PERGAMON

Journal of Neurolinguistics 17 (2004) 237–262

Journal of NEUROLINGUISTICS

www.elsevier.com/locate/jneuroling

Semantic priming in a first and second language: evidence from reaction time variability and event-related brain potentials

Natalie A. Phillips[a,b,c,*], Norman Segalowitz[a,d], Irena O'Brien[d,e], Naomi Yamasaki[a,d]

[a]*Department of Psychology, Concordia University, Montréal, Canada*
[b]*Centre for Research in Human Development, Montréal, Canada*
[c]*Lady Davis Institute for Medical Research, SMBD-Jewish General Hospital, Montréal, Canada*
[d]*Centre for the Study of Learning and Performance, Montréal, Canada*
[e]*Université du Québec à Montréal, Canada*

Abstract

We investigated individual differences in second language (L2) proficiency by looking at the efficiency or automaticity of semantic priming using behavioural and event-related brain potential (ERP) measures. In Experiment 1, 37 first language (L1) English speakers varying in second language (L2, French) proficiency made living/non-living judgments to English and French nouns in lists blocked by language. Sixty critical words were each presented twice, once primed by a semantic associate in the preceding trial (e.g. ADULT, CHILD) and once unprimed (e.g. RABBIT, CHILD). Measures of response time (RT) and intra-individual variability in response time (coefficient of variation, CV) were obtained. The CV provided an index of processing efficiency that has been related to automaticity. Participants performed faster and with lower CVs (i.e. with greater efficiency) in L1 than L2, and the more highly proficient bilinguals had lower CVs than the less proficient bilinguals. Experiment 2 replicated these results with 29 participants and provided an electrical brain activity measure of processing efficiency using the N400 ERP. The similar pattern of results obtained between the behavioural and N400 ERP CV measures supported the idea that the CV measure of electrical brain activity can provide useful information about the automaticity or efficiency of cognitive processing.

Keywords: Automaticity; Event-related brain potentials; Response variability; Semantic priming; Bilingualism; N400

---

* Corresponding author. Address: Department of Psychology, Concordia University, 7141 Sherbrooke Street West, Montreal, Québec, Canada H4B 1R6.

E-mail addresses: natalie.phillips@concordia.ca (N.A. Phillips), segalow@vax2.concordia.ca (N. Segalowitz).

0911-6044/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0911-6044(03)00055-1

A concept central to understanding individual differences in second language (L2) ability is that people can differ in terms of language processing efficiency or *automaticity* in the cognitive mechanisms underlying L2 functioning. Automaticity has been operationalized behaviourally in the bilingualism literature and in the general cognitive literature in a number of different ways, including in terms of processing speed (Lambert, 1955; Magiste, 1986), ballistic or unstoppable processing (Favreau & Segalowitz, 1983; Neely, 1977; Tzelgov, Henik, Sneg, & Baruch, 1996), effortless processing, and unconscious processing, among others (for bilingualism-related reviews see DeKeyser (2001), Segalowitz (2003) and Segalowitz and Hulstijn (2003)). However, there has been little research on the brain activity which might support automatic processing. In this paper we attempt to operationalize automaticity *electrophysiologically* within the context of studies of individual differences. The present study investigates one promising avenue in this regard, namely the relationship between behavioural response variability and event-related potential (ERP) variability.

Segalowitz and Segalowitz (1993) pointed out that individual differences in L2 proficiency could be reflected in faster responding for at least two different reasons. One reason may be that there are simple differences in the general speed of operation of underlying component processes, including differences in the speed of non-automatic or controlled processing. A second reason is that faster responding can be due to differences in the underlying structure of cognitive processing; that is, the *organization* of processes in one individual may result in more automatic—and hence faster—processing than in another individual, over and above any differences there might be between the two individuals in general speed of processing. Segalowitz and Segalowitz (1993) proposed an operational definition of automaticity that permitted one to experimentally distinguish between these two possibilities, that is, between individual differences due to differences in the structure of processing versus differences in simple speed of processing. This involves examining the *variability* of reaction time (RT) in cognitive tasks, and more specifically, the coefficient of variation (CV) of the RT, defined as an individual's standard deviation of response time divided by that person's mean response time.

Segalowitz (2003) and Segalowitz and Segalowitz, (1993) reasoned that when individual differences reflect only differences in speed of processing, then variability of the RT should differ, at most, proportionally to the change in RT itself. That is, if Subject X performs a task, on average, in half the time it takes Subject Y, and if they both carry out the cognitive operations in exactly the same way, then X's SD should be at best half of Y's SD. Put another way, the ratio SD/RT (or the CV) should be the same for the two individuals. On the other hand, X might perform faster than Y because X uses different and more efficient cognitive processing (e.g. X may access L2 meanings directly whereas Y does so via a translation process). That is, X accomplishes the task using fewer slow and temporally variable controlled processes such as those involved in self-monitoring, error correction, resolving signal-to-noise processing problems, etc. In this case, X will not only perform more quickly than Y, but the variability of his/her performance will be more than proportionally reduced, resulting in a change in the CV because change is mostly with the slow, highly variable component processes rather than with all components equally. In this sense, a significant difference in CV allows one to reject the null hypothesis that it is simply processing speed differences that account for the two differing levels of performance. Thus, the CV (the variability normalized to one millisecond of response time) indexes the *efficiency* of processing, which can be related to automaticity (Segalowitz & Segalowitz, 1993). Segalowitz and colleagues and others report a number of studies with tasks involving L2 and other stimuli supporting this CV interpretation of automaticity (Segalowitz and Freed, in press; Segalowitz, Poulsen, & Segalowitz, 1999; Segalowitz & Segalowitz, 1993; Segalowitz, Watson, & Segalowitz, 1995; Segalowitz, Segalowitz, & Wood, 1998; Vigneau, 1998). It is in this sense of efficient processing, reflected in low intra-individual variability, that we use the term 'automaticity' in this report. One goal of the present study was to further investigate the utility of the CV for understanding individual differences in L2 proficiency.

