ERPs Recorded during Early Second Language Exposure Predict Syntactic Learning

Laura Batterink[1] **and Helen J. Neville**[2]

Abstract

■ Millions of adults worldwide are faced with the task of learning a second language (L2). Understanding the neural mechanisms that support this learning process is an important area of scientific inquiry. However, most previous studies on the neural mechanisms underlying L2 acquisition have focused on characterizing the results of learning, relying upon end-state outcome measures in which learning is assessed after it has occurred, rather than on the learning process itself. In this study, we adopted a novel and more direct approach to investigate neural mechanisms engaged during L2 learning, in which we recorded ERPs from beginning adult learners as they were exposed to an unfamiliar L2 for the first time. Learners' proficiency in the L2 was then assessed behaviorally using a grammaticality judgment task, and ERP data acquired during initial L2 exposure were sorted as a function of performance on this task. High-proficiency learners showed a larger N100 effect to open-class content words compared with closed-class function words, whereas low-proficiency learners did not show a significant N100 difference between open- and closed-class words. In contrast, amplitude of the N400 word category effect correlated with learners' L2 comprehension, rather than predicting syntactic learning. Taken together, these results indicate that learners who spontaneously direct greater attention to open- rather than closed-class words when processing L2 input show better syntactic learning, suggesting a link between selective attention to open-class content words and acquisition of basic morphosyntactic rules. These findings highlight the importance of selective attention mechanisms for L2 acquisition. ■

INTRODUCTION

The ability to acquire a second language (L2) as an adult is becoming increasingly important in today's multicultural world, and the neural mechanisms that support this ability are an important topic of scientific inquiry. Most studies on the neural mechanisms supporting L2 acquisition have relied upon end-state outcome measures, in which the results of learning are assessed after learning has occurred. For example, McLaughlin, Osterhout, and Kim (2004) found that L2 learners show an N400 effect to L2 pseudowords after 14 hr of classroom instruction. Similarly, a number of studies have demonstrated that L2 learners elicit P600 effects to syntactic violations after relatively short periods of classroom instruction (McLaughlin et al., 2010; Osterhout et al., 2008) or after brief periods of laboratory training (e.g., Batterink & Neville, 2013; Davidson & Indefrey, 2008; Mueller, Hahne, Fujii, & Friederici, 2005). Although these findings have provided important insight into the neural mechanisms underlying L2 processing, such findings characterize the results of learning, rather than the learning process per se.

An alternative approach, which may be better suited to understanding the neural mechanisms involved in the learning process itself, involves the online recording of neural activity during learning. Such an approach has been most widely used in the memory literature by studies employing the subsequent memory paradigm (e.g., Schott, Richardson-Klavehn, Heinze, & Duzel, 2002; Röder, Rosler, & Neville, 2001; Paller, McCarthy, & Wood, 1988; Paller, Kutas, & Mayes, 1987; Neville, Kutas, Chesney, & Schmidt, 1986). In the most common variant of this paradigm, participants' ERPs are recorded while they encode a list of items. Participants are subsequently tested on these items, and ERP data from the initial encoding period are sorted as a function of participants' subsequent memory for each item. The difference in the neural response to remembered versus forgotten items is referred to as the Dm effect, operationally defined as the "difference based on later memory performance" (Paller et al., 1987). ERP studies using this procedure typically find that subsequently remembered items elicit a late positivity, usually occurring between 400 and 800 msec poststimulus and distributed over centroparietal recording sites (e.g., Schott et al., 2002; Röder et al., 2001; Paller, 1990; Paller et al., 1987, 1988; Neville et al., 1986). This positivity has been proposed to index elaborative encoding and consolidation processes engaged for the formation of distinctive memory traces (Mitchell, Andrews, & Ward, 1993; Besson, Kutas, & Van Petten, 1992; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). The Dm effect has also been shown to be sensitive to individual

---

[1]Northwestern University, [2]University of Oregon

Journal of Cognitive Neuroscience 26:9, pp. 2005–2020
doi:10.1162/jocn_a_00618

differences in encoding strategy, with dissociations occurring between participants who use elaborative rehearsal compared with the less effective rote rehearsal strategies (Karis, Fabiani, & Donchin, 1984).

More recently, online recordings of neural activity have also been used to investigate neural mechanisms involved in statistical learning. In this approach, learners' ERP (or fMRI) data are recorded while they are exposed to complex auditory or visual patterns, such as a continuous speech stream composed of nonsense words (e.g., Karuza et al., 2013; Abla & Okanoya, 2009; Cunillera et al., 2009; Abla, Katahira, & Okanoya, 2008; Cunillera, Toro, Sebastian-Galles, & Rodriguez-Fornells, 2006). After the initial exposure period, learning is assessed behaviorally, typically with a familiarity judgment task in which learners must discriminate between previously presented and novel sequences. ERP data are then sorted on the basis of participants' individual performance, for example, by dividing participants into groups of high and low performers. Studies following this approach have found that better statistical learning, as assessed by the subsequent familiarity task, is associated with enhanced N100 and/or N400 effects to sequence onsets during initial exposure (de Diego Balaguer, Fuentemilla, & Rodriguez-Fornells, 2011; Abla & Okanoya, 2009; Abla et al., 2008; Sanders, Newport, & Neville, 2002). Both the N100 and the N400 have been proposed to index online speech segmentation processes in this context, which occur earlier and more robustly in better learners (Abla et al., 2008; Cunillera et al., 2006; Sanders et al., 2002). Specifically, the N400 has been proposed to index lexical search, which depends upon the successful segmentation and identification of the recently learned words (Sanders et al., 2002). In contrast, the N100 may reflect learners' greater allocation of attention to the onsets of sequences (Sanders et al., 2002). At a domain-general level, the N100 is known to be a sensitive index of selective attention—the ability to attend to one type of input while suppressing distracting stimuli. Attended stimuli elicit larger N100 potentials than the same stimuli when unattended (e.g., Hansen & Hillyard, 1980; Hink, Hillyard, & Benson, 1978; van Voorhis & Hillyard, 1977; Hillyard, Hink, Schwent, & Picton, 1973), suggesting that the N100 indexes an early sensory gain control mechanism of attention (Luck, Woodman, & Vogel, 2000). Thus, the finding that the N100 predicts learning success suggests that selective attention to sequence onsets is an important component of statistical learning.

In summary, recordings of neural activity during learning have proven to be a powerful approach to understanding mechanisms involved in both long-term memory formation as well as statistical learning. However, this method has seldom been employed to examine other types of human learning. In the current study, we applied this approach to investigate the neural mechanisms involved in the early stages of adult L2 acquisition. Of the many different types of human learning, the study of L2 acquisition may be particularly likely to benefit from this type of approach. It has long been noted that there is a tremendous degree of individual variation in L2 acquisition (for reviews, see Dornyei, 2005; Ellis, 2004; Segalowitz, 1997). Adult L2 learners often attain drastically different levels of proficiency, even when factors such as age, native language, educational level, and L2 experience are similar. By using ERPs to index learning processes that occur during early L2 exposure, neural and cognitive mechanisms that contribute to high levels of L2 attainment may be identified. Identifying such mechanisms may yield new insight into why some L2 learners are more successful than others, even when they share comparable backgrounds and experience similar learning environments.

One candidate mechanism that may predict L2 acquisition success is the level of attention that learners direct toward particular linguistic elements in L2 input. This idea is central to a number of theoretical accounts of L2 acquisition, which propose that attention plays a key facilitative role in L2 development (e.g., Robinson, Mackey, Gass, & Schmidt, 2012; VanPatten, 1996, 2004, 2007; Schmidt, 1990, 2001; Robinson, 1996; Tomlin & Villa, 1994). For example, VanPatten's Input Processing model assumes that the processing capacity of L2 learners is limited and that only certain grammatical features receive attention during processing. Grammatical elements that are not attended will not be fully processed, preventing learners from making form-meaning connections and delaying acquisition of these features. Similarly, Schmidt's Noticing Hypothesis argues that learners must somehow notice elements in the L2 input for learning to occur. Schmidt views awareness and attention as essentially isomorphic, proposing that attention controls access to awareness and is thus necessary for subsequent learning. Tomlin and Villa (1994) also contend that attention is necessary for learning, although their view of the role of awareness diverges from Schmidt's account. According to their model, attention functions to enhance detection, defined as the cognitive registration of stimuli that does not require awareness, which in turn leads to further processing and learning. Finally, Robinson (Robinson et al., 2012; Robinson 1996) describes an attentional model that accommodates both Schmidt's Noticing Hypothesis and Tomlin and Villa's theory. Robinson argues that only a subset of input that is detected is subsequently focally attended and noticed; it is this focally attended input that is then available for learning. Of relevance to this study, Robinson also proposes that individual differences in memory and attentional capacity affect the extent of noticing, ultimately predicting subsequent L2 learning. In summary, all of these theories propose that the mere perception of certain L2 elements may not be sufficient to support learning and that attention plays a key role in facilitating L2 acquisition.

