

Language effects in second language learners and proficient bilinguals investigated with event-related potentials[☆]

Katherine J. Midgley^{a,b,*}, Phillip J. Holcomb^a, Jonathan Grainger^c

^a *Department of Psychology, Tufts University, Medford, MA, USA*

^b *Université de Provence - Aix-Marseille I, Marseille, France*

^c *CNRS, Marseille, France*

Received 30 March 2008; received in revised form 6 August 2008; accepted 11 August 2008

Abstract

The present study examines language effects in second language learners. In three experiments participants monitored a stream of words for occasional probes from one semantic category and ERPs were recorded to non-probe critical items. In Experiment 1 L1 English participants who were university learners of French saw two lists of words blocked by language, one in French and one in English. We observed a large effect of language that mostly affected amplitudes of the N400 component, but starting as early as 150 ms post-stimulus onset. A similar pattern was found in Experiment 2 with L1 French and L2 English, showing that the effect is due to language dominance and not language per se. Experiment 3 found that proficient French/English bilinguals exhibited a different pattern of language effects showing that these effects are modulated by proficiency. These results lend further support to the hypothesis that word recognition during the early phases of L2 acquisition in late learners of L2 involves a specific set of mechanisms compared with recognition of L1 words.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Visual word processing; Bilingualism; N400

[☆] This research was supported by grant numbers HD043251 & HD25889.

* Corresponding author. Department of Psychology, Tufts University, 490 Boston Avenue, Medford, MA 02155, USA. Tel.: +1 617 627 3521; fax: +1 617 627 3181.

E-mail address: kj.midgley@tufts.edu (K.J. Midgley).

What are the basic mechanisms underlying word recognition in a second language (L2) in late learners of a foreign language, and to what extent do they overlap with the mechanisms involved in word recognition in the first language (L1)? One possibility is that the mechanisms are basically the same, and L2 words are integrated into a common set of lexical representations much like newly acquired words in L1. This would appear to be an unlikely possibility for late learners of a second language, at least in the early phases of L2 acquisition, for several reasons. First, anecdotally, many L2 learners report using translation into L1 as a general heuristic for processing L2 words. Second, given the relatively large number of cognates (words that share form and meaning in the two languages) in languages such as English and French, it would be more economical to establish L2–L1 form–form associations rather than creating new form–meaning associations. This is the position adopted by the revised hierarchical model (RHM) of word recognition in bilinguals proposed by Kroll and Stewart (1994) (Fig. 1).

Nevertheless, the alternative hypothesis, that L2 words are basically processed in the same way as L1 words, would fit with an account of lexical representation in bilinguals according to which words from both languages are stored together. Indeed, there is an abundance of behavioral evidence in favor of a language non-selective, integrated lexicon view of written word comprehension in bilinguals, at least when the two languages share the same alphabet (e.g., Beauvillain & Grainger, 1987; Brysbaert, Van Dyck, & Van de Poel, 1999; De Groot, Delmaar, & Lupker, 2000; Dijkstra, Grainger, & van Heuven, 1999; Dijkstra, van Heuven, & Grainger, 1998; Dijkstra, Timmermans, & Schriefers, 2000; Dyer, 1973; Guttentag, Haith, Goodman, & Hauch, 1984; Lemhöfer et al., 2008; Nas, 1983). According to this view, bottom-up processing of a printed word proceeds independently of the language to which that word belongs, up to the level of whole-word representations stored in a word-form lexicon that is common to both languages, and possibly beyond. Perhaps the strongest evidence in favor of this non-selective approach to bilingual lexical access was provided by van Heuven, Dijkstra, and Grainger (1998) who showed that word recognition in one language was affected by the target word's orthographic neighbors in the other language. The major conclusion from this and

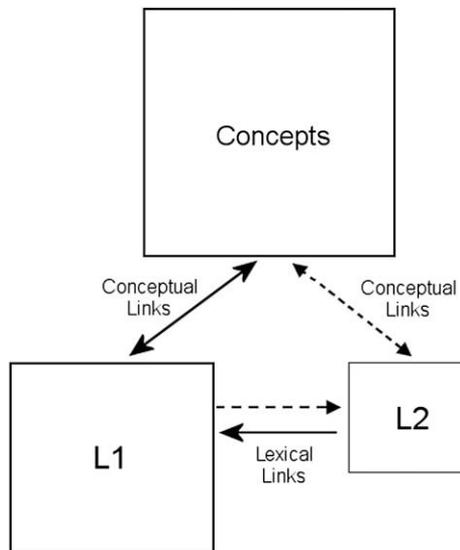


Fig. 1. The revised hierarchical model (Kroll & Stewart, 1994).

related behavioral research is that bilinguals cannot completely stop interference from the non-target language (see Midgley, Holcomb, van Heuven, & Grainger (in press), for further evidence from ERPs). These results counter the predictions of the input-switch mechanism first proposed by Macnamara and Kushnir (1972), according to which the words of the bilingual's two languages are kept apart (in separate stores) and the incoming sensory input is directed to the appropriate set of words as a function of context.

The non-selective access/integrated lexicon view of how languages are represented in the bilingual mind would appear to predict that L2 words should be processed much in the same way as L1 words. Therefore, according to this account, words in a second language learner's L2 should behave like low-frequency words in the L1. This therefore contrasts with the predictions of the RHM (Kroll & Stewart, 1994) that assigns very different mechanisms for processing words in L2 compared with L1. Of course, behavioral results show that performance in L2 is slower and less accurate than performance in L1, but this could be attributed to the frequency characteristics of the words in each language. Exposure to L2 words is generally much lower than exposure to words in L1, especially for beginning bilinguals. Therefore L2 words will have lower subjective frequencies than L1 words, and these differences in subjective frequency could be driving the behavioral effects. Furthermore, L2 words are learned after the majority of L1 words have already been learned, so age-of-acquisition (AoA) could be another factor driving observed differences in performance to L1 and L2 words. Finally, L2 words might be less interconnected to other L2 words (e.g., have smaller orthographic neighborhoods) than L1 words and these differences might also account for differences in performance.

Therefore, one key question that emerges from the above review of the behavioral literature on word recognition in bilinguals is whether L2 words show any processing specificities compared with L1 words that are quantitatively or qualitatively different than those attributable to subjective frequency, AoA or neighborhood density. In order to address this issue, the present study compared ERPs to words in L1 and L2 in lists that are blocked by language, hence avoiding issues of predictability and language-switching (see Chauncey, Grainger, & Holcomb, 2008, for a recent ERP study of language-switching). A number of previous studies have presented words in participants' L1 and L2 and compared various ERP effects such as priming or anomaly detection between languages (e.g., De Bruijn, Dijkstra, Chwilla, & Schriefers, 2001; Hahne & Friederici, 2001; Kerkhofs, Dijkstra, Chwilla, & de Bruijn, 2006; Kotz, 2001; Kotz & Elston-Guttler, 2004; Moreno & Kutas, 2005; Phillips, Klein, Mercier, & de Boysson, 2006; Phillips, Segalowitz, O'Brien, & Yamasaki, 2004; Weber-Fox & Neville, 1996). Surprisingly though, none of these studies has systematically compared ERPs to words in L1 and L2. ERPs would appear to be ideally suited for discerning word level differences between languages, both because of their excellent temporal resolution as well as their ability to differentiate multiple sensory and cognitive influences in a single experiment.