A potentially useful way of investigating the nature of individual differences in L2 proficiency is through the use of semantic priming tasks. In brief, semantic priming refers to the facilitation (e.g. faster RT during a lexical decision task) in the processing of a word (e.g. *dog*), which has been preceded by a related word (e.g. *cat*) versus an unrelated word (e.g. *table*). Although several possible mechanisms may underlie this phenomenon, two of the more commonly considered processes are automatic spread of activation through interconnected conceptual nodes or the generation of attentionally mediated expectancies (cf., Neely, 1991). Woltz (1999), in a general review of individual differences and priming, reported evidence to show that semantic priming, as opposed to repetition priming (priming due to an earlier occurrence of the exact same stimulus), correlates significantly with reading ability, because of the overlap in demands of semantic priming tasks and reading comprehension and in contrast to the more purely lexical effects underlying repetition priming. Presumably there is similar overlap in demands with L2 comprehension. Woltz concluded that 'priming differences may potentially explain learning variance that is unaccounted for by more commonly investigated cognitive ability constructs' (p. 153). Semantic tasks are also promising to use because there is a well-established literature on the electrophysiological correlates of semantic processing. For these reasons we made use of a semantic priming paradigm in the present study.

In bilingualism research, semantic priming studies have been principally concerned with cross-language rather than within-language priming, because of the way cross-language priming can shed light on how a bilingual's two mental lexicons are organized (Jiang & Forster, 2001; Kroll & De Groot, 1997; Kroll & Tokowicz, 2001). Those studies which have looked at within-language priming have generally found that more proficient bilinguals show greater facilitation effects with semantic priming than do less proficient bilinguals (Favreau & Segalowitz, 1983; Frenck-Mestre & Prince, 1997; Vasos, 1983, described in Segalowitz (1986)). Favreau and Segalowitz (1983) also demonstrated that these greater facilitation effects can reflect more automatic (more unstoppable or ballistic) activation of word meaning.

To our knowledge, studies of L2 within-language semantic priming have typically used lexical decision tasks (judging whether a letter string is a real word) or same/different judgments. None has used semantic classification (e.g. judging whether a target refers to something living or non-living), and none has looked at RT variability as an index of the efficiency or automaticity of performance. Semantic classification, in contrast to lexical decision or same-different judgments, would appear to involve task demands that overlap more closely with those of real world use of a second language. This is because people typically deal with words in terms of whether they mean one thing or another, not whether they are real words or nonwords. A second goal of the present study, therefore, was to investigate individual differences in the CV within the context of L2 semantic priming in a semantic classification task.

We also chose to examine these behavioural approaches to processing in conjunction with electrical brain activity associated with semantic processing in L1 and L2. Recordings of electrical brain activity (event-related brain potentials or ERPs) provide on-line measurement of cognitive processing. ERPs reflect voltage variations in electrical brain activity in response to a stimulus or cognitive process and can be extracted from the electroencephalogram (EEG) via signal averaging. ERP components are typically identified by their polarity (positive or negative), latency (occurrence after the eliciting stimulus, in ms), amplitude (in μV), and topographical distribution across the scalp (Coles & Rugg, 1995). The N400 ERP component has proven to be a very useful measure of on-line language processing. First described by Kutas and Hillyard (1980), the N400 is a centro-parietally distributed negative-going component, which peaks at approximately 400 ms after stimulus onset.

The amplitude of the N400 varies *inversely* with the amount of semantic activation a word has in memory; that is, N400 amplitude is reduced when a word is preceded by a semantically related context. This finding is robust and has been observed in a variety of contexts, including word-pair priming (Brown & Hagoort, 1993) and written (Connolly, Phillips, & Forbes, 1995; Kutas & Van Petten, 1988) and spoken (Connolly & Phillips, 1994) sentence processing. Despite the lack of unanimity as to whether the N400 reflects aspects of relatively early automatic lexical/semantic processes (Deacon, Hewitt, Yang, & Nagata, 2000; Kiefer, 2002; Kutas & Hillyard, 1989) or later post-lexical processes (Brown & Hagoort, 1993; Chwilla, Brown, & Hagoort, 1995; Holcomb, 1993), it is clear that N400 amplitude is modulated by semantic relations between words.

The number of studies investigating ERP measures of bilingual language processing is beginning to grow. Some researchers have investigated the processing of syntactic and/or semantic violations within sentences (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Hahne & Friederici, 2001; Moreno, Federmeier, & Kutas, 2002; Weber-Fox & Neville, 1996) or the consequences of switching between languages (Jackson et al., 2001; Moreno et al., 2002). Relatively fewer have looked at priming using word pairs or lists. Using a continuous lexical decision task, Kotz (2001) found that N400 priming effects were equivalent in L1 and L2 in early fluent Spanish-English bilinguals. de Bruijn, Dijkstra, Chwilla, and Schriefers (2001) also found equivalent priming for interlingual homographs regardless of whether the initial context was L1 or L2.

