

Growing *Random Forests* reveals that exposure and proficiency best account for individual variability in L2 (and L1) brain potentials for syntax and semantics

Lauren A. Fromont^{a,b,*}, *Phaedra Royle*^{a,b}, *Karsten Steinhauer*^{b,c}

^a École d'orthophonie et d'audiologie Université de Montréal, Montréal, Canada

^b Centre for Research on Brain, Language and Music, Montreal, Canada

^c School of Communication Sciences and Disorders, McGill University, Montreal, Canada

* Corresponding author:

Lauren A. Fromont, Ph.D.

Université de Montréal
École d'orthophonie et d'audiologie
7077 Av du Parc, 3ème étage
Montreal, Quebec H3N 1X7; Canada.
Email: lauren.fromont@umontreal.ca

Phaedra Royle, Ph.D.

Université de Montréal
École d'orthophonie et d'audiologie
7077 Av du Parc, 3ème étage
Montreal, Quebec H3N 1X7; Canada.
Email: phaedra.royle@umontreal.ca

Karsten Steinhauer, Ph.D.

McGill University
School of Communication Sciences & Disorders
2001 McGill College Ave. Unit 800
Montreal, Quebec H3A-1G1; Canada.
Email: karsten.steinhauer@mcgill.ca

Abstract

Late second language (L2) learners report difficulties in specific linguistic areas such as syntactic processing, presumably because brain plasticity declines with age (following the critical period hypothesis). While there is also evidence that L2 learners can achieve native-like online-processing with sufficient proficiency (following the convergence hypothesis), considering multiple mediating factors and their impact on language processing has proven challenging. We recorded EEG while native (n = 36) and L2-speakers of French (n = 40) read sentences that were either well-formed or contained a syntactic-category error or a lexical-semantic anomaly. Consistent with the critical period hypothesis, group differences revealed that while native speakers elicited a biphasic N400-P600 in response to ungrammatical sentences, L2 learners as a group only elicited an N400. However, individual data modeling using a Random Forests approach revealed that language exposure and proficiency are the most reliable predictors in explaining the ERP responses, with N400 and P600 effects becoming larger as exposure to French as well as proficiency increased, as predicted by the convergence hypothesis.

Keywords: Second language acquisition, Event-related potentials, Syntax, Semantics, Random Forests

1. Introduction

While native-like first-language (L1) attainment is considered to be the gold standard for second-language (L2) learners, there is considerable variability in how individuals process their L2. They may acquire their second language early in life or during adulthood, receive various amounts of exposure, and as a result their proficiency in different linguistic domains may vary. Many factors play a role in cognitive processes underlying sentence comprehension in a second language, but there is an ongoing debate on their relative importance (Birdsong, 2018; Steinhauer, White, & Drury, 2009; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013). Using event-related potentials (ERPs), the present study investigates online processing of French syntactic categories in L1-speakers and late (post-puberty) L2-learners of French. It addresses the relative strength of different predictors, including age of acquisition, proficiency, and exposure, in accounting for electrophysiological patterns.

Four ERP components are of interest. The ELAN (early left-anterior negativity) is an early negative shift over frontal electrodes found in response to syntactic-category violations (henceforth: SCVs) and is thought to reflect initial phrase-structure building (Friederici, 2002, 2011). The N400 is a central negativity that is generally observed in response to difficulties in lexical-semantic retrieval (Lau, Phillips, & Poeppel, 2008), integration (Steinhauer, Royle, Drury, & Fromont, 2017), or difficulties or failure to predict target words (Kutas & Federmeier, 2011). The LAN (a left-lateralized, often anterior or sometimes temporal, negativity) and the P600 (a parietal positive shift) are observed in response to (morpho-)syntactic anomalies, such as agreement errors (see Molinaro, Barber & Carreiras, 2011, for a review, but see Tanner, 2015) or SCVs (e.g., Bowden, Steinhauer, Sanz, Ullman, 2013). Although the P600 is far from reflecting *only* syntactic processes (Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer, Fitz, & Hoeks, 2012; Kuperberg, 2007; Meerendonk, Kolk, Vissers, & Chwilla, 2008) in L2 research, ELAN/LAN-P600 effects have long been thought to be a hallmark of native-like proficiency in morpho-syntactic processing (Rossi, Gugler, Friederici, & Hahne, 2005; Pakulak & Neville, 2011). Recently, this dominant view has been challenged by research focusing on individual differences in L1 and L2 speakers suggesting that a biphasic response is not necessarily a standard even in

native speakers (Tanner, 2019; Tanner, Goldshtein & Weissman, 2018). However, factors promoting native-likeness in L2 remain controversial.

Factors contributing to native-like attainment

There is a broad consensus that age of acquisition (AoA) is important when it comes to L2 learning. Later-life language learning is often associated with lower attainment attributed – at least partly – to biological factors such as decline in brain plasticity. One famous yet controversial framework, the critical period hypothesis (CPH; Johnson & Newport, 1989; Lenneberg, 1967; Newport et al., 2001), states that only during an early “critical” period in life is the brain able to establish optimal neural connections to process linguistic input. This would mean that early L2 experience is essential to reaching native-like mastery (Birdsong, 2006; VanHove, 2013), although not all linguistic domains may be equally affected by AoA. Lexical-semantic processing seems relatively unaffected (Clahsen & Felser, 2006), even though N400 effects may be somewhat delayed or smaller in late L2-learners (e.g. Sanders & Neville, 2003; see also Mueller’s 2005 review). Early ERP work on syntactic-category processing seemed to support the CPH. Weber-Fox and Neville (1996) investigated brain responses to SCVs (*Max’s *of proof the theorem* vs. *Max’s proof of the theorem*) and semantic violations (*Max’s *event of the theorem*), in L1-Mandarin Chinese L2-English speakers who were divided into five groups according to AoA. Native-like semantic N400 effects were reliably observed across groups, whereas ERP profiles for syntax violations appeared less native-like with increasing AoA. Participants who learned English after 16 did not even display a P600 component. These findings seem to suggest that late learners rely on different mechanisms than do native speakers to process syntactic structures, but show no qualitative differences in lexical-semantic processing. Similarly, auditory ERP studies that combined lexical-semantic anomalies and SCVs on the same target-word (Hahne & Friederici, 2001) reported that L2-learners displayed late semantic N400s in response to these combined anomalies, while L1-speakers displayed a “syntactic” ELAN but no N400, because syntactic violations should block any lexical-semantic processing in native speakers (the “semantic blocking” effect, Friederici, Steinhauer, & Frisch, 1999). Thus, late (adult) L2-learners appear to rely more on lexical-semantic cues than syntax-guided native speakers when processing

sentences. However, in many studies that have investigated AoA effects on syntactic processing (e.g. Weber-Fox & Neville, 1996; Hahne, 2001), AoA and L2 proficiency were correlated, which presents a challenging confound: one can question whether ERP group differences originally attributed to AoA were rather due to different proficiency levels (Caffarra, Molinaro, Davidson, & Carreiras, 2015; Steinhauer et al., 2009; Steinhauer, 2014).

The convergence hypothesis (Green, 2003; Steinhauer et al., 2009) proposes that different levels of language proficiency in morpho-syntax are characterized by qualitatively distinct neurocognitive processing mechanisms that are reflected by distinct ERP patterns which can ultimately converge on a native-like profile. It predicts that low proficiency L2-learners show no ERP responses if they do not process the error, or elicit an N400 in response to morpho-syntactic violations, reflecting the fact that they process them as lexical anomalies (see also Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; and Ullman's declarative/procedural model, 2001; 2004). As proficiency increases and linguistic rules become more grammaticalized, ERP signatures will mutate to broadly distributed P600-like effects, and finally converge on a native-like response (e.g., a biphasic LAN-P600 effect, if that is what is observed in L1-speakers). In a study focusing on L2-learners of an artificial language, Friederici, Steinhauer, and Pfeifer (2002) showed that participants who received grammar training elicited ERP responses similar to those usually found in native speakers (early negativities followed by a P600) in response to SCVs, while participants who only received vocabulary training did not (see also Morgan-Short, Steinhauer, Sanz, & Ullman, 2012). Steinhauer and colleagues (2009) also report a study where both native English speakers and high proficiency late learners elicit a LAN-P600 complex in response to SCVs, while lower proficiency participants only elicit a small P600. Proficiency also seems to play a role in native monolinguals. Pakulak and Neville (2010) suggest that L1-speakers' proficiency correlates with the magnitude of both anterior negativities and late positivity in response to syntactic errors (see White, Genesee, White, King, & Steinhauer, 2006, for similar findings). In a number-agreement processing study focusing on inter-individual variability in German L2-learners, Tanner et al. (2013)

observed that individuals varied along a continuum from an N400 to a P600 profile, and that only the P600 increased as participants became better able to detect anomalies.

The amount of exposure to a target language in a naturalistic environment has a positive influence on the processing of grammatical structures (e.g., Frenck-Mestre, 2002; 2001). Although few ERP studies have investigated specific effects of exposure on syntactic processing in cohorts of L2 adult speakers, some studies have included immersion, in terms of years of exposure to a target language, or length of residence in a country where the target language is spoken (Caffarra, Molinaro, Davidson, & Carreiras, 2015). For example, in an agreement study, Tanner, Inoue, and Osterhout (2014) found that proficiency – based on self-ratings and pencil-and-paper assessments – predicted overall response magnitude, reflected non-specifically by a larger N400 or a larger P600 component, while immersion (length of residence) and motivation to speak like a native speaker were associated with a P600-dominant response. Note, however, that immersion in years (or length of residence) does not necessarily reflect the amount of exposure L2 speakers receive in their daily lives. This is especially true for multilingual environments, where speakers may receive some input from their target language but not necessarily use it in their daily interactions.

As individual differences along an N400-P600 continuum are also observed in L1-speakers (Osterhout, 1997; Osterhout, McLaughlin, Kim, Greewald & Inoue, 2004; Tanner and Van Hell, 2014), response dominance may be a characteristic of sentence processing in general, reflecting individual preferences to rely on one processing stream over the other, e.g., memory-based heuristics (reflected by an N400 dominance) versus procedural or combinatorial information (indexed by a P600 dominance). In sum, a biphasic LAN-P600 should not be considered the gold standard for L1 attainment. In fact, the convergence hypothesis does not predict that highly proficient L2-speakers should specifically elicit a LAN-P600 profile, nor that this profile is consistently elicited in native speakers (Steinhauer, 2009, 2014). Rather, it predicts that at very high proficiency, L2-speakers' responses to syntactic violations would be indistinguishable from those of native speakers. We believe that this proficiency-based approach would benefit from considering inter-individual variability, not

only in L2-speakers, but also in L1-speakers. Tanner et al.'s (2013, 2014) studies consider L1- and L2-speakers separately, while we propose to group L1- and L2-speakers into the same analysis, without a priori assumptions.

Syntactic category identification and phrase-structure building

While the vast majority of studies observe a P600 effect in response to SCVs, some studies found an additional ELAN (starting with Friederici, Pfeifer, Hahne, 1993; and Friederici, 1995), others a LAN (Van den Brink & Hagoort, 2004), and more recently an N400 (Fromont, Steinhauer, & Royle, resubmitted; Nickels, Bokhari, & Steinhauer, 2014; Zhang et al., 2013). While the LAN-N400 discrepancy could be explained by inter-individual variability as explained above, we interpret with caution the much earlier ELAN effect as well as “semantic blocking”, due to serious methodological concerns with experimental designs (Steinhauer & Drury, 2012). Most studies investigating SCVs use “context manipulation”, whereby the pre-target context differs between conditions (e.g., Hahne, 2001; Hahne & Friederici, 2001; Isel, 2007; Mueller, Hahne, Fujii, & Friederici, 2005; Pakulak & Neville, 2011; Weber-Fox & Neville, 1996, for L2 processing). For example, in French, Isel (2007) manipulated syntactic-category by omitting the noun after a definite article (*L’homme qui est dans la *Ø/maison dort* ‘The man who is in the *Ø/house sleeps’), the context preceding the target verb *dort* ‘sleeps’ where the ERP is analyzed. He found ELANs for SCVs in both native speakers and proficient L2-learners. However, this finding is problematic because the differing pre-target contexts can elicit ERP artefacts that resemble ELANs but are not related to the SCV (Steinhauer & Drury, 2012, see also Lau, Stroud, Plesch, & Phillips, 2006). In fact, Isel (2007) found an “ELAN” 100 ms after target word onset, hundreds of milliseconds *before* the syntactic violation was present in the speech signal, clearly reflecting an artefact. Nichols and Joanisse (2019) observed that, in participants who learned French before puberty, proficiency, not AoA, modulated the LAN response to SCVs, while P600s were absent in both L1- and L2-speakers. However, they did not directly compare their conditions, so there is no statistical support for the presence or absence of ERP differences. In addition to these problems, many studies testing SCVs simply fail to create syntactic or semantic violations on the target word (Steinhauer & Drury, 2012).

Using a novel paradigm in French, Fromont et al. (resubmitted, 2019) systematically manipulated syntactic-category (correct/incorrect) and lexical-semantic anomalies (primed/unprimed), while using a balanced experimental design that avoided methodological issues present in previous studies. This is one of the few studies successfully creating outright SCVs *and* lexical-semantic anomalies on the target word, while deploying a balanced design where identical context *and* target words contributed equally to correct and violation conditions (see §2.2). This is a substantial improvement to context or target manipulation designs (e.g., Luo, Zhang, Feng, & Zhou, 2010; Zhang et al., 2013) and has only been implemented in a handful of studies (e.g., Bowden, Steinhauer, Sanz, & Ullman, 2013; Steinhauer et al., 2009). Fromont and collaborators reevaluated the time-course of syntactic-category identification in French native speakers ($n = 36$), and whether these processes interacted with lexical-semantic processing. They observed no ELAN in response to SCVs, which instead elicited a biphasic N400-P600. Lexical-semantic and syntactic manipulation effects on the N400 were additive, suggesting that the two types of error were processed in parallel but independently. Using the same design, we investigated ERP responses in L2-learners of French, while comparing them to the previously-described group of L1-French speakers, and explored effects of individual measures for AoA, exposure to French, and proficiency on ERP responses in both L1- and L2-speakers.

