the kids were the director's.*
    **IC:** De kinderen die naar de poedels <u>blaften</u> waren van de directeur.

11. De fietsers die tegen de voetgangers <u>aanreden</u> pasten niet op. *The bikers who <u>drove into</u> the pedestrians were not paying attention.*
    **IC:** De voetgangers die tegen de fietsers <u>aanreden</u> pasten niet op.

12. De alcoholist die naar de fles <u>keek</u> stond in de keuken. *The alcoholic who <u>looked</u> at the bottle was standing in the kitchen.*
    **IC:** De fles die naar de alcoholist <u>keek</u> stond in de keuken.

13. De familie die naar de muziek <u>luisterde</u> kwam uit Turkije. *The family that <u>listened</u> to the music was from Turkey.*
    **IC:** De muziek die naar de familie <u>luisterde</u> kwam uit Turkije.

14. De baby die naar de kat <u>glimlachte</u> had heerlijk geslapen. *The baby that <u>smiled</u> at the cat had slept very well.*
    **IC:** De kat die naar de baby <u>glimlachte</u> had heerlijk geslapen.

15. De toeristen die naar de feesten <u>vertrokken</u> waren er om Nieuwjaar te vieren. *The tourists who <u>left</u> for parties were there to celebrate New Year.*
    **IC:** De feesten die naar de toeristen <u>vertrokken</u> waren er om Nieuwjaar te vieren.

16. De bokser die in de ring <u>stapte</u> stond in de spotlights. *The boxer who <u>stepped</u> into the ring was stood in the spotlight.*
    **IC:** De ring die in de bokser <u>stapte</u> stond in de spotlights.

17. De haas die van de wolf <u>schrok</u> verborg zich snel in de struiken. *The hare that <u>was scared</u> by the wolf quickly hid in the bush.*
    **IC:** De wolf die van de haas <u>schrok</u> verborg zich snel in de struiken.

18. De studenten die voor de examens <u>slaagden</u> waren op de universiteit. *The students who <u>passed</u> the exams were at the university.*
    **IC:** De examens die voor de studenten <u>slaagden</u> waren op de universiteit.

19. De vrouw die door de heg <u>gluurde</u> stond er al een lange tijd. *The woman who <u>was peeping</u> trough the hedge already stood there for a long time.*
    **IC:** De heg die door de vrouw <u>gluurde</u> stond er al een lange tijd.

20. De kinderen die over de apen <u>vertelden</u> sprongen op en neer. *The children who <u>told</u> about the monkeys were jumping up and down.*
    **IC:** De apen die over de kinderen <u>vertelden</u> sprongen op en neer.

21. De vreemde die aan de deur <u>belde</u> werd nat van de regen. *The stranger who <u>ringed</u> at the door was getting wet by the rain.*
    **IC:** De deur die aan de vreemde <u>belde</u> werd nat van de regen.

22. Het schip dat op het water <u>dreef</u> ging wild op en neer. *The ship that <u>floated</u> on the water wildly went up and down.*
    **IC:** Het water dat op het schip <u>dreef</u> ging wild op en neer.

23. De acteur die vanaf de tribune <u>schreeuwde</u> was in het theater. *The actor who <u>yelled</u> from the stand was in the theater.*
    **IC:** De tribune die vanaf de acteur <u>schreeuwde</u> was in het theater.

24. De batterij die bij de camera <u>hoorde</u> was kapot gegaan. *The battery that <u>belonged</u> to the camera was broken.*
    **IC:** De camera die bij de batterij <u>hoorde</u> was kapot gegaan.

25. De dame die op de kleding <u>lette</u> was stijlvol en elegant. *The lady who <u>paid attention</u> to the clothing was stylish and elegant.*
    **IC:** De kleding die op de dame <u>lette</u> was stijlvol en elegant.

26. Het kevertje dat door het zand <u>kroop</u> was bruin van kleur. *The little beetle that <u>crawled</u> through the sand had the color brown.*
    **IC:** Het zand dat over het kevertje <u>kroop</u> was bruin van kleur.

27. De patiënt die tegen de griep <u>vocht</u> kwam uit Zuid-Afrika. *De patient who <u>fought</u> against the flu was from South-Africa.*
    **IC:** De griep die tegen de patiënt <u>vocht</u> kwam uit Zuid-Afrika.

28. Het meisje dat in het boek <u>tekende</u> was nog niet heel oud. *The girl that <u>drew</u> in the book was not very old yet.*
    **IC:** Het boek dat in het meisje <u>tekende</u> was nog niet heel oud.

29. Het mannetje dat tegen het standbeeld <u>mompelde</u> zag er vreemd uit. *The little man who <u>mumbled</u> to the statue was looking strange.*
    **IC:** Het standbeeld dat tegen het mannetje <u>mompelde</u> zag er vreemd uit.

30. De bezoeker die op de sofa <u>rustte</u> werd gekrabd door de kat. *The visitor who <u>rested</u> on the couch was scratched by the cat.*
    **IC:** De sofa die op de bezoeker <u>rustte</u> werd gekrabd door de kat.

31. De suiker die in de thee <u>oploste</u> smaakte maar een klein beetje zoet. *The sugar that <u>dissolved</u> in the tea tasted a bit sweet.*
    **IC:** De thee die in de suiker <u>oploste</u> smaakte maar een klein beetje zoet.

32. Het zonlicht dat op het water <u>schitterde</u> gaf de schrijver inspiratie. *The sun light that <u>shone</u> on the water gave the writer inspiration.*
    **IC:** Het water dat op het zonlicht <u>schitterde</u> gaf de schrijver inspiratie.

33. De producten die aan de regels <u>voldeden</u> waren goed voor de consumenten. *The products that <u>met</u> the regulations were good for the costumers.*
    **IC:** De regels die aan de producten <u>voldeden</u> waren goed voor de consumenten.