There are also comparatively few ERP studies that have examined second language learners in the process of formal classroom instruction. One notable exception is a study by McLaughlin, Osterhout, and Kim (2004). These authors showed that the N400 component to visually presented items in L2 (French) differentiated between words and pseudowords after only a very brief period of L2 learning (14 h) though semantic priming effects were not observed at this point in L2 learning. This result suggests that ERPs are sensitive to lexical representations laid down in the very initial phases of L2 learning. McLaughlin et al. did not report direct comparisons between L2 and L1, but a study by Alvarez, Grainger, and Holcomb (2003) did directly compare ERPs recorded to Spanish (L2) and English (L1) words in native English speakers enrolled in intermediate university Spanish classes. The finding of most

interest here was that starting as early as 150 ms and continuing on as late as 700 ms there were differences in the time course of ERP effects between L1 and L2. However, because their design included a repetition factor and all of their reported effects of language interacted with this factor, it is not clear whether there were pure differences between languages when participants read words in each language for the first time (i.e., prior to a repetition). Examination of their Figs. 2 and 3 does, however, suggest that L1 words produced somewhat larger N400s than L2 words.

In the current study we examined differences in the ERPs generated by L1 and L2 words for American students of French (Experiment 1), French students of English (Experiment 2) and proficient French–English bilinguals (Experiment 3). Table 1 summarizes the different level of expertise in L1 and L2 of these three groups of participants.

1. Experiment 1 – learners of French

In this experiment ERP language effects for L1 and L2 items were measured during a passive reading for meaning task. Participants were native speakers of American English selected from beginning and intermediate French courses at university and none had had a significant immersion experience. To better match these participants to the native French participants learning English in Experiment 2, an additional criterion for selection was a relatively early (by US educational standards) and ongoing exposure to French.

1.1. Methods

1.1.1. Participants

Twenty-two participants from Tufts University (17 female, mean age = 19.8, SD = 1.7) who were enrolled in French courses were paid for their participation. All were right handed (Edinburgh Handedness Inventory – Oldfield, 1971) and had normal or corrected-to-normal visual acuity with no history of neurological insult or language disability. English was reported to be the first language learned by all participants (L1) and French their primary second language (L2). Participants began their study of French on average at the age of 12.1 years (range 5–18 years, SD = 2.5).

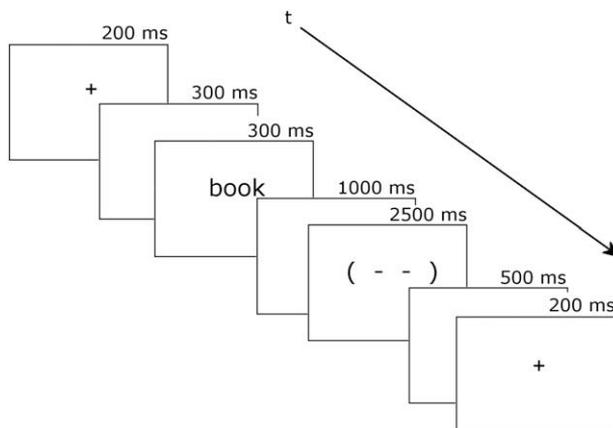


Fig. 2. A critical trial in the L1 block.

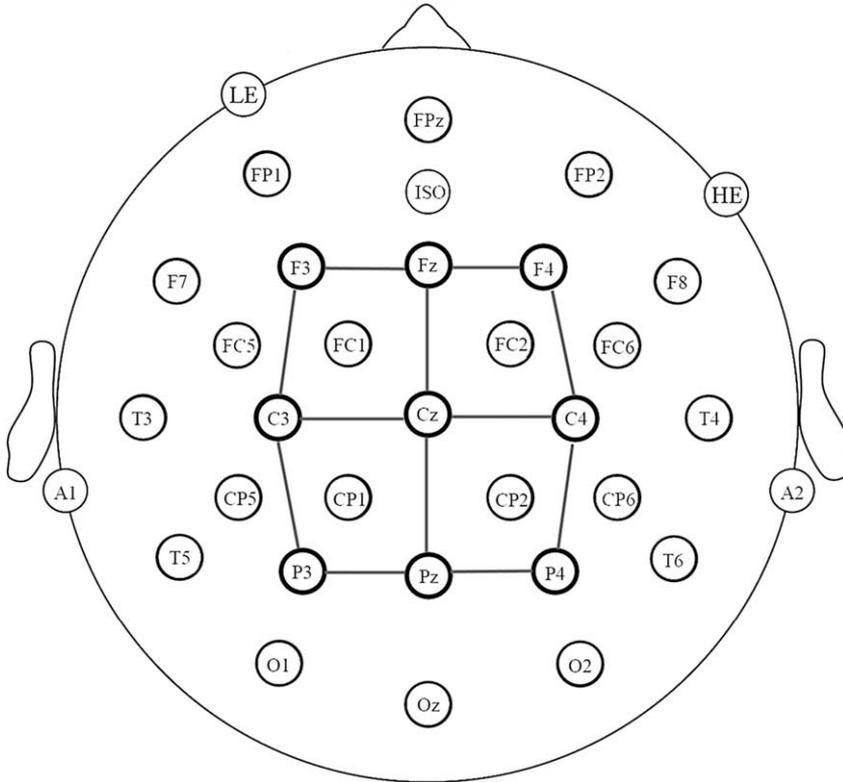


Fig. 3. Electrode montage and electrode analysis sites (connected by lines).

Participants’ auto-evaluation of English and French language skills were surveyed by questionnaire. On a seven-point Likert scale (1 = unable; 7 = expert) participants reported their abilities to read, speak and comprehend English and French as well as how frequently they read in both languages (1 = rarely; 7 = very frequently). The overall average of self-reported language skills in English was 7.0 (SD = 0.0) and in French was 4.2 (SD = 0.9). Our participants reported their average frequency of reading in English as 6.4 (SD = 0.9) and in French as 3.6 (SD = 1.1). See Table 1 for a comparison of participants across the 3 experiments.

1.1.2. Stimuli

The critical stimuli for this experiment were 80 four to seven letter non-cognate morphemically simple English words and their translations into French. The English items had a mean

Table 1

Average self-evaluation on a seven-point scale of overall language skills and reading frequency of the three groups of participants in Experiments 1–3

	Experiment 1		Experiment 2		Experiment 3	
	L1 – E	L2 – F	L1 – F	L2 – E	L1 – F	L2 – E
Language skills	7.0 (0.0)	4.2 (0.9)	6.8 (0.4)	3.9 (1.1)	6.9 (0.3)	5.7 (1.0)
Reading	6.4 (0.9)	3.6 (1.1)	6.4 (1.2)	2.7 (1.3)	6.3 (1.1)	5.8 (1.5)

log frequency (CELEX, 1993) of 1.73 (SD = 0.63, range 0.00–3.09). The French items had a mean log frequency (New, Pallier, Ferrand, & Matos, 2001) of 1.63 (SD = 0.53, range 0.24–2.65). The log frequencies of the items in the two languages were not found to be statistically different ($t(79) = 1.16$, $p = 0.25$). The log frequencies of the translation equivalents correlated highly ($r = 0.71$, $p < 0.01$). The average length of the English items was 4.7 letters (SD = 0.97) while the average length of the French items was 5.5 (SD = 1.09).