The current study

Previous SCV experiments yielded diverse results with respect to syntactic-category processing and its interaction with lexical-semantic processing. To better compare our data with previous findings, we will first use a traditional group design comparing L1-speakers with late L2-learners of French. While lexical-semantic anomalies should elicit similar N400s in both groups due to high levels of vocabulary mastery in the L2 group, we expect SCVs to elicit a biphasic N400-P600 response at high proficiency (based on Fromont et al. submitted) and a reduced or even absent P600 in lower L2-proficiency learners.

Second, we will investigate the effect of nine predictors related to AoA, proficiency, language exposure, and working memory abilities on the ERP responses. In order to (a) introduce a relatively large number of

correlated predictors in a single model and (b) evaluate their relative importance in explaining the modulation of ERP effects, we analyzed them using *Random Forests*, a statistical approach that has notable advantages over traditional ones but is novel to ERP research (see Tagliamonte & Baayen, 2012, and Tomaschek, Hendrix, & Baayen, 2018, for applications of Random Forests to linguistic data). The CPH would predict that AoA is the most important predictor explaining variability in the ERP responses, with native speakers showing a biphasic N400-P600 response, as in Fromont et al. (resubmitted, 2019), and L2-learners only an N400. Based on the convergence hypothesis, we predict that proficiency will modulate the ERP profiles of L2-learners to a larger extent than AoA, and may even account for some variability within native speakers.

2. Methods

2.1. Participants

Forty native French speakers (L1-speakers) and 45 late second-language learners of French (L2-speakers) who spoke Canadian or US dialects of English as a first language participated in the experiment. Data from eight participants (four per group) were excluded from the analyses due to excessive artifacts in the EEG, and one additional participant was excluded because of technical issues during the testing session. All participants were right-handed (Oldfield, 1971, adapted to French) and reported normal or corrected-to-normal vision. Age, education, and reading habits were obtained via a short demographic questionnaire, and their working memory was assessed using forward and backward digit span tests (Soylu, 2010). Critically, we ensured that the two groups only differed on language-related measures ($p < .001$, see Table 1). All L2-speakers reported an intermediate level in French: the B1 level on the Common European Framework of Reference for Languages (Council of Europe, 2011), and were not exposed to a third language more than 5% of the time. A language-background questionnaire established self-reports on age of acquisition (AoA, in the case of L2-speakers this would be their first French class), age of first regular exposure to French (AoE, the moment L2-speakers started living in a French-speaking environment). The amount of exposure was estimated using two variables. Immersion (in years) is an estimation of the amount of exposure across the lifespan and was calculated by subtracting AoE from a given participant's age. Daily usage was estimated by asking participants to what extent

they spoke French (as opposed to English) in their adult life in the following situations: at work or school, at home, and during social activities. Finally, participants were evaluated on their language proficiency using a C-Test (Tremblay & Garrison, 2010) and a lexical decision task (*LexTALE*, Brysbaert, 2013). Performance on each of the three tasks during the EEG experiment was used as independent and structure-specific proficiency measures, which according to Steinhauer et al. (2009) should be among the best predictors for ERP profiles. The internal consistency of individual responses to the linguistic tasks, estimated using Chronbach's alpha, suggests that the predictors were reliable (i.e. above the commonly accepted threshold of .7). Importantly, all measures related to AoA, exposure, and proficiency were correlated, as illustrated in Table 2.

Table 1. Descriptive statistics for participants demographics and individual differences measures. Significant differences are in bold.

Measure	L1 French (<i>n</i> = 36, 19F)		L2 French (<i>n</i> = 41, 26F)		<i>t</i> (<i>df</i>)	<i>p</i> -value	Internal consistency ^f
	<i>M</i> (<i>SD</i>)	Min – Max	<i>M</i> (<i>SD</i>)	Min – Max			
Age (years)	27.00 (5.47)	22 – 40	25.76 (4.08)	25 – 34	1.11 (76)	.270	
Education (years)	15.35 (1.92)	13 – 18	14.40 (2.86)	10 – 18	1.77 (76)	.081	
AoA (years)	0;0 (0;0)	–	12;5 (1;11)	10 – 18	43.09 (40)	< .001	
AoE (years)	0;0 (0;0)	–	17;11 (3;6)	12 – 27	34.39 (40)	< .001	
Daily usage^a	88;7 (10;8)	60 – 100	17;8 (13;11)	3 – 60	25.29 (76)	< .001	
Immersion (years)	27;0 (5;6)	22 – 40	7;10 (4;4)	2 – 21	16.83 (76)	< .001	
Reading Habits ^b	3.88 (0.81)	1 – 4	3.98 (0.83)	3 – 5	-0.51 (76)	.609	
Handedness	81.29 (16.01)	50 – 100	82.44 (16.88)	40 – 100	-0.31 (76)	.755	
LexTALE_FR^c	89.43 (4.04)	78 – 95	58.80 (11.03)	16 – 83	16.97 (57)	< .001	.87
Cloze test^c	68.33 (12.47)	33 – 91	39.65 (17.56)	11 – 89	8.09 (69)	< .001	.92
Performance SEM^d	0.28 (0.14)	0.03 – 0.57	0.12 (0.11)	-0.15 – 0.45	5.62 (76)	< .001	.87
Performance SYN^d	0.44 (0.17)	0.03 – 0.78	0.14 (0.13)	-0.23 – 0.42	8.51 (76)	< .001	.84
Performance SYNSEM^d	0.49 (0.18)	0.05 – 0.79	0.11 (0.25)	-0.52 – 0.51	7.84 (76)	< .001	.86
Reliable Digit Span score ^e	9.94 (2.41)	3 – 13	10.00 (2.65)	5 – 16	-0.09 (70)	.926	.89

^a Percentage of daily usage to French estimated since age 18

^b On a scale from 0 – never reads to 5 – reads a lot

^c Scores computed in percentages

^d Scores ranging from -1 to 1 (a score of zero indicates no discrimination between correct and incorrect sentences)

^e Task adapted from Soyulu (2010). The sum of the longest string of digits recalled correctly twice, under both forward and backward conditions (Greiffenstein, Baker, & Gola, 1994).

^f Estimated using Chronbach's alpha

Table 2. Correlation matrix between individual difference measures. All correlations except the ones involving digit span were significant ($p = .05$ or lower).

	AoA	AoE	C-test	Daily usage	Digit span	Immersion (in years)	LexTALE	Performance (Semantics)	Performance (Syntax)	Performance (Combined)
AoA										
AoE	.96***									
C-test	-.64***	-.60***								
Daily usage	-.86***	-.83***	.53***							
Digit span	.02	.04	.03	.01						
Immersion (in years)	-.80***	-.82***	.46***	.82***	-.03					
LexTALE	-.83***	-.81***	.54***	.80***	.06	.74***				
Performance (Semantics)	-.52***	-.50***	.31***	.54***	.09	.48***	.59***			
Performance (Syntax)	-.67***	-.65***	.45***	.72***	.05	.59***	.69***	.87***		
Performance (Combined)	-.60***	-.64***	.40**	.61***	.08	.55***	.64***	.73***	.83***	

2.2. Materials

Table 3 summarizes the conditions in our 2x2 design (SYNTAX; SEMANTICS: Correct/Anomalous). Each manipulation was based on item pairs (e.g., Condition 1 in Table 3). One item contains a verb as its target (hereafter underlined e.g., tackle)¹ because the preceding control² verb (*dare*) necessarily requires an infinitive verb as its complement. By contrast, the other item contains a noun as its target (e.g., toad), because the preceding transitive verb (*remove*) mandatorily selects for a noun phrase complement. Importantly, context words immediately preceding the target in both sentences were kept constant by using homographic clitic

¹ Note that the French form of the verb ‘tackle’ (*plaquer*) is not homographic or homophonous with the noun ‘tackle’ (*plaque*).

² We use the shortcuts “transitive verbs” and “control verbs” to distinguish the types of complements our main verbs take.

pronouns (e.g., *le* ‘him’) before verb targets, and definite determiners (e.g., *le* ‘the’) before noun targets. The implementation of homographic determiners/clitics allowed us to create SCVs by swapping target words across item pairs (Condition 2), while keeping the exact same pre-target words, thus avoiding any baseline problems during ERP analysis. Experimental sentences were preceded by context sentences that licensed the use of clitic pronouns and definite determiners. The context sentences contained a prime (e.g., *hockey* primes *tackle*), allowing us to manipulate semantic-context priming by swapping context sentences, such that the target was primed (Condition 1) or not (Condition 3) by its context. Note that swapping the context sentences *only* made the unprimed targets anomalous from a semantic-pragmatic point of view. We use the term “semantic anomaly” to describe this unprimed condition in order to avoid conflation with traditional priming paradigms. Finally, we created a combined anomaly condition where the target word belonged to the wrong syntactic-category and was also semantically anomalous (Condition 4).

Table 3. Sample stimuli for the eight experimental sub-conditions.

	SYNTAX	SEMANTICS	Introductory context sentence		Experimental sentence			
			PRIME		Subject pronoun + Verb	Clitic / Det.	Target	Prepositional phrase
1	✓	✓	<i>Marie et Jeanne jouent au hockey avec leur copain.</i>		<i>Elles osent</i>	<i>le</i>	<u><i>plaquer</i></u>	<i>sur le côté.</i>
			Mary and Jane are playing hockey with their friend _{MASC.}		They _{FEM} dare	him	<u>tackle</u>	on the side.
			<i>Marie et Jeanne vont au marais avec leur copain.</i>		<i>Elles ôtent</i>	<i>le</i>	<u><i>crapaud</i></u>	<i>sur le côté.</i>
			Mary and Jane are going to the swamp with their friend _{MASC.}		They _{FEM} remove	the _{MASC}	<u>toad_{MASC}</u>	on the side.
2	✗	✓	<i>Marie et Jeanne vont au marais avec leur copain.</i>		<i>Elles osent</i>	<i>le</i>	* <u><i>crapaud</i></u>	<i>sur le côté.</i>
			Mary and Jane are going to the swamp with their friend _{MASC.}		They _{FEM} dare	him	* <u>toad_{MASC}</u>	on the side.
			<i>Marie et Jeanne jouent au hockey avec leur copain.</i>		<i>Elles ôtent</i>	<i>le</i>	* <u><i>plaquer</i></u>	<i>sur le côté.</i>
			Mary and Jane are playing hockey with their friend _{MASC.}		They _{FEM} remove	the _{MASC}	* <u>tackle</u>	on the side.
3	✓	✗	<i>Marie et Jeanne vont au marais avec leur copain.</i>		<i>Elles osent</i>	<i>le</i>	? <u><i>plaquer</i></u>	<i>sur le côté.</i>
			Mary and Jane are going to the swamp with their friend _{MASC.}		They _{FEM} dare	him	? <u>tackle</u>	on the side.
			<i>Marie et Jeanne jouent au hockey avec leur copain.</i>		<i>Elles ôtent</i>	<i>le</i>	? <u><i>crapaud</i></u>	<i>sur le côté.</i>
			Mary and Jane are playing hockey with their friend _{MASC.}		They _{FEM} remove	the _{MASC}	? <u>toad_{MASC}</u>	on the side.
4	✗	✗	<i>Marie et Jeanne jouent au hockey avec leur copain.</i>		<i>Elles osent</i>	<i>le</i>	* <u><i>crapaud</i></u>	<i>sur le côté.</i>
			Mary and Jane are playing hockey with their friend _{MASC.}		They _{FEM} dare	him	* <u>toad_{MASC}</u>	on the side.
			<i>Marie et Jeanne vont au marais avec leur copain.</i>		<i>Elles ôtent</i>	<i>le</i>	* <u><i>plaquer</i></u>	<i>sur le côté.</i>
			Mary and Jane are going to the swamp with their friend _{MASC.}		They _{FEM} remove	the _{MASC}	* <u>tackle</u>	on the side.

Twenty verb pairs (transitive vs. control verbs) were selected, each with eight different target words (nouns vs verbs) and corresponding context sentences (Fromont, resubmitted, 2019). Our primary concern was to avoid pre-target and target-specific effects, as well as sentence-internal effects that can affect the ERPs. Therefore, we ensured minimal differences between the items of each matched-sentence pair in terms of phonological and lexical properties, acceptability ratings on correct and incorrect conditions, and degree of priming between prime and target word. A total of 1280 sentences were divided into four lists using a Latin-square design, such that each participant read 320 sentences (80 per condition), without repeating any context sentence or target word. Eighty filler sentences that were either correct or contained subject-verb agreement errors were added to each list (see Fromont et al., resubmitted, for details).

2.3. Experimental procedure

All procedures were approved by the Ethics Review Boards at McGill and University of Montreal faculties of Medicine. Participants were tested in a 2.5-hour session in the Neurocognition of Language Lab at McGill University. After completing background questionnaires, participants were seated in a chair 80 cm away from a computer monitor and read sentences in white 30-point Arial font on a black background. Each trial started with a “!!!” prompt for 1800 ms, where participants were encouraged to blink, followed by a fixation cross for 500 ms. Sentences were presented in rapid-serial-visual presentation mode: each word was presented on the screen for 300 ms followed by a 200 ms blank screen interval. At the end of every trial, 500 ms after the last word, a “???”-prompt remained on the screen until participants scored sentence acceptability by pressing a button between 1 to 5 (1: totally acceptable, 5: totally unacceptable). Their key press was immediately followed by an eye-blink prompt. After the EEG experiment, participants completed the Digit Span task, the C-test, and the lexical decision task (LexTALE).

2.4. EEG recording and data processing

EEG was recorded continuously from 21 Ag-Cl active-shield electrodes mounted on an EEG cap (Waveguard™ original, *ANT Neuro*, Netherlands) according to the 10-20 system (Jasper, 1958) at the following sites: FP1-FPZ-FP2-F7-F3-FZ-F4-F8-T3-C3-Cz-C4-T4-T5-P3-PZ-P4-T6-O1-OZ-O2, with a 512 Hz sampling rate and a 0.001–100 Hz online forward filter. All EEG electrodes were referenced online against the right mastoid. An electrode between FPZ and FZ served as ground. Impedances were kept below 5 k Ω .

Data were analyzed using *EEGLab* (Delorme & Makeig, 2004) and *ERPLab* (Lopez-Calderon & Luck, 2014). Continuous data were re-referenced offline to average mastoids, and high-pass – and low-pass – filtered separately with .1 and 40 Hz cut-off frequencies respectively (IIR Butterworth filters). After first epoching the data from -1000–2000 ms relative to target onset, and performing a baseline correction using a 200 ms pre-target baseline interval, we rejected data that exceeded a peak-to-peak threshold of 70 μ V (in 100 ms steps). We then visually inspected the remaining epochs and deleted ones still affected by artefacts. Analyses were then performed on shorter epochs of -200–1800 ms aligned to target word onset, and baseline corrected (-200–0 ms).