34. De hippie die op de muur <u>schilderde</u> stond in een vervallen deel van de stad. *The hippie who <u>painted</u> on the wall was standing in a dilapidated part of the city.*
    **IC:** De muur die op de hippie <u>schilderde</u> stond in een vervallen deel van de stad.

35. De man die tegen de boomstam <u>leunde</u> stond daar in de zomerhitte. *The man who <u>leaned</u> against the tree trunk was standing there in the summer heat.*
    **IC:** De boomstam die tegen de man <u>leunde</u> stond daar in de zomerhitte.
36. Het leeuwtje dat uit het circus <u>ontsnapte</u> werd gekocht door een rijke man. *The lion who <u>escaped</u> from the circus was bought by a rich man.*
    **IC:** Het circus dat uit het leeuwtje <u>ontsnapte</u> werd gekocht door een rijke man.
37. De pianist die met de improvisatie <u>doorging</u> fascineerde de toehoorders. *The pianist who <u>continued</u> the improvisation fascinated the listeners.*
    **IC:** De improvisatie die met de pianist <u>doorging</u> fascineerde de toehoorders.
38. De schoolkinderen die om de tekenfilms <u>grinnikten</u> irriteerden de leraar. *The school kids that <u>chuckled</u> about the cartoons were irritating the teacher.*
    **IC:** De tekenfilms die om de schoolkinderen <u>grinnikten</u> irriteerden de leraar.
39. De kameel die voor de zandstorm <u>schuilde</u> werd gefilmd door een cameraman. *The camel that <u>sought shelter</u> from the sand storm was filmed by a cameraman.*
    **IC:** De zandstorm die voor de kameel <u>schuilde</u> werd gefilmd door een cameraman.
40. De muis die in de val <u>trapte</u> was in de kelder van het huis. *The mouse that <u>stepped</u> into the trap was in the cellar of the house.*
    **IC:** De val die in de muis <u>trapte</u> was in de kelder van het huis.
41. De zon die boven de zee <u>opkwam</u> zag er prachtig uit. *The sun that <u>rose</u> over the sea looked beautiful.*
    **IC:** De zee die boven de zon <u>opkwam</u> zag er prachtig uit.
42. De masseur die over de rug <u>wreef</u> had veel tattoos. *The massager who <u>rubbed</u> over the back had a lot of tattoos.*
    **IC:** De rug die over de masseur <u>wreef</u> had veel tattoos.
43. De weg die naar de kust <u>voerde</u> was warm door de zon. *The road that <u>led</u> to the shore was warm from the sun.*
    **IC:** De kust die naar de weg <u>voerde</u> was warm door de zon.
44. De ambtenaren die over de financiën <u>beschikten</u> werden gecontroleerd door de minister. *The civil servants who <u>had</u> the finances <u>under their responsibility</u> were checked by the minister.*
    **IC:** De financiën die over de ambtenaren <u>beschikten</u> werden gecontroleerd door de minister.
45. De kinderen die naar de dolfijnen <u>zwaaiden</u> zagen er vrolijk uit. *The children that <u>waved</u> at the dolphins looked cheerful.*
    **IC:** De dolfijnen die naar de kinderen <u>zwaaiden</u> zagen er vrolijk uit.
46. De skiër die van de berg <u>afdaalde</u> was niet te zien door de mist. *The skier who <u>descended</u> from the mountain could not be seen because of the mist.*
    **IC:** De berg die van de skiër <u>afdaalde</u> was niet te zien door de mist.
47. De windstoten die aan de bomen <u>rukten</u> waren zeer sterk. *The blasts of wind that <u>pulled/jerked</u> on the trees were very strong.*
    **IC:** De bomen die aan de windstoten <u>rukten</u> waren zeer sterk.
48. De beroemdheid die in de hoofdstad <u>winkelde</u> werd in alle tijdschriften genoemd. *The celebrity who <u>was shopping</u> in the capital was named in every magazine.*
    **IC:** De hoofdstad die in de beroemdheid <u>winkelde</u> werd in alle tijdschriften genoemd.
49. De assistent die met de schaar <u>knipte</u> was nog nieuw. *The assistant who <u>cut</u> with the scissors was still new.*
    **IC:** De schaar die met de assistent <u>knipte</u> was nog nieuw.
50. De jager die over de steen <u>struikelde</u> bleef tussen de planten liggen. *The hunter who <u>stumbled</u> over the rock stayed down/lying between the plants.*
    **IC:** De steen die over de jager <u>struikelde</u> bleef tussen de planten liggen.
51. De inbreker die aan de deur <u>trok</u> werd door niemand bemerkt. *The burglar who <u>pulled</u> the door was not seen/noticed by anyone.*
    **IC:** De deur die aan de inbreker <u>trok</u> werd door niemand bemerkt.
52. De bruid die van de taart <u>smulde</u> werd bewonderd door de gasten. *The bride who <u>enjoyed</u> the cake was admired by the guests.*
    **IC:** De taart die van de bruid <u>smulde</u> werd bewonderd door de gasten.
53. De chauffeur die in de taxi <u>reed</u> bracht de reizigers snel thuis. *The driver who <u>drove</u> in the taxi brought the travelers home quickly.*
    **IC:** De taxi die in de chauffeur <u>reed</u> bracht de reizigers snel thuis.
54. Het bedrijf dat voor het project <u>koos</u> had de hulp van een architect nodig. *The company that <u>chose</u> for the project needed the help of an architect.*
    **IC:** Het project dat voor het bedrijf <u>koos</u> had de hulp van een architect nodig.
55. De politieagent die met de straf <u>dreigde</u> werd door alle mensen gerespecteerd.
    *The police officer who <u>threatened</u> with the punishment was respected by all people.*
    **IC:** De straf die met de politieagent <u>dreigde</u> werd door alle mensen gerespecteerd.
56. De bloemen die langs de straten <u>groeiden</u> werden verzorgd door de gemeente. *The flowers that <u>grew</u> alongside the streets were taken care of by the local authority.*
    **IC:** De straten die langs de bloemen <u>groeiden</u> werden verzorgd door de gemeente.
57. De demonstranten die met de stenen <u>gooiden</u> hinderden de bewoners. *The protesters who <u>threw</u> (with) the stones were in the way of the inhabitants.*
    **IC:** De stenen die met de demonstranten <u>gooiden</u> hinderden de bewoners.
58. De kleding die in de zon <u>droogde</u> voelde warm aan. *The clothes that <u>dried</u> in the sun felt warm.*
    **IC:** De zon die in de kleding <u>droogde</u> voelde warm aan.
59. De sneeuw die onder de schoen <u>kraakte</u> werd na enige tijd nat. *The snow that <u>cracked</u> under the shoe got wet after a while.*
    **IC:** De schoen die onder de sneeuw <u>kraakte</u> werd na enige tijd nat.
60. De vreemdeling die in de nacht <u>verdween</u> werd door de mensen nog lang herinnerd. *The stranger who <u>disappeared</u> into the night was remembered by the people for a long time.*
    **IC:** De nacht die in de vreemdeling <u>verdween</u> werd door de mensen nog lang herinnerd.