Forty animal names were selected as probe items and an additional 200 non-critical words were used as fillers for each language block. These non-critical items and the probe items were, like the critical items, four to seven letters in length (mean length: 5.5 letters for English items (SD = 1.00), 5.8 letters for French items (SD = 0.99)). These items had slightly lower log frequencies than the critical items (average log frequency for English items: 1.23 (SD = 0.52), French: 1.17 (SD = 0.50)).

The 80 English items were divided into two lists of 40 items and each participant saw one of the two lists. The same was done with the French items such that there was no repetition of translation equivalents. That is if a participant saw “tree” in the English block they would not see “*arbre*” in the French block. Both an English block and a French block were presented to each participant in a counterbalanced fashion across participants. The two blocks were thus comprised of 40 animal probes, 40 critical items and 200 fillers. All items in a block were in one language and there was no repetition of translation equivalents across blocks.

1.1.3. Procedure

Animal names served as *probe* items in a go/no-go semantic categorization task in which participants were instructed to rapidly press a single button whenever they detected an animal name. Participants were told to read all other words passively without responding (i.e., critical stimuli did not require an overt response). Stimulus lists were a pseudorandom mixture of critical trials, fillers and probe trials. A practice session was administered before the main experiment to familiarize the participant with the task.

The visual stimuli were presented on a 19" monitor located directly in front of the participant at a distance of approximately 150 cm. Stimuli were displayed at high contrast as white letters on a black background in the Verdana font (letter matrix 20 pixels wide \times 40 pixels tall). Each trial began with a fixation cross followed by a blank screen and then an item. The item was on screen for 300 ms, followed by 1000 ms of blank screen and then a 2500 ms blink symbol (see Fig. 2). The participants were instructed to blink only during this blink symbol. This symbol was followed by 500 ms of blank screen after which the next trial began with a fixation cross.

1.1.4. EEG recording procedure

Participants were seated in a comfortable chair in a sound attenuated darkened room. The electroencephalogram (EEG) was recorded from 29 active tin electrodes held in place on the scalp by an elastic cap (Electrode-Cap International – see Fig. 3). In addition to the 2 scalp sites, additional electrodes were attached to below the left eye (LE, to monitor for vertical eye movement/blinks), to the right of the right eye (VE, to monitor for horizontal eye movements), over the left mastoid bone (A1, reference) and over the right mastoid bone (recorded actively to monitor for differential mastoid activity). All EEG electrode impedances were maintained below 5 k Ω (impedance for eye electrodes was less than 10 k Ω and for the references electrodes less than 2 k Ω). The EEG was amplified by an SA Bioamplifier with a bandpass of 0.01 and 40 Hz and the EEG was continuously sampled at a rate of 200 Hz throughout the experiment.

1.1.5. Data analysis

Averaged ERPs were formed off-line from trials free of ocular and muscular artifact (9% of trials rejected for artifact) and were lowpass filtered at 15 Hz. The approach to data analysis involved the selection of a subset of the 29 scalp sites (see Fig. 3). Average waveforms were formed for the two levels of LANGUAGE (L1 vs. L2), three levels of POSTERIOR–ANTERIOR (posterior, central and anterior) and three levels of LATERALITY (right, midline and left). The main analysis approach involved measuring mean amplitudes in three temporal epochs; 150–300 ms, 300–500 ms, 600–800 ms capturing activity prior, during and after the typical N400. Separate repeated measures analyses of variance (ANOVAs) were used to analyze the data in each of these three epochs.¹ The Geisser and Greenhouse (1959) correction was applied to all repeated measures with more than one degree of freedom in the numerator.

1.2. Results

1.2.1. Visual inspection of ERPs

The ERPs time locked to critical target items are plotted in Fig. 4. Plotted in Fig. 5 are the voltage maps resulting from subtracting L2 from L1 ERPs (these plots reveal the scalp distribution of differences between languages). As can be seen in Fig. 4, early in the waveforms there is a small negativity (sharpest at anterior sites) peaking around 100 ms (N1), this is followed by a prominent positivity peaking near 200 ms (P2). The P2 is visible at all sites, but is larger at more anterior electrodes. Up to this point the ERPs for the two languages are quite similar. However, starting just after the peak of the P2 there are clear differences in the ERPs for the two languages. These can be seen most clearly in the voltage maps in Fig. 5. At posterior sites there are differences on the negativity that starts at about 250 ms and which peaks at about 400 ms (N400). Here the difference appears to be due to a prolonged attenuation of the N400 for L2 compared to L1 items starting as early as 200 ms and lasting through 500 ms. Also evident is a reduction and/or delay in the negativity which peaks at about 300 ms at anterior sites in L1 and at about 450 ms in L2 and a subsequent larger anterior sustained negativity for L2 compared to L1 words.

1.2.2. Analyses of ERP data

1.2.2.1. 150–300 ms epoch. In this epoch there was a main effect of LANGUAGE ($F(1,21) = 8.56$, $p = 0.008$) with L1 items being more negative-going than L2 items.

1.2.2.2. 300–500 ms epoch. In the traditional N400 epoch there were again differences between the languages, however, these effects differed as a function of scalp site (LANGUAGE \times POSTERIOR–ANTERIOR, $F(2,42) = 15.75$, $p = 0.0002$; LANGUAGE \times LATERALITY, $F(2,42) = 5.76$, $p = 0.009$). Examination of Fig. 4 suggests that the LANGUAGE \times POSTERIOR–ANTERIOR interaction reflects that at anterior sites L2 items are slightly only more negative-going than L1 items, but at posterior sites L1 items are substantially more negative-going than L2 items. The LANGUAGE \times LATERALITY interaction indicates that L1 was quite a bit more negative than L2 at midline and right hemisphere sites compared to left hemisphere sites.

¹ We performed a first pass analysis including the factor of order to test for differential effects of which target language block occurred first for all three epochs and for all three Experiments. There were no interactions involving the order and language (all $F_s < 2$). In all of the analyses reported we collapsed across this factor.