2.5. Statistical analyses

Behavioral and EEG data were analyzed using mixed effects models (*lme4*, Bates, Mächler, Bolker, & Walker, 2015; *lmerTest*, Kuznetsova, Brockhoff, & Christensen, 2016; *emmeans*, Lenth, 2018). The maximal converging random structure included random slopes for condition per participant. Model selection was performed by decrementally removing interactions and factors from this full model until we reached the optimal model, determined by comparing two minimally different models using ANOVAs.

As performance at the task, and in particular the ability to discriminate between correct and incorrect sentences, can be viewed as the most suitable predictor variable to explore inter-individual variability for structures under investigation (Nickels & Steinhauer, 2018; Steinhauer et al., 2009; Tanner et al., 2013), we analyzed online performance as follows: From the acceptability scores (on a 5-point scale, see above) we subtracted the rating for correct from anomalous conditions, and then divided the result by 4, thereby transforming scores to an index from -1 to 1 reflecting discrimination between incorrect and correct sentences (similar to d-prime values, but for scaled data). An index of 1 means that sentences with an error were always rated 5 and correct sentences were rated 1, an index of 0 indicates that sentences were rated the same regardless of their correctness, and an (unlikely) index below 0 means that correct sentences were accepted less than incorrect ones. This index, reflecting the average performance per participant and condition was used as the input for the mixed-effects model. It was used because it is analogous to calculating difference waves, which we used to explore interindividual variability in the ERP data. In this model, we assessed the effect of CONDITION (three levels: Syntax, Semantics, Combined) and GROUP (two levels: L1 and L2) on elicited ERP components.

ERP effects on the midline and lateral sites were also analyzed separately using mixed effects models. The input for the models was aggregated data³ with one average observation per participant per sub-condition. In this omnibus analysis we employed a 2x2 design following analyses in Friederici's studies (1999 and

³ In Fromont et al (resubmitted), analyses were performed on both single observations and aggregated data and did not show any differences. We opted for aggregated data in the present study for computational resource reasons.

following). We first calculated main effects and interactions for factors SYNTAX (two levels: Correct, Anomalous), SEMANTICS (two levels: Correct, Anomalous), GROUP, ANTERIORITY (two levels: F and C electrodes as Anterior sites, P and O electrodes as Posterior sites), and HEMISPHERE (two levels: Left, Right). Levels were coded such that the Intercept modeled the Correct levels of each experimental condition at the Left Anterior sites. We used ANOVA wrappers (Type III Wald chi-square test) with the *car* package (Fox & Weisberg, 2011) to present the outcomes.

We also wanted to assess which predictor best explained inter-individual variability in ERP components. Although mixed-effect models could incorporate random slopes that accounted for some inter-individual variability, it was not possible to integrate our nine individual variables into them, because we faced multicollinearity issues whenever variables were correlated. We therefore opted for *Random Forests* to explore inter-individual variability (Matsuki, Kuperman, & Van Dyke, 2016). Random Forests is a machine-learning algorithm that can incorporate many variables with relatively few cases, while taking interactions into account and remaining more robust in the presence of collinearity between variables. Importantly, it allowed us to rank the importance of predictor variables in explaining the data, which is a feature particularly suited to our research questions. Psycholinguistic studies have started to successfully use Random Forests (de Aguiar, Bastiaanse & Miceli, 2016; Strobl, Malley & Tutz, 2009), but to our knowledge this is the first time they are used to analyze linguistic ERP data.

3. Results

3.1. Online performance

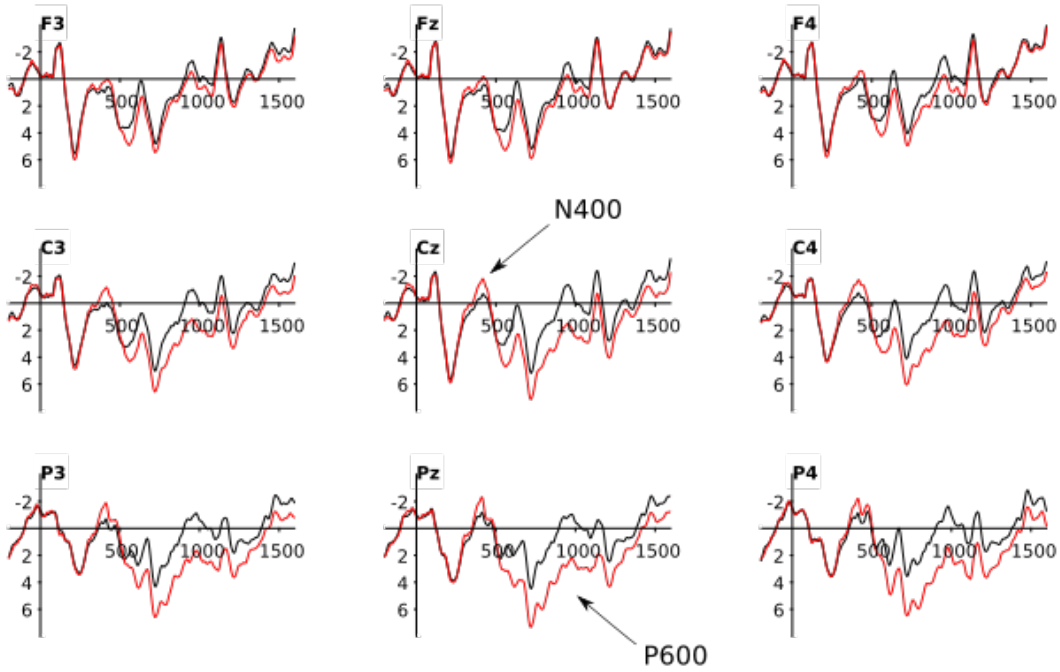
Overall, participants were accurate in judging the acceptability of sentences on the 5-point Likert scale (mean rating for correct items = 2.17, SD = 0.43; syntactic anomalies = 3.30, SD = 0.74; lexical-semantic anomalies = 2.95, SD = 0.53; combined anomalies = 3.47, SD = 0.78). Note however, that L2-speakers accepted correct sentences to a lesser extent (*Mean* = 2.35, *SD* = 0.38) than the L1-group (*Mean* = 1.93, *SD* = 0.39; *Mean difference* = 0.3; $X^2(1) = 9.88, p = 0.002$), which impacts their performance index, described above and presented in Table 1.

3.2.ERP effects: Group comparisons between L1 and L2

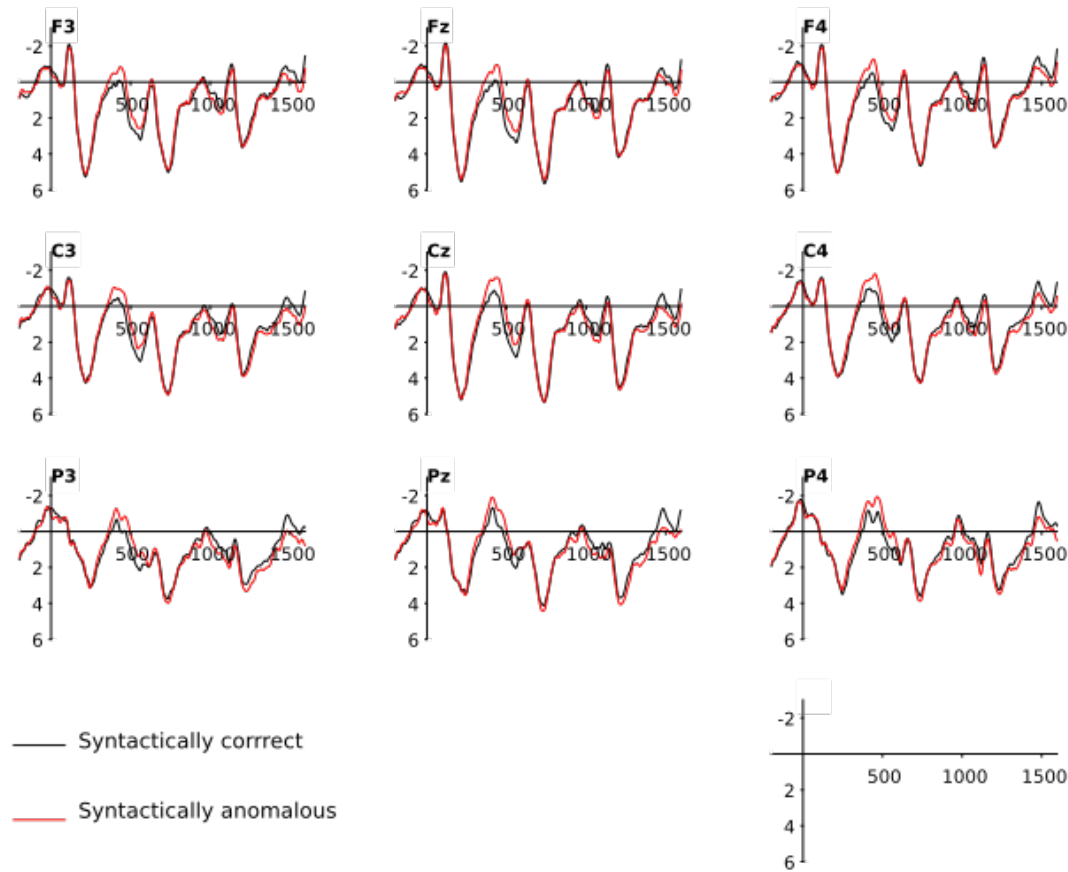
Following Friederici et al. (1999) and Nickels (2016), we first investigated the main effects of Syntax (SCVs and Combined anomalies vs. Correct and Lexical-semantic anomalies) and Semantics (Lexical-semantic and Combined anomalies vs. Correct and SCVs), as well as their potential interactions, comparing L1 and L2 groups (see Figure 1 and Table 4). Results showed that SCVs elicited a biphasic N400-P600 response in L1-speakers, but only an N400 in L2-speakers. In contrast, semantic anomalies yielded N400s in both groups with a similar central distribution. Our statistical analyses (Table 4) tested main effects and interactions for SYNTAX and SEMANTICS in two representative time-windows (250–500 ms for N400 effects, and 800–1200 ms for P600 effects). Non-adjacent time-windows were selected in order to reduce spatiotemporal component overlap between N400 and P600 waveforms that may affect latent N400 and P600 effects and inflate correlations between them.

