

Bilingualism: Language and Cognition**Date of delivery:****Journal and vol/article ref:** BIL 1200030**Number of pages (not including this page):** 16

This proof is sent to you on behalf of Cambridge University Press.

Authors are strongly advised to read these proofs thoroughly because any errors missed may appear in the final published paper. This will be your ONLY chance to correct your proof. Once published, either online or in print, no further changes can be made.

****THESE PROOFS SHOULD BE RETURNED WITHIN 2 WORKING DAYS****

HOW TO RETURN YOUR PROOFS

You can mark up proofs either on screen using the enabled electronic editing tools, or by hand on the hardcopy.

Marking up electronically. All proofs are enabled to allow electronic annotation in the freely available Adobe Reader software. Using your cursor select the text for correction and use the most appropriate single tool (i.e. 'Replace', 'Cross out', 'Insert' or 'Add note to text'). Please return the file as an attachment via email to **ewa@essex.ac.uk**

Marking up by hand. Alternatively, please print the PDF file, mark any amendments on the proofs and list the corrections on a separate page, scan, and email them to the address above. If either of these methods are unsuitable, you may post your corrections to:

Ewa Jaworska
Copy-editor
6 Bailey Dale
Colchester CO3 0LB, UK
Tel: + 44 (0)1206 331466

If sending the corrected proof by post, please use either Recorded Delivery, Airmail or Courier service if from overseas and either First Class Recorded Delivery or Special Delivery if from within the UK. Please inform the copy-editor in a brief e-mail message that your proof corrections will be coming by post.

Changes should be limited to the correction of editing and typographical errors and the answering of any author queries. Queries raised by the copy-editor and/or the typesetters are listed below; the text to which the queries refer is flagged in the margins of the proof. Any corrections that contradict journal style will not be transferred.

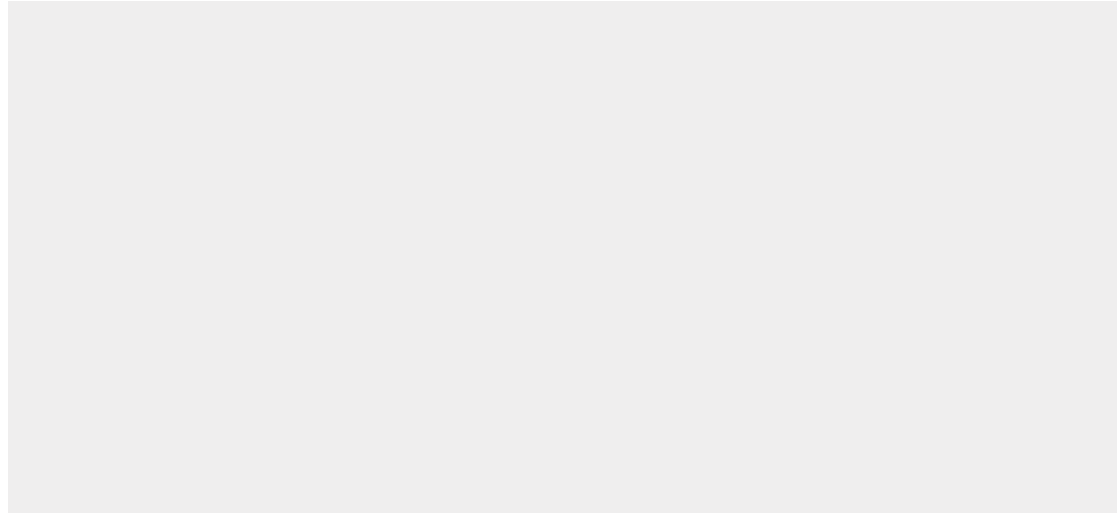
The quality of half-tones will be checked by the Editorial Office.

Copyright: If you have not already done so, please download a copyright form from:

http://journals.cambridge.org/images/fileUpload/documents/BIL_ctf.pdf

Please sign the form **by hand** and return by mail to the address shown on the form. Failure to send this form will delay the publication of your article.

Thank you for publishing in *Bilingualism: Language and Cognition*. You will automatically receive a link to the PDF version of your article once it has been published online.

Bilingualism: Language and Cognition

- The proof is sent to you for correction of typographical errors only. Revision of the substance of the text is not permitted, unless discussed with the editor of the journal. Only **one** set of corrections are permitted.
- Please answer carefully any author queries.
- Corrections which do NOT follow journal style will not be accepted.
- A new copy of a figure must be provided if correction of anything other than a typographical error introduced by the typesetter is required.
- If you have problems with the file please contact

mmochrie@cambridge.org

Please note that this pdf is for proof checking purposes only. It should not be distributed to third parties and may not represent the final published version.

Important: you must return any forms included with your proof.

Please do not reply to this email.

NOTE - for further information about **Journals Production** please consult our **FAQs** at http://journals.cambridge.org/production_faqs

Author queries:

Q1: Please insert the name of the university where the ms is held (perhaps first author's affiliation?). Also once more where thus indicated.

Q2: Please insert the name of the journal where the paper is due to appear. If still under review, perhaps omit this note altogether given that submissions to journals are usually confidential?

Typesetter queries:

Non-printed material:

Individual differences reveal stages of L2 grammatical acquisition: ERP evidence*

DARREN TANNER

Pennsylvania State University, Department of Psychology

JUDITH MCLAUGHLIN

University of Washington, Department of Psychology

JULIA HERSCHENSOHN

University of Washington, Department of Linguistics

LEE OSTERHOUT

University of Washington, Department of Psychology

(Received: 0, 0000)

Here we report findings from a cross-sectional study of morphosyntactic processing in native German speakers and native English speakers enrolled in college-level German courses. Event-related brain potentials were recorded while participants read sentences that were either well-formed or violated German subject–verb agreement. Results showed that grammatical violations elicited large P600 effects in the native Germans and learners enrolled in third-year courses. Grand mean waveforms for learners enrolled in first-year courses showed a biphasic N400–P600 response. However, subsequent correlation analyses revealed that most individuals showed either an N400 or a P600, but not both, and that brain response type was associated with behavioral measures of grammatical sensitivity. These results support models of second language acquisition which implicate qualitative changes in the neural substrates of second language grammar processing associated with learning. Importantly, we show that new insights into L2 learning result when the cross-subject variability is treated as a source of evidence rather than a source of noise.

Keywords: N400, P600, second language acquisition, individual differences, morphosyntax, ERP

Introduction

Over the last several decades, there has been an enormous increase in interest in the neural substrates of language processing. Accordingly, there has been a proliferation of studies using event-related brain potentials (ERPs) which have revealed a great deal about how and when different types of information are integrated during real-time comprehension in native (L1) speakers of a language. ERPs have also been used to study neurocognitive aspects of second language (L2) processing, as ERPs' multidimensional nature allows the investigation of fundamental questions about the cognitive processes subserving late-learned languages. In studies of

L2 learning, identifying whether or not learners' ERP waveforms approximate those of native speakers has sometimes been taken as a 'litmus test' for whether L2 processing is fundamentally similar to or different from native language processing. For example, an experimental effect size smaller than that found in native speakers is often taken to mean less robust processing in the learner population, while a qualitatively different ERP effect or the inability to detect some effect is often taken to reflect a fundamental difference in the neural substrates of L2 processing or a lack of that specific neurocognitive process in the group of learners, respectively (see e.g., Rossi, Gugler, Friederici & Hahne, 2006; Sabourin & Stowe, 2008, for examples of these types of inferences).

An important caveat, however, is that much of the published work has reported ERPs that represent averages over both trials and individuals. In order to achieve an adequate signal-to-noise ratio, voltages from the raw electroencephalogram (EEG) in a time epoch of interest are averaged over all trials in a given experimental condition within subjects, and then averaged again over subjects. These grand mean waveforms represent brainwave activity which is time- and phase-locked to the onset of the stimulus of interest and consistent across both trials and subjects (see Handy, 2005; Luck, 2005). In terms of L1 processing, researchers generally assume that monolingual native speakers of a language will exhibit

* We would like to thank the participants, as well as members of the Cognitive Neuroscience of Language Lab at the University of Washington, including Missy Takahashi, Ilona Pitkänen, and Geoff Valentine for their help in collecting the data for this study. We would especially like to thank Kayo Inoue for valuable discussion and insights into individual variation. This research was supported by grant R01DC01947 from the National Institute on Deafness and Other Communication Disorders to Lee Osterhout. Portions of this manuscript were prepared while Darren Tanner was supported by a Neurolinguistics Dissertation Fellowship from the William Orr Dingwall Foundation and by NSF OISE-0968369. We would also like to thank two anonymous reviewers who provided thoughtful comments on earlier versions of this article. Any remaining errors are, of course, our own.

Address for correspondence:

Darren Tanner, Pennsylvania State University, Center for Language Science, 4F Thomas Building, University Park, PA 16802, USA
dstanner@gmail.com

similar neural signatures of language processing. This assumption seems reasonable, as there is a remarkable consistency in ERP responses seen across experiments and languages. However, L2 learning is subject to significant individual variation, which in turn can lead to problems of interpretation for traditional ERP analyses. We show here that in certain circumstances this variability is highly systematic. We show further that new insights into L2 learning result when the cross-subject variability is treated as a source of evidence rather than a source of noise.