The recording of neural activity during L2 learning represents a valuable tool to test these ideas, providing an online index of where learners spontaneously direct their attention when processing L2 input. Attention- and language-related ERP effects can then be related to subsequent L2 performance, providing insight into optimal L2 processing strategies for beginning adult learners. Following this approach, we recorded ERPs from beginning adult learners as they were exposed to a novel L2 for the first time. We presented learners with a set of simple, syntactically correct L2 sentences, paired with pictures to illustrate their meaning. After the initial exposure period, we behaviorally assessed L2 syntactic learning using a grammaticality judgment task and then sorted participants' ERP data as a function of their subsequent performance. We used performance on the grammaticality judgment task as a measure of L2 proficiency because grammatical sensitivity is theorized to be an important component of language aptitude (e.g., Skehan, 2002; Carroll, 1981). Furthermore, the acquisition of L2 grammatical rules typically poses a particular challenge for L2 speakers (e.g., Weber-Fox & Neville, 1996; Johnson & Newport, 1989), and thus, assessing learning mechanisms that contribute to this ability is especially important. (Nonetheless, it is important to note that this assessment of proficiency is rather limited and does not include many other abilities known to be important for L2 attainment, such as verbal fluency, vocabulary, language comprehension, and communicative competence.)

To facilitate learning, no syntactic violations were presented during the initial exposure period, and thus, we focused on comparing the ERP response to words that primarily provide semantic information (open-class words) with those that primarily carry syntactic information (closed-class words). Previous studies have established that these two word categories elicit different ERPs in native speakers, with three major differences emerging (Munte et al., 2001; Weber-Fox & Neville, 2001; Brown, Hagoort, & ter Keurs, 1999; Neville, Mills, & Lawson, 1992). The N280 component, a negativity that is typically maximal over left anterior scalp, has been reported to be larger and/or to peak earlier for closed-class words relative to open-class words. In contrast, the N400 component, as a marker of semantic processing, is typically much larger for open-class words. Finally, from 400 to 700 msec, a slow negative potential is generally elicited by closed-class words, and not by open-class words. Because word category is typically confounded with other factors such as word length, word frequency, number of repetitions, and even low-level visual differences, there has been a great deal of debate regarding the functional interpretation of these word class differences (e.g., Brunelliere, Hoen, & Dominey, 2005; Osterhout, Allen, & McLaughlin, 2002; Munte et al., 2001; Brown et al., 1999; Osterhout, Bersick, & McKinnon, 1997; Neville et al., 1992). Nonetheless, the important finding for this study is that reliable differences are observed to open- and closed-class words. This feature allows us to examine word category ERP effects during normal canonical sentence processing in a group of L2 learners, without having to introduce syntactic violations that could slow or prevent language learning. In this study, our primary goal was to compare word class effects between learners to examine whether ERPs recorded during learning may predict learners' subsequent proficiency, rather than investigating the functional significance of the ERP word class effects per se.

On the basis of the findings from the Dm and statistical learning literatures, we hypothesized that learners' sensitivity to differences between open- and closed-class words—as indexed by their ERPs during early L2 exposure—would predict their subsequent proficiency. Guided by theoretical models of L2 acquisition, we further hypothesized that ERP differences between successful and unsuccessful learners would at least partially reflect differences in attentional mechanisms engaged during L2 learning. Regarding which specific ERP effects would be predictive of later performance, perhaps the most comparable studies have been those investigating statistical learning, in which learners' ERPs are recorded during initial exposure and then divided based on their individual performance on a subsequent test. Although L2 syntactic learning is quite different from statistical learning on the surface, these two types of learning both depend upon a learner's sensitivity to predictable sequences or patterns, and thus, L2 syntactic learning may rely upon some of the same mechanisms that have been implicated in statistical learning.

As described previously, statistical learning studies have indicated that the N100 is associated with statistical learning and often predicts subsequent performance, with good learners eliciting larger N100s to sequence onsets than poor learners (Abla et al., 2008; Cunillera et al., 2006; Sanders et al., 2002). This N100 effect is thought to reflect learners' greater allocation of attention to the onsets of sequences, which are the most unpredictable units that occur in the stimulus stream (Sanders et al., 2002). Based on these findings, we hypothesized that a differential N100 response to open- versus closed-class words may predict subsequent proficiency, with better learners showing a larger attentional enhancement to less predictable, more semantically meaningful open-class words compared with closed-class words. By directing greater attention to open-class words, learners may better encode the morphological endings of these words, leading to better syntactic learning. Such an effect would be consistent with theoretical accounts of L2 acquisition, described previously, which posit a key role of attention in L2 development (e.g., VanPatten, 2007; Schmidt, 2001; Robinson, 1996; Tomlin & Villa, 1994). Another effect commonly reported by studies of statistical learning is a larger N400 effect to sequence onsets in good learners compared with poor learners. This effect is thought to reflect the identification of recently segmented words and corresponding lexical search processes (Abla et al., 2008; Cunillera et al., 2006; Sanders et al., 2002). Given these results, we also hypothesized that better learners may show a larger N400 effect than poor learners to open- versus closed-class words, indexing better recognition of novel content words and corresponding efforts to link these words with their conceptual representations. We also considered the possibility that the N400, as a marker of semantic processing, may be more strongly related to semantic learning compared with syntactic learning and thus also examined the relationship between N400 amplitude and overall comprehension.

In addition to examining whether the N100 and N400 components were sensitive to subsequent syntactic learning, we also conducted exploratory analyses to investigate whether ERPs at any other time window predicted later proficiency. This analysis allowed us to broadly test our hypothesis that ERPs recorded during L2 learning would be sensitive to subsequent proficiency.

Finally, to confirm that learners processed the novel L2 stimuli in a language-like way, we examined previously reported word category effects (the N280, the N400, and the late negative shift; Munte et al., 2001; Weber-Fox & Neville, 2001; Brown et al., 1999; Neville et al., 1992), in both our group of learners as well as an additional group of native speakers.

METHODS

Participants

Sixty-seven native English speakers (33 women, mean age = 21.6 years) were recruited at the University of Oregon to participate in the experiment. All participants were carefully screened to ensure that they had never studied French or another Romance language in school or been otherwise exposed to a Romance language to a significant degree. Because the data described in this article were collected as part of a larger study on the effects of implicit and explicit L2 training conditions (Batterink & Neville, 2013), participants were randomly assigned to either an implicit ($n = 44$) or an explicit ($n = 23$) training condition, described in greater detail in the Procedure.[1] (A larger number of participants were assigned to the implicit condition because behavioral performance of implicitly trained participants was more variable.) All participants in both training conditions were right-handed. Implicit and explicit groups did not significantly differ in terms of age, sex, number of L2s studied, years of L2 study, prior study of a richly inflected language (e.g., German or Russian), or age of first exposure to an L2 (all ps > .2). On the basis of performance on the subsequent grammaticality judgment task (described in detail below), all learners were eventually divided by median-split into high-proficiency and low-proficiency groups. High- and low-proficiency groups also did not significantly differ in age, gender, number of L2s studied, years of L2 study, prior study of a richly inflected language, or age of first exposure to an L2 (all ps > .2; see Table 1). Significantly more implicitly trained learners constituted the low-proficiency group and significantly more explicitly trained learners contributed to the high-proficiency group, reflecting the finding that explicitly trained participants outperformed implicitly trained participants on the grammaticality judgment task ($p < .001$; Table 1).

In addition, data from 24 native French speakers (21 women, mean age = 26.4 years) run on the same paradigm were analyzed to directly compare word class effects in native speakers and in very early L2 learners. French-speaking participants' countries of origin included France ($n = 18$), Cameroon ($n = 3$), Belgium ($n = 1$), Burkina Faso ($n = 1$), and Italy ($n = 1$). All French speakers were born to at least one native French-speaking parent, spoke French in the home from infancy, and considered French to be their native language. Four French-speaking participants were left-handed.

All participants had normal or corrected-to-normal vision and had no history of neurological problems. Three

Table 1. Participant Demographics and Language Background by Group

	High-proficiency Group	*Low-proficiency Group*	*Statistical Difference between Groups*
Performance on GJT (d')	2.9 (1.0)	0.43 (0.42)	$t(62) = 12.9, p < .001$
Performance on comprehension questions	95%	93%	$t(62) = 1.35, p = .18$
Training condition (implicit/explicit)	13 I/19 E	29 I/3 E	$\chi^2(1) = 17.7, p < .001$
Sex	17 M/15 F	15 M/17 F	$\chi^2(1) = 0.25, p = .617$
Age (years)	22.1 (3.6)	21.2 (2.8)	$t(62) = 1.10, p = .28$
Number of L2s studied	1.1 (0.8)	0.9 (0.6)	$t(62) = 1.09, p = .28$
Years of L2 study	3.9 (3.5)	3.2 (3.7)	$t(62) = 0.76, p = .45$
Age of first exposure to an L2 (years)[a]	14.0 (4.1)	12.8 (4.5)	$t(50) = 1.03, p = .31$
Prior study of richly inflected language (yes/no)	16 Y/16 N	11 Y/21 N	$\chi^2(1) = 1.60, p = .21$

Standard deviations are provided in parentheses. GJT = grammaticality judgment task.

[a] Includes only participants who previously learned an L2.

participants in total (two from the implicit group and one from the explicit group) were excluded from EEG analyses because of excessive EEG artifact. Two additional native French speakers were excluded from all analyses because of abnormally poor performance on the grammaticality judgment task (both below 66% accuracy).