A- SYNTAX

L1 Group

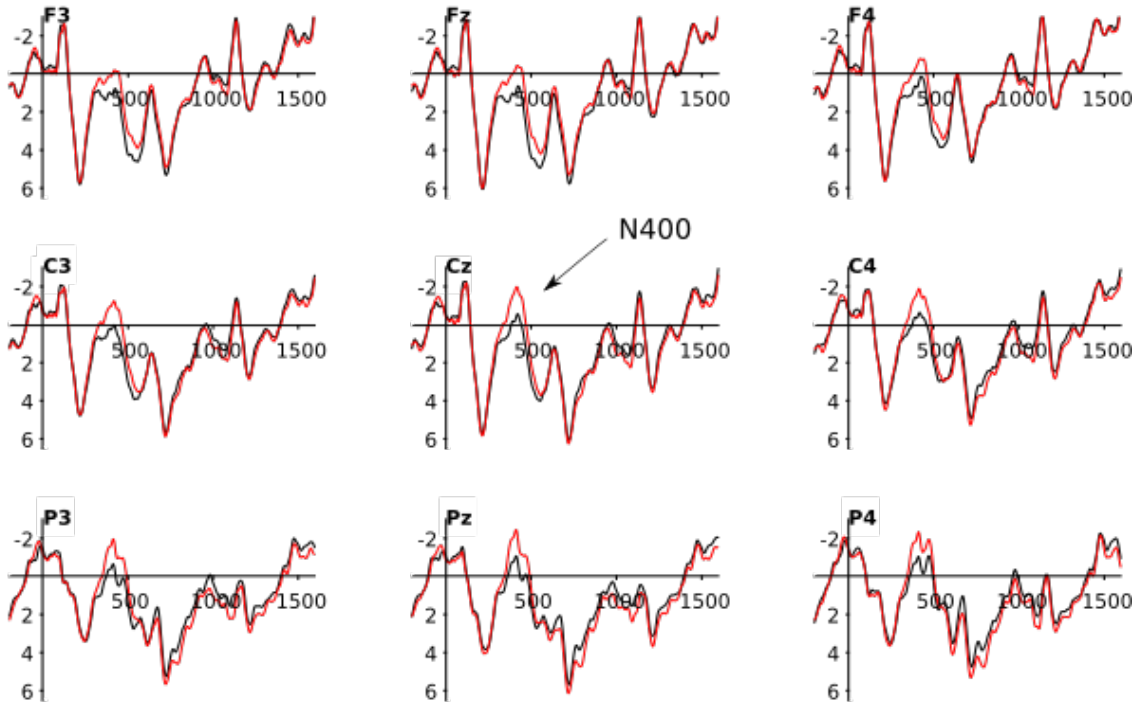


L2 Group

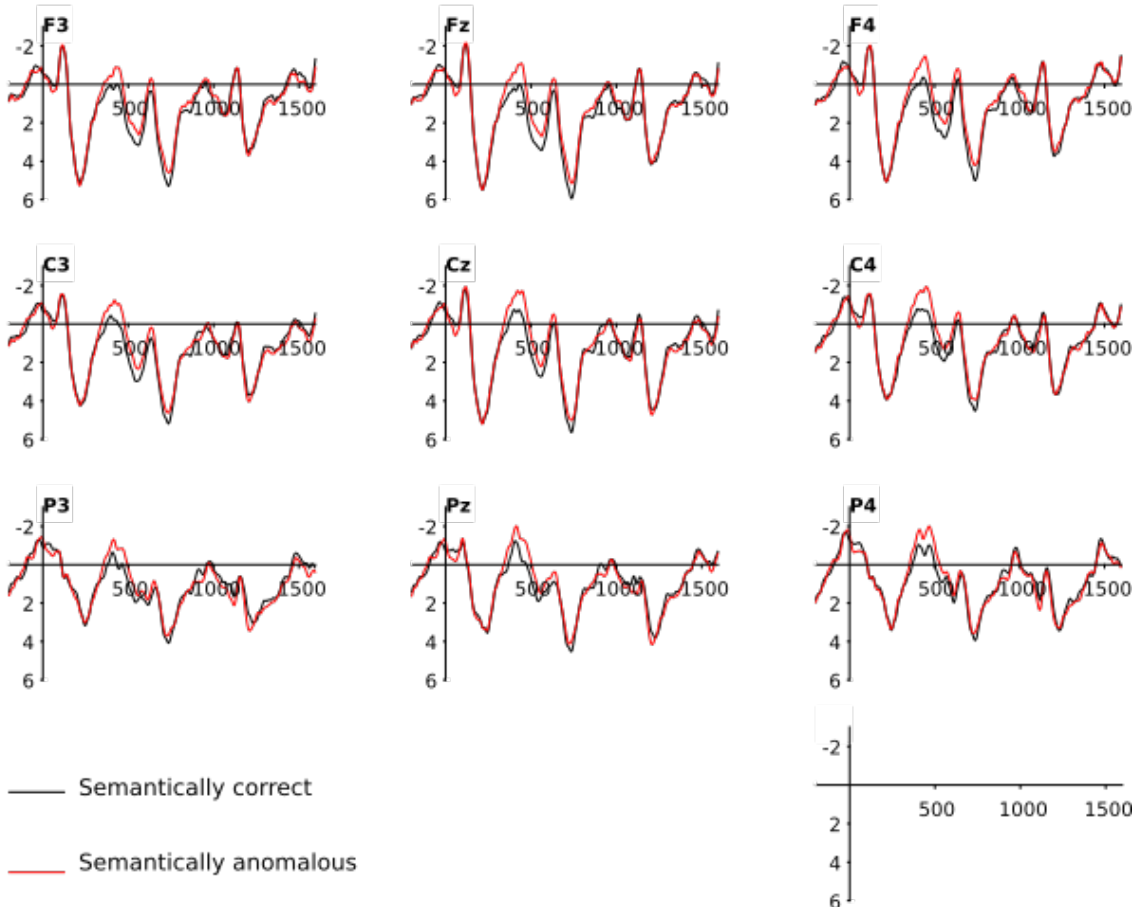


B- SEMANTICS

L1 Group



L2 Group



C- DIFFERENCE WAVES

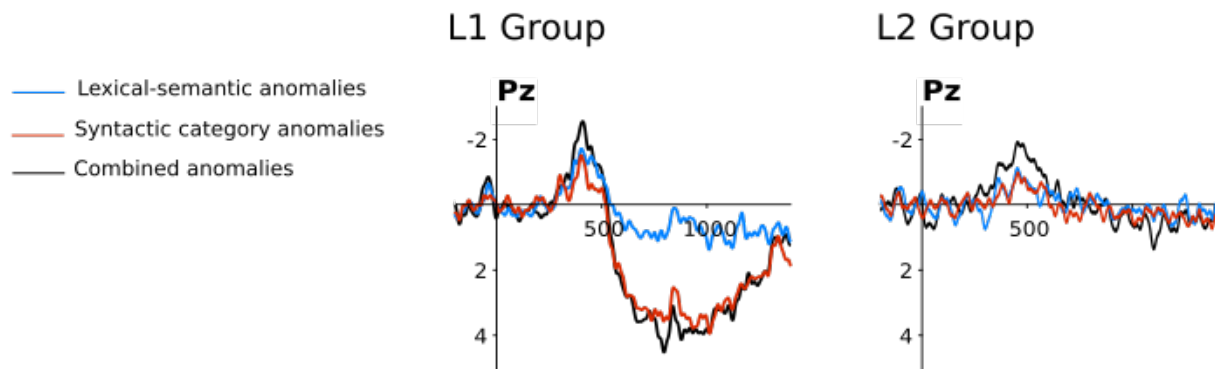


Figure 1. Grand average waveforms illustrating main effects of factors Syntax (A) and Semantics (B) in the 2x2 design in nine representative electrodes, and differences waves (C) between anomalous and correction conditions at Pz. Groups are plotted separately. Target presentation is indicated by the vertical bar, where tick bars represent 2 μ V of activity; tick marks on the horizontal line represent 500ms.

Table 4. Analysis of deviance table (Type III Wald chi-square tests) reporting significant effects corresponding to the main mixed-effect models on average amplitudes measured in the N400 and P600 time-windows, at midline electrodes and lateral sites.

Time-window	Site	Fixed effects	Chi-square	Df	p-value	
250–500 ms (N400 effect)	Midline $R^2 = .07^a$	(Intercept)	25.64	1	< 0.001	
		SYNTAX	6.34	1	0.012	
		SEMANTICS	8.82	1	0.003	
		SYNTAX×SEMANTICS×GROUP	3.94	1	0.047	
		Lateral sites $R^2 = .04$	(Intercept)	32.73	1	< 0.001
		SYNTAX	3.84	1	0.050	
		SEMANTICS	7.28	1	0.007	
	800–1200 ms (P600 effect)	Midline $R^2 = .05$	(Intercept)	0.46	1	0.493
			SYNTAX	17.22	1	< 0.001
			SEMANTICS	5.52	1	0.019
SYNTAX×GROUP			7.61	1	0.006	
SYNTAX×ANTERIORITY			5.64	1	0.018	
		SYNTAX×GROUP×ANTERIORITY	5.64	1	0.017	
Lateral sites $R^2 = .06$		(Intercept)	1.59	1	0.207	
		SYNTAX	7.38	1	0.007	
		SEMANTICS	6.91	1	0.009	
		SYNTAX×ANTERIORITY	20.68	1	< 0.001	
	SYNTAX×GROUP×ANTERIORITY	22.3103	1	< 0.001		

^a We report marginal R-squared values that are associated with fixed-effects.

In the 250–500 ms time-window, both L1- and L2-speakers showed broadly distributed N400s in response to lexical-semantic anomalies (i.e., a main effect of SEMANTICS), and a similar (albeit smaller) N400 effect for syntactic violations (main effect of SYNTAX). The absence of interactions between the two factors suggests that these N400 effects are additive. Additive semantic and syntactic N400 effects would predict the largest N400 in the combined violation condition, and this is exactly what was found in both groups (see Figure 1C). The SYNTAX×SEMANTICS×GROUP interaction at midline electrodes did not reveal any significant follow-up interactions by SYNTAX, SEMANTICS or GROUP (despite a marginal SYNTAX×SEMANTICS interaction in the L2 group, $p = .06$), and should be interpreted with caution. It primarily reflects larger N400s in L1 than L2 group for pure syntactic and semantic anomalies, but comparable N400s in both groups for the combined condition (as illustrated in Fig 2C). To support this conclusion, we ran further analyses focusing on the N400 at three

representative electrodes (C3-Cz-C4). We investigated effects of ELECTRODE, GROUP (2 levels: L1 vs. L2), and CONDITION (four levels). The first three levels were the difference waves between each of the three incorrect conditions (lexical-semantic incongruencies, syntactic category violations, and combined anomalies) minus the correct condition. We also included a fourth artificial condition by adding effects found on the lexical-semantic and pure syntactic N400s (i.e. the “additive” condition). If the observed effects are truly additive, the N400 effect observed in the combined condition should be greater than the N400 in the lexical-semantic and pure syntactic conditions, but should be indistinguishable from the artificial “additive” condition. This is exactly what we found. First, the mixed-effect model (including random slopes for CONDITION per PARTICIPANT) revealed a main effect of CONDITION ($X^2(1) = 20.75, p < .001$). Follow-up pairwise comparisons revealed that the estimated marginal means for lexical-semantic anomalies ($-0.69 \mu\text{V}$) and pure syntactic errors ($-0.57 \mu\text{V}$) were significantly smaller than those for combined anomalies ($-1.32 \mu\text{V}$; COMB-SEM: $t(78.05) = 3.928, p = .001$; COMB-SYN: $t(78.05) = 4.389, p < .001$) and also, importantly, than those for artificial “additive” anomalies ($-1.26 \mu\text{V}$; ADD-SEM: $t(78.05) = 3.929, p = .001$; ADD-SYN: $t(78.05) = 3.742, p < .001$). Conversely, the combined and “additive” conditions were statistically indistinguishable (COMB-ADD: $t(78.05) = -0.294, p = .991$).

In the 800–1200 ms time-window, we found a small but significant positivity (at Pz: $0.55 \mu\text{V}$ for L1 and $0.32 \mu\text{V}$ for L2) in response to lexical-semantic anomalies in both groups (main effect of SEMANTICS), and a large P600 effect in response to sentences with syntactic errors (main effect of SYNTAX) that was more posterior and mostly carried by native speakers, as supported by significant SYNTAX×GROUP×ANTERIORITY interactions especially at lateral sites. Follow-up analyses within each group confirmed this pattern: a SYNTAX×ANTERIORITY interaction was found in L1-speakers: Midline: $X^2(1) = 9.84, p = .002$, Lateral sites: $X^2(1) = 50.97, p < .001$, while no corresponding effect was observed in L2-speakers. In L1-speakers, the P600 was larger at posterior (Midline: SYN–COR = $2.56 \mu\text{V}$, $t(43.14) = 8.176, p < .001$; Lateral: SYN–COR = $2.27 \mu\text{V}$, $t(41.6) = 10.17, p < .001$) than anterior sites (Midline: SYN–COR = $1.16 \mu\text{V}$, $t(43.14) = 3.711, p < .001$; Lateral: SYN–COR = $0.60 \mu\text{V}$, $t(41.6) = 2.79, p = .008$).

As pointed out by one of the reviewers, observing a P600 in L2 learners for Semantics rather than Syntax seems rather surprising and, moreover, does not seem to be reflected in the ERP plots (Figure 1C). As we will see, this effect is related to individual differences, which we will address next.

3.3. Exploring inter-individual differences in ERP responses

Selecting a dependent variable

Previous work on inter-individual differences has pointed to a negative correlation between N400 and P600 effects, motivating computation of response dominance indices (N400 or P600 dominance) and response magnitude indices (component-independent response amplitudes) for individual participants (Tanner et al., 2014). Recall that a “negative correlation” means that if the N400 gets smaller (i.e., more positive), the P600 gets larger (also more positive), such that this pattern could be interpreted as (i) component overlap or (ii) a single slow wave modulating both ERP components (thus pointing to fundamentally different data interpretations). In fact, some studies analyze the entire N400-P600 complex as one unit, and incorporate time as a variable using generalized additive modeling, to better explore effects of AoA on inter-individual variability (Meulman, Wieling, Sprenger, Stowe, & Schmid, 2015). To determine the appropriateness of these approaches (and their conclusions), one would need to take topographic differences between the N400 and P600 into account. However, the aforementioned studies collapse across regions of interest to estimate the ERP patterns, running the risk that the correlation between the N400 and P600 simply reflects an overlap between these components. To determine whether we should consider P600 and N400 effects separately, or rather consider response dominance and magnitude indices as our dependent variable, we first calculated how our components correlated across the three incorrect conditions. Unlike Tanner et al. or Meulman et al., we selected C3-Cz-C4 for the central N400 effect, and P3-Pz-P4 for the posterior P600, as these recorded maximal effects for each component, respectively. In addition, we minimized the impact of (trivial) component overlap by using non-adjacent time-windows (250–500 ms for the N400, and 800–1200 ms for the P600). A significant

correlation between components was only found in the semantic manipulation condition ($r = -.46, p < 0.001$).

Considering the two components separately is therefore a more appropriate approach to investigate our data.

Random Forests methodology

We evaluated the strength of all our language-related variables, including online performance, in predicting ERP effects elicited by each incorrect condition. We grew six Random Forests models for each of the three difference waves (the 3 incorrect conditions minus the correct condition), in the above-mentioned time-windows for the N400 (at C3-Cz-C4) and P600 (at P3-Pz-P4). Our method was adapted from Tomaschek et al. (2018) and used unconditional variable importance with the *ranger* package (Wright & Ziegler, 2017). This method is deemed superior to both (i) the conditional variable importance implemented in the *party* package (Hothorn & Zeileis, 2015) which is heavier on resources, and (ii) the *randomForest* package (Liaw & Wiener, 2002), which inflates the importance of continuous variables and correlated data (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008; Strobl, Malley, & Tutz, 2009). The method implemented with *ranger* is not expected to be biased against Group (our categorical variable), or promote highly correlated variables (e.g. performance over digit span score)⁴. The number of variables randomly chosen at each node (*mtry*), as well as trees that are grown (*num.trees*) can both influence the outcome of Random Forests models. We thus determined the optimal values for these parameters using the *train()* function of the *caret* package (Kuhn, 2008). Prediction accuracy was evaluated under 10-fold cross-validation. A model with optimal values was then calculated. In order to ensure replicability, a random seed was set (at 19). To estimate variable importance, the algorithm randomly selected subsets of the data and modeled the effect of each predicting variable in every subset. Accuracy of each prediction was compared to the remaining observations. Strength of a predicting variable was calculated by randomly permuting its levels and thus erasing its importance: a predictor is deemed important if the model becomes worse after erasure (Breiman, 2001).

⁴ We thank the reviewer Darren Tanner for raising this concern with our initial analyses using the *randomForest* package.

Following suggestions made by the reviewers, we considered the possibility that even though *Random Forests* deal with multicollinearity better than linear regressions, the effects of predictors related to exposure and proficiency may be conflated with group effects. In order to further tease these effects apart, we (1) ran *Random Forests* for each group and condition and (2) used conditional inference trees to illustrate how the most important variables interact. As no tree function is implemented in the *ranger* package itself, we used the *ctree()* function from the *party* package (Hothorn, Hornik, & Zeileis, 2006; Hothorn & Zeileis, 2015). Trees predict the value of continuous variables (ERP amplitudes) from a set of continuous or categorical predictor variables, using recursive binary partitioning. Trees provide estimated “split points” at which the nodes separate between two groups with different outcomes. The splitting criteria are calculated using the permutation-test framework (Hothorn et al., 2006). For each possible split, the test-statistic value is calculated under a certain label rearrangement: if they are interchangeable, the splitting value is not relevant (and would not be reported by the software). Variables included were selected using a backward-elimination procedure that compared out-of-bag R-squared values. These values were above .7 (see supplementary materials), which suggests that the variance explained by Random Forests models is much higher than the variance explained by the best fitting mixed-effect model we ran (maximum R-squared: .07). Backward-elimination procedures have been adopted in gene selection research (Díaz-Uriarte & De Andres, 2006) and psycholinguistic studies (de Aguiar, Bastiaanse, & Miceli, 2016). Random Forests using *ranger* or *cforest* are superior to individual trees using *ctree* to account for data variance; nevertheless, trees provide useful insight on the complex interactions that characterize the data. All the following models and their outputs are available in the supplementary materials.