Within the context of native languages, the use of grand means has proven to be a useful tool for studying language processing. One of the most remarkably consistent and replicable results over 30 years of cross-linguistic language-related ERP research is that lexico-semantic and morphosyntactic manipulations elicit qualitatively different brain responses. All content words elicit a negative-going brain wave with a peak at around 400 ms after presentation (the N400), but the size of this peak can be modulated by numerous factors, such as a word's semantic relatedness to a preceding context, cloze probability, and corpus frequency (Bentin, 1987; Kutas & Federmeier, 2000; Kutas & Hillyard, 1980; Osterhout & Nicol, 1999). Larger peak amplitudes are thought to reflect greater difficulty with lexical access and integration (the N400 'effect'). On the other hand, relative to well-formed controls, a wide range of sentence-embedded morphosyntactic anomalies (such as violations of agreement, tense, case, and verb subcategorization) elicit a large positive-going wave with a peak around 600 ms poststimulus (the P600: Ainsworth-Darnell, Shulman & Boland, 1998; Friederici, Hahne & Mecklinger, 1996; Hagoort, Brown & Groothusen, 1993; Kaan, Harris, Gibson & Holcomb, 2000; Osterhout & Holcomb, 1992, 1995). Some studies of morphosyntactic processing have reported an additional negative-going wave with an onset of 100–400 ms poststimulus with a largely left anterior distribution preceding the P600 (the Left Anterior Negativity, or LAN: Friederici et al., 1996; Neville, Nicol, Barss, Forster & Garrett, 1991; Osterhout & Holcomb, 1992). Given the reliability of these results across languages, experimental manipulations, and task demands, it is clear that ERPs are differentially sensitive to distinct levels of processing, and that grand mean analyses capture this consistency.¹

¹ There are some exceptions to the generalization that semantic and syntactic violations always elicit N400 and P600 effects, respectively. Some studies have reported N400 effects to syntactic agreement violations (e.g., Bentin & Deutsch, 2001; Severens et al., 2008), and others have reported P600 effects to certain types of semantic violations (e.g., Kim & Osterhout, 2005; Kolk, Chwilla, van Herten & Oor, 2003; Nieuwland & Van Berkum, 2005; van de Meerendonk, Kolk, Vissers & Chwilla, 2010). However, the semantics/N400 and syntax/P600 correlation holds under the conditions investigated in

Other research has shown that ERPs are also sensitive to individual differences in L1 processing. For example, the amplitude and onset of ERP effects can be modulated by individuals' working memory capacity (King & Kutas, 1995; Vos, Gunter, Kolk & Mulder, 2001). More recent research has indicated that individuals' brain responses to syntactic anomalies can vary systematically with differences in language proficiency, even among monolingual native speakers of a language. Pakulak and Neville (2010) reported a correlation between waveform characteristics (the laterality of an early LAN component and the amplitude of the P600 component) and participants' L1 (English) proficiency. The anomalies elicited a more left-lateralized LAN and a larger-amplitude P600 in more proficient participants. Other researchers have shown that not only can quantitative aspects of ERP responses vary across individuals, but also the type of response. For example, some studies have demonstrated that under certain conditions, biphasic negative–positive responses to anomalies seen in grand mean waveforms may not represent true biphasic responses within individuals, but rather be an artifact of averaging across individuals, some of whom show an N400 and some of whom show a P600 (Nieuwland & Van Berkum, 2008; Osterhout, 1997; Osterhout, McLaughlin, Kim, Greewald & Inoue, 2004). More recently, results from Inoue & Osterhout (2012) indicate that within and across individuals, N400 and P600 effect magnitudes are negatively correlated, such that as one increases in magnitude, the other decreases to a similar degree. Furthermore, Nakano and colleagues (Nakano, Saron & Swaab, 2010) showed that working memory span can modulate type of response to verb–argument animacy violations. In their study, those with lower span measures showed N400 effects to animacy violations whereas those with higher span measures showed P600 effects. It therefore seems that systematic individual variation exists but that this variability is obscured by traditional grand mean ERP waveforms.

For L2 learners the assumption of homogeneity of responses across individuals may be even more tenuous. Unlike L1 acquisition, success in L2 learning has been shown to correlate with a number of individual factors such as general intelligence, specific language aptitude, learning strategy, and motivation (Dörnyei & Skehan, 2003; Naiman, Fröhlich, Stern & Todesco, 1996; Robinson, 2002; Skehan, 1989). McDonald (2006) has shown that L2 learners', but not native speakers', accuracy and reaction time in a grammaticality judgment task were correlated with working memory and lexical decoding measures. L2 learners also have shown variability in grammaticality judgment accuracy across testing

this study (i.e., the presentation of relatively simple sentences, see Kuperberg, 2007, for discussion).

sessions, even when identical items were used on both occasions (Johnson, Shenkman, Newport & Medin, 1996). Learners additionally can show knowledge of L2 grammatical information in offline tasks, but no sensitivity in online tasks, suggesting greater variability in the timing of access and integration of that knowledge relative to natives (Clahsen & Felser, 2006). This variability is made apparent in a study of reaction time and accuracy in a grammaticality judgment task by McDonald (2000): the reported standard deviations for late L2 learners were generally two and three times larger for reaction time and accuracy, respectively, than for native speaker controls. In terms of its implications for ERP research into L2 processing, this greater variation, both between individuals and between trials within individuals, means that there may be increased fluctuation in the timing and nature of neural responses to L2 stimuli, thus obscuring what may be true effects from surfacing in grand mean waveforms.

Indeed, ERP research into L2 processing has yielded somewhat mixed results regarding the nature and status of syntactic processes in non-native speakers. Several studies have shown that P600s can be reliably elicited in non-native speakers, suggesting some continuity between native and non-native syntactic processing systems, especially for grammatical features shared across the L1 and for novel L2 features for learners at high L2 proficiency (Foucart & Frenck-Mestre, 2011; Frenck-Mestre, Osterhout, McLaughlin & Foucart, 2008; Gillon Dowens, Guo, Guo, Barber & Carreiras, 2011; Hahne, Mueller & Clahsen, 2006; Morgan-Short, Sanz, Steinhauer & Ullman, 2010; Morgan-Short, Steinhauer, Sanz & Ullman, 2012; Rossi et al., 2006; Tokowicz & MacWhinney, 2005). Others have failed to find robust P600 effects to syntactic anomalies, usually when the L2 feature is not found or is realized differently in the L1 (Foucart & Frenck-Mestre, 2011; Hahne & Friederici, 2001; Ojima, Nakata & Kakigi, 2005; Sabourin & Haverkort, 2003; Sabourin & Stowe, 2008). Still others have reported that syntactic anomalies can elicit qualitatively different responses in L2 learners versus native speakers, usually in the form of a negativity rather than a positivity (Chen, Shu, Liu, Zhao & Li, 2007; Guo, Guo, Yan, Jiang & Peng, 2009; Sabourin & Stowe, 2008) or a biphasic negative–positive response (Weber & Lavric, 2008).

It should be noted that the studies mentioned above which have failed to find classic P600 effects in L2 learners used traditional grand averages to study ERP effects. However, as pointed out by Osterhout and colleagues (Osterhout, McLaughlin, Pitkänen, Frenck-Mestre & Molinaro, 2006) and McLaughlin and colleagues (McLaughlin, Tanner, Pitkänen, Frenck-Mestre, Inoue, Valentine & Osterhout, 2010), null results in L2 ERP research are especially problematic to interpret,

since a given electrophysiological effect may be present on most trials in a few individuals, or on a few trials in most individuals, but be obscured in the averaging process due to noise in the raw electroencephalogram. Variability in timing of the effect across trials and individuals can additionally reduce effect sizes in ERP grand means, even when the true amplitude of a given electrophysiological effect is consistent across trials (Luck, 2005). Nonetheless, given the findings of systematic variability in L1 processing reported above, it seems likely that at least some variability between L2 learners may also be systematic and therefore observable in analyses of individuals' ERPs.

Only a few studies have investigated individual differences in ERP correlates of L2 syntactic processing. These experiments have generally used grouped designs to investigate the impact of some individual-level variable, such as age of arrival (Weber-Fox & Neville, 1996) or L2 proficiency (Rossi et al., 2006), on learners' brain responses to syntactic violations. Using this approach could reduce problematic between-subject variability, as group members would be relatively homogenous with regard to some individual difference dimension (e.g., proficiency; see Steinhauer, White & Drury, 2009; van Hell & Tokowicz, 2010, for discussion). Others have further reduced between-subject variability by adopting within-subjects longitudinal designs to study changes in individuals' brain responses over time as L2 proficiency increases (McLaughlin et al., 2010; Morgan-Short et al., 2010; Morgan-Short et al., 2012; Osterhout et al., 2006), or artificial language training paradigms which allow learners to reach high proficiency in a very short amount of time (Friederici, Steinhauer & Pfeifer, 2002; Morgan-Short et al., 2010; Morgan-Short et al., 2012). Another possibility for studying individual differences in processing is to use regression-based statistical techniques, as regression models have the ability to capture potentially linear and graded effects of individual differences measures. However, only a few studies have used this approach, and nearly all of these have focused on modulations of the N400 component associated with semantic processing (Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville & Ullman, 2012; Ojima, Matsuba-Kurita, Nakamura, Hoshino & Hagiwara, 2011; though see Bond, Gabriele, Fiorentino & Alemán Bañón, 2011).