Stimuli

Training Task

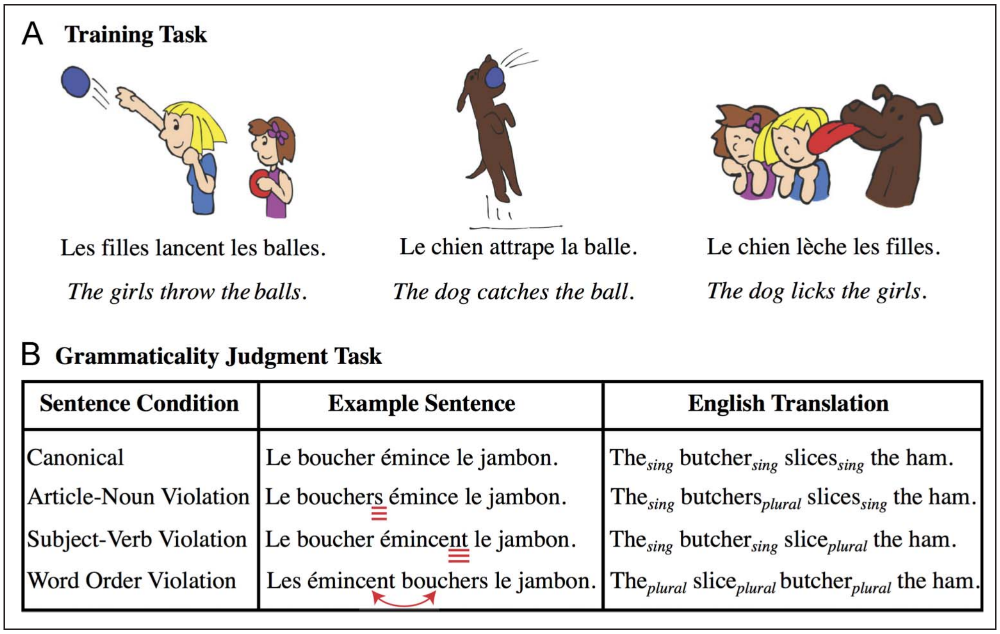
In the initial training paradigm, short narratives made up of simple French sentences that conformed to the same subject–verb–object grammatical structure were presented. Each sentence contained five words, consisting of an article, noun, direct verb, a second article, and a second noun. This "miniature" French language was intentionally designed to consist of only a limited number of words and a small set of syntactic rules to facilitate learning. Only three articles (definite articles *le*$_{\text{masc sing}}$, *la*$_{\text{fem sing}}$, and *les*$_{\text{plural}}$) and two verb conjugations (third-person singular and plural) were used throughout the training paradigm. A small pool of open-class words (98 nouns and 56 verbs) was frequently repeated across sentences and narratives (mean number of repetitions = 7). Nouns generally referred to people or common, everyday objects in the environment (e.g., boy, girl, dog, cat, bicycle, cookie). Verbs followed the regular French "-er" infinitive conjugation pattern and generally referred to simple actions that could be easily illustrated (e.g., eat, throw, catch). All sentences conformed to three grammatical rules: (a) article–noun agreement (e.g., le garçon/les garçons = "the$_{\text{sing}}$ boy/ the$_{\text{plural}}$ boys"), (b) subject–verb agreement (e.g., le garçon mange/les garçons mangent = "the boy eats$_{\text{sing}}$/the boys eat$_{\text{plural}}$"), and (c) correct word order (e.g., Le garçon mange le gateau = "The boy eats the cake").

To illustrate meaning, each sentence was paired with an accompanying picture. The picture was presented for 3 sec before the onset of the sentence. The picture was then faded out to reduce its visibility, and a fixation box was presented below the faded image. Next, the sentence was presented one word at a time in the center of the fixation box; each word had a duration of 400 msec with a 200-msec ISI. Both the faded image and fixation box stayed on the screen until 1500 msec after the onset of the final word. To ensure adequate attention, two alternative multiple-choice comprehension questions were presented after every narrative. The questions were in English, although the possible responses were French nouns. The questions focused on both subject and object nouns of the sentences; syntactic agreement was not probed. Examples of comprehension questions include "Who caught the ball?" (possible responses: *le chien* [the dog], *le chat* [the cat]) and "What did the girls buy?" (possible responses: *la robe* [the dress], *les souliers* [the shoes]). A total of 357 sentences, comprising 18 narratives, were presented. The number of sentences comprising each narrative ranged from 12 to 29 (average = 20 sentences). Examples of the sentences and pictures used in training are shown in Figure 1A. All stimuli were presented on a computer monitor placed approximately 140 cm away from the participant.

Grammaticality Judgment Task

A grammaticality judgment task was used to assess participants' learning of the syntactic rules of the novel language. In this task, new sentences that either conformed

Figure 1. Stimuli used in the (A) training task and (B) grammaticality judgment task. (A) In the training task, participants were presented with short narratives made up of simple sentences, paired with pictures to illustrate meaning. (B) Examples of syntactic violations tested in the grammaticality judgment task.



Sentence Condition	Example Sentence	English Translation
Canonical	Le boucher émince le jambon.	The$_{sing}$ butcher$_{sing}$ slices$_{sing}$ the ham.
Article-Noun Violation	Le bouchers émince le jambon.	The$_{sing}$ butchers$_{plural}$ slices$_{sing}$ the ham.
Subject-Verb Violation	Le boucher émincent le jambon.	The$_{sing}$ butcher$_{sing}$ slice$_{plural}$ the ham.
Word Order Violation	Les émincent bouchers le jambon.	The$_{plural}$ slice$_{plural}$ butcher$_{plural}$ the ham.

to or violated the grammatical rules established during the initial training phase were displayed. These sentences consisted of new verbs and nouns to which participants had not been exposed during training. Articles and verb conjugation endings did not differ from the forms that had been used in the training paradigm. As illustrated in Figure 1B, three grammatical constructions were tested: article–noun agreement, subject–verb agreement, and word order. Article–noun violations comprised number agreement mismatches between articles and nouns (e.g., Le garcons* = “the$_{\text{sing}}$ boys”). Subject–verb violations consisted of incorrectly inflected verbs that disagreed in number with the subject (e.g, Le garçon mangent* = “the boy eat$_{\text{plural}}$”). Word order violations were made up of sentences in which an article was immediately followed by an inflected third-person plural verb instead of a noun (e.g., Les mangent* = “the$_{\text{plural}}$ eat”) or in which a noun was immediately followed by another noun instead of a verb (e.g., Les livres garçons = “the$_{\text{plural}}$ books boys”). Each sentence containing a grammatical violation was matched with a grammatically correct control sentence. A total of 240 sentences were presented, with 40 sentences in each of the six conditions (article–noun violation, article–noun canonical control, subject–verb violation, subject–verb canonical control, word order violation, word order canonical control). Sentences were presented visually one word at a time, and participants made offline grammaticality judgments of each sentence.

Procedure

Participants were tested individually in a single 2.5-hr session. After application of an electrode cap, participants were seated in a comfortable chair in a dimly lit, acoustically and electrically shielded booth. Participants assigned to the implicit group were told that they would be reading stories in a foreign language that were paired with pictures to aid comprehension. They were instructed to read the sentences carefully, to follow each story as well as they could, and to learn as many of the new foreign language words as possible. No mention was made of grammar or of the upcoming grammaticality judgment task. Participants assigned to the explicit group were also informed that they would be reading stories in a foreign language. However, before the training task began, they were given a sheet of paper with a list and description of each of the grammatical rules of mini-French. They were told to read these rules carefully and informed that they would be tested on these rules in the second part of the experiment. They were further instructed that all of the sentences in the exposure task would conform to these grammatical rules and were asked to focus on both the grammar and meaning of these sentences. Additional details concerning the two training conditions have been published (Batterink & Neville, 2013). The training task took approximately 1 hr to complete.

The grammaticality judgment task was performed immediately after the training task. Participants in the implicit group were told that all of the sentences that they had read were examples of grammatically correct sentences and that, based on the knowledge that they had acquired during the training phase, they would now need to decide whether new sentences in the same language were grammatically correct or incorrect. Participants in the explicit group were instructed to judge whether each sentence was correct or incorrect based on the rules that they had learned in the training phase.

Native French speakers performed the same paradigm. They were instructed simply to read the sentences in the training task for comprehension and to judge each sentence in the grammaticality judgment task as grammatically correct or incorrect.

ERP Recording and Analysis

EEG data were collected throughout the training task and the grammaticality judgment task. Only data from the training task are reported here; data from the grammaticality judgment were reported in a previous publication (Batterink & Neville, 2013). EEG was recorded at a sampling rate of 512 Hz from 64 Ag-AgCl- tipped electrodes attached to an electrode cap using the 10–20 system. Recordings were made with the Active-Two system (Biosemi, Amsterdam, the Netherlands), which does not require impedance measurements, an online reference, or gain adjustments. Additional electrodes were placed on the left and right mastoid, at the outer canthi of both eyes, and below the right eye. Scalp signals were recorded relative to the Common Mode Sense active electrode and then re-referenced offline to the algebraic average of the left and right mastoid. To facilitate recognition of eye artifacts during data cleaning, a separate HEOG channel was computed by referencing the left and right horizontal eye channels to one another and a separate VEOG channel was computed by referencing the vertical eye channel to FP1.

ERP analyses were carried out using EEGLAB (Delorme & Makeig, 2004). Data were band-pass filtered from 0.1 to 20 Hz. Next, epochs beginning 300 msec before the onset of the first word in each sentence and continuing until 1000 msec after the final word of the sentence (total epoch length = 3700 msec) were extracted. Data containing large or paroxysmal artifacts or movement artifacts were identified by visual inspection and removed from further analysis. Data (excluding the computed HEOG and VEOG channels) were then submitted to an independent component analysis (ICA) using the extended runica routine of EEGLAB software. Ocular artifacts were identified from component scalp topographies and the component time series and removed, and ICA-cleaned data were then subjected to a manual artifact correction step to detect any residual or atypical ocular artifacts not removed completely with ICA. Finally, epochs time-locked to the onset of the first, second, third, and fourth word of each sentence were extracted from −100 to 600 msec.