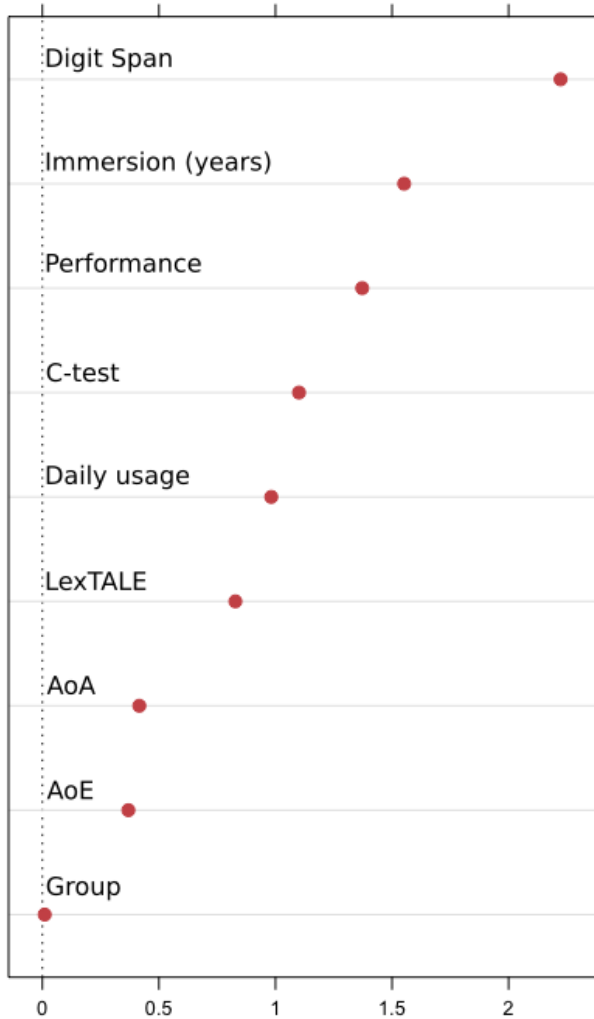
Predictor variable importance

To make comparisons across conditions easier to interpret, we present variable importance for each ERP response and condition in Figure 2. Group, AoA, and AoE were the least important in almost all conditions, except for the P600 response to combined anomalies (in this case they were mildly important). Immersion (in years) was a very important predictor for ERPs to Semantic anomalies, and mildly important in the Combined condition. Importantly, daily usage (in percentage) strongly predicted the P600 effect in all conditions.

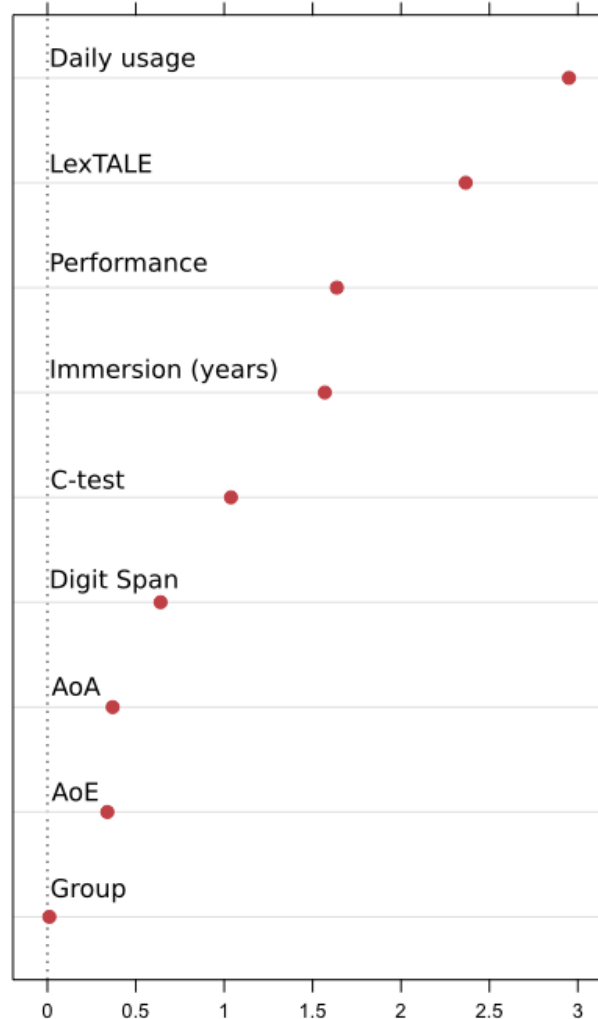
Proficiency measures (online performance, LexTALE, and C-test scores) were also very important predictors for all conditions. Finally, digit span score was an important predictor for the N400 effect in the Semantics condition.

Lexical-semantic anomalies

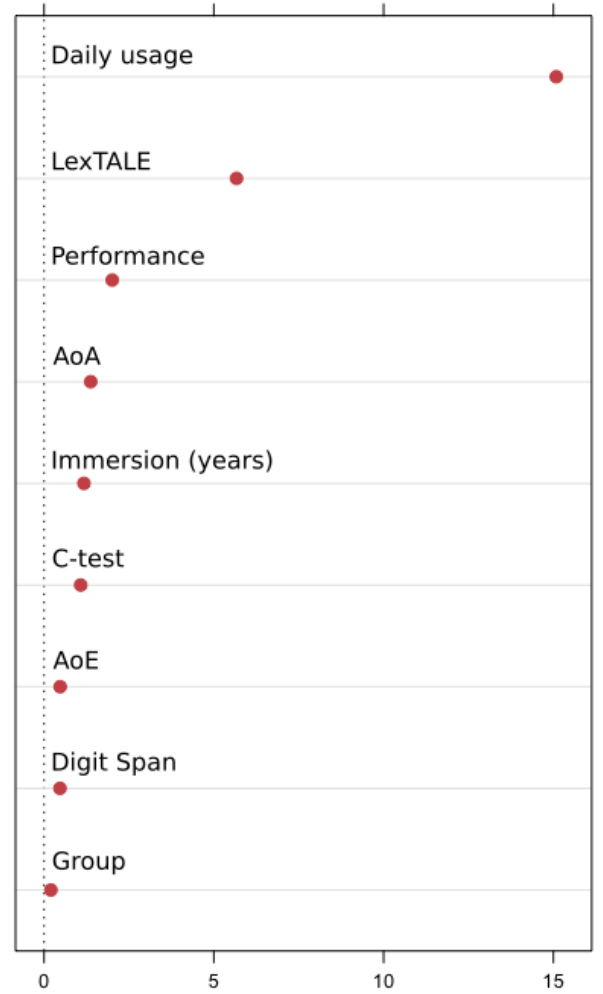
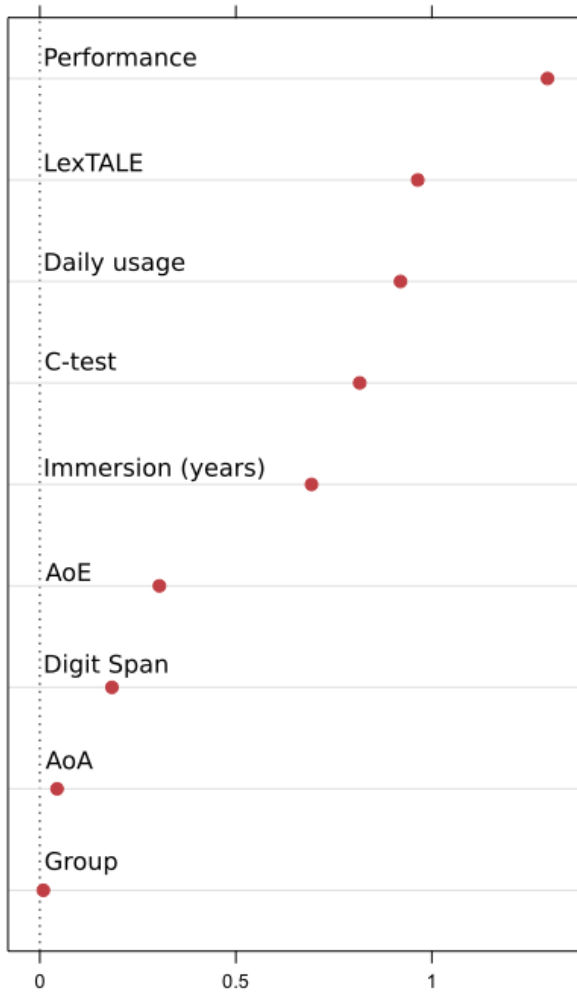
N400



P600



Syntactic category anomalies



Combined anomalies

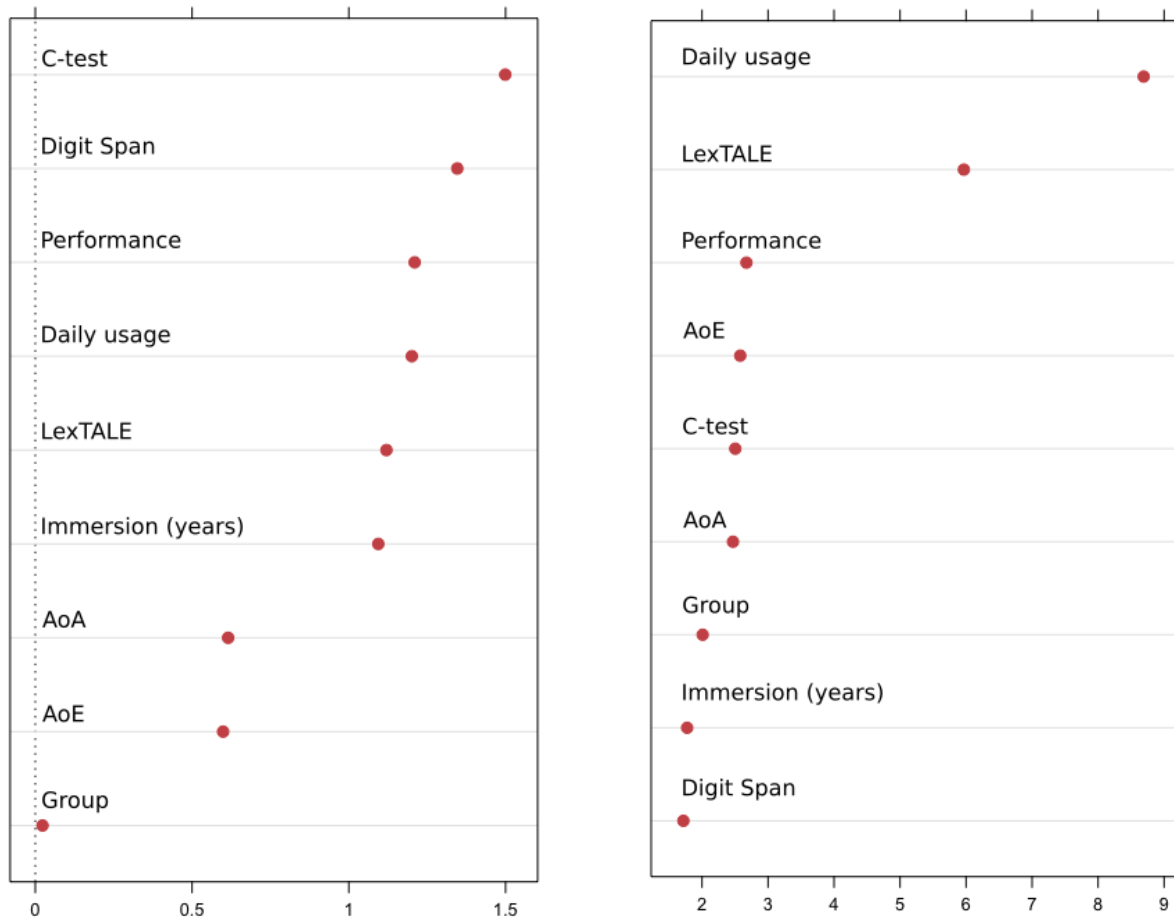


Figure 2. Predictor importance for N400 and P600 effects in all conditions.