To our knowledge, no studies have employed ERP measures as a means for examining intra-individual variation in L2 fluency. It has been argued that ERPs are subject to large inter-subject variability which may limit their clinical utility (Polich & Herbst, 2000).[1] We sought to determine whether intra-subject variability could in fact be used as a *dependent* measure of the efficiency or automaticity of cognitive processing in one's first and second language by examining the CV of the N400 to obtain an electrophysiological brain measure of automaticity.

In the experiments presented here, we used a sequential list requiring a semantic decision on each word. This task should be sensitive to automatic spread of activation and should be a useful paradigm with which to examine ERP priming effects in the L1 and L2. Thus, a third goal of the present study was to investigate whether and how aspects of the N400 ERP component might reflect the efficiency or automaticity of processing, along with RT and CV-based behavioural measures.

Two experiments are reported below using a speeded semantic classification (living/non-living) in which target words were either semantically primed by the immediately preceding target stimulus (a high semantic associate) or unprimed (a non-associate). The determination of subjects' L2 proficiency levels was based on RT performance on baseline trials intermixed with the primed and unprimed trials. In Experiment 1, the relationship between CV on primed and unprimed trials and L2 proficiency was examined. Experiment 2 involved a replication of Experiment 1 plus the collection of ERP measures.

1. Experiment 1

The goal of this first experiment was to demonstrate basic priming effects with the present set of stimuli and to establish if expected patterns would obtain concerning second language (L2) and first language (L1) coefficients of variation (CV) of response time (RT) in more proficient versus less proficient bilinguals.

2. Method

2.1. Participants

Thirty-seven students (10 male, 27 female; median age = 22 years; mean age = 23.11 years; SD = 4.99; range = 19–46 years), native speakers of English (L1) who varied with respect to their proficiency in speaking French (L2), were recruited from Concordia University. The participants received course credit or payment for their participation. Important information regarding the participants' history and use of both languages is summarized in Table 1.

2.2. Stimuli

In each language, English (L1) and French (L2), the set of critical experimental words were 60 animate and inanimate nouns, presented twice each. In the presentation list, each critical word was preceded once by a semantic associate

---

[1] Polich and Herbst (2000) examined the CV of the P300 waveform as a measure of *inter*-subject variability.

Table 1
Participant characteristics from Experiments 1 and 2 as a function of Proficiency Group

Proficiency Group	Experiment 1		Experiment 2	
	High ($n = 19$)	Low ($n = 18$)	High ($n = 15$)	Low ($n = 14$)
Sex	13 F, 6 M	14 F, 4 M	9 F, 6 M	12 F, 2 M
Parent with French as L1: n	7	2	3	3
Age (year) began to learn L2 (SD)	5.11 (2.26)	6.94 (4.47)	5.47 (2.51)	7.31 (4.35)
French spoken at home: *n*	2	1	1	3
Attended French school: *n*				
Primary	10	5	6	6
Secondary	7	4	9	3
Post-secondary	4	0	0	1
Ability: self-rating[a] (SD)				
Speaking L2	3.79 (0.79)	3.22 (0.73)	3.87 (0.83)	3.14 (0.53)
Reading L2	4.11 (0.66)	3.39 (0.92)	4.20 (0.68)	3.43 (0.65)
Writing L2	3.37 (0.76)	2.89 (0.68)	3.40 (0.63)	3.07 (0.62)
Time spent: self-rating[b] (SD)				
Speaking L2	3.21 (1.23)	2.94 (1.21)	3.67 (1.05)	2.92 (0.95)
Reading L2	2.21 (1.03)	1.94 (0.94)	2.73 (0.70)	2.38 (0.77)
Writing L2	1.37 (1.01)	1.22 (0.43)	2.13 (0.92)	1.69 (0.85)

[a] Self-rated ability on a five point Likert-type scale, with 1 being no ability at all and 5 being native-like ability. All participants rated their L1 abilities at or near 5. Participants rated their L1 ability on all three measures significantly higher than the corresponding L2 measures (Z between -3.49 and -4.02, $p < 0.001$).

[b] Number of times per week spent on language activities rated by participant on a five-point Likert-type scale, with 1 being never/almost never and 5 being main language used. All participants reported their time spent on L1 activities at or near 5. Participants rated time spent on all three L1 activities significantly higher than the corresponding L2 activities (Z between -2.97 and -3.49, $p < 0.003$).