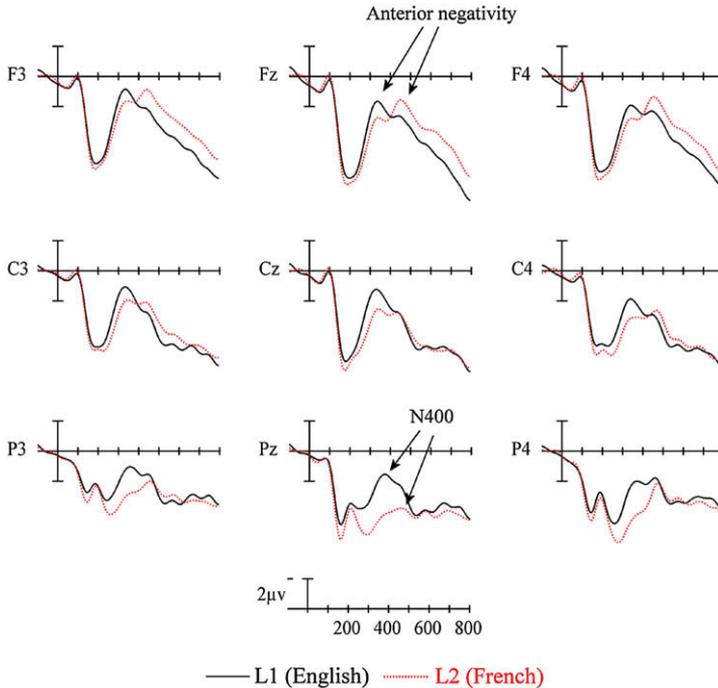


Fig. 4. ERP language effects from nine sites (see Fig. 3 for electrode locations) for native English speakers reading critical words in their L1 (English) and L2 (French) in Experiment 1.

1.2.2.3. 600–800 ms epoch. In the final epoch there were again language effects that differed as a function of scalp site (LANGUAGE \times POSTERIOR–ANTERIOR, $F(2,42) = 11.64$, $p = 0.001$). Fig. 4 suggests that this interaction reflects that L2 was more negative-going than L1 at anterior sites while at posterior sites L1 was still slightly more negative-going than L2.

1.2.3. Behavioral data

Participants detected on average 98.2% (SD = 2.7%) of probes in the L1 block. In the L2 block the participants detected 78.9% of probes (SD = 10.5%). Participants produced false alarms on an average of 0.5 items (SD = 1.1) in L1 (0.2%) and on 7.6 items (SD = 7.8) in L2 (3.8%).

1.3. Discussion

Experiment 1 tested learners of French still at a relatively early point in acquiring their second language. ERPs were time locked to passively read words in L1 (English) and L2 (French). Both languages elicited a similar pattern of early ERP components. However, in the time frame of the N400 component there were clear effects of language dominance. At posterior electrode sites L1 items were associated with larger N400-like negativities starting as early as 200 ms and continuing on through 500 ms. At anterior sites L1 items started off more negative-going than L2 items (between 200 and 400 ms), but after 400 ms L2 items became more negative than L1 items. This latter effect suggests that the anterior negativity produced by words in both languages is delayed by some 150 ms in L2.

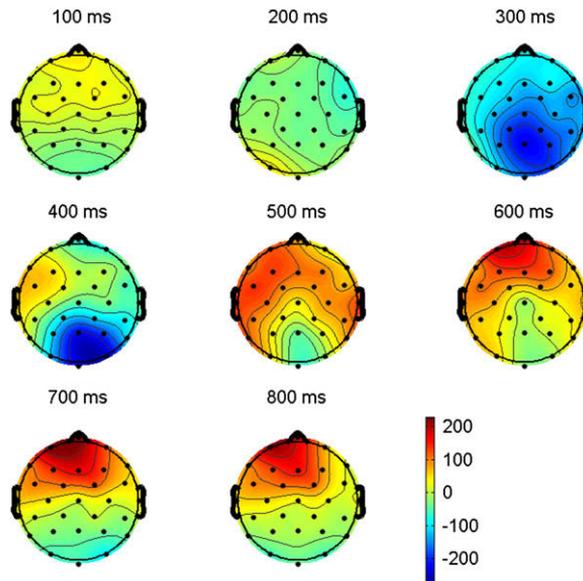


Fig. 5. Voltage maps computed by subtracting ERPs recorded to L2 words (French) from ERPs recorded from L1 words (English) in Experiment 1. Note that computing the subtraction in this direction (L1–L2) results in the larger early posterior negativity from L1 ERPs (see Fig. 4) producing a larger effect at posterior sites (blue areas at 300 and 400 ms) and the larger late anterior negativity from L2 ERPs (see Fig. 4) producing an effect at anterior sites (red areas at 600, 700 and 800 ms).

2. Experiment 2

In Experiment 1, L2 speakers of French showed smaller posterior negativities and delayed/prolonged anterior negativities to words read in their L2 compared to their L1. It seems likely that these effects are due to these participants less competent language status in L2, although it is possible that they are due entirely or in part to inherent differences between French and English words. Neville et al. have demonstrated that there are both similarities and differences in ERP effects for users of different languages (English and ASL) and that some of these effects can be attributed to language competence but that some effects reflect basic differences in the languages themselves (Neville et al., 1997; Neville, Mills, & Lawson, 1992). In order to test if the language effects seen in Experiment 1 are due to specific characteristics of English and French or reflect the different level of participants' competence in English and French, in Experiment 2 we tested native speakers of French that are learners of English in a similar paradigm with different items.

2.1. Methods

2.1.1. Participants

Eighteen participants (14 female, mean age = 22.1, SD = 4.7) were recruited from the University of Provence and paid for their participation. All were right handed (Edinburgh Handedness Inventory – Oldfield, 1971) and had normal or corrected-to-normal visual acuity with no history of neurological insult or language disability. French was reported to be the first language learned by all participants (L1) and English their primary second language (L2).

Participants began their study of English in their sixth year of primary school at approximately the age of 12 years, as is customary in the French school system.

Participants' auto-evaluation of English and French language skills were surveyed by questionnaire. On a seven-point Likert scale (1 = unable; 7 = expert) participants reported their abilities to read, speak and comprehend English and French as well as how frequently they read in both languages (1 = rarely; 7 = very frequently). The overall average of self-reported language skills in French was 6.8 (SD = 0.4) and in English was 3.9 (SD = 1.1). Our participants reported their average frequency of reading in French as 6.4 (SD = 1.2) and in English as 2.7 (SD = 1.3). See Table 1 for a comparison of participants across the 3 experiments.

2.1.2. Stimuli

The critical stimuli for this experiment were 74 four and five letter English words and 74 four and five letter French words. The English items had a mean log frequency (CELEX, 1993) of 1.08 (SD = 0.489, range 0.301–2.158). The French items had a mean log frequency (New et al., 2001) of 1.16 (SD = 0.611, range 0.000–2.615). These log frequencies were not found to be statistically different ($t(73) = 0.654, p = 0.51$). The average length of the English items was 4.34 letters (SD = 0.48) while the average length of the French items was 4.43 (SD = 0.50). The critical stimuli were morphemically simple items. Stimulus lists were formed by mixing the above critical items with probe words which were all members of the semantic category of "body parts" (20% of trials). These probe items were, like the critical items, four and five letters in length (mean length: 4.5 letters (SD = 0.51) for both English and French items). These items had slightly higher log frequencies than the critical items (average log frequency for English items: 1.53 (SD = 0.80), French: 1.69 (SD = 0.61)).

2.1.3. Procedure

The procedure was the same as Experiment 1. Two blocks were presented in counter-balanced order. In the L1 block only French items were presented and in the L2 block only English items were presented.

The laboratory in France was designed to be as similar as possible to the laboratory at Tufts University. The same EEG recording system is used and the same software is used in Experiments 1–3 as well as the same experimenters.