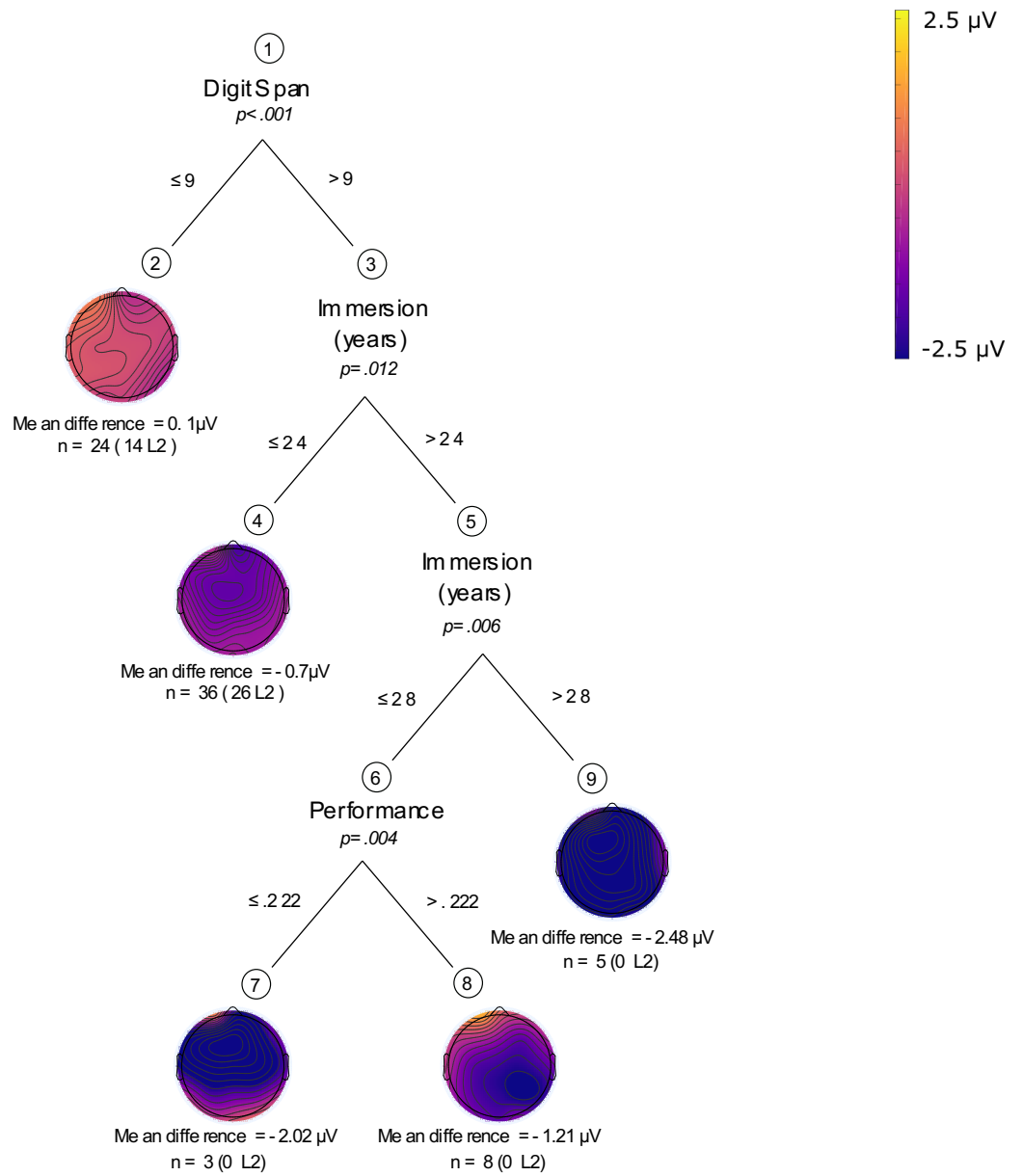
Variable importance for the N400 and P600 effects across all conditions are illustrated for each group in Supplementary materials. The purpose of these analyses was to investigate whether variable importance differed between groups, and in particular whether AoA would be important for L2 speakers. In fact, even when focusing the analyses on the L2 group alone, AoA does not emerge as an important predictor. Not only in the L2 group, but also in L1 participants, the N400 effect was modulated by daily usage and proficiency measures (i.e., LexTALE score and online performance). In the L2 group, C-test and digit span scores were also among the most important measures. For the P600 effect, however, there is a contrast between the L1 group, where LexTALE matters most, and the L2 group, where daily usage is the most important predictor.

Illustration of effects using decision trees

To illustrate effects expressed by the individual trees, we drew scalp maps based on split points determined by the inference trees in Figure 3. At split points, we indicate how many speakers (and how many L2 learners) fall into each group. In the Semantics condition, the N400 effect is first determined by digit span (scores below 9 do not show any effect). Then, it is split by immersion: the lowest immersion cluster has a mix of L1 and L2 speakers, but the higher clusters (> 24, nodes ⑤ to ⑨) only have L1 speakers. Finally, performance had an effect in a higher-immersion subgroup (node ⑥), but higher performance was actually associated with smaller N400s (node ⑦ vs. ⑧). No effect was apparent for the P600. In the Syntax condition the N400 effect was best explained by performance: the split point revealed that a few L2 participants who were at chance level did not display an N400 effect (node ②), but a frontal positivity instead. The P600 effect was first split by daily usage: only L1 participants fall into the very high group (node ⑦). Among them, surprisingly, participants who had higher LexTALE scores show a *smaller* P600 (node ⑧ vs. ⑨). Participants who were exposed to French between 22% and 90% of the time also display a P600 (node ⑥). For participants with the lowest daily usage (L2 speakers exclusively) online performance determined the presence or absence of a (small) P600 — lower performance is once again associated with a frontal positivity (node ④ vs. ⑤). In the Combined condition no split was observed for the N400, while daily usage significantly accounted for the P600, with participants above 31% daily usage displaying the largest effect (node ⑤). Surprisingly, digit span, a less important variable, showed a split among participants who had lower exposure, with larger digit span reflected by a negativity (node ③ vs. ④).

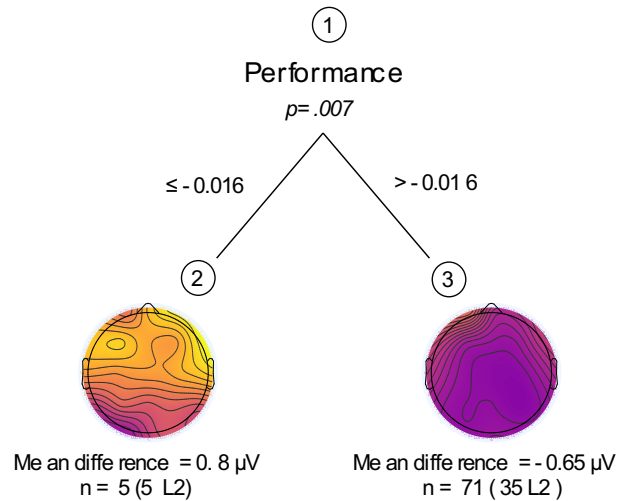
Lexical-semantic anomalies

-
N400



Syntactic category anomalies

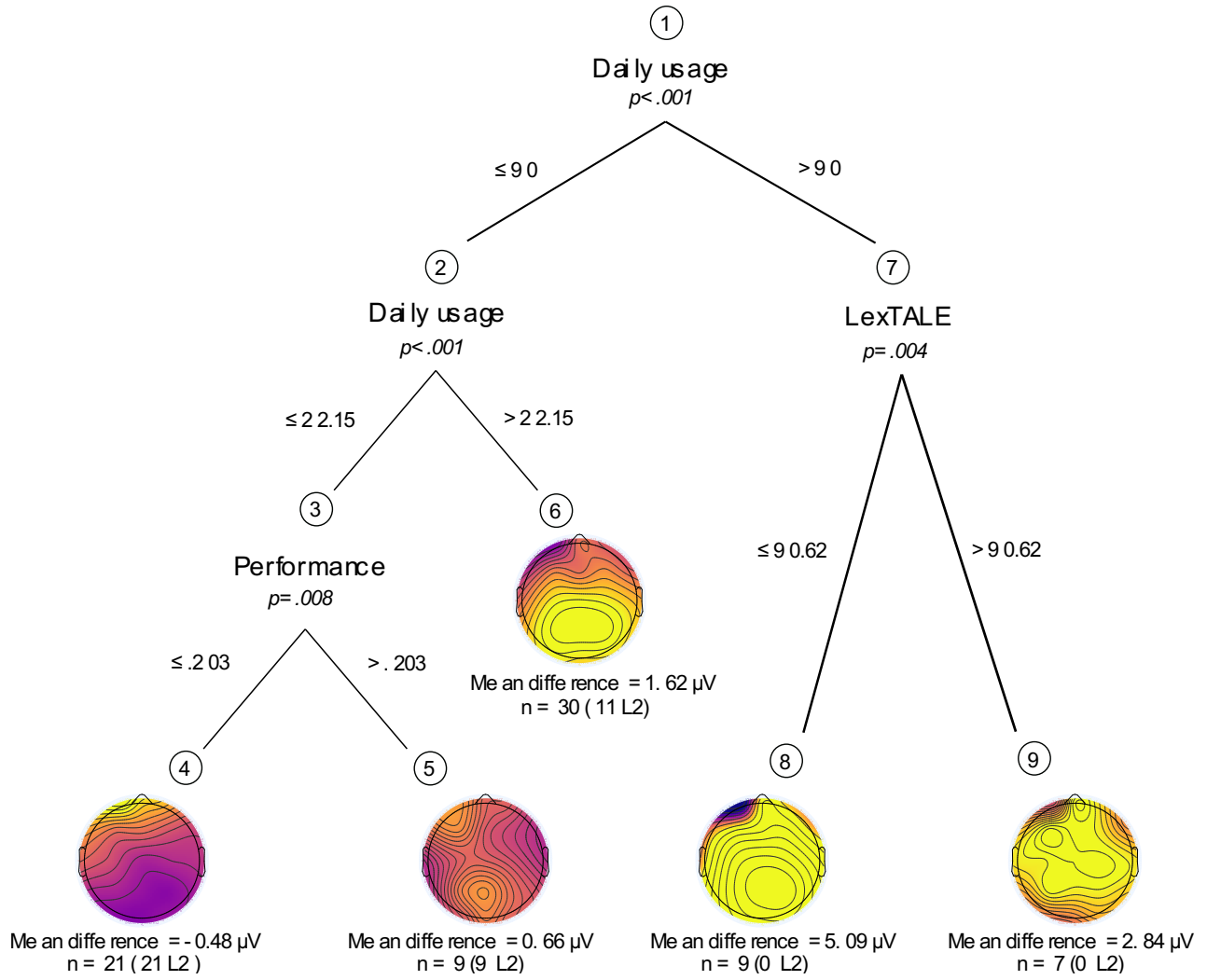
-
N400



Syntactic category anomalies

-

P600



Combined anomalies

-

P600

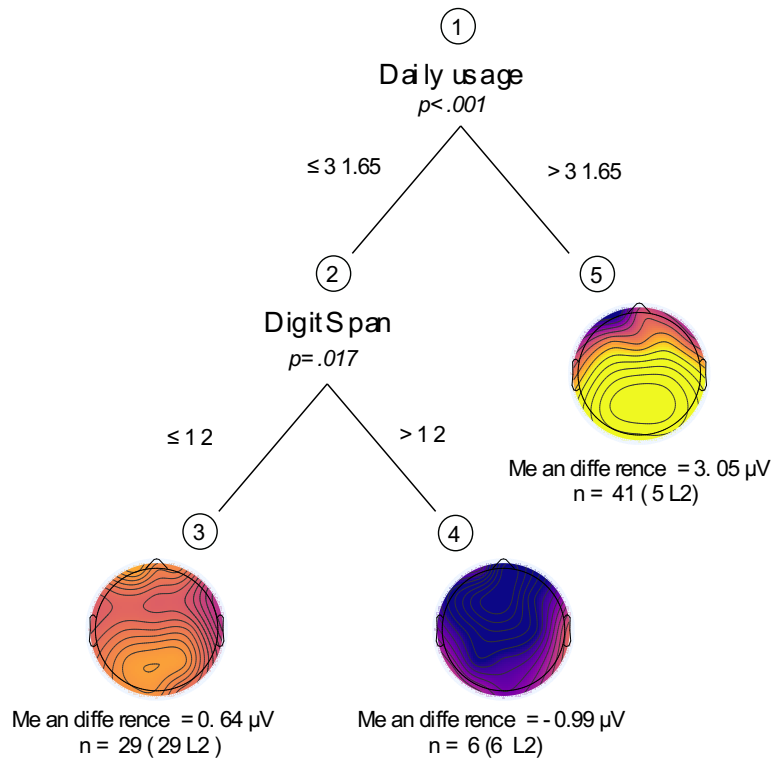


Figure 3. Maps illustrating inference tree outputs for Semantic, Syntactic, and Combined anomalies in N400 and P600 time-windows. Split points were determined using inference-tree calculations (the circled numbers are the nodes). Under each map are mean amplitude values estimated by the models, as well the number of participants who fell into each cluster (in parenthesis the number of L2-speakers). Yellow depicts a positivity and blue a negativity.

4. Discussion

The present study used ERPs to investigate the time course of cognitive processes in response to syntactic category and semantic priming manipulations as well as their interactions, in both first and second language speakers of French. Since most previous studies investigating syntactic-category processing used unbalanced designs (Hahne, 2001; Hahne & Friederici, 2001; Isel, 2007; Mueller et al. 2005; Pakulak & Neville, 2011; Weber-Fox & Neville, 1996) that could potentially lead to artefacts (Steinhauer & Drury, 2012), our first goal was to reevaluate this issue at a group level. We thus employed a design in French that manipulated both contexts and targets while systematically controlling for them. We observed that both L1- and L2-speakers elicited similar N400s to syntactic and semantic manipulations. While only L1-speakers elicited a reliable P600 to ungrammatical sentences as a group, semantic anomalies elicited small P600s in both groups. However, this pattern was not observed when focusing on individual differences in our data: the observed pattern will be expanded below. Second, we investigated what factors related to language learning could account for the range of observed ERP profiles. We estimated the relative importance of our predictors using *Random Forest*, as well as decision trees to assess the significance of the most important factors identified. We found that daily usage, immersion, and performance on the online acceptability-rating task were the most reliable predictors explaining our data. Specifically, these factors predicted the amplitude of both the N400 and the P600 effects, in L1- and L2-speakers *alike*. Immersion (but also, importantly, digit span) predicted the semantic N400, while daily usage and performance were better predictors for the syntactic N400 and the P600 effect.

Group effects

Our finding that SCVs elicit an N400 instead of an ELAN in both groups differs from a large body of literature in L2-processing (Hahne, 2001; Hahne & Friederici, 2001; Isel, 2007; Mueller et al. 2005; Pakulak &

Neville, 2011) as well as L1 studies (see Friederici, 2011, for a review, and Fromont et al., resubmitted for additional discussion of sentence-processing results). We attribute the absence of an early effect to our balanced⁵ design, which avoids baseline issues. Together with Steinhauer and Drury (2012), this is evidence that the ELAN cannot be viewed as a reliable index of native-like linguistic achievement in L2-speakers of French. The absence of a LAN also differs from previous studies (Nichols & Joanisse, 2019; Steinhauer et al., 2009), although some studies focusing on L1-processing have observed an N400, and no LAN, in response to SCVs (Nickels et al., 2014; Zhang et al., 2013). In morphosyntactic studies focusing on agreement, there is currently a debate on whether the LAN is, in fact, an N400 that adopts a frontal, left-lateralized topographic distribution when it is superimposed by a right-posterior P600 (Guajardo & Wicha, 2014; Tanner et al., 2014; but see discussion in Courteau et al., 2019). Since some LANs have been observed in the absence of a P600 (Hasting & Kotz, 2008), and considering recent studies suggesting that at least some LAN effects are not a product of averaging (Caffarra, Mendoza, & Davidson, 2019), one possibility is that LAN and N400 are not quite distinct but represent a continuum reflecting a mismatch between predicted features and the target, with a topography becoming more N400-like when these features belong to the lexical-semantic domain (Bornkessel-Schlesewsky & Schlesewsky, 2018; Molinaro, Barber, Caffarra, & Carreiras, 2015). In the context of our experiment, the N400 could indicate that syntactic-category identification in French relies more on word stem information than morphological regularities.