(e.g. ADULT–CHILD) and once by a low semantic associate (e.g. RABBIT–CHILD). With the exception of one case (MILK–COW and LETTER–COW), the preceding word was drawn from the same animacy category (living or non-living) as the target word. The concreteness, familiarity, and written frequency values (Coltheart, 1981) for the critical English words are presented in Table 2. The semantic associates used on primed trials were selected for the strength of their forward association with the critical experimental words ($M = 0.454$, $SD = 0.213$; Nelson, McEvoy, & Schreiber, 1998). The words used as ‘non-associates’ for unprimed trials were selected for not appearing as associates to the critical experimental words in the association norms (Nelson et al., 1998). Following the testing session, 24 of the participants rated the strength of the associations between the critical target words and the preceding associate or non-associate on a 5-point Likert scale (0–4) and found it to be significantly greater for semantic associates ($M = 3.44$, $SD = 0.38$) than for non-associates ($M = 0.47$, $SD = 0.33$; $t(118) = 45.86$, $p < 0.001$). The French words were translations of the English words and the same semantic associates and non-associates were used in the French list. Since French was a second language for all the participants, it was reasoned that the association values

Table 2
Concreteness, familiarity, and frequency values for the English critical and baseline words used in Experiments 1 and 2

Words	Concreteness		Familiarity		Frequency	
	M	SD	M	SD	M	SD
Critical words	586	26	580	46	117	178
Baseline words	590	37	543	43	40	55

Values are from the MRC Psycholinguistic Database, where the maximum possible score on the concreteness and familiarity scales is 700 (Coltheart, 1981) and frequency is number of occurrences per million words of text.

attributable to the English words would be valid for the French words. As expected, the mean association strength rated by the 24 participants was significantly greater for the French semantic associates ($M = 3.14$, $\mathrm{SD} = 0.70$) than for the French non-associates ($M = 0.53$, $\mathrm{SD} = 0.32$; $t(118) = 25.15$, $p < 0.001$).

The 120 critical experimental words (60 words presented twice) were combined with 118 filler words (59 words presented twice) to create the full list of 238 words with the critical words appropriately distributed across the list. The filler words were inserted to ensure that each critical word was preceded by both a semantic associate and a non-associate and to dilute the obviousness of priming. These filler words also served as baseline words for determining proficiency levels in the analyses reported below. These filler words were themselves not primed by semantic associates, but the sequence of words in the list was arranged in such a way that both the critical words and the filler words could serve as semantic associate or non-associate to the immediately following word. For example, in the sequence '…PERSON, BENCH, SEAT, COMPUTER, SHIRT…', PERSON was a filler word, BENCH was a filler word and also a semantic associate of SEAT, SEAT was a critical word which was also used as a non-associate of COMPUTER, and COMPUTER was a critical word and a non-associate of SHIRT. The concreteness, familiarity, and written frequency values (Coltheart, 1981) for the filler words are presented in Table 2. Each complete list contained 55 nouns referring to living things and 64 concrete nouns referring to nonliving things[2] (Critical words: 26 living and 34 nonliving; Filler words: 29 living and 30 nonliving), each word presented twice with an interval of at least 15 words before the second presentation. For each critical target word, one of the presentations was preceded by a trial containing a semantically related associate (primed trial) and the other by a trial containing a semantically unrelated associate (unprimed trial). Six warm-up trials consisting of three living and three nonliving words were added to the beginning of each list for a total list length of 244 word presentations in each language. All words were preceded by the articles THE, A, or AN in English or LE,

---

[2] The imbalance between animate and inanimate nouns was inadvertent and discovered only after the data had been collected. The imbalance was not overwhelming and since the presentation of the two types of nouns was randomized, at any point in the study there would have been an imbalance slightly in favor of one noun type over the other. This factor could not have differentially affected primed versus unprimed trials since every target word occurred both with and without a semantic prime. Nevertheless, the imbalance was corrected in Experiment 2.

LA, UN, or UNE in French to ensure that English target words were read as nouns and not verbs (e.g. HAMMER can be both noun and verb) and to highlight the English or French nature of the word.

2.3. Procedure

The stimuli were presented one word at a time in sequence together with the related article on a computer monitor in separate language blocks. Working as quickly and accurately as possible, the participants pressed the right key of a keypad with their right index finger if the word displayed referred to something living and the left key with their left index finger if the word displayed referred to something not living. A tone was sounded immediately after an incorrect response. Each word was presented until the participant responded or until a deadline of 5000 ms had elapsed. The response-stimulus interval was set at 0 ms in order to optimize the priming effect.

2.4. Design

There were two quasi-random presentation orders of the lists, one for each language. Within each list, the critical experimental words were counterbalanced so that half the critical words were preceded on their first occurrence with a trial containing a semantic associate (primed) and on their second occurrence with a trial containing a non-associate (unprimed), and the reverse for the other half of the critical words. The strength of this design is that the same words were used in both the primed and unprimed conditions. Over the entire list of 244 words, half the trials required an animacy response which matched the previous required response, and half required a response which did not match the previous required response, in order to avoid creating response priming biases. The use of the various definite and indefinite articles was distributed across the animate and inanimate nouns in an equal and counterbalanced way. The order of the language blocks was alternated across participants.

2.5. Data analysis

On each trial, a participant's RT and performance accuracy were recorded. All correct RTs were winsorized at the 10% level, that is, for each individual the fastest 10% and slowest 10% of RT trials in a given condition were replaced with the values of the next fastest and slowest trials, respectively, to remove the impact of outliers within each participant's data set. From the set of retained correct responses for a given participant, the mean and the CV of the RTs were calculated. For all the analyses reported below, the alpha level selected for significance was 0.05.