In the study reported below we cross-sectionally investigated grammatical processing in English-speaking learners of L2 German who were enrolled in classroom-based university German courses. Our participants are therefore representative of a common L2 learner population in the United States. We recorded participants' brain responses as they read sentences that were either well-formed or contained violations of German subject–verb agreement. Verb agreement is a grammatical feature

shared by both English and German.² Shared features are often transferred from the L1 to the L2, and should thus be acquired early during the L2 learning process (MacWhinney, 2005; Sabourin, Stowe & de Haan, 2006; Schwartz & Sprouse, 1996). We quantified learners' brain responses first using grand mean analyses, and then analyzed individual variation among learners' ERPs responses with regression-based models. We demonstrate that although grand mean analyses showed statistically robust findings in L2 learners of all levels, they obscured systematic, qualitative and quantitative differences among learners' brain responses to L2 grammatical anomalies.

Method

Participants

Our participants included 13 native speakers of German (mean age: 28 years; range: 18–51; eight female) and 33 native English-speaking students enrolled in university-level second language German courses. Twenty were novice learners enrolled in the final course of the first-year German sequence (mean hours of instruction = 123.8, $SD = 10.0$; mean age: 20 years; range: 18–25; 10 female) and 13 were enrolled in third-year German courses (mean age: 20 years; range: 19–24; six female). All participants were healthy and had normal or corrected-to-normal vision and gave their informed consent after the nature and possible consequences of the study were explained. Participants received a small monetary compensation for taking part in the study.

Materials

Stimuli were sentences in German consisting of lexical items chosen from the first seven chapters of the textbook used in first-year German courses at the University of Washington. Sixty sentence pairs were created, with one member of each pair being semantically coherent and grammatical and the second member being identical, except for showing incorrect agreement between the subject pronoun and verb (e.g., *Ich wohne/*wohnt in Berlin*, "I live/*lives in Berlin"). All person/number combinations in German are marked with overt, phonologically realized morphemes. Grammatical

and ungrammatical sentence pairs were distributed across two lists in a Latin-square design, such that each list contained only one version of each sentence. Experimental sentences were randomized among 140 filler sentences (70 ungrammatical) containing other types of syntactic anomalies. Sixty sentences contained violations of number agreement between a determiner or quantifier and noun (e.g., *Viele/*ein Bücher liegen auf dem Tisch*, "Many/*a books are on the table") and 10 sentences contained an extra auxiliary verb (e.g., **Mein Bruder macht sind seine Arbeit*, "My brother does are his work"). Each list contained a total of 200 sentences, half of which were ungrammatical.

Procedure

Participants were tested in a single session lasting approximately 85 minutes (including about 30 minutes of experimental preparation). Upon arrival in the laboratory, each participant was asked to fill out an abridged version of the Edinburgh Handedness Questionnaire and a language history questionnaire. Each participant was randomly assigned to one of the stimulus lists and was seated in a comfortable recliner in front of a CRT monitor. Participants were instructed to relax and minimize movements while reading and to read each sentence as normally as possible. Each trial consisted of the following events: each sentence was preceded by a blank screen for 1000 ms, followed by a fixation cross, followed by a stimulus sentence, presented one word at a time. The fixation cross and each word appeared on the screen for 475 ms followed by a 250 ms blank screen between words. Sentence-ending words appeared with a full stop followed by a "Good/Bad" response prompt. Participants were instructed to respond "good" if they felt it was a well-formed, grammatical sentence in German and "bad" if they felt it was ungrammatical or violated some rule of German. Participants were randomly assigned to use either their left or right hand for the "good" response.

Data acquisition and analysis

Continuous EEG was recorded from 19 tin electrodes attached to an elastic cap (Eletro-cap International) in accordance with the 10–20 system (Jasper, 1958). Eye movements and blinks were monitored by two electrodes, one placed beneath the left eye and one placed to the right of the right eye. The electrodes were referenced to an electrode placed over the left mastoid and were amplified with a bandpass of 0.01–100 Hz (3 dB cutoff) by an SA Instruments bioamplifier system. EEG was recorded from an additional electrode placed on the right mastoid to identify if there were any experimental effects detectable over the mastoids; no such effects were found. Impedances

² There are, however, differences between English and German in how they realize this agreement. English marks agreement on the copula *be* and marks 3rd person singular agreement on lexical verbs (with *-s*), but only in the present tense. German, on the other hand, marks four unique person–number combinations in both present and past tense, but the morphological expression of agreement in German is highly regular (Durrell, 2006). Nonetheless, given the similarity of the grammatical contrast across the two languages and the regularity of the German rule, we expect English learners of German to show little difficulty with German agreement.

at scalp and mastoid electrodes were held below 5 k Ω and below 15 k Ω at eye electrodes.

Continuous analog-to digital conversion of the EEG and stimulus trigger codes was performed at a sampling frequency of 200 Hz. ERPs, time-locked to the onset of the critical word, were averaged off-line for each participant at each electrode site in each condition. A digital low-pass filter of 30 Hz was applied to individuals' averaged waveforms prior to analysis. Grand average waveforms were created by averaging over participants. Trials characterized by eye blinks, excessive muscle artifact, or amplifier blocking were not included in the averages; 11.8% of trials overall were removed due to artifacts. The number of rejections did not differ significantly between conditions or groups.

Behavioral results were quantified both using d-prime scores (Wickens, 2002) and proportion correct in the grammatical and ungrammatical sentence conditions. Behavioral results were analyzed with ANOVAs using group (native, third year, first year) as a between-subjects factor; ANOVAs on proportion correct contained grammaticality (grammatical, ungrammatical) as an additional repeated-measures factor. ERP components of interest were quantified by computer as mean voltage within a window of activity. In accordance with previous literature and visual inspection of the data, the following time windows were chosen: 50–150 ms (N1), 150–300 ms (P2), 300–500 ms (N400), and 500–800 ms (P600), relative to a 100 ms prestimulus baseline. Within each time window ANOVAs were calculated with grammaticality (grammatical, ungrammatical) as a within-subjects factor. Data from midline (Fz, Cz, Pz), medial–lateral (right hemisphere: Fp2, F4, C4, P4, O2; left hemisphere: Fp1, F3, C3, P3, O1), and lateral–lateral (right hemisphere: F8, T8, P8; left hemisphere: F7, T7, P7) electrode sites were treated separately in order to identify topographic and hemispheric differences. ANOVAs on midline electrodes included electrode as an additional within-subjects factor (three levels), ANOVAs on medial–lateral electrodes included hemisphere (two levels) and electrode pair (five levels) as additional within-subjects factors, and ANOVAs over lateral–lateral electrodes included hemisphere (two levels) and electrode pair (three levels) as additional within-subjects factors. The Greenhouse–Geisser correction for inhomogeneity of variance was applied to all repeated measures on ERP data with greater than one degree of freedom in the numerator. In such cases, the corrected *p*-value is reported.

Results

Behavioral results

Mean d-prime scores, proportions judged correctly, and standard deviations are reported in Table 1. On average,

Table 1. Mean d-prime scores and proportion of sentences judged correctly for native speakers, third-year learners, and first-year learners. Standard deviations are reported in parentheses.

Group	d-prime	Proportion correct	
		Grammatical	Ungrammatical
Native speakers	4.85 (0.93)	0.95 (0.05)	0.98 (0.05)
Third year	3.61 (1.40)	0.93 (0.05)	0.90 (0.14)
First year	3.20 (1.34)	0.91 (0.74)	0.88 (0.13)
Total	3.78 (1.41)	0.93 (0.06)	0.92 (0.12)

Note: A d-prime of 0 indicates chance performance on the acceptability judgment task; a d-prime of 4 indicates near-perfect discrimination between well-formed and ill-formed sentences.

all participants, including first-year learners, performed very well in the acceptability judgment task. Statistical analyses for d-prime scores showed a main effect of group, $F(2,43) = 6.991$, $MSE = 1.578$, $p = .002$. A Tukey's HSD post-hoc test showed significant differences between the first-year learners and native speakers, $p = .002$, and between the third-year learners and native speakers, $p = .040$. There were no differences between the first and third-year learners, $p = .637$. An ANOVA on proportion judged correctly showed a main effect of group, $F(2,43) = 4.216$, $MSE = 0.010$, $p = .021$, but no effect of grammaticality, $F < 1$, and no grammaticality by group interaction, $F(2,43) = 1.185$, $MSE = 0.007$, $p = .316$. A Tukey's HSD post-hoc test showed a significant difference between native speakers and first-year learners, $p = .016$, but no differences between the other groups, p s $> .167$.