To avoid sentence wrap-up effects (e.g., Hagoort, 2003; Osterhout, Holcomb, & Swinney, 1994), sentence-final words were not included in open-class word averages (closed-class words never occurred in sentence-final positions). Thus, open-class averages included the first noun and verb from every sentence (but not the last noun), whereas closed-class averages included both the first and second article. After artifact rejection, there were a total of 41,518 open-class word trials and 41,606 closed-class word trials across all learners. The number of trials identified as artifacts and rejected from analysis did not differ between the high-proficiency (mean = 118 trials of 1428) and low-proficiency participants (mean = 140 trials of 1428; $t(62) = 1.05$, $p = .30$).

To increase the signal-to-noise ratio over the 64 channels, amplitudes were averaged across neighboring electrodes to form nine channel groups of interest for statistical analysis (left anterior region: AF7, AF3, F7, F5, F3; left central region: FT7, FC5, FC3, T7, C5, C3; left posterior region: TP7, CP5, CP3, P7, P5, P3, PO7, PO3; midline anterior region: AFZ, F1, FZ, F2; midline central region: FC1, FCZ, FC2, C1, CZ, C2; midline posterior region: CP1, CPZ, CP2, P1, PZ, P2, POZ; right anterior region: AF4, AF8, F4, F6, F8; right central region: FC4, FC6, FT8, C4, C6, T8; right posterior region: CP4, CP6, TP8, P4, P6, P8, PO4, PO8).

First, general word class effects—the N280, the N400, and the late negative shift—were examined in both L2 learners and in native speakers to examine whether learners processed the L2 stimuli in a language-like way. Mean amplitudes for these components were computed using time windows selected on the basis of visual inspection of the waveforms and on previously published findings. The N280 time window was selected from 200 to 380 msec poststimulus, the N400 time window was selected from 300 to 400 msec poststimulus, and the late negative shift was captured from 400 to 600 msec poststimulus. All measurements were computed relative to a 100 msec prestimulus baseline. Repeated-measures ANOVAs were conducted with Group (native, nonnative) included as a between-subject factor and with the three distributional factors described above as within-subject factors.

To address the main hypothesis of the study, that N100 and N400 effects to open- versus closed-class words presented during initial L2 exposure would predict subsequent proficiency, mean amplitudes for the N100 and N400 to open- and closed-class words were computed for each learner. Mean N100 amplitude was computed from 100 to 120 msec poststimulus for open-class words and from 130 to 150 msec poststimulus for closed-class words, as the N100 effect peaked earlier for open-class words than closed-class words. Mean N400 amplitude was again computed from 300 to 400 msec poststimulus for both open- and closed-class words. The N100 and N400 word class effects were calculated by subtracting the response to closed-class words from the response elicited by open-class words within each time window. Multiple regression was then used to test whether N100 and N400 differences between open- and closed-class words predicted subsequent proficiency. Separate multiple regression analyses were conducted for both time windows, using d' on the grammaticality judgment task as the dependent variable and mean amplitude of the word class effect at each of the nine electrode regions as the predictors. If a model was found to significantly predict subsequent performance on the grammaticality judgment task, a series of simple linear regression analyses were used to examine which electrode regions individually predicted performance, using d' on the grammaticality judgment task as the dependent variable and mean ERP amplitude at a single electrode region as the predictor variable.

Group analyses were also conducted to examine N100 and N400 grand averages as a function of proficiency. Learners were divided by median-split into high- and low-proficiency groups. Group differences in mean amplitude of the N100 and N400 were analyzed using separate repeated-measures ANOVAs, including Word Class (open, closed), Left/Right (left-hemisphere, midline, right-hemisphere), and Anterior/Posterior (anterior, central posterior) as within-subject factors and Group (high proficiency, low proficiency) as a between-subject factor. Where significant group differences were found, effects within each group were quantified in follow-up analyses using separate ANOVAs for each group. An additional follow-up analysis examined whether significant group effects changed during the training session. In this analysis, fthe training session was divided into nine equal blocks (two stories per block). A repeated-measures ANOVA was then conducted with Group (high proficiency, low proficiency) as a between-subject factor and block (1–9), word class, and the three distributional factors as within-subject factors.

In addition, as a broad test of whether ERPs during any time window were sensitive to subsequent performance on the grammaticality judgment task, we conducted an exploratory analysis designed to examine potential proficiency effects over the entire averaging epoch (−100 to 600 msec). A running independent samples t test comparing high- and low-proficiency groups at each time point and electrode location was conducted. Significant p values ($p < .05$) were then plotted superimposed over a grand average of the group difference waves, allowing us to visualize any ERP components that were potentially sensitive to subsequent proficiency. Any effect that persisted for at least 50 msec and was evident across at least four neighboring channels was identified as a candidate for further analysis.

Finally, we examined whether N100 and N400 effects correlated with online comprehension, with the caveat that our two-alternative forced-choice comprehension questions were primarily designed to ensure that learners paid adequate attention to the L2 stories and represent a rather crude measure of comprehension.

RESULTS

Behavioral Results

Training Task

All learners performed well on the comprehension task of the training paradigm (mean percentage correct = 94% [5%]), indicating that they paid adequate attention to the presented stimuli. Native French speakers achieved significantly higher comprehension scores than learners (mean = 99% [2%]; $t(84) = 5.89$, $p < .001$).

Grammaticality Judgment Task

As expected, performance on the subsequent grammaticality judgment task was highly variable across participants, with scores ranging from at-chance to ceiling levels (mean d' score = 1.66 ($SD = 1.45$), range in d' scores = −0.1 to 4.52; mean accuracy = 72% (18%), range in accuracy = 48–99%). As a group, implicitly trained participants performed more poorly than explicitly trained participants, $t(62) = 7.20$, $p < .001$, although both groups scored significantly above chance (implicit group: mean accuracy = 64% (14%), mean d' score = 0.96 [1.0], $t(41) = 6.18$, $p < .001$; explicit group: mean accuracy = 89% [13%], mean d' score = 3.00 [1.2], $t(21) = 11.56$, $p < .001$).

Across all learners, performance was significantly lower in the verb agreement condition ($d' = 1.42$ (1.72)) relative to the other two conditions (omnibus ANOVA: $F(2, 126) = 7.01$, $p = .002$; contrast: $F(1, 63) = 10.1$, $p = .002$). Performance did not significantly differ between the noun agreement condition ($d' = 1.73$ [1.49]) and the word order condition ($d' = 1.85$ [1.48]; contrast: $F(1, 63) = 1.44$, $p = .235$). Training condition did not interact with differences in performance on the three grammatical conditions, $F(2, 124) = 0.95$, $p = .38$. Performance on the three grammatical conditions was highly correlated across learners (all rs > .77, all ps < .001), indicating that learners who successfully learned one grammatical rule were also more likely to acquire the other two rules.

Native French speakers performed near ceiling on this task, significantly outperforming the nonnative learners (mean accuracy = 97% [3%]; mean d' score = 3.96 [0.56]; $t(84) = 10.57$, $p < .001$).

High- versus Low-proficiency Groups

As described in the Methods, participants were divided by median-split into high- and low-proficiency groups based on overall performance on the grammaticality judgment task. Comprehension scores during the initial training task did not significantly differ between learners who were subsequently assigned to the high- and low-proficiency groups (high-proficiency group mean = 95% [6%]; low-proficiency group mean = 93% [5%], $t(62) = 1.35$, $p = .18$). Similarly, there was no significant correlation between performance on the comprehension task and on the subsequent grammaticality judgment task across learners ($r = .20$, $p = .11$).