The visual stimuli were presented on a 15" monitor located directly in front of the participant at a distance of approximately 150 cm. Stimuli were displayed at high contrast as white letters on a black background in the Verdana font (letter matrix 30 pixels wide \times 60 pixels tall). All else was as in Experiment 1. See Fig. 2 for a typical L2 trial.

2.1.4. Data analysis

Data analysis was identical to Experiment 1. Trials containing ocular and muscular artifact were excluded from analysis (13% of trials rejected due to artifact).

2.2. Results

2.2.1. Visual inspection of ERPs

The ERPs time locked to critical target items are plotted in Fig. 6. Plotted in Fig. 7 are the voltage maps resulting from subtracting L2 from L1. As can be seen, and similar to Experiment 1, early in the waveforms there is a small negativity (sharpest at anterior sites) peaking around 100 ms (N1). This is followed by a prominent positivity peaking near 200 ms (P2). The P2 is

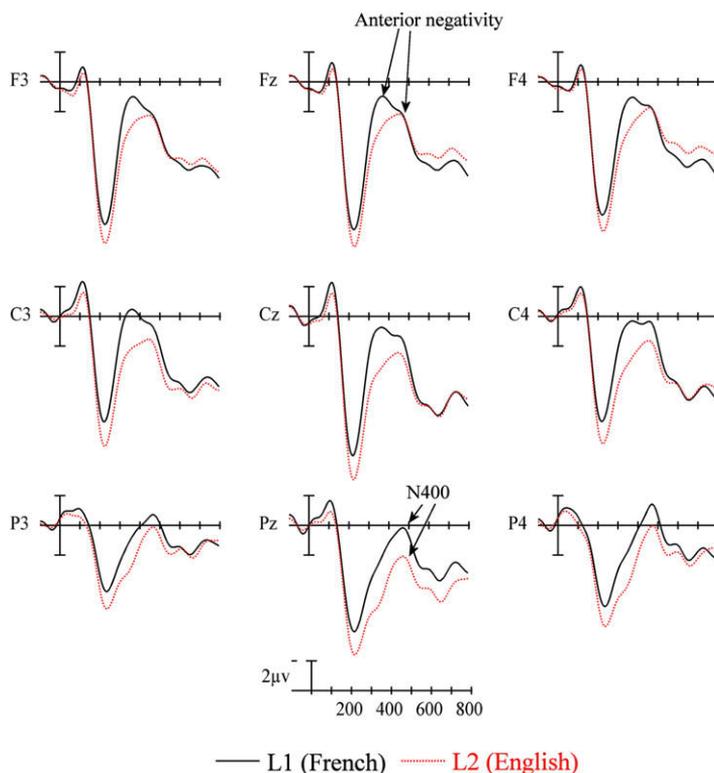


Fig. 6. ERP language effects from nine sites (see Fig. 2 for electrode locations) for native French speakers reading critical words in their L1 (French) and L2 (English) in Experiment 2.

visible at all sites, but is larger at more anterior electrodes. Again, as in Experiment 1, up to the peak of the P2 the ERPs for the two languages are quite similar. However, starting near the peak of the P2 there are clear differences in the ERPs for the two languages. Most evident across the scalp, but most notable at posterior sites there are differences on the negativity starting as early as 200 ms and peaking at about 450 ms (N400). This difference appears to be due to a prolonged attenuation of the N400 for L2 compared to L1 items (see the large central/posterior blue area in Fig. 7 between 200 and 600 ms). Another difference is a reduction in the anterior negativity in L2 compared to L1 especially after 400 ms (the red area in Fig. 7 between 500 and 800 ms).

2.2.2. Analyses of ERP data

2.2.2.1. 150–300 ms epoch. In this epoch there was a main effect of LANGUAGE ($F(1,21) = 7.42$, $p = 0.014$) with L1 items being more negative-going than L2 items.

2.2.2.2. 300–500 ms epoch. In this epoch there were differences between the languages ($F(1,17) = 17.83$, $p = 0.001$) as well as an interaction between LANGUAGE \times POSTERIOR–ANTERIOR \times LATERALITY ($F(4,68) = 5.45$, $p = 0.002$). Fig. 6 suggests that this interaction reflects that the difference between languages (L1 more negative than L2) tended to be larger at more posterior and midline sites.

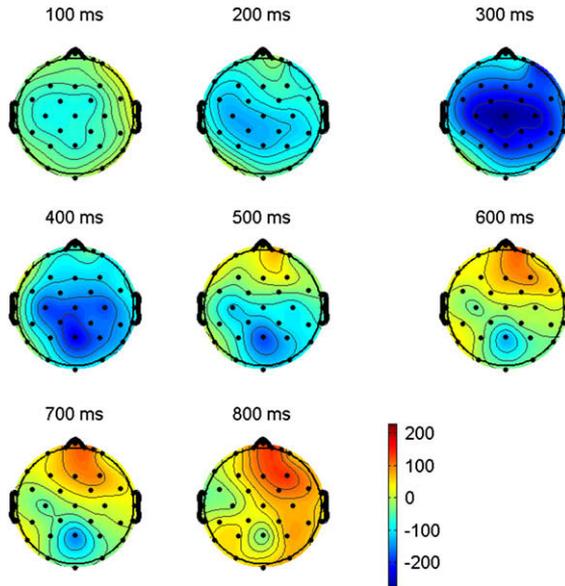


Fig. 7. Voltage maps computed by subtracting ERPs recorded to L2 words (English) from ERPs recorded from L1 words (French) in Experiment 2. Note that as in Fig. 4 there is a large early central/posterior effect resulting from L1 ERPs being more negative-going (blue areas at 200, 300, 400 and 500 ms) and a large late anterior effect resulting from L2 ERPs being more negative at anterior sites (red areas at 500, 600, 700 and 800 ms).

2.2.2.3. *600–800 ms epoch.* In the final epoch there were again language effects that differed as a function of scalp site (LANGUAGE \times POSTERIOR–ANTERIOR \times LATERALITY ($F(4,68) = 8.60$, $p = 0.0002$)). Fig. 6 suggests that at anterior sites, especially along the midline and over the right hemisphere, L2 was more negative-going than L1, while at posterior sites, especially along the midline, L1 was still more negative than L2.

2.2.3. Behavioral data

Participants detected on average 87.0% (SD = 10.9%) of probes in the L1 block. In the L2 block the participants detected 52.9% of probes (SD = 21.7%). Participants produced false alarms on an average of 1.4 items (SD = 1.1) in L1 (1.9%) and on 1.0 items (SD = 1.0) in L2 (1.4%).

2.3. Discussion

Experiment 2 tested French learners of English still at a relatively early point in acquiring their second language. As in Experiment 1, ERPs were time locked to passively read words but now French is the participants' L1 and English their L2. Again, both languages elicited a similar pattern of early ERP components. Just after the peak of the P2, at the beginning of the time frame of the N400 component there were again effects of language dominance. At central and posterior electrode sites L1 items were associated with larger N400-like negativities starting as early as 200 ms and continuing on through 700 ms. At anterior sites L1 items started off more negative-going than L2 items (between 200 and 500 ms) but after 500 ms L2 items became more negative than L1 items especially at midline and right hemisphere sites. This

latter effect is consistent with the possibility that the anterior negativity produced by words in both languages is delayed by some 150 ms in L2. The overall pattern of effects in Experiments 1 and 2 was quite similar suggesting that the observed differences in L1 and L2 are not due to specific attributes of either language, but rather are more general effects of language competence in L1 and L2. This hypothesis is further tested in Experiment 3 with more balanced French–English bilinguals.