Event-related potentials results

Grand mean analyses

Grand mean waveforms for native speakers are plotted in Figure 1. In these and all subsequent waveforms, the general shapes of the waveforms were consistent with previous data using visually presented language stimuli (e.g., Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995). Statistical analyses of native speakers' ERP responses showed that there were no reliable effects in the early time windows; however, there was a trend toward a main effect of grammaticality in the 300–500 ms time window [midline: $F(1,12) = 3.651$, $MSE = 6.284$, $p = .080$; medial–lateral: $F(1,12) = 4.047$, $MSE = 9.992$, $p = .067$], suggesting the onset of a positivity to ungrammatical verbs. In the 500–800 ms time window there was a significant main effect of grammaticality, indicating a P600 effect to ungrammatical verbs [midline:

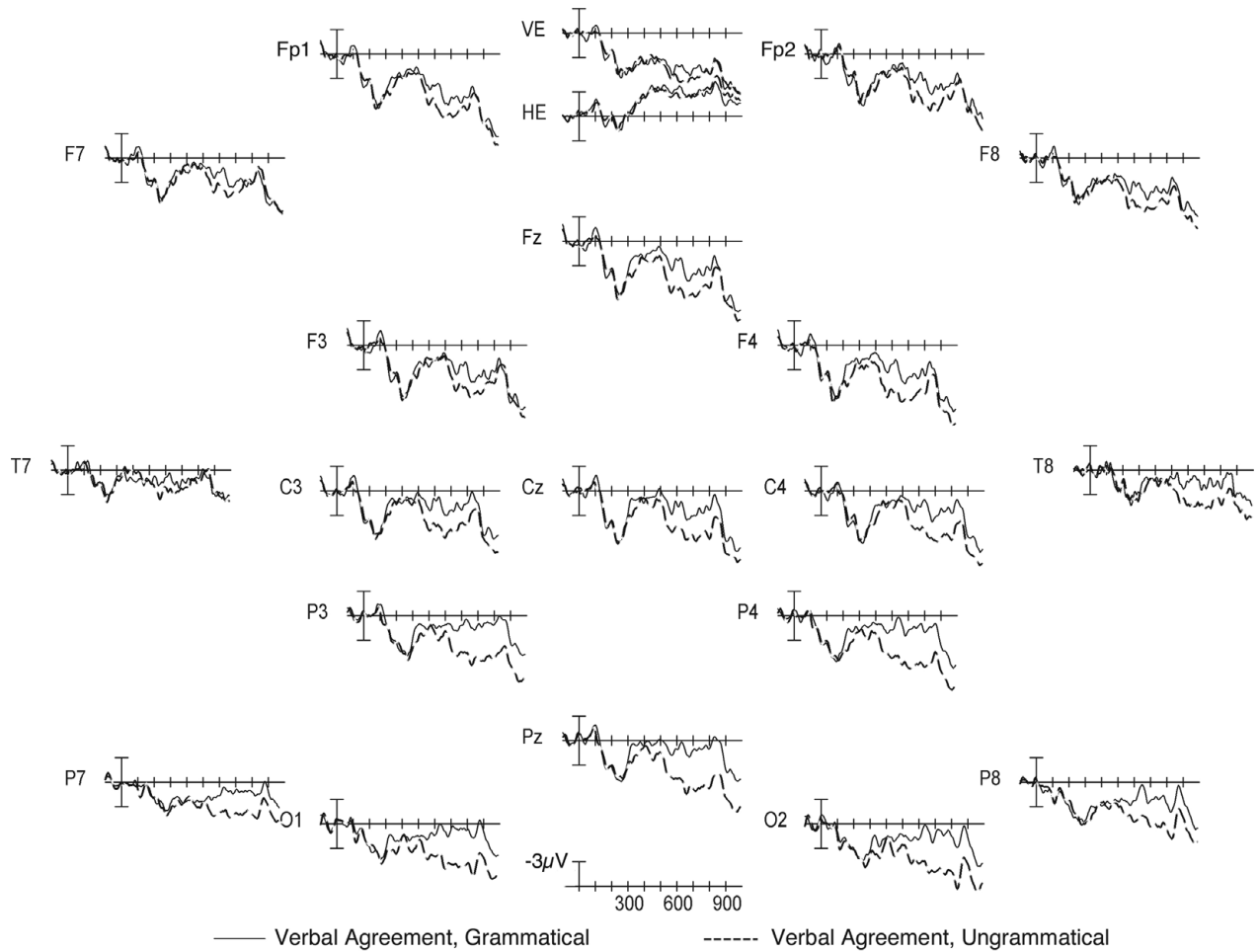


Figure 1. Grand average ERP waveforms for native German speakers ($n = 13$) to grammatical (solid line) and ungrammatical (dashed line) verbs. Onset of the verb is indicated by the vertical bar. Calibration bar shows $3\mu\text{V}$ of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

$F(1,12) = 26.407$, $MSE = 6.956$, $p = .0003$; medial-lateral: $F(1,12) = 31.163$, $MSE = 16.473$, $p = .0001$; lateral-lateral: $F(1,12) = 19.302$, $MSE = 7.075$, $p = .0009$] that was largest over posterior electrodes [grammaticality \times electrode interaction, midline: $F(2,24) = 3.766$, $MSE = 1.291$, $p = .045$; medial-lateral: $F(4,48) = 5.249$, $MSE = 3.329$, $p = .014$; lateral-lateral: $F(2,24) = 5.098$, $MSE = 0.949$, $p = .024$]. The P600 additionally showed a slight right-hemisphere bias over lateral-lateral electrodes [grammaticality \times hemisphere interaction: $F(1,12) = 6.273$, $MSE = 1.044$, $p = .028$].

Third-year learners' brain responses (Figure 2) showed no significant effects in the N1, P2, or N400 time windows. In the 500–800 ms time window there was a main effect of grammaticality, indicating a reliable P600 effect to violations of subject-verb agreement [midline: $F(1,12) = 18.316$, $MSE = 9.731$, $p = .001$; medial-lateral: $F(1,12) = 22.103$, $MSE = 18.826$, $p < .0005$; lateral-lateral: $F(1,12) = 17.804$, $MSE = 5.566$,

$p = .001$]. However, there were no significant interactions with electrode or hemisphere in this time window.

ERPs from first-year learners (Figure 3) showed no significant effects in the N1 and P2 time windows, but there was a significant main effect of grammaticality over midline electrodes and a near-significant effect over lateral sites in the 300–500 ms window, indicating an N400-like negativity to disagreeing verbs [midline: $F(1,19) = 5.776$, $MSE = 7.251$, $p = .027$; medial-lateral: $F(1,19) = 3.759$, $MSE = 16.772$, $p = .068$; lateral-lateral: $F(1,19) = 3.751$, $MSE = 5.412$, $p = .068$]. There were no interactions with electrode or hemisphere. This N400 was followed by a trend toward a P600 effect over midline electrodes in the 500–800 ms time window [main effect of grammaticality: $F(1,19) = 3.156$, $MSE = 13.493$, $p = .092$]; there were no significant or near-significant effects over medial-lateral or lateral-lateral sites. Thus, grand mean waveforms to disagreeing verbs showed a small biphasic response: ungrammatical verbs elicited a