ERP Results

General Word Class Effects in Native Speakers and Early L2 Learners

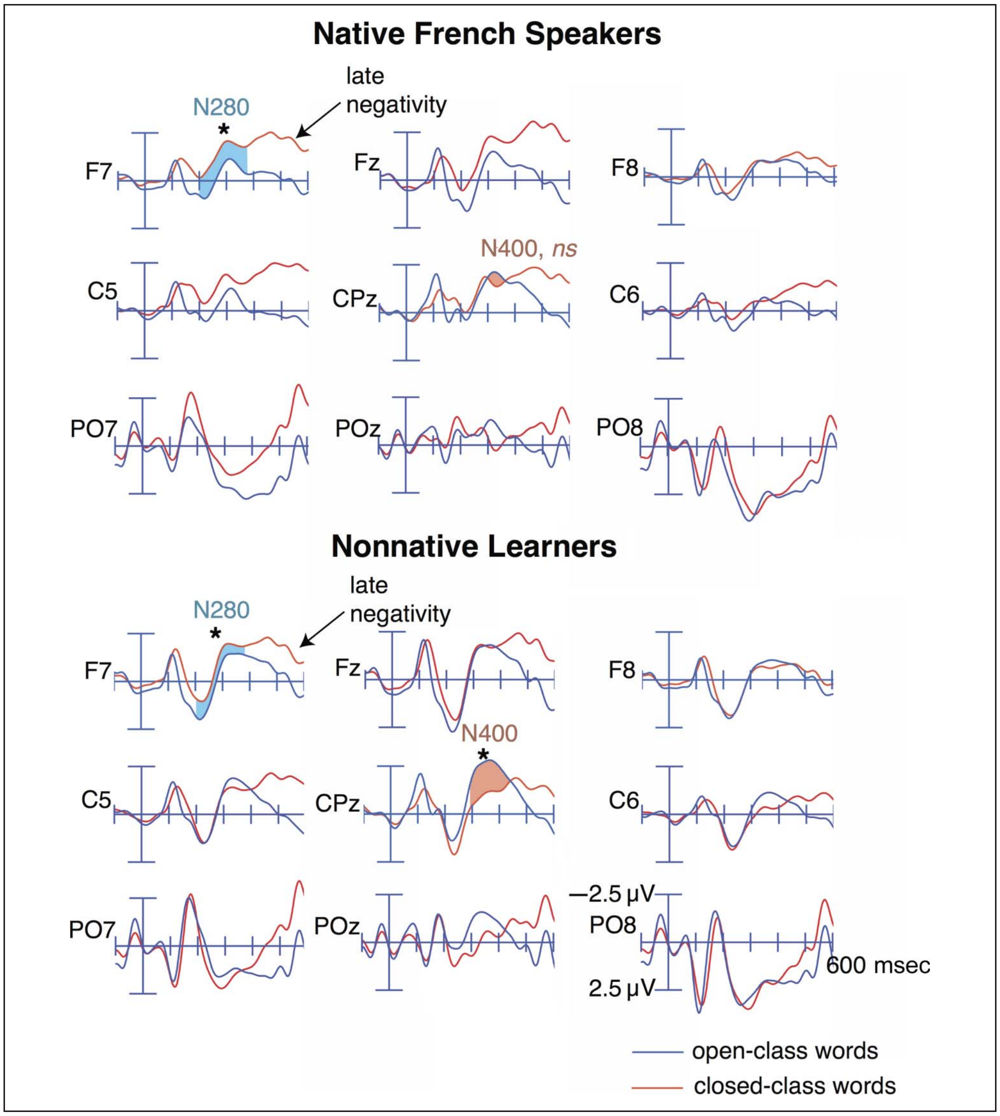
Characteristic word class effects that are typically observed in native speakers—the N280, the N400, and the late negative shift—were also observed in our L2 learners, establishing that learners processed the novel L2 stimuli in a language-like way (Figure 2). Nonetheless, some differences between groups were also observed, described in more detail below.

N280 time window (200–380 msec). Native French speakers showed a significantly larger N280 effect, reflecting a greater negativity to closed-class words relative to open-class words, than L2 learners (NS vs. L2 Group: $F(1, 84) = 8.06$, $p = .006$). Step-down analyses revealed that the N280 was nonetheless significant in both native French speakers as well as nonnative learners, with similar distributions in both groups, maximal over left anterior sites (Native group: Word Class: $F(1, 21) = 11.91$, $p = .002$; Word Class × Left/Right: $F(2, 42) = 6.74$, $p = .004$; Word Class × Ant/Post: $F(2, 42) = 6.03$, $p = .016$; Nonnative group: Word Class × Left/Right: $F(2, 126) = 7.82$, $p = .001$; Word Class × Ant/Post: $F(2, 126) = 21.1$, $p < .001$; follow-up ANOVA over left anterior region: Word Class: $F(1, 63) = 12.63$, $p = .001$; Figure 2).

N400 time window (300–400 msec). Nonnative learners showed a significantly larger N400 effect than native French speakers (NS vs. L2 Group × Word Class: $F(1, 84) = 6.93$, $p = .010$). During this time window, native French speakers showed an extended negativity to closed-class words that was maximal over left anterior sites (Word Class: $F(1, 21) = 4.39$, $p = .049$; Word Class × Left/Right: $F(2, 42) = 13.99$, $p < .001$; Word Class × Ant/Post: $F(2, 42) = 6.79$, $p = .012$). Although a hint of an N400 effect can be seen over midline posterior electrodes (Figure 2), this effect did not reach significance, even when the analysis was restricted to this region ($p > .3$). In contrast, nonnative learners showed a robust N400 effect with a typical distribution, largest over midline posterior electrodes (Word Class: $F(1, 63) = 4.71$, $p = .034$; Word Class × Left/Right: $F(2, 126) = 10.67$, $p < .001$; Word Class × Ant/Post: $F(2, 126) = 16.62$, $p < .001$; Figure 2).

Late negative shift (400–600 msec). Both native French speakers and nonnative learners showed a similar late negative effect to closed-class words, maximal over left anterior/central electrodes sites (Word Class [across groups]: $F(1, 84) = 65.1$, $p < .001$; Word Class × Left/Right: $F(2, 168) = 39.7$, $p < .001$; Word Class × Left/Right × Ant/Post: $F(4, 336) = 23.9$, $p < .001$). There were no significant differences in either the amplitude or the distribution of this effect between the two groups (NS vs. L2 Group × Word Class: $F(1, 84) = 2.06$, $p = .16$; all distribution interactions $p > .14$). Unlike most previous studies

Figure 2. ERPs time-locked to open- and closed-class words. ERPs demonstrate word class effects (the N280, the N400, and the late negative shift) in both native speakers and in L2 learners, providing a general picture of processing.



(e.g., Weber-Fox & Neville, 2001; Brown et al., 1999; Neville et al., 1992), ERPs to closed-class words did not show a clear return to baseline between the N280 and the late negative shift (Figure 2).

Hypothesis-driven Analyses

We now turn to analyses designed to test the main hypothesis of the study, that N100 and N400 differences between open- and closed-class words predict subsequent proficiency.

Effects of Training Condition on N100 and N400 Word Class Effects

Training condition (implicit or explicit) did not have a significant effect on word class effects during either the N100 or N400 time window (Training × Word Class: N100: $F(1, 62) = 0.97$, $p = .33$, all interactions *ns*, $p > .4$; N400: $F(1, 62) = 0.95$, $p = .33$, all interactions *ns*, $p > .3$). Thus, we collapsed across implicitly and explicitly trained learners, excluding training condition as a factor, in all subsequent ERP analyses.

ERP Word Class Effects Predictive of Subsequent Proficiency in L2 Learners

N100 time window. REGRESSION ANALYSES. Across all learners, N100 word class amplitude significantly predicted subsequent performance on the grammaticality judgment task, $F(9, 54) = 2.31$, $p = .028$; $R^2 = 0.28$. Follow-up analyses conducted over individual electrode regions showed that N100 mean amplitude significantly predicted proficiency at the midline anterior ($R = 0.259$, $p = .039$), midline central ($R = 0.313$, $p = .012$), midline posterior ($R = 0.285$, $p = .023$), right anterior ($R = 0.330$, $p = .008$), and right central ($R = 0.306$, $p = .014$) regions (Figure 3). To confirm that these relationships were not simply the result of individual differences in overall engagement with the task, we examined whether the N100 word class effect

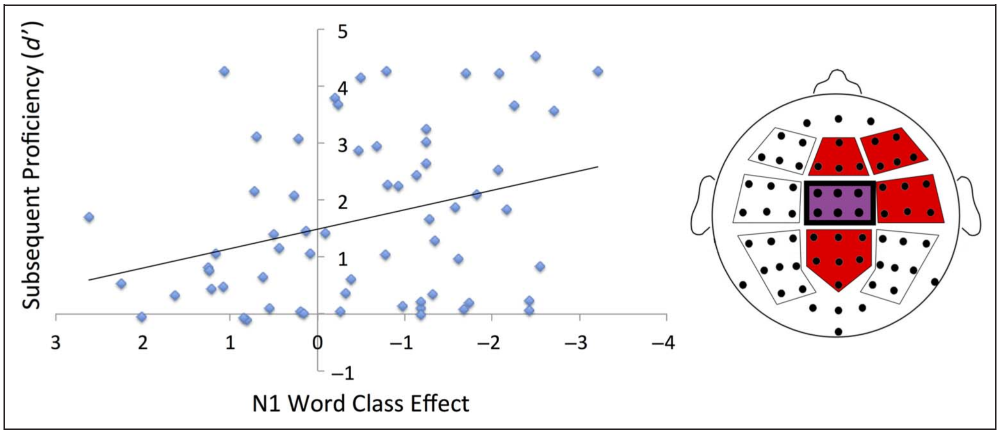


Figure 3. Scatterplot showing correlations between amplitude of the N100 word class effect, during initial L2 exposure, and subsequent performance on the grammaticality judgment task (d'), at the midline central electrode region, across all learners (n = 64). The midline central electrode region is outlined in black and colored purple. Additional electrode regions that also showed significant N100 proficiency correlations are indicated in red on the montage.

continued to predict performance on the grammaticality judgment task after controlling for comprehension performance. Results from this analysis changed very little from the original analysis (midline anterior: $R = 0.282$, $p = .025$, midline central: $R = 0.315$, $p = .012$, midline posterior: $R = 0.281$, $p = .025$, right anterior: $R = 0.325$, $p = .009$, and right central: $R = 0.284$, $p = .024$).

High- versus low-proficiency group analysis. As shown in Figure 4, high- and low-proficiency learners (divided by median split) showed significant differences in the N100 word class effect, measured as the difference in N100 amplitude elicited by open-class versus closed-class words (Proficiency Group × Word Class: $F(1, 62) = 5.13$, $p = .027$). High-proficiency learners showed a significantly larger N100 response to open-class words compared with closed-class words, an effect that was maximal over midline posterior and midline central sites (Word Class: $F(1, 31) = 6.74$, $p = .014$; Word Class × Left/Right: $F(2, 62) = 13.09$, $p < .001$; Word Class × Left/Right × Ant/Post: $F(4, 124) = 12.99$, $p < .001$). In contrast, low-proficiency learners did not show a significant word class N1 effect at any electrode region (Word Class: $F(1, 31) = 0.49$, $p = .49$; Word Class × Left/Right $F(2, 62) = 7.60$, $p = .003$; follow-up ANOVA over midline sites: Word Class effect *ns*, $p > .4$). This Word Class × Left/Right interaction indicates that different electrode regions show different ERPs from one another as a function of word class, but that the word class effect in low-proficiency learners is not reliable at any one particular region.