*A.2 Control condition*

(50 sentences, **IC:** incorrect version with semantic anomalies, the critical words are underlined. They were the ones where participants could detect the anomaly in the implausible version, i.e. they were not always the same as the replaced words.)

*A.2.1 Conjunction*

1. De Britten heersten over de <u>zee</u> en bouwden een groot imperium op. *The British ruled the <u>sea (sun)</u> and built a large empire.*
   **IC:** De Britten heersten over de <u>zon</u> en bouwden een groot imperium op.
2. De bandieten profiteerden van de <u>paniek</u> en holden weg met de juwelen. *The bandits took advantage of the <u>panic (lip)</u> and ran away with the jewels.*
   **IC:** De bandieten profiteerden van de <u>lip</u> en holden weg met de juwelen.
3. De bruid schreed langs haar <u>familie</u> en zag er stralend uit. *The bride strode past her <u>family (face)</u> and looked shining.*
   **IC:** De bruid schreed langs haar <u>gezicht</u> en zag er stralend uit.
4. De timmerman sloeg op zijn <u>vinger</u> en schreeuwde hard. *The carpenter hit on his <u>finger (street)</u> and screamed out loud.*
   **IC:** De timmerman sloeg op zijn <u>weg</u> en schreeuwde hard.
5. De pastoor knikte naar de <u>dames</u> en liep haastig verder. *The priest nodded to the <u>ladies (cities)</u> and hastily walked along.*
   **IC:** De pastoor knikte naar de <u>steden</u> en liep haastig verder.
6. De gangsters sprongen in de <u>wagen</u> en verlieten de stad. *The gangsters jumped into the <u>car (question)</u> and left the city.*
   **IC:** De gangsters sprongen in de <u>vraag</u> en verlieten de stad.
7. De ambtenaar belde met zijn <u>chef</u> en besprak het project. *The officer called his <u>boss (giraffe)</u> and discussed the project.*
   **IC:** De ambtenaar belde met zijn <u>giraffe</u> en besprak het project.
8. De voetballer kwam langs de <u>keeper</u> en trof het doel. *The football player came past the <u>keeper (cream cheese)</u> and hit the goal.*
   **IC:** De voetballer kwam langs de <u>roomkaas</u> en trof het doel.
9. De partijen streden over het <u>budget</u> en konden geen compromis vinden. *The parties argued about the <u>budget (eye)</u> and did not come to an agreement.*
   **IC:** De partijen streden over het <u>oog</u> en konden geen compromis vinden.
10. De kat sprong op de <u>vensterbank</u> en snuffelde aan het vliegje. *The cat jumped on the <u>windowsill (cloud)</u> and sniffed at the fly.*
    **IC:** De kat sprong op de <u>sterrenhemel</u> en snuffelde aan het vliegje.
11. De eendjes zwommen in het <u>water</u> en kregen brood van de bezoekers. *The ducks were swimming in the <u>water (menu)</u> and got some bread from the visitors.*
    **IC:** De eendjes zwommen in het <u>menu</u> en kregen brood van de bezoekers.
12. De moeder had de kinderen in <u>bed</u> gelegd en keek naar de televisie. *The mother put the children (chairs) to <u>bed</u> and watched television.*
    **IC:** De moeder had de stoelen in <u>bed</u> gelegd en keek naar de televisie.
13. De dirigent wees naar het orkest en kreeg applaus van het <u>publiek</u>. *The director pointed at the orchestra and was applauded by the <u>audience (beach)</u>.*
    **IC:** De dirigent wees naar het orkest en kreeg applaus van het <u>strand</u>.
14. De matrozen kwamen in de problemen en werden boos op de <u>kapitein</u>. *The sailors got into trouble and got mad at the <u>captain (attic)</u>.*
    **IC:** De matrozen kwamen in de problemen en werden boos op de <u>zolder</u>.
15. De conducteur liep door de wagon en controleerde de <u>tickets</u>. *The conductor walked through the wagon and checked the <u>tickets (planets)</u>.*
    **IC:** De conducteur liep door de wagon en controleerde de <u>planeten</u>.
16. De directeur klaagde over de arbeiders en gaf ze geen <u>loon</u>. *The manager complained about the workers and did not give them <u>salary (winter)</u>.*
    **IC:** De directeur klaagde over de arbeiders en gaf ze geen <u>winter</u>.
17. De prins deed zijn plicht en volgde de <u>koning</u> op. *The prince fulfilled his duty and succeeded the <u>king (tree)</u>.*
    **IC:** De prins deed zijn plicht en volgde de <u>boom</u> op.
18. De chauffeur wachtte op zijn klanten en stak een <u>sigaret</u> aan. *The driver was waiting for his clients and lighted a <u>cigarette (rabbit)</u>.*
    **IC:** De chauffeur wachtte op zijn klanten en stak een <u>konijn</u> aan.
19. De avonturier vertelde over de roofdieren en maakte veel <u>indruk</u>. *The adventurer told about the beasts of prey and made a great <u>impression (bread)</u>.*
    **IC:** De avonturier vertelde over de roofdieren en maakte veel <u>brood</u>.
20. De ijsbeer dook in het water en ving de <u>vissen</u>. *The polar bear dived into the water and caught the <u>fish (pistols)</u>.*
    **IC:** De ijsbeer dook in het water en ving de <u>pistolen</u>.