Even without considering word class markers, processing the word stem of a verb (e.g., *tackle*) replacing a noun (e.g., *toad*) – and *vice versa* – can be expected to result in a lexical-semantic mismatch and a corresponding N400, unless semantic processing is assumed to be “blocked” by a SCV on the same word (Friederici, 2002; Friederici et al., 1999; 2004). Crucially, our finding that enhanced N400s for both syntactic and semantic anomalies were additive (in both L1 and L2) clearly contradicts the “semantic blocking”

⁵ One reviewer suggested that analyzing the sub-conditions separately (i.e., breaking up the balanced design) could reveal qualitative (Nieuwland, Martin, & Carreiras, 2013) or quantitative (Mehravari, Tanner, Wampler, Valentine, & Osterhout, 2015) differences that may help understand the data. As splitting between sub-conditions revealed differences between the *correct* sub-conditions (i.e., control vs. transitive verbs), it is impossible to interpret whether the ERP effects of syntactic anomalies in separate sub-conditions are driven by context or lexical effects. Extended discussion can be found in Fromont (resubmitted, 2019) and Steinhauer, Drury, Portner, Walenski, and Ullman (2010).

hypothesis stipulating that semantic processing becomes moot once a syntactic error is identified, as already predicted by Steinhauer and Drury (2012) and partly confirmed by Nickels et al. (2014) and Zhang et al. (2013). Studies that seemed to support semantic blocking (e.g., Friederici et al., 1999; Hahne & Friederici, 2002) invariably used a paradigm in which the lack of an N400 in the “combined violation” condition could also be explained without any reference to the syntactic violation (see Steinhauer & Drury, 2012, for details). In contrast, our semantic manipulation using contextual priming is immune to this concern. The presence of a clear semantic N400 whose amplitude was *enhanced* rather than reduced by a simultaneous category violation, provides very strong evidence that lexical-semantic processing is not blocked by SCVs. From this perspective, the absence of N400s in certain “double-violation” conditions from Friederici’s group was likely due to a consistent methodological flaw in creating that condition, and not to semantic blocking. Moreover, since our paradigm was a reading study in which all word information became available at once and did not unfold over time, as in auditory research, we can also rule out Friederici and Kotz’s (2003) creative hypothesis that the word stem’s status as a potentially free or bound morpheme may determine whether semantic blocking takes place or not. Instead, we conclude that semantic blocking is a myth, based on misinterpreted data. Native speakers process semantic and syntactic information in parallel, and late L2-learners converge toward the same pattern at high exposure and proficiency levels.

The finding that lexical-semantic anomalies elicit an N400 is in line with a very prolific body of literature (reviewed by Kutas & Federmeier, 2011). The observed priming effects on the N400 are likely to reflect distinct processes (namely, automatic spreading activation and post-lexical integration, see Steinhauer et al., 2017, for a recent account), but our design does not allow us to tease these apart, and we consider priming effects as a whole. We found no interactions for syntactic category errors and priming, suggesting additive effects of the lexical-semantic and syntactic manipulations. Additive ERP effects in turn suggest that the underlying cognitive mechanisms used to process these different types of information are neurally and functionally distinct. This finding coheres with literature arguing that different neurocognitive and neurolinguistic mechanisms reflected by modulations of the N400 effect can be independent. For example,

Chow et al. (2014), found additive effects on the N400 when manipulating word repetition and predictability, given a sentence context (*Brian looked all over the house for his missing keys/watch before leaving for work*, where target words were expected or not (*watch*), and old versus new). We could therefore interpret the N400 in response to SCVs as a mismatch between predicted features (e.g. *-er, -ir, -re* endings on French verbs) and the target (as proposed by Molinaro et al., 2014; or Tanner et al., 2014), while the N400 in response to lexical-semantic anomalies could reflect the absence of priming. Note that the additive effects should, however, be interpreted with caution: as the amplitude of the N400 in every condition is relatively small, it could mean that not all available resources are recruited to process either type of anomaly, and that they therefore do not need to compete when the two anomalies are combined (see Fromont et al, resubmitted, for further discussion).

Consistent with a majority of the literature on L1 and L2 processing, SCVs elicited a larger P600 in L1- than in L2-speakers (e.g. Hahne & Friederici, 2001; Rossi et al., 2006; Steinhauer et al., 2009; Weber-Fox and Neville, 1996). This P600 effect is compatible with virtually all previous P600 accounts, ranging from task-related well-formedness judgments (Sassenhagen et al., 2014), conflict resolution in a monitoring context (Vissers et al., 2008), and sentence diagnosis and reanalysis (e.g. Friederici, Mecklinger, Spencer, Steinhauer, & Donchin, 2001). A P600 effect was also found for pure lexical-semantic anomalies, even though the ERP plots suggest that the effect is small (therefore the degree of additivity of the P600s should be viewed with caution). This finding is in line with previous studies involving judgments (Diaz & Swaab, 2007; Royle et al, 2013; Steinhauer, Drury, Portner, Walenski, & Ullman, 2010). Using the same materials but focusing only on the L1 participants, Fromont et al (resubmitted) observed a weak but significant interaction between Syntax and Semantics. Given the large P600 observed in response to ungrammatical sentences, they concluded that the cognitive resources reflected by this component were used up by syntactic processing, so that no additional increase in amplitude was observed in the combined condition. In the present study we observed additive effects, which apparently contradict this interpretation. As it will become clear from individual-response analyses, the weak effect observed in L2 speakers can be explained by inter-individual variability in their ERP

responses. Therefore, exploratory analyses of individual differences that we will discuss next turned out as a very helpful way of elucidating underlying patterns that are at play.

Individual differences

Inspection of individual data revealed some biases toward N400 or P600 profiles (Kim, Oines, & Miyake, 2018; Tanner et al., 2014; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013), however in our data only lexical-semantic anomalies showed a significant correlation between N400 and P600 amplitudes. We propose two explanations for the discrepancy between our results and Tanner and colleagues' (see also van Hell & Abdollahi, 2017). First, they selected the *same* electrodes (at central and parietal sites) for both components and correlation measures. While central and parietal electrodes are representative for both the N400 and P600 effects, it is expected that the two components highly impact each other, especially when considering their amplitude at the same site and adjacent time windows. Brouwer and Crocker (2017) point out that peaks observed in waveform-based component structures are only epiphenomena of latent components. That N400 and P600 effects correlate when measured at the same sites does not mean that the latent N400 and P600 truly correlate among participants. Our decision to quantify N400 and P600 effects at distinct electrodes and in non-adjacent time intervals may not have completely circumvented these issues, but should certainly have reduced the risk of finding a correlation that was simply due to overlap between the two components. Second, Tanner et al (2013) focused on one dimension of language processing (morpho-syntactic agreement). We show that there is indeed a correlation between the N400 and P600 when investigating lexical-semantic anomalies alone, but not in conditions involving SCVs.

In this study, we considered aspects that have been argued to be either sensitive to AoA or not (i.e. syntax vs semantics), and investigated the effects of AoA, proficiency, and exposure on these. Regression-based approaches have been argued to be appropriate to model continuous variables (van Hell & Tanner, 2012; Meulman et al., 2015), but we have seen that they are difficult to implement with a relatively large number of correlated variables. Using *Random Forests* allowed us to rank the most relevant predictors in order of importance, before using decision trees to establish their effects on ERPs.

Despite using an analysis package (i.e., *ranger*) that did not favor continuous over categorical predictor variables, GROUP (L1 vs L2) always failed to emerge as a relevant factor in all conditions. Further, AoA and AoE were usually irrelevant with one exception: the P600 effect in the Combined condition. Even when considering the L2 group separately, AoA is systematically the least important variable, and AoE is always less important than daily usage and proficiency variables. In contrast, daily usage, immersion, and online performance were generally the most reliable factors explaining our data. Note that this finding does not deny AoA's role in language learning, which has been well documented (e.g. Hernandez & Li, 2007; Steinhauer, 2014): there are more L2-learners in the lower clusters and more L1-speakers in the higher ones. Rather, it suggests that exposure and proficiency may be more appropriate measures accounting for variability in the data, and that a priori splitting between L1 and L2 groups may not be the most suitable way to analyze data. Thus, our results are globally inconsistent with the CPH.

Percentage of daily usage was a determining factor accounting for the P600 response in the Syntax and Combined conditions: the more speakers were exposed to French, the larger their P600 responses to these errors. Daily usage has not been the focus of L2 studies of sentence processing in adults; research on bilingual language acquisition and developmental language disorder, however, has shown that regular exposure (between 20-30 % daily) is essential to L2 mastery (Ojima, Nakata, & Kakigi, 2005) including morphosyntactic development (Elin Thordardottir, 2015; Marquis & Royle, 2019). Interestingly, we observe that participants with this amount of exposure or more elicit a P600 effect. It seems that for L1 and L2 speakers alike, regular exposure over a given threshold enhances the ability to recruit mechanisms that relate to rule-based, procedural processing (Ullman, 2004) and conflict monitoring (Vissers et al., 2008). This effect is not categorical: participants who receive the highest exposure (over 90%) show the largest P600 effect in response to 'pure' SCVs. Daily usage was also an important variable accounting for the P600 effect in the Semantics condition, but inspection of a sample tree revealed no significant split.

As predicted by the convergence hypothesis, another central variable accounting for both the N400 and P600 effects was proficiency. Information provided by the cut-off values suggests different interpretations. The

link between P600 effects and higher online performance in the Syntax condition suggests that this component is associated with more stringent categorization of unacceptable sentences. On the other hand, the N400 seems to be predicated on above-chance levels in the online task: it could tentatively be interpreted either as a byproduct of participants paying attention to the task or as a lexical-semantic bottleneck (Hopp, 2015). However, a bottleneck interpretation would be strongly supported by LexTALE effects, which we do not observe. In the Combined condition, LexTALE split the high-exposure cluster: participants with a better LexTALE score showed a smaller P600 effect. We are presently unsure what explanation could account for this surprising effect. Online performance effects could reflect attention effects in two (not mutually exclusive) ways: higher performance at the task could be associated with less noisy data in participants who are paying attention, or participants who score better on average display larger ERP effects in general.

Duration of French immersion accounts for variability in the Semantics condition. It may reflect how the size of the lexicon increases with experience as seen in ERP studies with children (Ojima et al., 2005) and behavioral studies with aging adults (Cohen-Shikora & Balota, 2016; Robert & Rico Duarte, 2016; Royle, Steinhauer, Dessureault, Herbay, & Brambati, 2019). Digit-span score was the most important predictor in that condition, and tree inspection showed that participants with a lower score on the working memory digit span task (below 9) did not elicit any N400 effect in response to lexical-semantic anomalies, suggesting that participants may need to recall the sentence context to detect anomalies in the experimental sentence. More surprisingly, participants with a high digit-span score (above 12) elicited a negativity instead of a P600 in response to the Combined condition. The relationship between working memory load and sustained negativities using similar experimental materials in the auditory modality is currently under evaluation (see Fromont, Royle, Herbay, Misirliyan & Steinhauer, 2019) and may shed light on this effect.

Inspection of individual differences using *Random Forests* and decision trees helps us understand the seemingly small P600s (or absent P600s in the case of SCVs) in the L2 group. Data partitioning shows that participants with less exposure who are L2ers do not display any P600 effect, while L2ers who are more exposed to French do display one. Further, some participants even display *opposite* effects in the late time-

window: subgroups show a relative parietal negativity (at sites where the P600 is measured) in the SCV condition and a large negativity in the Combined condition. These negativities likely attenuated or canceled out the group effects, suggesting that *Random Forests* is an appropriate method to identify the variables that truly contribute to the variability in our data (i.e. daily usage, performance, and digit span score, rather than group). Further, as even the best fitting mixed-effects models fail to capture high portions of variance in the ERP data (see also Fromont et al, resubmitted), using *Random Forests* may prove to be a fitting way to complement these analyses by allowing us to take more factors into account and capture more variability.

5. Conclusion

Our results demonstrate that both native speakers and proficient late L2-learners process syntactic information in parallel with lexical-semantic information. At a group level, native speakers elicited a biphasic N400-P600 response to SCVs – and no ELAN or LAN – while L2-learners only displayed a significant N400 effect. However, when investigating interindividual variability and the relative importance of predictor variables related to AoA, L2 exposure, and proficiency, we found that daily usage, immersion, and proficiency – not AoA – were the most important predictors for the observed ERP components. Both L1- and L2-speakers display larger N400 and P600 responses to syntactic errors as their daily usage, immersion, and proficiency increase. This evidence lends further support to the convergence hypothesis, and suggests that higher language exposure and proficiency are associated with both memory-based heuristics and rule-based processes.

Acknowledgements

The authors acknowledge funding from the FQRSC – *Team grant* (2016-SE-188196) to Lydia White et al (2016–2019) – and the SSHRC – CRSH - *Insight* grant (435-2013-0583) to Steinhauer, Royle et al (2013–2018). The first author acknowledges funding from the Excellence Scholarship Program for International Students (FRQSC), the Excellence Scholarship from the Biomedical Sciences Program (University of Montréal) the Graduate Student Stipend (Centre for Research on Brain, Language and Music in Montréal). We thank an anonymous reviewer and Darren Tanner for insightful comments on previous versions of this manuscript.