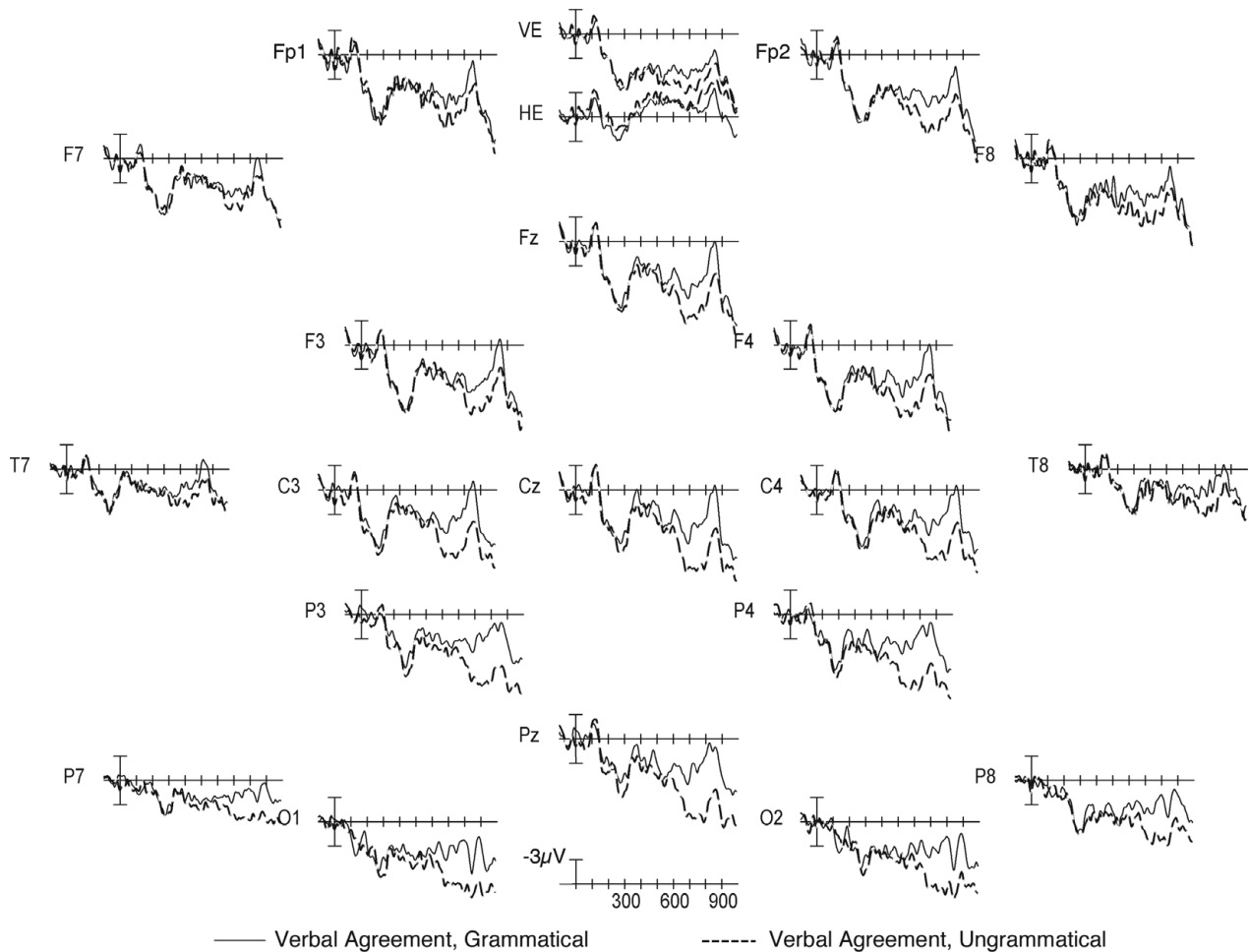


Figure 2. Grand average ERP waveforms for learners enrolled in third-year German courses ($n = 13$) to grammatical (solid line) and ungrammatical (dashed line) verbs. Onset of the verb is indicated by the vertical bar. Calibration bar shows $3\mu\text{V}$ of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

broadly distributed negativity in the 300–500 ms time window, but a small positivity in the 500–800 ms time window that did not reach full significance.

Analyses of individuals' ERP responses

As noted above, first-year learners' grand mean waveforms showed a small biphasic response to disagreeing verbs. However, inspection of individuals' waveforms showed that most learners did not show this biphasic response. Rather, for most subjects the response to these words was either dominated by an enhanced N400 or by the later positivity. Following Inoue and Osterhout (2012), we further investigated this by first computing the magnitude of the N400 and P600 effects for each individual, and then regressing the N400 effect magnitude onto that of the P600 effect for first-year learners. N400 effect magnitude was computed as mean amplitude in the 300–500 ms window in the grammatical condition minus mean amplitude in the ungrammatical

condition, averaged over midline electrodes; P600 effect magnitude was computed as mean amplitude in the 500–800 ms window in the ungrammatical condition minus the mean amplitude in the grammatical condition, again averaged over midline electrodes. The two effects were significantly negatively correlated, $r = -.616$, $p = .004$. As can be seen in Figure 4, learners' brain responses showed a similar function to that reported by Inoue and Osterhout for native speakers of Japanese processing case violations: brain responses varied along an N400–P600 continuum such that as one response increased, the other decreased. First-year learners were divided into N400 ($n = 9$) and P600 ($n = 11$) groups, based on whether the individual's response showed an N400 or P600 dominance. Grand mean waveforms for learners in the N400 group showed mild differences in the prestimulus baseline, so a corrected 50 ms prestimulus to 50 ms poststimulus baseline was used for this group. ERP responses over midline electrodes for these separate

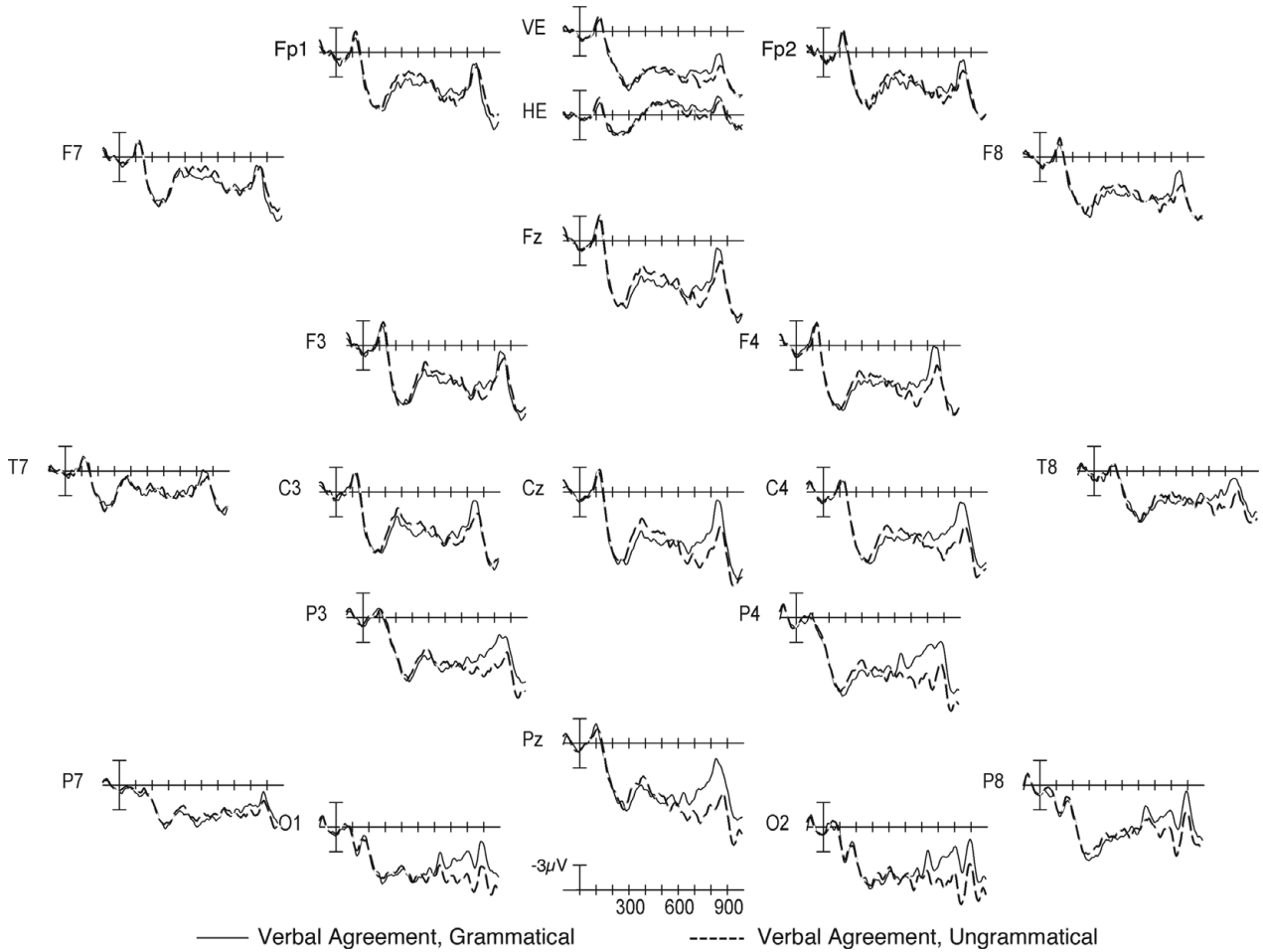


Figure 3. Grand average ERP waveforms for learners enrolled in first-year German courses ($n = 20$) to grammatical (solid line) and ungrammatical (dashed line) verbs. Onset of the verb is indicated by the vertical bar. Calibration bar shows $3\mu V$ of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

groups are shown in Figure 5. Learners in the N400 group showed a significant effect of grammaticality in the 300–500 ms time window, indicating a reliable N400 effect, $F(1,8) = 10.020$, $MSE = 6.940$, $p = .013$, but no significant effects in the P600 window, $F_s < 1$. Learners in the P600 group showed no effects over midline electrodes in the N400 time window, $F_s < 1.5$; however, there was a significant effect of grammaticality in the later time window, $F(1,10) = 37.290$, $MSE = 4.609$, $p = .0001$. Thus, the biphasic response seen in the grand mean waveform was in fact an artifact of averaging over individuals who showed qualitatively different brain responses to disagreeing verbs (see Nieuwland & Van Berkum, 2008; Osterhout, 1997; Osterhout & Inoue, 2012).