General L2 proficiency. As an index of *general L2 proficiency* (Proficiency), we computed a measure of speed of lexical access in the L2, using the corresponding speed in L1 as baseline, as follows. For each participant, we partialled out the RT on the baseline words for L1 (English) trials from the RT on the baseline words for L2 (French) trials and saved the residual. This residual provided an index of the degree to which performance in the L2 was better or worse than expected after taking into account L1 performance. This index controls for individual differences in motor responses, general reading abilities, and in fulfilling other demand characteristics that would be reflected in the RTs on both L1 and L2 trials. These residualized measures of general L2 proficiency allowed us to divide the participants into higher and lower proficiency groups by median split on the scores (referred to below, for convenience, as High and Low Proficiency Groups although the terms should be interpreted as indicating relative, not absolute, levels of proficiency). Segalowitz and Freed (in press) found that similarly residualized measures of RT obtained in a living-nonliving judgment task correlated significantly with oral fluency in learners of Spanish as a second language (i.e. the extent to which they spoke the second language without filled pauses (ums, ahs, etc.) in a standardized interview).

Automaticity. A measure of L2-specific automaticity was calculated using intra-individual CVs, following the operational definition described earlier in the introduction (i.e. SD/mean). For each participant, we calculated a residualized measure by partialling out the CV for the L1 trials from the CV for the L2 trials. This provided an index of the degree to which efficiency or automaticity of performance in the L2 was better or worse than expected after taking into account L1 performance. This index controls for individual differences in general variability of motor responses, in general automaticity for reading, and in variability in fulfilling other demand characteristics that would be reflected in the CVs on both L1 and L2 trials. L2-specific automaticity indices were calculated separately for trials with baseline, primed, and unprimed words. Segalowitz and Freed (in press) also found that similarly residualized measures of CV correlated significantly with oral fluency in learners of Spanish as a second language.

3. Results and discussion

Mean percent error on L1 trials was 3.7% and on L2 trials was 5.5%, across baseline, primed, and unprimed trials.

For the sample as a whole, the L2-specific automaticity measure and the general L2 proficiency measure derived from the CVs and RTs on baseline trials, respectively, correlated significantly ($r = 0.751$, $n = 37$, $p < 0.0001$), indicating that individual differences in general proficiency were strongly related to individual differences in automaticity. This result was consistent with previously reported findings in Segalowitz and Segalowitz (1993) and Segalowitz et al. (1998), which used an unprimed lexical decision paradigm.

The baseline RTs were also submitted to a 2 × 2 mixed analysis of variance (ANOVA) with Language (L1, L2) as the within factor and Proficiency Group (High, Low) as the between factor. The analysis yielded a significant Proficiency Group by Language interaction ($F(1,35) = 60.20$, $\text{MSE} = 1872.90$, $p < 0.0001$, $\text{eta}^2 = 0.632$), indicating that while the two groups did not differ from each other on L1 (High L1: $M = 634$ ms, $\text{SE} = 20.44$; Low L1: $M = 621$ ms, $\text{SE} = 21.00$), the L2-L1 RT difference was greater for the less proficient bilinguals, as expected (Difference: $M = 234$ ms, $\text{SE} = 15.35$) than for the more proficient bilinguals (Difference: $M = 77$ ms, $\mathrm{SE} = 13.13$). These results confirm that the groups differed significantly with respect to their L2 performance as reflected in the residualized baseline measure.

To test for basic priming effects, the RT data were submitted to a $2 \times 2 \times 2$ mixed ANOVA with the between factor being Proficiency Group (High, Low), and the within factors being Language (L1, L2) and Priming (Primed, Unprimed). The analysis yielded expected significant main effects for Priming ($F(1,35) = 59.78$, $\mathrm{MSE} = 2651.53$, $p < 0.0001$, $\mathrm{eta}^2 = 0.631$), and for Language ($F(1,35) = 121.78$, $\mathrm{MSE} = 3446.73$, $p < 0.0001$, $\mathrm{eta}^2 = 0.778$). There was a significant Proficiency Group by Language interaction ($F(1,35) = 15.85$, $\mathrm{MSE} = 3446.73$, $p = 0.0003$, $\mathrm{eta}^2 = 0.312$), again reflecting the fact that the L2-L1 RT difference in the less proficient bilinguals (Difference: $M = 145$ ms, $\mathrm{SE} = 16.47$) was greater than in the more proficient bilinguals (Difference: $M = 68$ ms, $\mathrm{SE} = 10.49$; see Table 3). The result also confirms that basic priming effects were obtained with this paradigm and that task performance, therefore, was sensitive to semantic processing.

To test whether the high and low proficient bilinguals differed in their level of automaticity on primed and unprimed trials in their two languages, we submitted the CV data to a $2 \times 2 \times 2$ mixed ANOVA with the within factors being Priming (Primed, Unprimed) and Language (L1, L2) and the between factor being Proficiency Group (High,

Table 3
Means (and standard errors) of RT (milliseconds) and CV data from Experiments 1 and 2 on baseline, primed, and unprimed trials as a function of Proficiency Group and Language

Proficiency group *n*		Experiment 1		Experiment 2	
		High 19	Low 18	High 15	Low 14
Baseline trials					
RT	L1	634 (22.90)	620 (17.94)	660 (25.89)	662 (39.09)
	L2	711 (26.22)	854 (28.43)	699 (19.62)	808 (28.97)
CV	L1	0.208 (0.021)	0.193 (0.012)	0.202 (0.017)	0.198 (0.028)
	L2	0.234 (0.016)	0.321 (0.018)	0.209 (0.015)	0.267 (0.023)
Primed trials					
RT	L1	561 (21.05)	568 (21.35)	599 (21.83)	618 (39.63)
	L2	624 (20.73)	703 (27.17)	612 (16.37)	733 (37.45)
CV	L1	0.230 (0.024)	0.220 (0.016)	0.242 (0.023)	0.220 (0.031)
	L2	0.232 (0.014)	0.281 (0.017)	0.223 (0.017)	0.283 (0.035)
Unprimed trials					
RT	L1	629 (21.80)	616 (19.29)	654 (26.24)	685 (50.27)
	L2	703 (27.30)	770 (24.70)	663 (21.80)	787 (36.18)
CV	L1	0.222 (0.017)	0.197 (0.013)	0.218 (0.022)	0.228 (0.031)
	L2	0.253 (0.025)	0.269 (0.020)	0.208 (0.017)	0.267 (0.026)