3. Experiment 3

In order to provide a further test of the hypothesis that the language effects obtained in Experiments 1 and 2 are due to different levels of proficiency in each language, Experiment 3 tests a group of proficient French–English bilinguals in the same paradigm as Experiment 2.

3.1. Methods

3.1.1. Participants

Twenty participants (13 female, mean age = 23.0, SD = 4.7) were recruited as proficient bilinguals and paid for their participation. All were right handed (Edinburgh Handedness Inventory – Oldfield, 1971) and had normal or corrected-to-normal visual acuity with no history of neurological insult or language disability. French was reported to be the first language learned by all participants (L1) and English their primary second language (L2). Participants began their study of English in their sixth year of primary school at approximately the age of 12 years, as is customary in the French school system.

Participants' auto-evaluation of French and English language skills were surveyed by questionnaire. On a seven-point Likert scale (1 = unable; 7 = expert) participants reported their abilities to read, speak and comprehend English and French as well as how frequently they read in both languages (1 = rarely; 7 = very frequently). The overall average of self-reported language skills in French was 6.9 (SD = 0.3) and in English was 5.7 (SD = 1.0). Our participants reported their average frequency of reading in French as 6.3 (SD = 1.1) and in English as 5.8 (SD = 1.5). See Table 1 for a comparison of participants across the 3 experiments.

3.1.2. Stimuli and procedure

The stimuli and procedure for this experiment were the same as in Experiment 2.

3.1.3. Data analysis

Data analysis was identical to Experiment 1. Trials containing ocular and muscular artifact were excluded from analysis (7% of trials rejected).

3.2. Results

3.2.1. Visual inspection of ERPs

The ERPs time locked to critical target items are plotted in Fig. 8. Plotted in Fig. 9 are the voltage maps resulting from subtracting L2 (English) from L1 (French) ERPs. As can be seen, and similar to Experiments 1 and 2, early in the waveforms there is a small negativity (sharpest at anterior sites) peaking around 100 ms (N1). This is followed by a prominent positivity peaking near 200 ms (P2). In Experiments 1 and 2 there were prominent differences between languages that appeared to begin just after the P2 (Figs. 4 and 6). These effects are not so

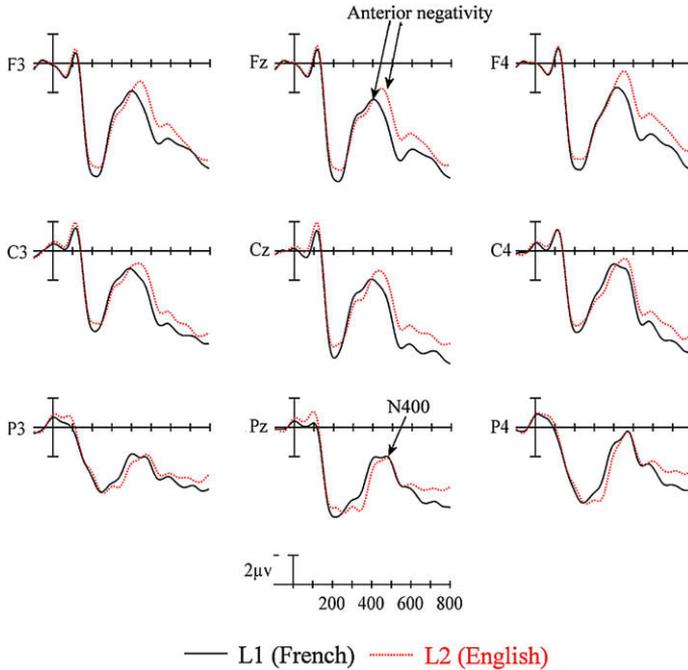


Fig. 8. ERP language effects from nine sites (see Fig. 3 for electrode locations) for native French speakers reading critical words in their L1 (French) and L2 (English) in Experiment 3.

apparent in Experiment 3 (Fig. 8). Still evident, however, is the delay in the anterior negativity which peaks after 350 ms at the frontal sites in L1 and just after 450 ms in L2. Moreover, this negativity appears to now be larger in L2 than L1. Another difference between the ERPs in Experiment 3 and those in Experiments 1 and 2 is that in Experiment 3 there are no longer large differences on the negativity that peaks between 400 and 500 ms (N400) at posterior sites. Here the peak of the negativity is roughly equivalent for the two languages. It is only in the epoch following the N400 where there are small differences between languages (L2 more negative than L1).

3.2.2. Analyses of ERP data

3.2.2.1. 150–300 ms epoch. Consistent with the visual observation of Fig. 8 there were no significant effects of LANGUAGE in this epoch.

3.2.2.2. 300–500 ms epoch. In this epoch there were differences between the languages as a function of scalp site (LANGUAGE \times POSTERIOR–ANTERIOR ($F(2,38) = 5.21, p = 0.022$)). Examination of Fig. 8 suggests that this interaction was sensitive to the difference between languages from the front to the back of the head.

3.2.2.3. 600–800 ms epoch. In the final epoch there were again language effects that differed as a function of scalp site (LANGUAGE \times LATERALITY ($F(2,38) = 4.20, p = 0.031$)). Fig. 8 suggests that differences between the languages were greater over midline and right hemisphere sites.

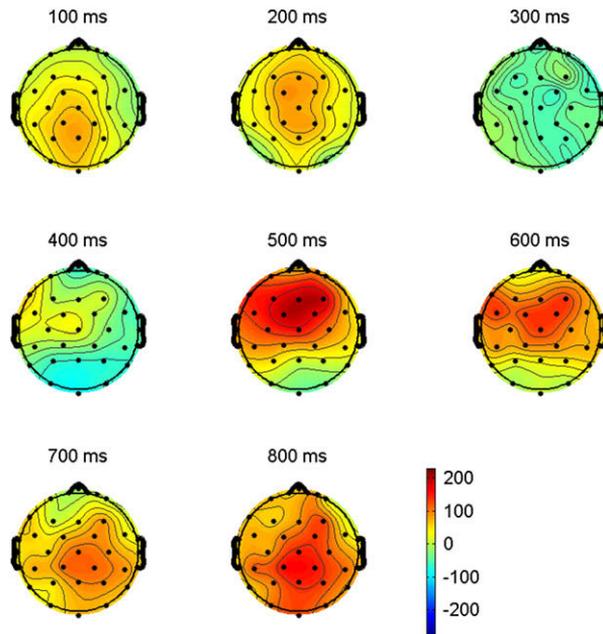


Fig. 9. Voltage maps computed by subtracting ERPs recorded to L2 words (English) from ERPs recorded from L1 words (French) in Experiment 3. Note that the early posterior is much smaller than in Figs. 5 and 7 reflecting the fact that L1 ERPs are only slightly more negative-going than L2 ERPs (pale blue area at 400 ms), but that there is still a large late anterior effect resulting from L2 ERPs being more negative at anterior sites (red areas at 500, 600, 700 and 800 ms).