*A.2.2 Object Clause*

21. De dorpsbewoners proostten op de <u>vissers</u> die een goede vangst hadden. *The villagers toasted to the <u>fishermen (shoes)</u> who had a good catching.*
    **IC:** De dorpsbewoners proostten op de <u>schoenen</u> die een goede vangst hadden.
22. De leraar lette op het meisje dat hard <u>werkte</u> in de klas. *The teacher noticed the girl (book) who <u>worked</u> hard in class.*
    **IC:** De leraar lette op het boek dat hard <u>werkte</u> in de klas.
23. De indianen liepen in de <u>val</u> die de kolonisten gelegd hadden. *The Indians walked into the <u>trap (friend)</u> that the colonists had set.*
    **IC:** De indianen liepen in de <u>vriend</u> die de kolonisten gelegd hadden.
24. De arts flirtte met de <u>zusters</u> die hem assisteerden. *The physician flirted with the <u>nurses (cars)</u> who assisted him.*
    **IC:** De arts flirtte met de <u>auto's</u> die hem assisteerden.
25. De tijger sloop naar de <u>bok</u> die het gevaar niet opmerkte. *The tiger crept to the <u>goat (star)</u> which did not notice the danger.*
    **IC:** De tijger sloop naar de <u>ster</u> die het gevaar niet opmerkte.
26. De lifters stopten bij de <u>herbergier</u> die altijd vriendelijk was. *The lifters stopped at the <u>innkeeper (form)</u> who was always welcoming.*
    **IC:** De lifters stopten bij de <u>vorm</u> die altijd vriendelijk was.
27. De agent zocht naar de inbrekers die in de stad <u>actief</u> waren. *The officer sought for the burglars (bills) who were <u>active</u> in the city.*
    **IC:** De agent zocht naar de rekeningen die in de stad <u>actief</u> waren.
28. Het meisje was bang voor de spin die over haar been <u>liep</u>. *The girl was afraid of the spider (computer) that <u>walked/crept</u> over her leg.*
    **IC:** Het meisje was bang voor de computer die over haar been <u>liep</u>.
29. De studenten demonstreerden tegen de <u>minister</u> die geld op het onderwijs bespaarde. *The students demonstrated against the <u>minister (coffee)</u> who saved money on education.*
    **IC:** De studenten demonstreerden tegen de <u>koffie</u> die geld op het onderwijs bespaarde.

30. De postbode reed op de <u>fiets</u> die onder de auto kwam. *The postman was driving the <u>bicycle (glue)</u> that came under the car.*
    **IC:** De postbode reed op de <u>lijm</u> die onder de auto kwam.
31. De reporters luisterden naar de <u>advocaat</u> die de aangeklaagde verdedigde. *The reporters listened to the <u>lawyer (plant)</u> who defended the defendant.*
    **IC:** De reporters luisterden naar de <u>plant</u> die de aangeklaagde verdedigde.
32. De jager staarde naar het ree dat in de <u>struiken</u> verdween. *The hunter stared at the roe that disappeared in the <u>bushes (stairs)</u>.*
    **IC:** De jager staarde naar het ree dat in de <u>trappen</u> verdween.
33. De burgemeester sprak over de schurken die de bank hadden <u>beroofd</u>. *The mayor spoke about the thugs who had <u>robbed</u> the bank (wind).*
    **IC:** De burgemeester sprak over de schurken die de <u>wind</u> hadden <u>beroofd</u>.
34. De broer speelde met de poppen die van zijn <u>zusje</u> waren. *The brother played with the dolls that belonged to his <u>sister (brick)</u>.*
    **IC:** De broer speelde met de poppen die van zijn <u>baksteen</u> waren.
35. De jongen klom in de boom waar een <u>vogel</u> een nest in maakte. *The boy climbed into the tree in which a <u>bird (door)</u> made a nest.*
    **IC:** De jongen klom in de boom waar een <u>deur</u> een nest in maakte.
36. De rector stond achter de leraar die de leerlingen <u>straf</u> gaf. *The rector stood behind the teacher who <u>punished</u> the students (photographs).*
    **IC:** De rector stond achter de leraar die de foto's <u>straf</u> gaf.
37. De duikers zwommen naar het schip dat aan het <u>anker</u> vast lag. *The divers swam to the ship that was moored with the <u>anchor (leaf)</u>.*
    **IC:** De duikers zwommen naar het schip dat aan het <u>blad</u> vast lag.
38. De gokkers scholden op de jockey die van zijn <u>paard</u> was gevallen. *The gamblers shouted at jockey who fell from his <u>horse (hair)</u>.*
    **IC:** De gokkers scholden op de jockey die van zijn <u>haar</u> was gevallen.
39. De advocaten spraken met de verdachte die een portemonnee had <u>gestolen</u>. *The lawyers spoke with the suspect who had <u>stolen</u> a wallet (building).*
    **IC:** De advocaten spraken met de verdachte die een gebouw had <u>gestolen</u>.
40. De journalisten schreven over de burgers die op de <u>politici</u> hadden gestemd. *The journalists wrote about the civilians who voted for the <u>politicians (cables)</u>.*
    **IC:** De journalisten schreven over de burgers die op de <u>kabels</u> hadden gestemd.