Low). The Language effect was significant ($F(1,35) = 17.07$, MSE $= 0.0037$, $p = 0.001$, eta$^2 = 0.328$), indicating lower CVs (greater automaticity) in L1 than in L2. More importantly for the present study was the significant Proficiency Group by Language interaction ($F(1,35) = 6.21$, MSE $= 0.0037$, $p = 0.018$, eta$^2 = 0.151$) indicating that the more proficient bilinguals had similar CVs in L1 ($M = 0.226$, SE $= 0.017$) and L2 ($M = 0.243$, SE $= 0.016$; i.e. similar levels of automaticity) whereas the less proficient bilinguals did not (L1: $M = 0.208$, SE $= 0.017$; L2: $M = 0.275$, SE $= 0.016$; see Table 3).

Thus, overall, the CV index of automaticity, whether obtained from baseline or critical trials, proved to be significantly related to the general L2 proficiency measure. This result is consistent with the idea that higher levels of L2 proficiency involve, among other things, the ability to make semantic links in an automatic fashion (see also Favreau and Segalowitz (1983) who demonstrated, using a different measure of automaticity, that bilinguals with greater reading proficiency were more automatic in word recognition than were bilinguals with lower reading proficiency).

4. Experiment 2

In Experiment 1, we found that the participants' RTs were sensitive to priming conditions, and their CVs were sensitive to their skill with the target language. The goal of Experiment 2 was to replicate these behavioural findings and to investigate electrophysiological correlates of semantic priming by examining N400 amplitude, latency, and its variability (as reflected in a CV-analysis of the N400 waveform).

5. Method

5.1. Participants

Thirty participants were tested; however, one (S15) was excluded because she showed morphologically atypical waveforms in her L1. The remaining sample consisted of 29 native speakers of English (L1) who varied with respect to their proficiency in speaking French (L2; 8 males, 21 females; median age $= 23$ years, mean age $= 25.28$ years, SD $= 5.85$, range $= 19-38$ years; 27 participants were right-handed and two were left-handed). Twenty-six of these participants were recruited from the Concordia University student population; the remaining three were recruited through word of mouth. The participants received course credit or payment for their participation. The right-hand columns of Table 1 summarize the history and use of both languages, subdivided for the high and low proficiency participants in Experiment 2.

5.2. Stimuli

The 60 critical experimental words and their high and low semantic associates were the same as in Study 1, except for the following minor changes: seven French words and one

English word were replaced after native French speakers informed us that these would be more appropriate for a Québec population. Also, the inadvertent imbalance in the Experiment 1 stimulus list regarding the number of animate and inanimate nouns was corrected by adding 16 new words to the list to make them equal. These changes did not alter the reported concreteness, familiarity, or frequency values, nor did they significantly alter the reported association strength (new $M = 0.451$; new $\mathrm{SD} = 0.217$) for primed trials. Upon completion of the testing session, 23 of the participants were available to rate the strength of association between the critical experimental words and their semantic associates on a 5-point Likert scale (0–4) and found the association to be significantly greater for high semantic associates (English: $M = 3.48$, $\mathrm{SD} = 0.37$; French: $M = 3.36$, $\mathrm{SD} = 0.48$) than for low semantic associates (English: $M = 0.40$, $\mathrm{SD} = 0.42$; French: $M = 0.42$, $\mathrm{SD} = 0.35$) in each language (English: $t(118) = 42.70$, $p < 0.001$; French: $t(118) = 37.99$, $p < 0.001$).

The total number of words in the stimulus list was 276 in each language. The first 12 trials were warm-up trials and data from these were not analyzed. All words were preceded by definite and indefinite articles as in Experiment 1 and in all other respects the stimulus list was the same as in that experiment.

5.3. Procedure and design

The procedure and design adopted in Experiment 2 was essentially the same as that used in Experiment 1, with the exception that the response-stimulus interval used was 400 ms, instead of immediate presentation after the participant's response. We imposed this delay to ensure a period of time in which the N400 could be clearly examined, without the risk of contamination from sensory responses evoked by the next stimulus presentation. Participants were tested individually in a single session, which required approximately 30–45 min to administer the priming task and the health, language background, and word association questionnaires. Short breaks were given when required. Following the application of the EEG electrodes, participants were seated in front of a computer monitor approximately 1 m away, where they viewed yellow words preceded by their article presented individually on a black background.

Behavioural measures. The behavioural measures of automaticity were based on CVs and the behavioural indices of L2 proficiency were based on RTs obtained exactly as in Experiment 1. Participants were required to make a living-nonliving judgment to stimulus words that were either semantically primed (word on the previous trial was semantically related to the current stimulus word) or unprimed (word on the previous trial was semantically unrelated to the stimulus word). As in Experiment 1, a tone was sounded immediately after an incorrect response. The task yielded, for each participant, a mean RT and a CV in L1 and L2 for primed and unprimed critical trials, and for baseline trials. Data were winsorized as in Experiment 1.