3.2.3. Behavioral data

Participants detected on average 92.8% (SD = 5.7%) of probes in the L1 block. In the L2 block the participants detected 93.0% of probes (SD = 10.7%). Participants produced false alarms on an average of 1.8 items (SD = 1.1) in L1 (2.4%) and on 2.2 items (SD = 3.9) in L2 (2.9%).

3.3. Discussion

This experiment examined proficient bilinguals in both their L1 (French) and L2 (English). As in Experiments 1 and 2, ERPs were time locked to passively read words in both languages. Both languages elicited a similar pattern of early ERP components and this similarity in early effects extended through 300 ms. It was not until the middle portion of the N400 (around 350 ms) that the somewhat larger posterior N400 for L1 items could be seen in this Experiment. At anterior sites the negativity which peaked around 350 ms for L1 was clearly delayed in L2, peaking around 50–100 ms later. Also the late negativity that follows the N400 especially at anterior sites was larger in L2 than in L1. However, the latency shift in anterior N400 was certainly smaller than seen in the less proficient participants of Experiments 1 and 2.

4. General discussion

Across three experiments we compared ERPs to words passively read for meaning in bilingual participants who were comparatively late learners of their L2 (after age 12). In Experiment 1, participants were native speakers of English (L1 English) and were in the process of learning

French (L2). In Experiment 2 participants were native speakers of French (L1 French) and were in the process of learning English (L2). In Experiment 3 participants were again native speakers of French (L1 French) but were also competent speakers of English (L2).

In all three experiments words in L1 and L2 produced a series of early ERP components (N1 and P2) prior to 200 ms that were quite similar in morphology (shape) scalp distribution and amplitude (see Figs. 4, 6 and 8). This is not surprising because these early exogenous ERP components are believed to reflect sensory and perceptual characteristics of the eliciting stimulus (Luck, 2005) and in the current study we carefully matched these characteristics across the words used in the two languages (e.g., no French words with accents were used).

The most prominent feature of ERPs to passively read content words after 200 ms is the N400 (Osterhout & Holcomb, 1995). The N400 in single word paradigms is believed to be sensitive to word processing as early as the word-form/meaning interface (e.g., Holcomb, O'Rourke, & Grainger, 2002) and therefore is likely to be the first ERP component that would carry any language effects in a passive reading paradigm. As expected words presented in L1 were associated with a large broadly distributed N400, and this was true for both English and French L1 speakers. Turning first to the two less competent L2 groups (Experiment 1 and 2), perhaps most notable is the observation that both experiments produced a very similar pattern of L1–L2 ERP effects. Because the language of L1 and L2 were reversed across the experiments, this suggests that language-specific differences were not at the root of the effects observed. In both experiments the L2 N400 was attenuated relative to L1 especially at centro-posterior sites. However, at anterior sites while the N400 started out smaller in L2 than L1, this difference appeared to be due in part to the anterior N400 being somewhat delayed in L2 compared to L1. In fact, looking at the epoch just after the traditional N400 at anterior sites, L2 words are more negative-going than L1 words and this difference continued all the way to the end of the recording epoch. Moreno et al (2005) also found delayed N400s as well as a posterior N400 negativity in competent L2 speakers in a sentence reading task. They found that this late negativity was also associated with language dominance, with ERPs to the less dominant language having the larger late negativity.

Experiment 3 examined the impact of L2 proficiency level on these language differences by testing participants with roughly comparable proficiency in their L1 and L2 (in Experiments 1 and 2 participants were significantly less competent in L2), but who had begun learning their L2 at a similarly late age. In this experiment the difference between L1 and L2 in centro-posterior N400 amplitude that was seen previously in Experiments 1 and 2 was virtually non-existent. The L2 words in Experiment 3 generated a centro-posterior N400 of similar size to words in L1. This pattern of similar N400 amplitude for L1 and L2 has been reported in previous studies of competent L2 speakers in other language paradigms (e.g., Moreno et al., 2005; Neville et al., 1992) and is consistent with the hypothesis that the posterior N400 effect reflects competence in L2, less competent users producing smaller N400s. The anterior N400 latency shift seen clearly in Experiments 1 and 2 was still evident in Experiment 3, but the latency difference was much smaller. This would suggest that the anterior N400 latency shift reflects competence in L2, but contrary to the posterior N400 effect continues to reflect differences in L1 and L2 processing even in relatively proficient bilinguals.

These large differences in the ERPs generated by words in L1 and L2 and their modulation by proficiency in L2 are in line with the hypothesis that word recognition in L2 involves distinct mechanisms compared with the first language, at least in the relatively early phases of L2 acquisition in late learners of L2. One specific model, the RHM (Kroll & Stewart, 1994), predicted such differences in L1 and L2 word recognition over and above possible differences

due to subjective frequency, AoA, and other correlated factors. In the RHM, it is the process of translation from an L2 word-form to its equivalent in L1 that is the key distinguishing characteristic of L2 vocabulary acquisition in beginning bilinguals. That is, word recognition in L2 is achieved mainly by translation of the L2 word-form into its equivalent word-form in L1, which then leads to the appropriate semantic activation. The delayed anterior N400 latency to L2 words might well reflect this specific process of L2 word recognition, that is inevitably delayed compared with L1 word recognition for which the translation route is non-dominant (see Fig. 1). The fact that an albeit smaller latency shift was still observed in the proficient bilinguals of Experiment 3 suggests that the translation route is still operational in these participants, but to a lesser degree.

Do these results therefore force us to reject an integrated lexicon account of bilingual lexical representation, such as implemented in the bilingual interactive-activation model (BIA-model, Grainger & Dijkstra, 1992; van Heuven et al., 1998)? One way of saving the BIA-model is to assume that it better reflects processing in relatively proficient bilinguals, while the RHM is a better model of lexical processing in beginning bilinguals. The idea would be that during L2 acquisition, the L2–L1 translation route of the RHM would be gradually replaced by L2 lexical representations that become part of an integrated network along with L1 representations. This could be achieved by gradually weakening the L2–L1 connections between translation equivalents, and gradually strengthening the connections between L2 word-forms and semantics. Thus as proficiency develops in L2, lexical processing in L2 becomes more and more akin to lexical processing in L1.

But why is the posterior N400 smaller for L2 items in less competent speakers? Indeed, one might have expected a greater N400 amplitude to L2 words compared with L1 words, given that N400 amplitude is often associated with processing effort or cost. One possibility is that in less competent speakers exposure to L2 words is generally much lower than exposure to words in L1. Therefore L2 words will have lower subjective frequencies than L1 words. Differences in subjective frequency could be the source of the smaller posterior N400 for L2 items. However, the effects of word frequency on the N400 have typically been reported to be just the opposite of those obtained here (i.e., larger N400s for less frequent words – Van Petten & Kutas, 1990), so a mechanism like subjective word frequency is an unlikely source.