In order to investigate what factors may have been important in predicting the type and magnitude of response learners showed to disagreeing verbs, we conducted a series of correlation analyses using individuals' N400 and P600 effect magnitudes over

electrode Pz, where ERP effects were the largest. First-year learners' P600 effect magnitudes were reliably correlated with d-prime scores, $r = .532$, $p = .016$; for third-year learners the correlation neared significance, $r = .504$, $p = .079$; and for all learners (first- and third-year) combined the correlation was highly significant, $r = .534$, $p = .001$ (Figure 6). Thus, learners' P600 effect magnitudes increased linearly with their ability to detect agreement anomalies. For native speakers there was no relationship between d-prime and P600 amplitude, $r = .274$, $p = .344$. Since learners' P600 responses were associated with better performance in the acceptability judgment task, it is also possible that N400 responses were associated with poorer performance. Correlations did not reach significance for first-year learners, $r = -.297$, $p = .204$; third-year learners, $r = -.406$, $p = .169$; or native speakers, $r = -.322$, $p = .261$. However, the correlation for all learners combined was weak, but did reach statistical significance, $r = .367$, $p = .036$ (Figure 7). In this time window enhanced negativities to ungrammatical

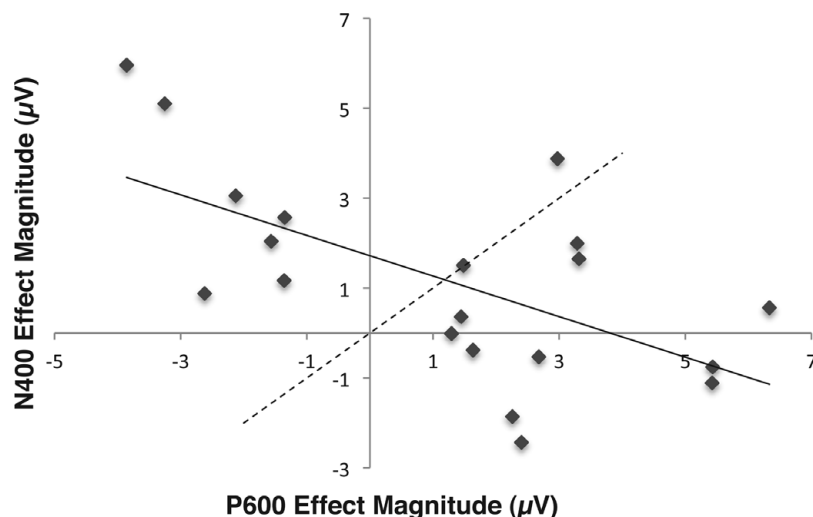


Figure 4. Scatterplot showing the distribution of N400 and P600 effect magnitudes across first-year learners, averaged across three midline electrodes (Fz, Cz, and Pz). Each dot represents a data point from a single learner. The solid line shows the best-fit line for the data from the regression analysis. The dashed line represents equal N400 and P600 effect magnitudes and shows where learners were divided into groups: individuals above/to the left of the dashed line showed primarily an N400 effect to German verb agreement violations, while individuals below/to the right of the dashed line showed primarily a P600 effect.

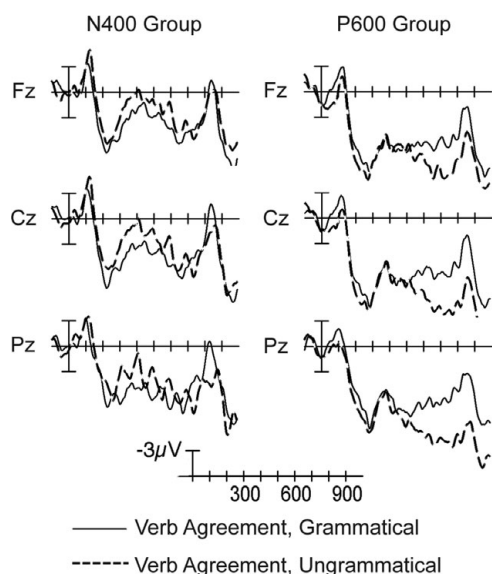


Figure 5. ERPs over midline electrodes to grammatical (solid line) and ungrammatical (dashed line) verbs for first-year learners who showed either N400-dominant (left panel; $n = 9$) or P600-dominant (right panel; $n = 11$) brain responses. Onset of the verb is indicated by the vertical bar. Calibration bar shows $3 \mu\text{V}$ of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

verbs were associated with poorer performance in the acceptability judgment task.

In a study of word learning in L2 French, McLaughlin and colleagues (McLaughlin, Osterhout & Kim, 2004)

found that learners' individual N400 amplitudes to French-like pseudowords were highly correlated with the number of hours of instruction the subjects had been exposed to during the first quarter of classroom French instruction. In order to test for a similar correlation in the current data, the number of hours of classroom exposure was computed for all first-year learners. There was no correlation between hours of exposure and amplitude difference over any of the midline electrodes in the N400 or in the P600 time window. Moreover a regression model including d-prime score as an independent variable and P600 magnitude at Pz as the dependent variable was significant, $R^2_{\text{Adjusted}} = .243$, $F(1,18) = 7.098$, $p = .016$; a model including both d-prime score and hours of instruction only neared significance, $R^2_{\text{Adjusted}} = .198$, $F(2,17) = 3.352$, $p = .059$. Whereas d-prime scores alone account for approximately 24% of the variance in P600 effect magnitude, including hours of instruction as an independent variable actually removed predictive power from the overall model. Partial correlations in the second model show that after controlling for effects of d-prime score, there was no relationship between hours of instruction and P600 magnitude, $r = .004$, while d-prime remained significant after controlling for hours of instruction, $r = .514$, $p = .02$. No regression model including d-prime score, hours of instruction, or a combination of the two accounted for a significant portion of the variance in N400 amplitudes. It therefore seems that the variation in individuals' brain responses is more a function of grammatical learning than pure classroom exposure.

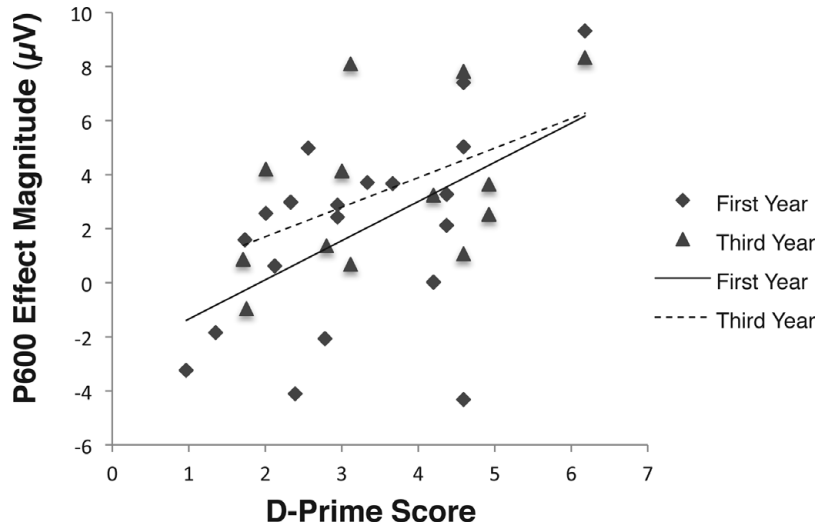


Figure 6. Correlation between P600 effect magnitude and d-prime scores from the acceptability judgment task for all learners. P600 effect magnitude is quantified as mean amplitude in 500–800 ms time window over Pz in the ungrammatical minus grammatical condition. More positive values on the y-axis reflect larger P600 effects.

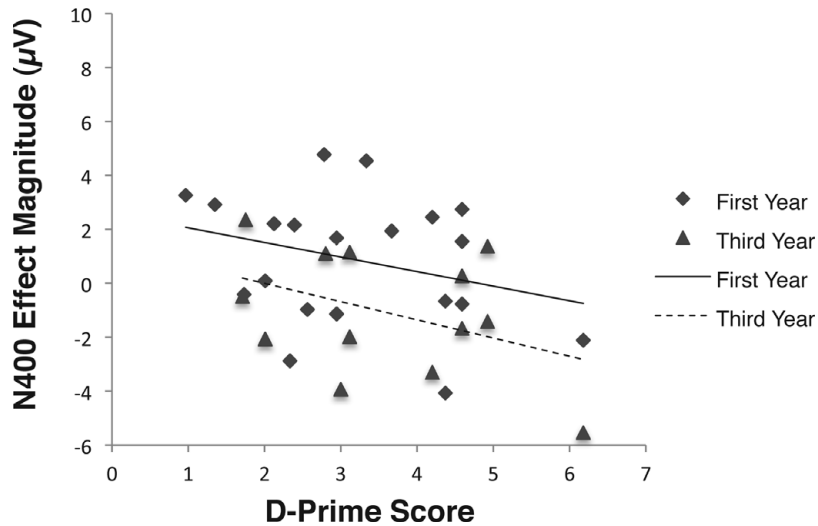


Figure 7. Correlation between N400 effect magnitude and d-prime scores from the acceptability judgment task for all learners. N400 effect size is quantified as mean amplitude between 300–500 ms time window over Pz in the grammatical minus ungrammatical condition. More positive values on the y-axis reflect larger N400 effects.

A further issue is the relationship between participants’ end-of-sentence judgments and their online brain responses. One possibility is that the correlation between d-prime scores and P600 magnitude might reflect a scenario where individuals who performed more poorly on the grammaticality judgment task showed a smaller P600 effect on any given trial than those who performed better. Alternately, participants might show a full P600 on any trial when they recognized the agreement error, but no P600 on the other trials; the result after averaging would then be that those who recognized fewer errors (and who had lower d-prime scores) would show smaller average P600 effects. Moreover, it is also possible

that correctly- and incorrectly-judged trials elicited qualitatively different ERP effects, such that incorrectly judged ungrammatical trials would elicit an N400, while correctly judged ungrammatical trials would elicit a P600 effect. The net result would be that those who show poorer judgment performance would show an N400-dominant brain response, whereas those who show better judgment would show a P600-dominant response.