*A.2.3 Subject Clause*

41. De pan waarin de soep zat was van <u>metaal</u>. *The pan in which the soup was made of <u>metal (wool)</u>.*
    **IC:** De pan waarin de soep zat was van <u>wol</u>.
42. Het dorp waarin de villa stond bood plaats aan veel <u>toeristen</u>. *The village in which the villa was hosted a lot of <u>tourists (theories)</u>.*
    **IC:** Het dorp waarin de villa stond bood plaats aan veel <u>theorieën</u>.
43. De bloem waarin de wesp zich verstopte werd ontdekt door de <u>bioloog</u>. *The flower in which the wasp hid was discovered by the <u>biologist (bag)</u>.*
    **IC:** De bloem waarin de wesp zich verstopte werd ontdekt door de <u>zak</u>.
44. De film waarin de actrice meespeelde was geliefd bij het grote <u>publiek</u>. *The film in which the actress participated was loved by the big <u>audience (building)</u>.*
    **IC:** De film waarin de actrice meespeelde was geliefd bij het grote <u>gebouw</u>.
45. Het café waarin de toeristen feest vierden was versierd met <u>ballonnen</u>. *The café in which the tourists celebrated a party was decorated with <u>balloons (shoulders)</u>.*
    **IC:** Het café waarin de toeristen feest vierden was versierd met <u>schouders</u>.
46. De richting waarin de ontdekker vertrok was aangewezen door een <u>dorpsbewoner</u>. *The direction to which the explorer went was advised by a <u>villager (tiger)</u>.*
    **IC:** De richting waarin de ontdekker vertrok was aangewezen door een <u>tijger</u>.
47. De weduwe tegen wie de priester sprak had een beeldschone <u>villa</u>. *The widow to whom the priest spoke to had a beautiful <u>villa (chocolate)</u>.*
    **IC:** De weduwe tegen wie de priester sprak had een beeldschone <u>chocolade</u>.
48. De reiziger met wie de familie dineerde kwam van een onbekende <u>plaats</u>. *The traveler with whom the family had dinner came from an unknown <u>place (foot)</u>.*
    **IC:** De reiziger met wie de familie dineerde kwam van een onbekende <u>voet</u>.
49. De bankier van wie de dame advies kreeg had net een nieuwe <u>hond</u>. *The banker from whom the lady got advice just had a new <u>dog (moon)</u>.*
    **IC:** De bankier van wie de dame advies kreeg had net een nieuwe <u>maan</u>.
50. De held over wie de dichter schreef redde het hele <u>land</u>. *The hero about whom the poet wrote saved the whole <u>country (paper)</u>.*
    **IC:** De held over wie de dichter schreef redde het hele <u>papier</u>.

*A.3 Filler sentences*

(10 sentences, IC: incorrect version with semantic reversal anomalies, the critical words are underlined)

1. De wortels waarvan de hamsters <u>aten</u> waren klein. *The carrots (hamsters) from which the hamsters (carrots) <u>were eating</u> were small.*
   **IC:** De hamsters waarvan de wortels <u>aten</u> waren klein.
2. De stal waarvan de koe <u>wegrende</u> stond in brand. *The stable (cow) from which the cow (stable) <u>ran away</u> was on fire.*
   **IC:** De koe waarvan de stal <u>wegrende</u> stond in brand.
3. De deur waarop de man <u>sloeg</u> zag er heel oud uit. *The door (man) on which the man (door) <u>hit</u> looked very old.*
   **IC:** De man op wie de deur <u>sloeg</u> zag er heel oud uit.
4. Het penseel waarmee de kunstenaar <u>verfde</u> viel van het balkon. *The pencil (painter) with which the painter (pencil) <u>was painting</u> fell of the balcony.*
   **IC:** De kunstenaar met wie het penseel <u>verfde</u> viel van het balkon.
5. Het matras waarop de kat <u>lag</u> voelde zacht aan. *The mattress (cat) on which the cat (mattress) <u>lay</u> felt soft.*
   **IC:** De kat waarop het matras <u>lag</u> voelde zacht aan.

6. Het idee waarover de wetenschapper <u>nadacht</u> was geniaal. *The idea (scientist) of which the scientist (idea) <u>thought</u> was genius.*
   **IC:** De wetenschapper over wie het idee <u>nadacht</u> was geniaal.
7. Het water waarvan de stier <u>dronk</u> dampte in de Spaanse hitte. *The water (bull) from which the bull (water) <u>drank</u> was damping in the Spanish heat.*
   **IC:** De stier waarvan het water <u>dronk</u> dampte in de Spaanse hitte.
8. Het circus waarin de beer <u>danste</u> lokte veel toeschouwers. *The circus (bear) in which the bear (circus) <u>danced</u> enticed many spectators.*
   **IC:** De beer waarin het circus <u>danste</u> lokte veel toeschouwers.
9. De muziek waarnaar de componist <u>luisterde</u> was beïnvloed door de romantiek. *The music (composer) to which the composer (music) <u>listened</u> was affected by romanticism.*
   **IC:** De componist naar wie de muziek <u>luisterde</u> was beïnvloed door de romantiek.
10. Het schip waarop de matroos <u>stond</u> voer naar de grote oceaan. *The ship (sailor) on which the sailor (ship) <u>stood</u> on was sailing to the ocean.*
    **IC:** De matroos op wie het schip <u>stond</u> voer naar de grote oceaan.