Another possible explanation for these effects is that they reflect differential activity in a mechanism similar to the age-of-acquisition effect for words in L1. That is, that words acquired younger in life are afforded some special privilege (Zevin & Seidenberg, 2004). Plotted in Fig. 10 are ERPs recorded to words in the same passive reading paradigm of the current study in adult monolingual speakers of English. These waves have been segregated according to whether they were learned before or after the age of eight years. As can be seen, the very small effects seen at anterior sites for the two word categories are in the opposite direction to those predicted by the late negative effects seen for words in L2 in Figs. 4, 6 and 8. So AoA appears to be an unlikely source of the posterior N400 language effect.

Finally, another possibility is that L2 words, especially in a less competent speaker may be less interconnected to other L2 words – that is, on average, they would tend to have smaller orthographic neighborhoods. At least one previous study has shown that words from dense orthographic neighborhoods have larger N400s than words from sparse neighborhoods (Holcomb et al., 2002). Therefore, the larger posterior N400 amplitude to L1 words might be a reflection of the greater number of other words that are co-activated during target word recognition. A similar argument could be made at the level of semantic representations, with the semantic representations of L1 words being more richly interconnected within the semantic network than L2 words. Such an analysis would fit with one account of the effects of

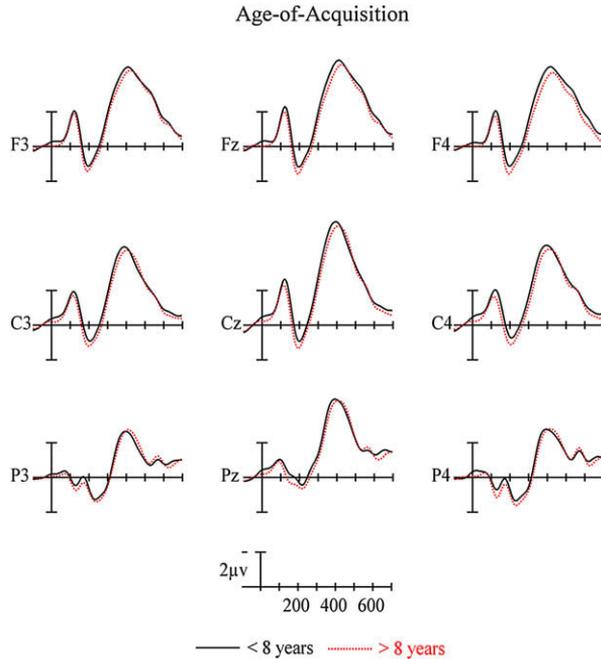


Fig. 10. ERPs in the same passive reading task and at the same 9 scalp sites as the current study. ERPs are based on 250 words divided at the median on age-of-acquisition ratings.

concreteness on N400 amplitude (concrete words generate larger N400 amplitudes than abstract words, e.g., Kounios & Holcomb, 1994) according to which this would be due to the greater semantic interconnectivity for concrete words.

5. Conclusions

The present study provides evidence that the processing of L1 and L2 words in late L2 learners diverge in two distinct ways as reflected in the ERP waveforms generated by these words during silent reading for meaning. An anterior part of the N400 component showed a distinct latency shift with L2 amplitudes peaking later than L1 amplitudes, and although the latency shift was still present in proficient bilinguals, it was smaller in magnitude. This latency shift might well reflect differences in processing difficulty associated with words in L1 and L2. The posterior part of the N400 revealed larger amplitudes to L1 compared with L2 words in our low-proficiency participants, and very little difference in the proficient bilinguals. It is suggested that these amplitude differences might reflect a lower level of interconnectivity of L2 lexical and semantic representations in beginning bilinguals, that disappears with increasing competence in L2 and greater integration of L2 words in a common lexical-semantic network.

References

- Alvarez, R., Grainger, J., & Holcomb, P. J. (2003). Accessing word meaning in primary and secondary languages: a comparison using event-related brain potentials. *Brain & Language*, 87, 290–304.

- Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: some limitations of a language-selective access. *Journal of Memory and Language*, 26(6), 658–672.
- Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 137–148.
- CELEX English database (Release E25). (1993). Available from. <http://www.mpi.nl/world/celex>.
- Chauncey, K., Grainger, J., & Holcomb, P. J. (2008). Code-switching effects in bilingual word recognition: a masked priming study with event-related potentials. *Brain & Language*.
- De Bruijn, E. R. A., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition*, 4, 155–168.
- De Groot, A. M., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 53A(2), 397–428.
- Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: the neglected role of phonology. *Journal of Memory and Language*, 41, 496–518.
- Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, 38(3–4), 177–196.
- Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: effects of task demands on interlingual homograph recognition. *Journal of Memory and Language*, 42(4), 445–464.
- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1(2), 106–120.
- Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 207–220). Amsterdam: North-Holland.
- Guttentag, R. E., Haith, M. M., Goodman, G. S., & Hauch, J. (1984). Semantic processing of unattended words by bilinguals: a test of the input switch mechanism. *Journal of Verbal Learning & Verbal Behavior*, 23(2), 178–188.
- Hahne, A., & Friederici, A. D. (2001). Processing a second language: late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458–483.
- Holcomb, P. J., O'Rourke, T., & Grainger, J. (2002). An event-related brain potential study of orthographic similarity. *Journal of Cognitive Neuroscience*, 14, 938–950.
- Kerkhofs, R., Dijkstra, T., Chwilla, D. J., & de Bruijn, E. R. A. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research*, 1068, 170–183.
- Kotz, S. A. (2001). Neurolinguistic evidence for bilingual language representation: a comparison of reaction times and event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 143–154.
- Kotz, S. A., & Elston-Guttler, E. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, 17, 215–235.
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: event-related brain potential evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 804–823.
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: a mega-study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 12–31.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Macnamara, J., & Kushnir, S. L. (1972). Linguistic independence of bilinguals: the input switch. *Journal of Verbal Learning & Verbal Behavior*, 10(5).
- McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, 7(7), 703–704.
- Midgley, K. J., Holcomb, P. J., van Heuven, W. J. B., & Grainger, J. An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research*, in press.
- Moreno, E., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, 22, 205–220.
- Nas, G. (1983). Visual word recognition in bilinguals: evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning & Verbal Behavior*, 22(5), 526–534.

- Neville, H. J., Coffey, S. A., Lawson, D. S., & Fischer, A. (1997). Neural systems mediating American sign language: effects of sensory experience and age of acquisition. *Brain & Language*, *57*(3), 285–308.
- Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*(3), 244–258.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: Lexique. *L'Année Psychologique*, *101*, 447–462.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.
- Osterhout, L., & Holcomb, P. J. (1995). The electrophysiology of language comprehension. In M. D. Rugg, & M. G. H. Coles (Eds.), *Electrophysiology of mind* (pp. 171–215). Oxford University Press.
- Phillips, N. A., Klein, D., Mercier, J., & de Boysson, C. (2006). ERP measures of auditory word repetition and translation priming in bilinguals. *Brain Research*, *1125*, 116–131.
- Phillips, N. A., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, *17*, 237–262.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393.
- 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.
- Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-acquisition effects in reading aloud: test of cumulative frequency and frequency trajectory. *Memory & Cognition*, *32*, 31–38.