Block analysis. To examine when high- and low-proficiency participants began to show different N100 word class effects, we divided the training block into nine equal blocks (two stories per block). In the first block, there were no significant proficiency group differences in the N100 effect (Proficiency Group × Word Class: $F(1, 62) = 1.26$, $p = .27$, p value for all interactions > .2). However, by the second block, high-proficiency participants showed a significantly larger N100 effect to open-class versus closed-class words than low-proficiency participants (Proficiency Group × Word Class: $F(1, 62) = 7.87$, $p = .007$). These group differences persisted across the remaining seven blocks (Proficiency Group × Word Class Effect across Blocks 3–9: $F(1, 62) = 3.46$, $p = .067$). These findings indicate that the two proficiency groups began to

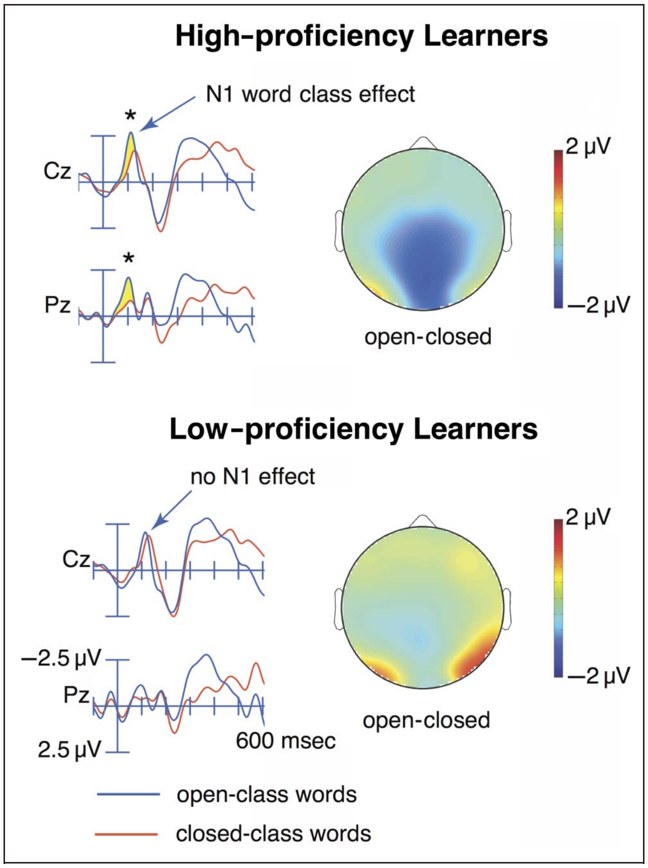


Figure 4. ERPs demonstrating the N100 word class effect in high- and low-proficiency learners (n = 32 per group, divided by median split). ERPs are time-locked to open-class (blue) and closed-class (red) words. Voltage maps show the distribution of the N100 word class effect, computed by subtracting the mean ERP amplitude to closed-class words (from 130 to 150 msec poststimulus) from the mean ERP amplitude to open-class words (from 100 to 120 msec poststimulus).

show different N100 word class effects soon after—though not immediately after—the start of the training session.

N400 time window. REGRESSION ANALYSES. N400 word class amplitude did not significantly predict subsequent performance on the grammaticality judgment task, $F(9, 54) = 0.84$, $p = .58$; $R^2 = 0.12$.

HIGH- VERSUS LOW-PROFICIENCY GROUP ANALYSIS. High- and low-proficiency participants did not show significant amplitude differences for the N400 word class effect (Proficiency Group × Word Class: $F(1, 62) = 0.048$, $p = .83$; p value for all interactions $> .16$).

Exploratory analysis across entire averaging epoch. Our exploratory analysis, designed to test whether any ERP word class effects other the N100 and N400 predicted subsequent proficiency, confirmed that only the N100 was sensitive to performance on the grammaticality judgment task ($p < .05$). No other time windows showed significant differences between high- and low-proficiency participants (all ps $> .05$), and thus, we did not subject any other ERP word class effects to further analysis.

ERP Word Class Effects Correlating with Comprehension in L2 Learners

The amplitude of the N100 word class effect did not correlate with comprehension performance at any electrode region (all ps $> .11$). In contrast, the N400 word class effect and comprehension performance correlated significantly at left posterior electrodes ($r = .31$, $p = .012$) and showed marginal correlations at midline posterior ($r = .25$, $p = .051$) and right posterior sites ($r = .22$, $p = .087$). This correlation appears to be primarily driven by the N400 amplitude to open-class words, which significantly or marginally significantly correlated with comprehension performance at left posterior ($r = .21$, $p = .097$), midline central ($r = .22$, $p = .091$), midline posterior ($r = .26$, $p = .036$), and right posterior sites ($r = .24$, $p = .056$), such that learners who showed a larger N400 effect to open-class words performed more accurately on the comprehension questions.

DISCUSSION

We found that the difference in the N100 response to open- and closed-class words during early L2 exposure predicts learners' success in acquiring the language's syntax, as assessed by a separate grammaticality judgment task. Overall, L2 learners who showed an enhanced N100 response to open-class words relative to closed-class words subsequently discriminated between grammatically correct and incorrect sentences more accurately than learners who did not show an N100 word class effect. The amplitude of the N100 effect predicted subsequent performance on the grammaticality judgment task in a linear fashion. In contrast, the N400 word class effect did not predict syntactic learning but showed significant correlations with L2 comprehension.

ERPs during L2 Exposure as a Predictor of Acquisition Success

The major novel finding of our study is that ERPs recorded during early L2 exposure predict subsequent performance on a separate language measure in adult learners. We applied an approach that has been used previously to investigate long-term memory and statistical learning—the online recording of neural activity during learning—to the study of adult L2 acquisition. This approach capitalizes on the large degree of individual variation between L2 learners to identify neural mechanisms involved in the learning process that are likely to promote successful L2 acquisition. This approach also has the advantage of yielding insight directly on the learning process per se, rather than simply on the results of learning, as in most previous ERP L2 learning studies (e.g., Batterink & Neville, 2013; McLaughlin et al., 2004, 2010; Davidson & Indefrey, 2008; Osterhout et al., 2008; Mueller et al., 2005). At a very general level, our results build upon findings from the subsequent memory ("Dm") and statistical learning literatures, demonstrating that ERPs acquired during learning can be used to predict subsequent performance in a novel cognitive domain—the acquisition of L2 syntax.

Given that the N100 is a sensitive index of selective attention (e.g., Luck et al., 2000), our finding that the N100 predicts L2 acquisition success provides evidence that selective attention plays an important role in the acquisition of L2 syntax in adults. These results converge with a number of current theoretical models of L2 acquisition (e.g., Robinson et al., 2012; VanPatten, 1996, 2004, 2007; Schmidt, 1990, 2001; Robinson, 1996; Tomlin & Villa, 1994), which argue that grammatical features in L2 input must be adequately attended to be acquired. These theories are based primarily upon behavioral evidence, for example, from studies that measure attention during L2 processing using concurrent or retrospective verbal reports (e.g., Leow, 1997, 1998, 2001). Although such studies have contributed important evidence for the development of L2 acquisition theories, verbal reports may clearly fail to reflect all cognitive processes that occur during L2 processing; many aspects of processing may not be accessible to learners' awareness and thus not reflected in their verbal reports. A second concern is that metalinguistic verbalization represents an additional task that may alter normal L2 processing (cf. Robinson et al., 2012; Leow & Bowles, 2005). Because ERPs do not suffer from these limitations, our finding that the N100 effect predicts L2 proficiency provides important converging data for the facilitative role of selective attention in L2 acquisition.

Our results also have implications for theoretical models of L2 acquisition at a more fine-grained level. Our N100 data indicate that learners who spontaneously direct greater attention to open-class words rather than closed-class words when processing L2 input show better syntactic learning. This finding suggests a link between selective attention to open-class content words and morphosyntactic learning, a relationship that may be better understood in the context of VanPatten's Input Processing model (VanPatten, 1996, 2004, 2007). This theory consists of a set of principles that describe how learners initially perceive and process incoming linguistic data to make connections between form and meaning/function (e.g., recognizing that -ed indicates past tense). Central to this model are the ideas that L2 learners have limited working memory resources for processing input and that learners process input for communicative meaning before processing it for grammatical form. The theory states that learners tend to seek out and preferentially process content words, the principal source of referential meaning, whereas grammatical items (such as closed-class words and inflections) are skipped over or only partially processed before being dumped from working memory. It also proposes that learners will rely on lexical items rather than grammatical items for meaning when both encode the same semantic information. Finally, it posits learners will process more meaningful morphology, defined as items that carry nonredundant semantic information, over less meaningful morphology.