To investigate these possibilities, we computed response-contingent averages, including only those trials which were ultimately judged correctly by the participants. Difference waveforms comparing all-trial averages and response-contingent averages for third-year

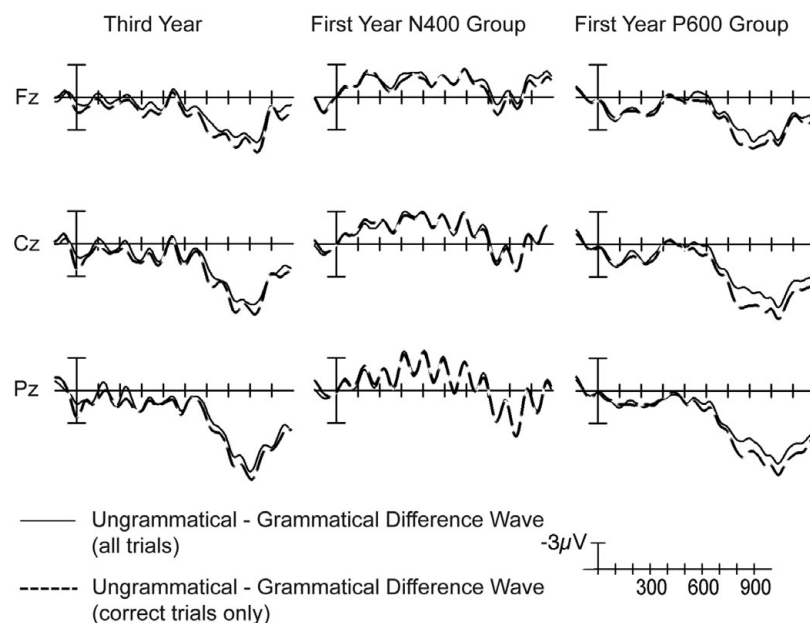


Figure 8. Grand average difference waves for the ungrammatical minus grammatical conditions, comparing effect sizes for the all-trial and response-contingent analyses. Positive or negative deviations from zero indicate a positivity or negativity in the ungrammatical condition relative to the grammatical condition, respectively. Difference waveforms were filtered with a 10 Hz low-pass filter for presentation purposes. Onset of the verb is indicated by the vertical bar. Calibration bar shows 3 μ V of activity; each tick mark represents 100 ms of time. Positive voltage is plotted down.

learners and first-year learners in the N400- and P600-dominant groups are shown in Figure 8. As can be seen, the two sets of averages show very similar effects of the grammaticality manipulations. N400 and P600 effect magnitudes in the all-trial and response-contingent averages were nearly perfectly correlated within learners (N400 effect magnitude: $r = .907$, $p < .000001$; P600 effect magnitude: $r = .969$, $p < .000001$). Additionally, the correlation between d-prime scores and P600 magnitude for all learners remained significant even when including only correctly-judged trials in the ERP averages, $r = .494$, $p = .004$. Overall this indicates that N400 effects were not driven only by incorrectly-judged trials, as the N400 effects remained robust even when considering only correctly-judged trials. The remaining correlation between d-prime and P600 magnitude also indicates that the full P600 on correctly-judged trials/no P600 on incorrectly-judged trials account is incorrect. Nonetheless, this does not unequivocally show that P600 effects on any given trial were consistently smaller across all trials in those with lower d-prime scores. Indeed, the linear change in P600 magnitude may still have been driven by cross-trial differences in effect amplitude. The present results simply indicate that these differences were not directly associated with an individual's eventual judgment about a given sentence (see McLaughlin et al., 2004; Tokowicz & MacWhinney, 2005, for evidence of dissociations between on-line brain responses and

off-line judgments). Future research on trial-level modeling of brain responses may shed light on this issue (see Zayas, Greenwald & Osterhout, 2010).

Discussion

The study reported here investigated morphosyntactic processing in native German speakers and in English-speaking university students enrolled in their first or third year of German instruction. Our most striking finding was the existence of systematic individual differences in the learners' ERP responses to subject-verb agreement anomalies. These anomalies elicited an N400 effect in some learners and a P600 effect in others. The amplitudes of these effects were negatively correlated across learners, and accuracy in the sentence-acceptability judgment task predicted the amplitude of the ERP response to ungrammatical stimuli, with greater accuracy being associated with more positive-going brain activity throughout the N400 and P600 windows.

Prior work has shown that individuals differ with respect to working memory capacity, vocabulary knowledge, neural efficiency, and in many other ways that could impact language processing (Prat, 2011). One possibility, therefore, is that the individual differences among the German learners reflect durable subject variables (i.e., "traits") that persist over time. If so, then the individual differences observed here might be expected

to persist even as the learner becomes more proficient in the L2. An alternative possibility, however, is that learners were progressing between two distinct processing stages (as manifested in the N400 and P600 responses to morphosyntactic anomalies), and that individual learners varied with respect to the rate of transition between the two stages. A compelling test of these different interpretations requires a longitudinal design that tracks learners over an extended period of L2 instruction. Some relevant evidence is provided by a longitudinal ERP study of first-year French learners (Osterhout et al., 2012; see McLaughlin et al., 2010; Osterhout et al., 2006, for preliminary reports; see also Morgan-Short et al., 2010; Morgan-Short et al., 2012, for similar findings from an artificial language learning study). ERPs were recorded to violations of French subject–verb agreement. Most learners responded to these anomalies with an N400 effect after about one month of L2 instruction and a P600 effect after about seven months of instruction. When tested during the middle of the instructional period (after about four months of instruction), the grand average ERP revealed a small-amplitude biphasic N400–P600 effect. Inspection of individual subjects' ERPs showed that the grand average obscured robust individual differences, such that some learners showed an N400 effect and others a P600 effect to the same set of agreement anomalies. Learners' N400 and P600 effect magnitudes were negatively correlated at each testing session.

Collectively, the evidence seems to indicate that individual learners progress through distinct stages of learning, but that the rate of progression varies across learners. An important goal is to characterize the functional significance of the developmental stages. Whereas the current state of the field does not allow one to draw a direct link between a given ERP effect and a specific underlying cognitive or linguistic process, some parallels exist between claims in the broader psycholinguistics literature and the pattern of results obtained here. For example, some have argued that although native speakers typically compute detailed syntactic representations, they may sometimes use shallow (or 'good enough') processing heuristics instead of full syntactic parses during language comprehension in complex syntactic situations, such as passive constructions and garden path sentences (Christianson, 2008; Christianson, Hollingworth, Halliwell & Ferreira, 2001; Ferreira, Bailey & Ferraro, 2002). Theorists have proposed a link between the use of a shallower, heuristic or lexical processing stream and a deeper, rule-based or combinatorial processing stream, and the N400 and P600 components, respectively (Kuperberg, 2007; Severens, Jansma & Hartsuiker, 2008; Tanner, 2011). One possibility is that novice L2 learners were more reliant on these shallower lexical or probabilistic processing heuristics than native speakers and more

advanced learners for even simple grammatical relations, like agreement. The shift to a P600-dominant response might reflect the gradual development of a more abstract, rule-based processing stream for L2 grammar, as is typically employed by L1 speakers in these constructions. The additional negative correlation between the N400 and P600 effect magnitudes might be explainable in terms of processing models which posit a "competitive dynamic" between the two streams (Jackendoff, 2007; Kim & Osterhout, 2005; MacWhinney, Bates & Kliegl, 1984).

The current data also share some features with predictions made by Ullman's Declarative/Procedural (D/P) model (Ullman, 2001, 2004, 2005). For example, the N400 effect elicited by morphosyntactic violations in early-stage L2 acquisition is compatible with the D/P model's prediction that both grammatical and lexical processing in novice learners will show heavy reliance on the declarative memory system. With increasing proficiency, grammatical processing should then show increased reliance on the procedural memory system. However, Ullman argues that use of procedural memory will be indexed by a LAN effect in response to grammatical anomalies. LANs were not found in any group (learners or native) in this study. Moreover, LAN effects are missing in native speakers in many studies of syntactic processing (e.g., Ainsworth-Darnell et al., 1998; Allen, Badecker & Osterhout, 2003; Frenck-Mestre et al., 2008; Hagoort, 2003; Hagoort & Brown, 1999; Hagoort et al., 1993; Kaan, 2002; Nevins, Dillon, Malhotra & Phillips, 2007; Severens et al., 2008), so it is difficult to interpret the absence of this effect in the current study as reflecting incomplete grammatical acquisition or deficient processing in our more advanced learners or native speakers. More research is needed to precisely identify the experimental conditions under which LAN effects are reliably elicited.

The qualitative change in processing seen in early-stage L2 learners is incompatible with some recent proposals about L2 processing (Clahsen & Felser, 2006; Clahsen, Felser, Neubauer, Sato & Silva, 2010). Clahsen and colleagues argue that L2 learners are restricted to the shallower, 'good enough' parses that are sometimes available to native speakers, regardless of L2 proficiency or L1–L2 pairing. However, the current data indicate that adult L2 learners can move beyond shallow processing heuristics and develop deeper grammatical processing strategies within only a few months of classroom instruction. Moreover, L2 proficiency can have an effect on a learner's depth of processing, as a relative reliance on the shallower lexical/heuristic or deeper grammatical/combinatorial processing stream was associated with behavioral measures of grammatical learning in the current study. The data reported here provide strong evidence that learners at different stages

of development use qualitatively different processing streams to deal with L2 grammatical information, and are consistent with longitudinal findings that individuals may shift dominance from one stream to the other as their L2 competence increases over time (see McLaughlin et al., 2010; Steinhauer et al., 2009, for further discussion about the possible functional significance of the N400–P600 shift).