Electrophysiological measures. A commercially available nylon EEG cap containing tin electrodes (Electro-Cap International) was used for EEG recording. The EEG was recorded from five midline sites (Fz, FCz, Cz, CPz, Pz) and 22 lateral sites (frontal left: FP1, F3, F7, FT7; frontal right: FP2, F4, F8, FT8; central left: FC3, C3; central right: FC4, C4; temporal left: T5, TP7; temporal right: T6, TP8; parietal left: CP3, P3; parietal right:

CP4, P4; occipital: O1, O2). A cephalic (forehead) location was used as ground. All sites were referenced to the left ear during acquisition and were re-referenced off-line to a linked ear reference. Electro-oculogram activity (EOG) was recorded from electrodes placed at the outer canthi of both eyes (horizontal EOG) and above and below the left eye (vertical EOG). Vertical EOG artifacts exceeding ± 50 Φμ and trials with EEG artifacts exceeding ± 100 Φμ were rejected off-line for all participants. EEG was sampled continuously with critical EEG epochs time-locked to the onset of each word. EEG data were amplified using Neuroscan Synamps in a DC-30 Hz bandwidth, sampled at 100 Hz for 1100 ms (100 ms pre-stimulus), and processed off-line using Neuroscan Edit 4.2 software. ERPs were averaged for each subject for each language and prime condition yielding a maximum of 60 trials for each of the four experimental cells. Only trials on which the subject performed correctly were included in the averages. The mean amplitude of the − 100–0 ms period of each averaged waveform was calculated and served as the 0 Φμ baseline for post-stimulus activity. N400 amplitude for each individual waveform was quantified by computing the averaged amplitude in four 50 ms intervals (300–350, 350–400, 400–450, 450–500 ms post-stimulus baseline). The 300–500 ms window was chosen for two reasons. First, it is the classical time period for N400 effects. Second, the end of this window was earlier than the mean reaction time of our fastest performing participants and, thus, allowed us to quantify the waveforms in a time epoch uncontaminated by behavioural responses.

6. Results

We first present a brief summary of the results of the behavioural analysis and then we discuss the electrophysiological analyses.

Behavioural analysis. Mean percent error on L1 trials was 4.1% and on L2 trials was 5.3%, across baseline, primed, and unprimed trials.

As in Experiment 1, the measure of general L2 proficiency, based on residualized RTs obtained on baseline trials, correlated significantly with the residualized automaticity measure, also based on baseline trials ($r = 0.739$, $n = 29$, $p < 0.0001$), indicating that individual differences in general L2 proficiency were strongly related to individual differences in L2-specific automaticity (see Fig. 1).

The baseline RTs were submitted to a 2 × 2 mixed ANOVA with Language (L1, L2) as the within factor and Proficiency Group (High, Low) as the between factor. As in Experiment 1, the analysis yielded a significant Language effect ($F(1,27) = 68.51$, $\mathrm{MSE} = 1815.39$, $p < 0.0001$, $\mathrm{eta}^2 = 0.716$) and a significant Proficiency Group by Language interaction ($F(1,27) = 22.72$, $\mathrm{MSE} = 1815.39$, $p < 0.0001$, $\mathrm{eta}^2 = 0.457$). This interaction indicated that the High and Low Proficiency groups did not differ from each other in L1 (High Proficiency: $M = 660$ ms, SE = 25.89; Low Proficiency: $M = 662$ ms, SE = 39.09) whereas the High Proficiency group performed faster in L2 than did the Low Proficiency group (High Proficiency: $M = 699$ ms, SE = 19.62; Low Proficiency: $M = 808$ ms, SE = 28.97). The interaction effect in the present experiment also reflected a greater L2-L1 RT difference for the low proficiency bilinguals (Difference: $M = 146$ ms, SE = 19.74) than for the high proficiency bilinguals (Difference: $M = 39$ ms,

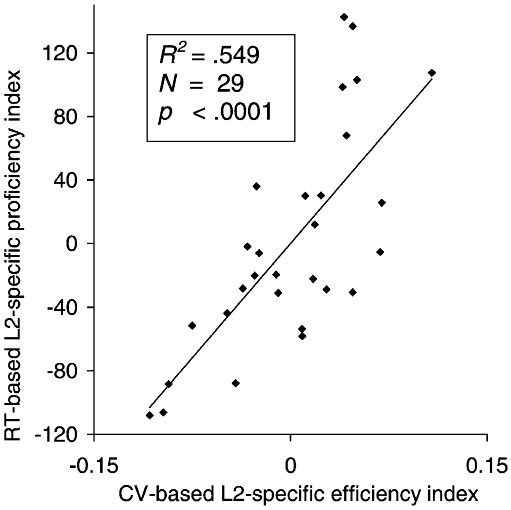


Fig. 1. Plot of L2-specific proficiency index based on residualized RTs as a function of L2-specific automaticity based on residualized CVs. Data are from baseline trials in Experiment 2.

SE = 11.37; see Table 3). These results confirm that the groups differed with respect to their L2 performance as reflected in their baseline RTs.