## Appendix B. Individual difference measurements

### C.1 L2 (Dutch) proficiency test

The L2 (Dutch) proficiency test was specifically tailored to German learners of Dutch and was adapted from the Dutch Proficiency Test (Transparent Language, Inc, 2015) and the Placement Test in the Dutch Language (Ernst Klett Sprachen, GmbH, 2015). The proficiency test consisted of 48 multiple-choice questions with four alternatives, measuring both grammar and vocabulary. The test was computer-based and took about 15 min to complete.

The Cronbach's alpha for the vocabulary part (22 items) is .68, and for the grammar part (26 items) is .64.

### C.2 Working memory tests

We used the working memory tests developed by Klaus and Schriefers (2016), which consisted of a reading span test and an operation span test.

#### Reading span test

Participants were instructed to read sentences (e.g., "Zodra ik klaar ben met deze afgunst, ga ik naar huis." *[Once I finish this envy, I will go home.]* in the Dutch version, equivalent sentences in the German version) presented on the screen and decide whether they made sense or not by pressing a button. After each sentence, a word was presented that the participants had to remember. After two to six trials, three question marks "???" were presented on the screen and the words had to be recalled vocally. Participants' responses were coded by the experimenter as correct or incorrect. After the participant named all the remembered words, the experimenter initiated the presentation of the next trial. In the beginning, two practice runs with a memory set size of three words were presented to familiarize participants with the task. Then, 15 runs with memory set sizes ranging from two to six words were randomly presented (the same randomization for all participants), with three runs of each length. The final score of each participant was calculated by first calculating the percentage of correct answers per memory set and then averaging the percentages across sets (Miyake and Friedman, 1998). Results were presented as decimals between 0 and 1.

#### Operation span test

Similar to the reading span test, participants were instructed to read equations (e.g., "(2 × 6)–3 = 5") presented on the screen and make judgments about whether the equation was correct or not by pressing a button. After four to seven trials, three question marks "???" were presented on the screen, and the numbers in previous trials had to be recalled vocally. In the beginning, three practice runs with a memory set size of four numbers were presented to familiarize participants with the task. Then, twelve runs with memory set sizes ranging from four to seven words were presented in random order (the same randomization for all participants), three runs of each length. The rest of the procedure was the same as the reading span test. The scores were calculated in the same way as the reading span test.

The two tests in total took about 25 min to complete. Both tests were performed in participants' native language (i.e., Dutch and German).

### C.2 Length of exposure

As a part of the language background questionnaire, all the participants from the L2 group were asked how long they had been exposed to the Dutch language ("*Wie viele Jahre hast du insgesamt Erfahrung mit Niederländisch?*"). This number had been originally entered in years but was converted to months.

## Appendix C. Factor analysis

### B.1 Factor extraction

We used a principal component analysis to extract latent factors from the five original predictors (vocabulary proficiency, grammar proficiency, length of exposure, operation span, reading span). Two components were extracted, which in combination explained 73% of the total variance. Table B.1 shows the factor loadings.

**Table B.1**
Summary of the exploratory factor analysis results (N = 61).

| Item | Factor loadings | |
|------|----------|----------|
| | Factor 1 | Factor 2 |
| Vocabulary proficiency | **.80** | − .36 |
| Grammar proficiency | **.80** | − .25 |
| LOE | **.70** | − .38 |
| OST | **.56** | **.70** |
| RST | **.62** | **.60** |
| Eigenvalues | 2.46 | 1.18 |
| % of variance | 49.14 | 23.53 |

Note: OST = *operation span test*; RST = *reading span test*; LOE = *length of exposure*. Factor loadings over .40 appear in bold.

### B.2 Linear regression

We used the two latent factors instead of the five original predictors in the linear regression, with *d'* and P600 amplitude as dependent variables, separately.

Results showed that the two latent factors explained a significant 35% of the variance of *d'* scores ($F_{change}(2, 58) = 15.61$, $p < .001$), with both Factor 1 and 2 being a significant predictor ($\beta = 0.53$, $t = 5.01$, $p < .001$; Factor 2: $\beta = 0.36$, $t = 2.47$, $p = .017$). On the contrary but similar to the original analysis reported in 3.3, the two factors again failed to predict P600 amplitude ($R^2_{adjusted} = .02$, $F_{change} (2, 58) = 0.64$, $p = .531$), with neither Factor 1 ($\beta = 0.15$, $t = 1.12$, $p = .268$) nor Factor 2 ($\beta = 0.02$, $t = 0.16$, $p = .872$) being a significant predictor.

### References

Ardal, S., Donald, M.W., Meuter, R., Muldrew, S., Luce, M., 1990. Brain responses to semantic incongruity in bilinguals. Brain Lang. 39, 187–205. https://doi.org/10.1016/0093-934X(90)90011-5.

Bialystok, E., Hakuta, K., 1994. In Other Words: The Science and Psychology of Second-Language Acquisition. Basic Books, New York, NY.

Bornkessel-Schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on "semantic P600" effects in language comprehension. Brain Res. Rev. 59, 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., ... Schlesewsky, M., 2011. Think globally: cross-linguistic variation in electro-physiological activity during sentence comprehension. Brain Lang. 117, 133–152. https://doi.org/10.1016/j.bandl.2010.09.010.

Brouwer, H., Fitz, H., Hoeks, J.C.J., 2012. Getting real about Semantic Illusions: rethinking the functional role of the P600 in language comprehension. Brain Res. 1446, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055.