In the present study, effect magnitude correlated with accuracy in the grammaticality judgment task for L2 learners. The relationship between effect magnitude and d-prime is reminiscent of that reported by Pakulak and Neville (2010), who found that participants' P600 effect magnitudes were linearly related to their proficiency in L1 English. Our findings are also consistent with suggestions made by Steinhauer and colleagues (Steinhauer et al., 2009; see also van Hell & Tokowicz, 2010) that increasing L2 proficiency co-occurs with a more L1-like profile of ERP responses to L2 anomalies. However, this result might not generalize to all situations or populations of language learners. Recent results from our lab show a similar profile of brain responses in high-proficiency late L1-Spanish–L2-English bilinguals with long-term L2 immersion as seen here in novice L2 learners (Tanner, Inoue & Osterhout, 2012; see also Tanner, 2011). Instead of proficiency, motivation provided the strongest correlate of brain response type in that study. However, there are several demographic differences between the Spanish–English bilinguals and the novice learners in the current study, including age of acquisition and amount of L2 exposure. Also, the types of linguistic input received by immersed versus classroom L2 learners may have had an impact on brain response profiles. More research is needed in order to identify how input and other individual-level variables interact in shaping L2 learning and processing (see e.g., Morgan-Short, Faretta, Brill, Wong & Wong, 2012). Language proficiency may therefore be one of many factors responsible for determining the neural substrates of syntactic processing (Prat, 2011).

Finally, the present findings compellingly illustrate the dangers inherent in the exclusive use of grand-average ERPs to characterize L2 sentence processing. In some cases a thorough investigation of between-learner variability can be more informative than inspection of grand mean waveforms. Our results add to results from the lexico-semantic processing domain showing that regression-based statistical methods can be used on ERP data to model individual difference profiles (Moreno & Kutas, 2005; Newman et al., 2012; Ojima et al., 2011). However, our strongest predictor of brain responses, namely d-prime scores, was an experiment-internal variable. A logical next question is what other variables may be at play in predicting how quickly novice learners grammaticalize L2 features (i.e., how quickly they move from N400 to P600 responses to subject–

verb agreement violations), or what factors predict the magnitude of ERP effects at higher levels of proficiency. Behavioral research has long noted correlations between learning measures and certain cognitive and affective variables. It remains to be seen how these variables map onto the neurocognitive correlates of learning that we report here (see Bond et al., 2011, for a first attempt to link individuals' specific language aptitude and non-verbal reasoning ability with L2 ERP effects). Nonetheless, the approach taken here demonstrates that ERPs provide a valuable tool for understanding individual factors in L2 grammatical learning, and encourage us to hope that future research will elucidate determinants of the rate and success of L2 acquisition.

References

- Ainsworth-Darnell, K., Shulman, R., & Boland, J. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related brain potentials. *Journal of Memory and Language*, *38*, 112–130.
- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, *18*, 405–430.
- Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain and Language*, *31*, 308–327.
- Bentin, S., & Deutsch, A. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language*, *45*, 200–224.
- Bond, K., Gabriele, A., Fiorentino, R., & Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In D. Tanner & J. Herschensohn (eds.), *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)*, pp. 17–29. Somerville, MA: Cascadia Proceedings Project.
- Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*, 161–174.
- Christianson, K. (2008). Sensitivity to syntactic changes in garden path sentences. *Journal of Psycholinguistic Research*, *37*, 391–403.
- Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.
- Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.
- Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, *60*, 21–43.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (eds.), *The handbook of second language acquisition*, pp. 589–630. Malden, MA: Blackwell.
- Durrell, M. (2006). *Hammer's German grammar and usage*. London: Arnold.

- Ferreira, F., Bailey, K., G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.
- Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*, 379–399.
- Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, *128*, 528–536.
- Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic processing: Early and late event-related potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1219–1248.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences USA*, *99*, 529–534.
- Gillon Dowens, M., Guo, T., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from event related potentials. *Neuropsychologia*, *49*, 1651–1659.
- Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, *22*, 123–134.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*, 883–899.
- Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research*, *28*, 715–728.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*, 439–484.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*, 123–141.
- Hahne, A., Mueller, J. L., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related brain potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, *18*, 121–134.
- Handy, T. C. (2005). *Event-related potentials: A methods handbook*. Cambridge, MA: MIT Press.
- Inoue, K., & Osterhout, L. (2012). Sentence processing as a neural seesaw. Ms., University of Washington, Seattle, WA.
- Jackendoff, R. (2007). A parallel architecture perspective on language processing. *Brain Research*, *1146*, 2–22.
- Jasper, H. H. (1958). The ten–twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.
- Johnson, J., Shenkman, K., Newport, E., & Medin, D. (1996). Indeterminacy in the grammar of adult language learners. *Journal of Memory and Language*, *35*, 335–352.
- Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject–verb dependencies: An ERP study. *Journal of Psycholinguistic Research*, *31*, 165–193.
- Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, *52*, 205–225.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 376–395.
- Kolk, H. 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 and Language*, *85*, 1–36.
- Kuperberg, G. (2007). Neural mechanisms of language comprehension: Vhallenges to syntax. *Brain Research*, *1146*, 23–49.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science*, *207*, 203–205.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (eds.), *Hanbook of bilingualism: Psycholinguistic approaches*, pp. 49–67. Oxford: Oxford University Press.
- MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*, 127–150.
- McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*, 395–423.
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401.
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704.
- McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*, 123–150.
- Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, *22*, 205–220.
- Morgan-Short, K., Faretta, M., Brill, K., Wong, F., & Wong, P. (2012). Declarative and procedural memory as individual

- Q1 differences in second language acquisition. Ms., University of [■■■■■](#).
- Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potentials study. *Language Learning, 60*, 154–193.
- Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*, 933–947.
- Naiman, N., Fröhlich, M., Stern, H. H., & Todesco, A. (1996). *The good language learner*. Philadelphia, PA: Multilingual Matters.
- 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. *Journal of Cognitive Neuroscience, 22*, 2886–2898.
- Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.
- Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164*, 81–94.
- 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. *Journal of Cognitive Neuroscience, 24*, 1205–1223.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research, 24*, 691–701.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphor noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119–131.
- Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and the amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research, 70*, 197–205.
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of Proficiency. *Journal of Cognitive Neuroscience, 17*, 1212–1228.
- Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*, 494–522.
- Osterhout, L., Frenck-Mestre, C., Inoue, K., McLaughlin, J., Tanner, D., & Herschensohn, J. (2012). Morphosyntactic learning and second language acquisition: Evidence from event-related potentials. Ms., University of [■■■■■](#).
- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 785–806.
- Osterhout, L., & Holcomb, P. J. (1995). Event-related brain potentials and language comprehension. In M. D. Rugg & M. G. H. Coles (eds.), *Electrophysiology of mind: Event-related brain potentials and cognition*, Oxford: Oxford University Press.
- Osterhout, L., McLaughlin, J., Kim, A., Greewald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton (eds.), *The on-line study of sentence comprehension: Eyetracking-ERPs, and beyond*, pp. 271–308. New York: Psychology Press.
- Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck-Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning, 56*, 199–230.
- Osterhout, L., & Mobley, L. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.
- Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes, 14*, 282–317.
- Pakulak, E., & Neville, H. J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience, 22*, 2728–2744.
- Prat, C. S. (2011). The brain basis of individual differences in language comprehension abilities. *Language and Linguistic Compass, 5*, 635–649.
- Robinson, P. (2002). *Individual differences and instructed language learning*. Amsterdam: Benjamins.
- Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 18*, 2030–2048.
- Sabourin, L., & Haverkort, M. (2003). Neural substrates of representation and processing of a second language. In R. van Hout, A. Hulk, F. Kuiken & R. Towell (eds.), *The lexicon–syntax interface in second language acquisition*, pp. 175–195. Amsterdam & Philadelphia: John Benjamins.
- Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*, 397–430.
- Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research, 22*, 1–29.
- Schwartz, B. D., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research, 12*, 40–72.
- Severens, E., Jansma, B. M., & Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject–verb agreement: An ERP study. *Brain Research, 1228*, 135–144.
- Skehan, P. (1989). *Individual differences in second-language learning*. New York: Arnold.
- Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence

from event-related brain potentials. *Second Language Research*, 25, 13–41.

Tanner, D. (2011). Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference. Ph.D. dissertation, University of Washington.

Tanner, D., Inoue, K., & Osterhout, L. (2012). Brain-based individual differences in on-line L2 sentence comprehension. Ms., University of Washington, Seattle, WA. [Submitted for publication in ■■■■.]

Q2

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar – An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173–204.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105–122.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270.

Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (ed.), *Mind and context in adult second language acquisition*, pp. 141–178. Washington, DC: Georgetown University Press.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22, 67–82.

van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, 26, 43–74.

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41–63.

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

Weber-Fox, C. M., & Neville, H. J. (1996). Maturation constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256.

Wickens, T. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.

Zayas, V., Greenwald, A., & Osterhout, L. (2010). Unintentional covert motor activations predict behavioral effects: Multilevel modeling of trial-level electrophysiological motor activations. *Psychophysiology*, 48, 208–217.