Our finding that high-proficiency learners show a greater N100 to open-class compared with closed-class words is consistent with VanPatten's proposal that learners preferentially process content words to obtain meaning from L2 input. Our findings further suggest that greater attention to content words not only facilitates online comprehension but may also contribute to the acquisition of basic syntactic rules, possibly via a chunking mechanism. VanPatten's model proposes that morphological inflections may be processed but not in isolation; rather, they may be fused with the content words with which they occur (VanPatten, 2004). Through this type of chunking, learners who focus on content words to obtain meaningful semantic information may be more likely to simultaneously process the morphological endings of these content words. These learners may then make a connection between the morphological form and its corresponding function (e.g., linking −s to pluralization), contributing to the acquisition of basic syntactic agreement and word order rules. In contrast, because L2 comprehension and processing is extremely effortful for beginning learners, learners who do not strategically locate and process the major "units of meaning" (VanPatten, 2004) in L2 input are likely to overload their working memory capacities. When working memory resources are exhausted, morphological endings may be skipped over, preventing the acquisition of related syntactic rules. Thus, for beginning L2 learners, focusing primarily on the extraction of meaning from the input may be adaptive not only for online comprehension but also for the acquisition of certain basic syntactic rules. To be sure, although this strategy may be effective during very early stages of L2 acquisition, it may become counterproductive during later acquisition stages. If learners preferentially attend to content words, skipping closed-class function words because they are not needed to comprehend an utterance, this will likely delay acquisition of certain syntactic properties of the L2. For example, the acquisition of grammatical gender, a syntactic feature that does not carry inherent semantic value, may be delayed when learners process L2 input for meaning rather than for form (VanPatten, 2004).

In contrast to the N100 word category effect, the N400 effect did not predict syntactic learning in our group of learners. However, the N400 did moderately correlate with learner's online comprehension, with learners who showed larger N400 effects to open-class content words performing more accurately on the comprehension task. Thus, we found a double dissociation, with the N100 effect predicting syntactic learning but not comprehension and the N400 correlating with comprehension but not syntactic learning. At a general level, the N400 is thought to index lexical access or semantic integration processes (Lau, Phillips, & Poeppel, 2008) and has been shown to be a sensitive index of semantic word learning (Batterink & Neville, 2011; Mestres-Misse, Rodriguez-Fornells, & Munte, 2007; McLaughlin et al., 2004). Modulation of the N400 has also been reported during speech segmentation tasks that involve extracting new nonsense words from a continuous speech stream (Cunillera et al., 2006, 2009; de Diego Balaguer et al., 2007; Sanders et al., 2002). In the context of this study, a larger N400 to open-class content items may reflect learners' efforts to map the lexical form of these words onto their conceptual representations, with learners who engage in this process to a larger extent showing better semantic learning. It is important to note, however, that our study was not specifically designed to assess word learning, with our comprehension questions representing a fairly crude measure of semantic acquisition. Future studies may use a more fine-grained measure of semantic word learning to more carefully investigate this possible double dissociation between the N100 and N400 and syntactic and semantic aspects of L2 learning.

Interestingly, training condition (implicit vs. explicit) did not have a significant impact on the amplitude of either the N100 or N400 word class effects. In other words, training condition and N100 amplitude represent uncorrelated variables that both made independent predictions about syntactic proficiency. Although the effect of training condition predicted a much greater proportion of the variance in proficiency than the effect of N100 amplitude (accounting for approximately 46% of the variance in performance on the grammaticality judgment task as compared with 28%), this result shows that using a direct neural measure as a predictor can explain additional variability in performance beyond what is accounted for by the initial set of independent variables established by the study design. The finding that training condition and N100 amplitude are uncorrelated also has implications for the theoretical significance of the N100 effect, suggesting that this effect reflects a mechanism that operates independently of training condition. It appears that better selective attention to more relevant content words cannot be preferentially engaged simply by giving participants a particular set of instructions but may instead relate more to a learner's set of individual, internal characteristics, such as working memory capacity or general attention skills.

Predicting L2 Learning Success: Practical Implications

These findings contribute to a broad field of research concerned with predicting L2 learning success. Why some learners do better than others is a central question in L2 acquisition research, one of both scientific and practical interest. A longstanding observation is that adult L2 learners often attain very different levels of proficiency even when important factors such as age, native language, educational background, and L2 experience are similar (Dornyei, 2005; Ellis, 2004; Segalowitz, 1997). For many decades, behavioral studies of L2 learners have identified and described a large number of learner characteristics that contribute to success in mastering an L2, ranging from affective factors such as personality and motivation to cognitive factors such as intelligence and language aptitude (cf. Dornyei, 2005). Now, with advances in cognitive neuroscience methods, these behavioral findings may be complemented by examining individual differences in neural activity during L2 learning, which can provide direct insight into cognitive and neural mechanisms engaged by successful L2 learners.

The main novel finding of this study—that allocating greater attention to more meaningful open-class words predicts successful acquisition of basic L2 syntactic rules—may have future practical implications. The variability between L2 learners that has been so commonly noted in the behavioral L2 acquisition literature was clearly evident in our sample of participants, who showed drastically different levels of performance on the grammaticality judgment task after the same amount of L2 exposure. The present results suggest that, in addition to learner characteristics such as age of acquisition and motivation (e.g., Lightbown & Spada, 2006), individual differences in selective attention may play an important role in predicting L2 attainment. Better attention skills may allow L2 learners to more easily focus on the most relevant or useful aspects of language, facilitating language processing and ultimately leading to higher levels of L2 attainment.

Data from training studies in children support this idea, demonstrating that better selective attention skills often confer advantages in other cognitive domains. For example, children with specific language impairment who underwent 6 weeks of intensive computerized training showed improvements in both a neural measure of selective auditory attention as well as receptive language, suggesting that the training program improved language in part by targeting attention (Stevens, Fanning, Coch, Sanders, & Neville, 2008). In a second study, kindergarten children at risk for reading difficulties who received supplemental reading instruction showed enhancements in a neural index of selective auditory attention, accompanied by behavioral improvements in preliteracy skills (Stevens et al., 2011). Similarly, attention training in typically developing preschoolers led to improvements in behavioral and neural measures of attention as well as standardized measures of intelligence (Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005). These studies suggest that selective attention has powerful effects that may spill over into other domains, including language acquisition.

The present findings may be used as a springboard for future research designed to assess whether L2 training methods that explicitly target selective attention are effective. Even in the somewhat artificial context of this study, in which L2 stimuli were presented visually and at a relatively slow rate, selective attention to open-class words was found to predict L2 acquisition. This type of attentional mechanism is likely to play an even more important role in a real-world L2 acquisition environment, in which beginning L2 learners are required to rapidly process continuous speech streams made up of mostly unknown words. Given that temporal selective attention also contributes to normal speech processing in native speakers (Astheimer & Sanders, 2009, 2011), one speculation is that training that targets temporal selective attention skills in beginning L2 learners may help to improve L2 acquisition outcomes in adults. Importantly, previous studies have established that selective attention can be trained and improved even in adults (e.g., Tang et al., 2007; Green & Bavelier, 2003, 2006), providing a proof of principle for the malleability of attention in adulthood. Future research could evaluate whether providing adult L2 learners with supplementary auditory selective attention training, in addition to a standard L2 curriculum, leads to gains in language comprehension and syntactic learning.

ERPs Elicited by Open- and Closed-class Words in Native Speakers and L2 Learners

Characteristic word class effects typically observed in native speakers—the N280, the N400, and the late negative shift—were present in our L2 learners, providing confirmation that learners processed the novel L2 stimuli in a language-like way. However, although L2 learners showed a robust N400 effect to open-class words, as has been reported previously for both native and nonnative speakers (Weber-Fox & Neville, 2001; Neville et al., 1992), the N400 word class effect in native speakers was not statistically significant. Group comparisons confirmed that the N400 effect in L2 learners was significantly larger than in native speakers. Because L2 learners were generally unfamiliar with the L2 open-class words, they likely experienced greater difficulty mapping the lexical form of these items onto their conceptual representations, consequently eliciting larger N400 effects (Lau et al., 2008). In contrast, these same open-class words were likely processed quite differently by native French speakers. Because these words consisted of very common French nouns and verbs, were highly expected given the prior picture and story context, and were repeated frequently throughout the exposure period, native speakers would have encountered very little difficulty with lexical access and semantic integration. These factors—frequency of usage, degree of semantic expectancy, and repetition—are known to dramatically alter the overall amplitude of the N400 (Besson et al., 1992; Van Petten et al., 1991; Van Petten & Kutas, 1990; Kutas & Hillyard, 1980, 1984) and likely contributed to the lack of N400 effect observed in our native speakers.

Limitations

There are two limitations to this study that should be discussed. First, as all ERP effects reported here involve contrasting between open- and closed-class words, it is important to note that any ERP differences between these two types of words may be driven not only by word class but also by confounding factors such as word length, overall word frequency, number of repetitions within the experiment, and even low-level differences in visual characteristics. Given prior evidence that other factors such as word length can have a large impact on observed word class effects (e.g., Osterhout et al., 1997, 2002), we cannot make any claims about which factors may be contributing to main effects of word category across participants. Fortunately, the main findings of this study involve comparisons of effects between participants and thus can still be meaningfully interpreted. Because all participants were exposed to identical stimuli, factors such as word length and word repetition would contribute to effects in both high- and low-proficiency learners, and thus differences between learners must necessarily drive our main between-participant effects of interest.