Brouwer, H., Crocker, M.W., Venhuizen, N.J., Hoeks, J.C.J., 2017. A neurocomputational model of the N400 and the P600 in language processing. Cogn. Sci. 41, 1318–1352. https://doi.org/10.1111/cogs.12461.

Brysbaert, M., Lagrou, E., Stevens, M., 2017. Visual word recognition in a second language: a test of the lexical entrenchment hypothesis with lexical decision times. Biling.: Lang. Cogn. 20, 530–548. https://doi.org/10.1017/S1366728916000353.

Caffarra, S., Molinaro, N., Davidson, D., Carreiras, M., 2015. Second language syntactic processing revealed through event-related potentials: an empirical review. Neurosci. Biobehav. Rev. 51, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010.

Clahsen, H., Felser, C., 2006. Continuity and shallow structures in language processing. Appl. Psycholinguist. 27, 107–126. https://doi.org/10.1017/S0142716406060206.

Clahsen, H., Felser, C., 2017. Some notes on the shallow structure hypothesis. Stud. Second Lang. Acquis. 1–14. https://doi.org/10.1017/S0272263117000250.

Cop, U., Drieghe, D., Duyck, W., 2015. Eye movement patterns in natural reading: a comparison of monolingual and bilingual reading of a novel. PLoS One 10, e0134008. https://doi.org/10.1371/journal.pone.0134008.

Dąbrowska, E., Street, J., 2006. Individual differences in language attainment: comprehension of passive sentences by native and non-native English speakers. Lang. Sci. 28, 604–615. https://doi.org/10.1016/j.langsci.2005.11.014.

Dowens, M.G., Vergara, M., Barber, H.A., Carreiras, M., 2010. Morphosyntactic processing in late second-language learners. J. Cogn. Neurosci. 22, 1870–1887. https://doi.org/10.1162/jocn.2009.21304.

Ernst Klett Sprachen, GmbH, 2015. The Placement Test in Dutch Language. Retrieved from ⟨http://www2.klett.de/⟩.

Ferreira, F., 2003. The misinterpretation of noncanonical sentences. Cogn. Psychol. 47, 164–203. https://doi.org/10.1016/S0010-0285(03)00005-7.

Ferreira, F., Patson, N.D., 2007. The "Good Enough" approach to language comprehension. Lang. Linguist. Compass 1, 71–83. https://doi.org/10.1111/j.1749-818X.2007.00007.x.

Ferreira, F., Bailey, K.G.D., Ferraro, V., 2002. Good-enough representations in language comprehension. Curr. Dir. Psychol. Sci. 11, 11–15. https://doi.org/10.1111/1467-8721.00158.

Foucart, A., Frenck-Mestre, C., 2011. Grammatical gender processing in L2: electrophysiological evidence of the effect of L1-L2 syntactic similarity. Biling.: Lang. Cogn.

14, 379–399. https://doi.org/10.1017/S136672891000012X.

Frenzel, S., Schlesewsky, M., Bornkessel-Schlesewsky, I., 2011. Conflicts in language processing: a new perspective on the N400-P600 distinction. Neuropsychologia 49, 574–579. https://doi.org/10.1016/j.neuropsychologia.2010.12.003.

Friederici, A.D., Pfeifer, E., Hahne, A., 1993. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cogn. Brain Res. 1, 183–192. https://doi.org/10.1016/0926-6410(93)90026-2.

Gunter, T.C., Friederici, A.D., 1999. Concerning the automaticity of syntactic processing. Psychophysiology 36, 126–137. https://doi.org/10.1017/S004857729997155X.

Guo, J., Guo, T., Yan, Y., Jiang, N., Peng, D., 2009. ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. J. Neurolinguist. 22, 123–134. https://doi.org/10.1016/j.jneuroling.2008.09.001.

Hagoort, P., 2003. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. J. Cogn. Neurosci. 15 (6), 883–899. https://doi.org/10.1162/089892903322370807.

Hahne, A., 2001. What's different in second-language processing? Evidence from event-related potentials. J. Psycholinguist. Res. 30, 251–266. https://doi.org/10.1023/A:1010490917575.

Hahne, A., Friederici, A.D., 2002. Differential task effects on semantic and syntactic processes as revealed by ERPs. Cogn. Brain Res. 13, 339–356. https://doi.org/10.1016/S0926-6410(01)00127-6.

Hoeks, J.C.J., Stowe, L.A., Doedens, G., 2004. Seeing words in context: the interaction of lexical and sentence level information during reading. Cogn. Brain Res. 19, 59–73. https://doi.org/10.1016/j.cogbrainres.2003.10.022.

Hopp, H., 2014. Working memory effects in the L2 processing of ambiguous relative clauses. Lang. Acquis. 21, 250–278. https://doi.org/10.1080/10489223.2014.892943.

Hopp, H., 2018. The bilingual mental lexicon in L2 sentence processing. Retrieved from ⟨http://site.uit.no/lava/files/2018/01/20171230_J_SLA_Yearbook_Hopp_4.pdf⟩.

Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: evidence from event-related potentials. J. Mem. Lang. 52, 205–225. https://doi.org/10.1016/j.jml.2004.10.002.

Kim, A., Oines, L., Miyake, A., 2018. Individual differences in verbal working memory underlie a tradeoff between semantic and structural processing difficulty during language comprehension: an ERP investigation. J. Exp. Psychol.: Learn. Mem. Cogn. 44, 406–420. https://doi.org/10.1037/xlm0000457.

Klaus, J., Schriefers, H., 2016. December 7. Measuring verbal working memory capacity: A reading span task for laboratory and web-based use. Retrieved from osf.io/nj48x.

Kolk, H.J., Chwilla, D.J., van Herten, M., Oor, P.J.W., 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. Brain Lang. 85, 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5.