References

- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, 17, 137–153. doi:10.1162/0898929052880101
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56, 9–49. doi:10.1111/j.1467-9922.2006.00353.x
- Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, 9, 81. doi:10.3389/fpsyg.2018.00081
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73. <https://doi.org/10.1016/j.brainresrev.2008.05.003>
- Bornkessel-Schlesewsky, I., & Schlewsky, M. (2018, November 20). Towards a neurobiologically plausible model of language-related, negative event-related potentials. OSF Preprints. doi:10.31219/osf.io/aynsp
- Bowden, H., Steinhauer K., Sanz C., Ullman M.T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, 51, 2492–2511. doi:10.1016/j.neuropsychologia.2013.09.004
- Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brouwer, H., & Crocker, M. W. (2017). On the proper treatment of the N400 and P600 in language comprehension. *Frontiers in Psychology*, 8, 1327. doi:10.3389/fpsyg.2017.01327
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143. <https://doi.org/10.1016/j.brainres.2012.01.055>
- Brysbaert, M. (2013). LexTALE_Fr a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, 531, 23–37. doi:10.5334/pb-53-1-23
- Caffarra, S., Mendoza, M., & Davidson, D. (November, 2017). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language*, 191, 9–16. doi:10.1016/j.bandl.2019.01.003.
- Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews*, 51, 31–47. doi:10.1016/j.neubiorev.2015.01.010
- Chow, W. Y., Lago, S., Barrios, S., Parker, D., Morini, G., & Lau, E. (2014). Additive effects of repetition and predictability during comprehension: Evidence from event-related potentials. *PLoS ONE*, 9. doi:10.1371/journal.pone.0099199
- Clahsen, H. & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3–42. doi:[10.1017.S0142716406060024](https://doi.org/10.1017/S0142716406060024)
- Cohen-Shikora, E. R., & Balota, D. A. (2016). Visual word recognition across the adult lifespan. *Psychology of Aging*, 31(5), 488–502. <https://doi.org/doi:10.1037/pag0000100>

- Courteau, É., Martignetti, L., Steinhauer, K., & Royle, P. (2019). Eliciting ERP components of morphosyntactic agreement mismatches in perfectly grammatical sentences. *Frontiers in Psychology*, 10, 1152. doi: 10.3389/fpsyg.2019.01152
- de Aguiar, V., Bastiaanse, R., & Miceli, G. (2016). Improving production of treated and untreated verbs in aphasia: A meta-analysis. *Frontiers in Human Neuroscience*, 10, 468. doi:10.3389/fnhum.2016.00468
- Diaz, M. T., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, 1146, 85–100. doi:10.1016/j.brainres.2006.07.034
- Elin Thordardottir. (2015). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 17(2), 97–114. <https://doi.org/10.3109/17549507.2014.923509>
- Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23(4), 527–552.
- Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression*, Second Edition. Thousand Oaks, CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- French-Mestre, C. (2002). An on-line look at sentence processing in the second language. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing* (pp. 217–236). Elsevier.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *TRENDS in Cognitive Sciences*, 6(2), 78–84. doi:10.1016/S1364-6613(00)01839-8
- Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, 91(4), 1357–1392. doi:10.1152/physrev.00006.2011
- Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *NeuroReport*, 15, 165–169. doi:10.1097/00001756-200401190-00032
- Friederici, A., & Kotz, S. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage*, 20, 8–17. doi:10.1016/j.neuroimage.2003.09.003
- Friederici, A.D., Mecklinger, A., Spencer, K., M., Steinhauer, K., & Donchin, E. (2001). Syntactic parsing preferences and their on-line revisions: A spatio-temporal analysis of event-related brain potentials. *Cognitive Brain Research*, 11, 305–323. doi:10.1016/S0926-6410(00)00065-3
- Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory and Cognition*, 27(3), 438–453. doi:10.3758/BF03211539
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis, *PNAS* 99(1), 529–534. doi:10.1073/pnas.012611199
- Fromont, L. A. (2019). *Verbing and nouning in French: toward an ecologically valid approach to sentence processing*. Université de Montréal.
- Fromont, L. A, Royle, P, Herbay, A., Misirliyan, C, Steinhauer, K. (2019, August) Reevaluating the dynamics of auditory sentence processing: an ERP study in French. Poster presented at the Annual Meeting of the Society for the Neurobiology of Language, Helsinki, Finland.

- Fromont, L. A., Steinhauer, K., & Royle, P. (resubmitted). *Verbing nouns and nouning verbs: Using a balanced design provides ERP evidence against “syntax-first” approaches to sentence processing.*
- Green, D. W. (2003). Neural basis of lexicon and grammar in L2 acquisition. In R. van Hout, A. Hulk, F. Kuiken, & T. R.J. (Eds.), *The Lexicon–Syntax Interface in Second Language Acquisition* (pp. 197–218). Amsterdam, the Netherlands: Benjamins.
- Guajardo, L. F., & Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component : event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262–272. <https://doi.org/10.1016/j.neuroimage.2013.09.077>
- Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*, 556–568. doi:10.1162/089892900562336
- Hahne, A. (2001). What’s different in second-language processing? Evidence from event-related potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266. doi:10.1023/A:1010490917575
- 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*(2), 123–141. doi:10.1017/S1366728901000232
- Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*, 339–356. doi:10.1016/S0926-6410(01)00127-6
- Hasting, A. S., & Kotz, S. A. (2008). Speeding up syntax: On the relative timing and automaticity of local phrase structure and morphosyntactic processing as reflected in event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(7), 1207–1219. doi:10.1162/jocn.2008.20083
- Hernandez A.E., Li P. (2007). Age of acquisition: its neural and computational mechanisms. *Psychological Bulletin*, *133*(4), 638–650. doi:10.1037/0033-2909.133.4.638
- Hopp, H. (2015). The timing of lexical and syntactic processes in L2 sentence comprehension, *Applied Psycholinguistics*, *37*(5), 1253–1280.
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. doi:10.1198/106186006X133933
- Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909.
- Isel, F. (2007). Syntactic and referential processes in second-language learners: Event-related brain potential evidence. *NeuroReport*, *18*, 1885–1889. doi:10.1097/WNR.0b013e3282f1d518
- Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology*, *74*, 337–346. doi:10.1016/j.biopsycho.2006.09.003
- Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375. doi:10.1080/00029238.1961.11080571
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

doi:10.1016/0010-0285(89)90003-0

- 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(3), 406–420. <https://doi.org/10.1037/xlm0000457>
- Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205–225. doi:10.1016/j.jml.2004.10.002
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146(1), 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). *Annual Review of Psychology*, 62(1), 621–647. doi:10.1146/annurev.psych.093008.131123
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933. doi:10.1038/nrn2532
- Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, 98, 74–88. doi:10.1016/j.bandl.2006.02.003
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Lenth, R. (2018). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.1.3. Retrieved from <https://cran.r-project.org/package=emmeans>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22. Retrieved from <https://CRAN.R-project.org/doc/Rnews/>.
- Luo, Y., Zhang, Y., Feng, X., & Zhou, X. (2010). Electroencephalogram oscillations differentiate semantic and prosodic processes during sentence reading. *Neuroscience*, 169(2), 654–664. doi:10.1016/j.neuroscience.2010.05.032
- Marquis, A., & Royle, P. (2019). Verb acquisition in monolingual and multilingual children and adults. In P. Guijarro-Fuentes & C. Suárez-Gómez (Eds.), *Proceedings of GALA 2017: Language Acquisition and Development* (pp. 307–324). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The Random Forests statistical technique: an examination of its value for the study of reading. *Scientific Studies of Reading*, 20, 20–33. doi:10.1080/10888438.2015.1107073
- Meerendonk, N. Van De, Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2008). Monitoring in language perception : mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience*, 22(1), 67–82.
- Mehravari, A. S., Tanner, D., Wampler, E. K., Valentine, G. D., & Osterhout, L. (2015). Effects of

- grammaticality and morphological complexity on the P600 event-related potential component. *PLoS ONE*, 10(10), 1–16. <https://doi.org/10.1371/journal.pone.0140850>
- Meulman N., Wieling M., Sprenger S.A., Stowe L.A., Schmid M.S. (2015). Age Effects in L2 Grammar Processing as Revealed by ERPs and How (Not) to Study Them. *PLoS ONE*, 10(12), e0143328. doi:10.1371/journal.pone.0143328
- Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2014). On the left anterior negativity (LAN): The case of morphosyntactic agreement: A Reply to Tanner et al. *Cortex*, 66, 156–159. doi:10.1016/j.cortex.2014.06.009
- 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(4), 933–947. doi:10.1162/jocn_a_00119
- Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research*, 21(2), 152–174. doi:10.1191/0267658305sr256oa
- Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal of Cognitive Neuroscience*, 17(8), 1229–1244. doi:10.1162/0898929055002463
- Newport, E. L., Bavelier, D., & Neville, H. J. (2001). Perspectives on a critical period for language acquisition. In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development* (pp. 481–502). Cambridge, MA: MIT Press. doi:10.1067/mhn.2001.115372
- Nichols, E. S., & Joanisse, M. F. (2019). Individual differences predict ERP signatures of second language learning of novel grammatical rules. *Bilingualism: Language and Cognition*, 22, 78–92. doi:10.1017/S1366728917000566
- Nickels, S. (2016). *A critical re-evaluation of Friederici's (2002) sentence processing model with particular consideration of task effects, item effects, and inter-individual differences*. McGill University, Montreal.
- Nickels, S., Bokhari, F. S., & Steinhauer, K. (2014, November). 'Semantic blocking' solved at last: ERP evidence challenging syntax-first models. Poster presented at the Annual Meeting of the Society for the Neurobiology of Language. doi:10.13140/2.1.4809.6640
- Nickels, S., & Steinhauer, K. (2018). Prosody-syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models. *Second Language Research*, 34, 9–37. doi:10.1177/0267658316649998
- Nieuwland, M. S., Martin, A. E., & Carreiras, M. (2013). Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension. *Brain and Language*, 126(2), 151–158. <https://doi.org/10.1016/j.bandl.2013.04.005>
- Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: effects of proficiency. *Journal of Cognitive Neuroscience*, 17(8), 1212–1228.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113. doi:10.1016/0028-3932(71)90067-4
- 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. doi:10.1006/brln.1997.1793

- Osterhout, L. (2012). Computational approaches to figurative language. *Metaphor and Symbolic Activity*, 11, 85–100. doi:10.1207/s15327868ms1101_5
- Osterhout, L., McLaughlin, J., Kim, A., Greenwald, 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: Eye-Tracking, ERPs and Beyond* (pp. 271–308). Psychology Press: New York.
- 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. doi:10.1111/j.1467-9922.2006.00361.x
- Pakulak, E., & Neville, H. J. (2011). Maturational constraints on the recruitment of early processes for syntactic processing. *Journal of Cognitive Neuroscience*, 23, 2752–2765. doi:10.1162/jocn.2010.21586.Maturational
- Robert, C., & Rico Duarte, L. (2016). Semantic richness and aging: the effect of number of features in the lexical decision task. *Journal of Psycholinguistic Research*, 45(2), 359–365. <https://doi.org/10.1007/s10936-015-9352-8>
- Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: An ERP study. *Neuroscience Letters*, 384, 228–233. doi:10.1016/j.neulet.2005.04.077
- Royle, P., Drury, J. E., & Steinhauer, K. (2013). ERPs and task effects in the auditory processing of gender agreement and semantics in French. *The Mental Lexicon*, 8, 216–244. doi:10.1075/ml.8.2.05roy
- Royle, P., Steinhauer, K., Dessureault, É., Herbay, A. C., & Brambati, S. M. (2019). Aging and language: maintenance of morphological representations in older adults. *Frontiers in Communication*, 4, 16. <https://doi.org/10.3389/fcomm.2019.00016>
- Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, 15, 228–240. doi:10.1016/S0926-6410(02)00195-7
- Sassenhagen, J., Schlesewsky, M., Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned, *Brain and Language*, 137, 29–39. doi:10.1016/j.bandl.2014.07.010.
- Soylu, F. (2010). Forward/Backward Digit-Span Task. *Archives of Neurobehavioral Experiments and Stimuli*, 218. Retrieved from http://www.neurobs.com/ex_files/expt_view?id=218
- Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, 35, 393–417. doi:10.1093/applin/amu028
- Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, 120, 135–162. doi:10.1016/j.bandl.2011.07.001
- Steinhauer, K., Drury, J. E., Portner, P., Walenski, M., and Ullman, M. T. (2010). Syntax, concepts, and logic in the temporal dynamics of language comprehension: Evidence from event-related potentials. *Neuropsychologia*, 48, 1525–1542. doi:10.1016/j.neuropsychologia.2010.01.013.
- Steinhauer, K., Royle, P., Drury, J. E., & Fromont, L. A. (2017). The priming of priming : Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience Letters*, 651, 192–197. doi:10.1016/j.neulet.2017.05.007

- 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. doi:10.1177/0267658308098995
- Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for Random Forests, *11*, 1–11. <https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and decision trees, Bagging, and Random Forests. *Psychological methods*, 14, 323–348. doi:10.1037/a0016973
- Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24, 135–178. doi:10.1017/S0954394512000129
- Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: a commentary on “Grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex*, 66, 149–155. <https://doi.org/10.1016/j.cortex.2014.04.007>
- Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, 111, 210–237. <https://doi.org/10.1016/j.cortex.2018.10.011>
- Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of Learning and Motivation*, Volume 68 (pp. 229–335). Cambridge, MA: Academic Press.
- Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition: Language and Cognition*, 17, 277–293. doi:10.1017/S1366728913000370
- Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism*, 16(2), 367–382. doi:10.1017/S1366728912000302
- Tanner, D., & Van Hell, J. G. (2014). Neuropsychologia ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, 56, 289–301. doi:10.1016/j.neuropsychologia.2014.02.002
- Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics*, 71, 249–267. <https://doi.org/10.1016/j.wocn.2018.09.004>
- Tremblay, A., & Garrison, M. D. (2010). Cloze tests: a tool for proficiency assessment in research on L2 French. In M. T. Prior, Y. Watanabe, & S.-K. Lee (Eds.), *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 73–88). Somerville, MA: Cascadilla Proceedings Project.
- Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews Neuroscience*, 2, 717–726. doi:10.1038/35094573
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92, 231–270. doi:10.1016/j.cognition.2003.10.008
- Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62, 148–171. doi:10.1111/j.1467-9922.2012.00710.x

- Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. *PLoS ONE*, 8(7). doi:10.1371/journal.pone.0069172
- Vissers, C., Kolk, H. H. J., van de Meerendonk, N., & Chwilla, D. J. (2008). Monitoring in language perception: Evidence from ERPs in a picture-sentence matching task. *Neuropsychologia*, 46, 967–982. doi:10.1016/j.neuropsychologia.2007.11.027
- 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. doi:10.1162/jocn.1996.8.3.231
- White, E. J., Genesee, F., White, L., King, E., & Steinhauer, K. (2006, November). Factors contributing to native-like processing of a late-acquired second language: Evidence from event-related brain potentials. Poster presented at the Annual Boston University Conference on Language Development. Abstract available at: <https://www.bu.edu/buclid/files/2011/06/handbook-312006.pdf>.
- Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R, 77(1), 1–17. <https://doi.org/10.18637/jss.v077.i01>
- Zhang, Y., Li, P., Piao, Q., Liu, Y., Huang, Y., & Shu, H. (2013). Syntax does not necessarily precede semantics in sentence processing: ERP evidence from Chinese. *Brain and Language*, 126, 8–19. doi:10.1016/j.bandl.2013.04.001