A second issue concerns the generalizability of our paradigm. As noted previously, our L2 sentences were intentionally designed to consist of a very limited number of open- and closed-class words and to follow a very simple, predictable structure to facilitate L2 learning in a short laboratory session. In contrast, L2 learners in the real world typically encounter linguistic input that is far more variable and that consists of a vastly larger vocabulary pool. Thus, one question is whether the N1 effect that we report here would also be elicited by L2 learners processing more complex, ecologically valid input. Although this question remains to be addressed empirically by future research, our working hypothesis is that such an effect would indeed continue to be observed in a more natural context. Although the sentences in our paradigm were highly repetitive and predictable, many aspects of natural language are also highly predictable, containing a large number of cues that can be used to predict certain features of an upcoming word. For example, whether a word is open- or closed-class can frequently be predicted based on the preceding context, with open-class words being very likely to follow certain categories of words such as articles. Native speakers are frequently able to predict the part of speech of an upcoming word on the basis of previous words (Kimball, 1975), and there is strong evidence that the brain engages in implicit, probabilistic anticipatory language processing to optimize language comprehension (Kutas, DeLong, & Smith, 2001). When processing spoken language, L2 learners can also use prosody to help them locate meaningful lexical items in L2 input, as open-class content words tend to receive stronger stress than closed-class items (VanPatten, 2004). Thus, when tasked with comprehending an unfamiliar language, beginning L2 learners may take advantage of such cues to predict whether an upcoming word will require additional processing or be relevant for comprehension. Good language learners with better selective attention skills may allocate greater attention to these words than less proficient learners, facilitating online processing and allowing them to acquire a novel L2 more quickly.

Conclusions

Our data indicate that adult learners who show a larger N100 effect to open-class versus closed-class words during initial exposure to a novel L2 attain higher levels of proficiency in the novel language, as assessed by performance on a subsequent grammaticality judgment task. This finding indicates that greater selective attention to more relevant content words when processing a novel L2 is associated with better language outcomes, suggesting that selective attention may be an important component of successful L2 acquisition. At a more general level, this study also represents a methodological advance, providing evidence that the online recording of neural activity during learning can provide new insight into neural and cognitive mechanisms contributing to L2 acquisition.

Reprint requests should be sent to Laura Batterink, Psychology Department, Northwestern University, 2029 Sheridan Road, Evanston, IL 60208, or via e-mail: lbatterink@northwestern.edu.

Note

1. Because the hypotheses for this study did not relate to training condition and there were no significant effects of training condition on any of our ERP effects, we collapse across training condition in all analyses to increase our statistical power.

REFERENCES

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20,* 952–964.

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience Research, 64,* 185–190.

Astheimer, L. B., & Sanders, L. D. (2009). Listeners modulate temporally selective attention during natural speech processing. *Biological Psychology, 80,* 23–34.

Astheimer, L. B., & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia, 49,* 3512–3516.

Batterink, L., & Neville, H. J. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience, 23,* 3181–3196.

Batterink, L., & Neville, H. J. (2013). Implicit and explicit second language training recruit common neural mechanisms for syntactic processing. *Journal of Cognitive Neuroscience, 25,* 936–951.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4,* 132.

Brown, C. M., Hagoort, P., & ter Keurs, M. (1999). Electrophysiological signatures of visual lexical processing: Open- and closed-class words. *Journal of Cognitive Neuroscience, 11,* 261–281.

Brunelliere, A., Hoen, M., & Dominey, P. F. (2005). ERP correlates of lexical analysis: N280 reflects processing complexity rather than category or frequency effects. *NeuroReport, 16,* 1435–1438.

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Cunillera, T., Camara, E., Toro, J. M., Marco-Pallares, J., Sebastian-Galles, N., Ortiz, H., et al. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage, 48,* 541–553.

Cunillera, T., Toro, J. M., Sebastian-Galles, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research, 1123,* 168–178.

Davidson, D. J., & Indefrey, P. (2008). An event-related potential study on changes of violation and error responses during morphosyntactic learning. *Journal of Cognitive Neuroscience, 21,* 433–446.

de Diego Balaguer, R., Fuentemilla, L., & Rodriguez-Fornells, A. (2011). Brain dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience, 23,* 3105–3120.

de Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Levi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE, 11,* e1175.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21.

Dornyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah, NJ: Erlbaum.

Ellis, R. (2004). Individual differences in second language learning. In C. Elder & A. Davies (Eds.), *Handbook of applied linguistics* (pp. 525–550). Oxford, UK: Blackwell.

Green, C., & Bavelier, D. (2003). Action video game modifies visual attention. *Nature, 423,* 534–537.

Green, C., & Bavelier, D. (2006). Effects of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1465–1478.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience, 15,* 883–899.

Hansen, J. C., & Hillyard, S. A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroencephalography and Clinical Neurophysiology, 49,* 277–290.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182,* 177–180.

Hink, R. F., Hillyard, S. A., & Benson, P. J. (1978). Event-related brain potentials and selective attention to acoustic and phonetic cues. *Biological Psychology, 6,* 1–16.

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 langauge. *Cognitive Psychology, 21,* 60–99.

Karis, D., Fabiani, M., & Donchin, E. (1984). “P300” and memory: Individual differences in the Von Restorff effect. *Cognitive Psychology, 16,* 177–216.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127,* 46–54.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. In J. Kimball (Ed.), *Syntax and semantics* (Vol. 4, pp. 155–179). New York: Academic Press.

Kutas, M., DeLong, K. A., & Smith, N. J. (2001). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). New York: Oxford University Press.

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203.

Kutas, M., & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307,* 161.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9,* 920–933.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning, 47,* 467–505.

Leow, R. P. (1998). Toward operationalizing the process of attention in SLA: Evidence for Tomlin and Villa’s (1994) fine-grained analysis of attention. *Applied Psycholinguistics, 19,* 133–159.

Leow, R. P. (2001). Attention, awareness, and foreign language behavior. *Language Learning, 51,* 113–155.

Leow, R. P., & Bowles, M. A. (2005). Attention and awareness in SLA. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 179–203). Washington, DC: Georgetown University Press.

Lightbown, P. M., & Spada, N. (2006). Individual differences in second language learning. In *How languages are learned* (pp. 53–76). Oxford, UK: Oxford University Press.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences, 4,* 432–440.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7,* 703–704.

McLaughlin, J., Tanner, D., Pitkanen, I., Frenck-Metre, C., Inoue, K., Valentine, G., et al. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning, 60,* 123–150.

Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17,* 1858–1866.

Mitchell, P., Andrews, S., & Ward, P. (1993). An event-related potential study of semantic congruity and repetition in a sentence-reading task: Effects of context change. *Psychophysiology, 30,* 496–509.

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,* 1229–1244.

Munte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed-class words: Class and frequency effects. *Neuropsychologia, 39,* 91–102.

Neville, H. J., Kutas, M., Chesney, G., & Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25,* 75–92.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2,* 244–258.

Osterhout, L., Allen, M., & McLaughlin, J. (2002). Words in the brain: Lexical determinants of word-induced brain activity. *Journal of Neurolinguistics, 15,* 171–187.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology, 46,* 143–168.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 786–803.

Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., et al. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics, 21,* 509–512.

Paller, K. A. (1990). Recall and stem-completion priming have different electrophysiological correlates and are modified differentially by directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1021–1032.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography & Clinical Neurophysiology, 67,* 360–371.

Paller, K. A., McCarthy, G., & Wood, C. C. (1988). ERPs predictive of subsequent recall and recognition performance. *Biological Psychology, 26,* 269–276.

Robinson, P. (1996). *Consciousness, rules, and instructed second language acquisition*. New York: Peter Lang.

Robinson, P., Mackey, A., Gass, S., & Schmidt, R. (2012). Attention and awareness in second language acquisition. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 247–267). New York: Routledge.

Röder, B., Rosler, F., & Neville, H. (2001). Auditory memory in congenitally blind adults: A behavioral-electrophysiological investigation. *Cognitive Brain Research, 11,* 289.

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 14931–14936.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5,* 700–703.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11,* 129–158.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, UK: Cambridge University Press.

Schott, B., Richardson-Klavehn, A., Heinze, H.-J., & Duzel, E. (2002). Perceptual priming versus explicit memory: Dissociable neural correlates at encoding. *Journal of Cognitive Neuroscience, 14,* 578–592.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. de Groot & J. Kroll (Eds.), *Tutorials in bilingualism* (pp. 85–112). Hillsdale, NJ: Erlbaum.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam: John Benjamins.

Stevens, C., Fanning, J., Coch, D., Sanders, L., & Neville, H. J. (2008). Neural mechanisms of selective auditory attention are enhanced by computerized training: Electrophysiological evidence from language-impaired and typically developing children. *Brain Research, 1205,* 55–69.

Stevens, C., Harn, B., Chard, D. J., Currin, J., Parisi, D., & Neville, H. J. (2011). Examining the role of attention and instruction in at-risk kindergarteners: Electrophysiological measures of selective auditory attention before and after an early literacy intervention. *Journal of Learning Disabilities, 46,* 73–86.

Tang, Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., et al. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Science, 104,* 17152–17156.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 12,* 287–301.

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition, 18,* 380–393.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3,* 131–150.

van Voorhis, S., & Hillyard, S. (1977). Visual evoked potentials and selective attention to points in space. *Perception and Psychophysics, 22,* 54–62.

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Procesing instruction: Theory, research, and commentary* (pp. 5–31). Mahwah, NJ: Erlbaum.

VanPatten, B. (2007). Input procesing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 115–135). Mahwah, NJ: Lawrence Erlbaum Associates.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research, 44,* 1338–1353.

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