Kotz, S.A., 2009. A critical review of ERP and fMRI evidence on L2 syntactic processing. Brain Lang. 109, 68–74. https://doi.org/10.1016/j.bandl.2008.06.002.

Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Res. 1146, 23–49. https://doi.org/10.1016/j.brainres.2006.12.063.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205. https://doi.org/10.1126/science.7350657.

Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: a quick and valid lexical test for

advanced learners of English. Behav. Res. Methods 44, 325–343. https://doi.org/10.3758/s13428-011-0146-0.

Lim, J.H., Christianson, K., 2013. Second language sentence processing in reading for comprehension and translation. Biling.: Lang. Cogn. 16, 518–537. https://doi.org/10.1017/S1366728912000351.

Macmillan, N.A., Kaplan, H.L., 1985. Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. Psychol. Bull. 98, 185–199. https://doi.org/10.1037/0033-2909.98.1.185.

McDonald, J.L., 2006. Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. J. Mem. Lang. 55, 381–401. https://doi.org/10.1016/j.jml.2006.06.006.

Meschyan, G., Hemandez, A.E., 2006. Impact of language proficiency and orthographic transparency on bilingual word reading: an fMRI investigation. NeuroImage 29, 1135–1140. https://doi.org/10.1016/j.neuroimage.2005.08.055.

Miyake, A., Friedman, N.P., 1998. Individual differences in second language proficiency: working memory as language aptitude. In: Healy, A.F., Bourne, Jr, L.E. (Eds.), Foreign Language Learning: Psycholinguistic Studies on Training and Retention. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 339–364.

Morishima, Y., 2013. Allocation of limited cognitive resources during text comprehension in a second language. Discourse Process. 50, 577–597. https://doi.org/10.1080/0163853X.2013.846964.

Nakano, H., Saron, C., Swaab, T.Y., 2010. Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. J. Cogn. Neurosci. 22, 2886–2898. https://doi.org/10.1162/jocn.2009.21400.

Newman, A.J., Tremblay, A., Nichols, E.S., Neville, H.J., Ullman, M.T., 2012. The influence of language proficiency on lexical semantic processing in native and late learners of English. J. Cogn. Neurosci. 24, 1205–1223. https://doi.org/10.1162/jocn_a_00143.

Ojima, S., Nakata, H., Kakigi, R., 2005. An ERP study of second language learning after childhood: Effects of proficiency. J. Cogn. Neurosci. 17, 1212–1228. https://doi.org/10.1162/0898929055002436.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9. https://doi.org/10.1155/2011/156869.

Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. J. Mem. Lang. 31, 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z.

Osterhout, L., Nicol, J., 1999. On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. Lang. Cogn. Process. 14, 283–317. https://doi.org/10.1080/016909699386310.

Paczynski, M., Kuperberg, G.R., 2012. Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world

event/state knowledge and animacy selection restrictions. J. Mem. Lang. 67, 426–448. https://doi.org/10.1016/j.jml.2012.07.003.

Roberts, L., 2012. Individual differences in second language sentence processing. Lang. Learn. 62, 172–188. https://doi.org/10.1111/j.1467-9922.2012.00711.x.

Schlesewsky, M., Bornkessel-Schlesewsky, I., 2009. When semantic P600s turn into N400s: On cross-linguistic differences in online verb-argument linking. In Alter.K., M. Horne, M. Lindgren, M. Roll, & J. v. K. Torkildsen (Eds.), Brain Talk. In: Proceedings of the 1st Birgit Rausing Language Program Conference in Linguistics Lund: Lund University, Media Tryck, pp. 75–97. Retrieved from ⟨http://hdl.handle.net/11858/00-001M-0000-0010-EB01-D⟩.

Shen, W., Fiori-Duharcourt, N., Isel, F., 2016. Functional significance of the semantic P600: evidence from the event-related brain potential source localization. NeuroReport 27, 548–558. https://doi.org/10.1097/WNR.0000000000000583.

Steinhauer, K., White, E.J., Drury, J.E., 2009. Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. Second Lang. Res. 25, 13–41. https://doi.org/10.1177/0267658308098995.

Tanner, D., McLaughlin, J., Herschensohn, J., Osterhout, L., 2013. Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. Biling.: Lang. Cogn. 16, 367–382. https://doi.org/10.1017/S1366728912000302.

Tanner, D., Inoue, K., Osterhout, L., 2014. Brain-based individual differences in online L2 grammatical comprehension. Biling.: Lang. Cogn. 17, 277–293. https://doi.org/10.1017/S1366728913000370.

Tolentino, L.C., Tokowicz, N., 2011. Across languages, space, and time: a review of the role of cross-language similarity in L2 (morpho)syntactic processing as revealed by fMRI and ERP methods. Stud. Second Lang. Acquis. 33, 91–125. https://doi.org/10.1017/S0272263110000549.

Transparent Language, Inc, 2015. Dutch Proficiency Test. Retrieved from ⟨http://www.transparent.com/learn-dutch/proficiency-test.html⟩.

Traxler, M.J., 2014. Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. Trends Cogn. Sci. 18, 605–611. https://doi.org/10.1016/j.tics.2014.08.001.

van Herten, M., Kolk, H.H.J., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. Cogn. Brain Res. 22, 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002.

van Herten, M., Chwilla, D.J., Kolk, H.J., 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. J. Cogn. Neurosci. 18, 1181–1197. https://doi.org/10.1162/jocn.2006.18.7.1181.

VanRullen, R., 2011. Four common conceptual fallacies in mapping the time course of recognition. Front. Psychol. 2, 365. https://doi.org/10.3389/fpsyg.2011.00365.

Weber-Fox, C.M., Neville, H.J., 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. J. Cogn. Neurosci. 8, 231–256. https://doi.org/10.1162/jocn.1996.8.3.231.