To test for basic priming effects, the RT data were submitted to a $2 \times 2 \times 2$ mixed ANOVA with the between factor being Proficiency Group (High, Low), and the within factors being Language (L1, L2) and Priming (Primed, Unprimed). The analysis yielded a significant main effect for Priming ($F(1,27) = 99.04$, MSE = 936.11, $p < 0.0001$, eta^2 = 0.786) indicating faster responding on Primed trials ($M = 640$ ms, SE = 20.44) than on Unprimed trials ($M = 697$ ms, SE = 23.56) and a significant Language main effect ($F(1,27) = 30.56$, MSE = 3383.70, $p < 0.0001$, eta^2 = 0.531) indicating faster responding in L1 ($M = 639$ ms, SE = 24.77) than in L2 ($M = 699$ ms, SE = 20.04). There was also a significant Proficiency Group by Language interaction ($F(1,27) = 20.32$, MSE = 3383.70, $p < 0.0001$, eta^2 = 0.429), reflecting, as was the case with baseline RTs, a greater L2-L1 RT difference for the low proficiency bilinguals (Difference: $M = 109$ ms, SE = 14.23) than for the high proficiency bilinguals (Difference: $M = 11$ ms, SE = 16.11; see Table 3).

Again, to test whether the high and low proficient bilinguals differed in their level of automaticity on primed and unprimed critical trials, we submitted the CV data to a $2 \times 2 \times 2$ mixed ANOVA with the within factor being Priming (Primed, Unprimed) and Language (L1, L2) and the between factor being Proficiency Group (High, Low; see Table 3). There was a significant Priming effect ($F(1,27) = 4.61$, MSE = 0.00087, $p = 0.041$, eta^2 = 0.146) indicating that CVs on primed trials were slightly higher ($M = 0.242$, SE = 0.018) than on unprimed trials ($M = 0.230$, SE = 0.016; mean primed-unprimed difference = 0.012, SE = 0.0056). There was also a significant Proficiency Group by Language interaction ($F(1,27) = 8.47$, MSE = 0.0037, $p = 0.007$, eta^2 = 0.239). This interaction reflected the fact that the Low Proficient Group had a significantly higher CV in the L2 compared to L1 (L1: $M = 0.224$; L2: $M = 0.275$), in contrast to the High

Proficiency group whose L2 CVs were, if anything, marginally lower (L1: $M = 0.230$; L2: $M = 0.216$).

Together, the behavioural results of Experiment 2 replicated the main patterns found in Experiment 1. The baseline CV measure correlated with general L2 proficiency in this experiment, and the RT measure was sensitive to priming conditions.

Electrophysiological analyses. We now turn to the ERP data recorded in this task, for which we had two goals. The first was to evaluate in these bilingual participants the classic electrophysiological priming effects, namely effects modulating N400 amplitude which typically vary over time epoch and across the scalp. Our second goal was to develop a method for assessing the N400 waveform for evidence of intra-subject variability within a specific priming condition, which might be related to proficiency in language performance. We did this in the CV analyses reported below.

Fig. 2 shows the grand waveforms, averaged across all participants, for primed and unprimed trials in L1 and L2 conditions. The waveforms were characterised by well-defined N1-P2 components followed by a negative deflection in the 300–500 ms window, which was smaller for primed than unprimed words. This latter activity, the N400, was observed in both L1 and L2 conditions, but differed in L2 as a function of proficiency group. These observations are also illustrated in Fig. 3 (high L2 proficient bilinguals) and Fig. 4 (low L2 proficient bilinguals) and were confirmed statistically (below). Finally, a large positive-going component was apparent after the N400 waveform. This late positivity is likely related to the fact that participants were required to make a button press to each trial. Thus, we restricted our analysis of the N400 component to a time window (300–500 ms), which minimized any influence of this subsequent response-related activity.

The ERP analyses reported below used SPSS v.11.0 statistical software and employed the Greenhouse and Geisser (1959) non-sphericity correction for effects with more than one degree of freedom in the numerator. Following convention, unadjusted degrees of freedom are reported, along with the Greenhouse-Geisser epsilon value (ε) and adjusted p-value. Mean square error values reported are those corresponding to the Greenhouse-Geisser correction. All significant main effects are reported first, followed by the highest order interaction effects involving Language and/or Priming. Unless otherwise stated, within-group interactions were further assessed using simple effects analyses and pairwise comparisons with alpha $= 0.05$. Significant interactions involving the between-subjects factor of L2 proficiency were followed with ANOVAs conducted separately for each proficiency group.

N400 Mean Amplitude and Time Interval Analyses. An initial mixed factors repeated measures ANOVA was conducted, including the within factors Language (L1, L2), Priming (Primed, Unprimed), Time interval (300–350, 350–400, 400–450, 450–500 ms post-stimulus), and two levels of Electrode Scalp Region, namely Coronal (frontal, fronto-central, central, centro-parietal, parietal) and Laterality (midline, left hemisphere, right hemisphere). Proficiency in L2 (High, Low) was included as a between-subjects factor. Although N400 amplitude did not differ overall between the two Proficiency Groups ($F(1,27) = 0.01$, MSE $= 2188.49$, $p = 0.94$), there were significant interactions between Language, Prime, Laterality, and Proficiency Group ($F(2,54) = 3.32$, MSE $= 3.20$, $p < 0.05$, $\epsilon = 0.923$), and between Language, Prime, Time, and Proficiency Group

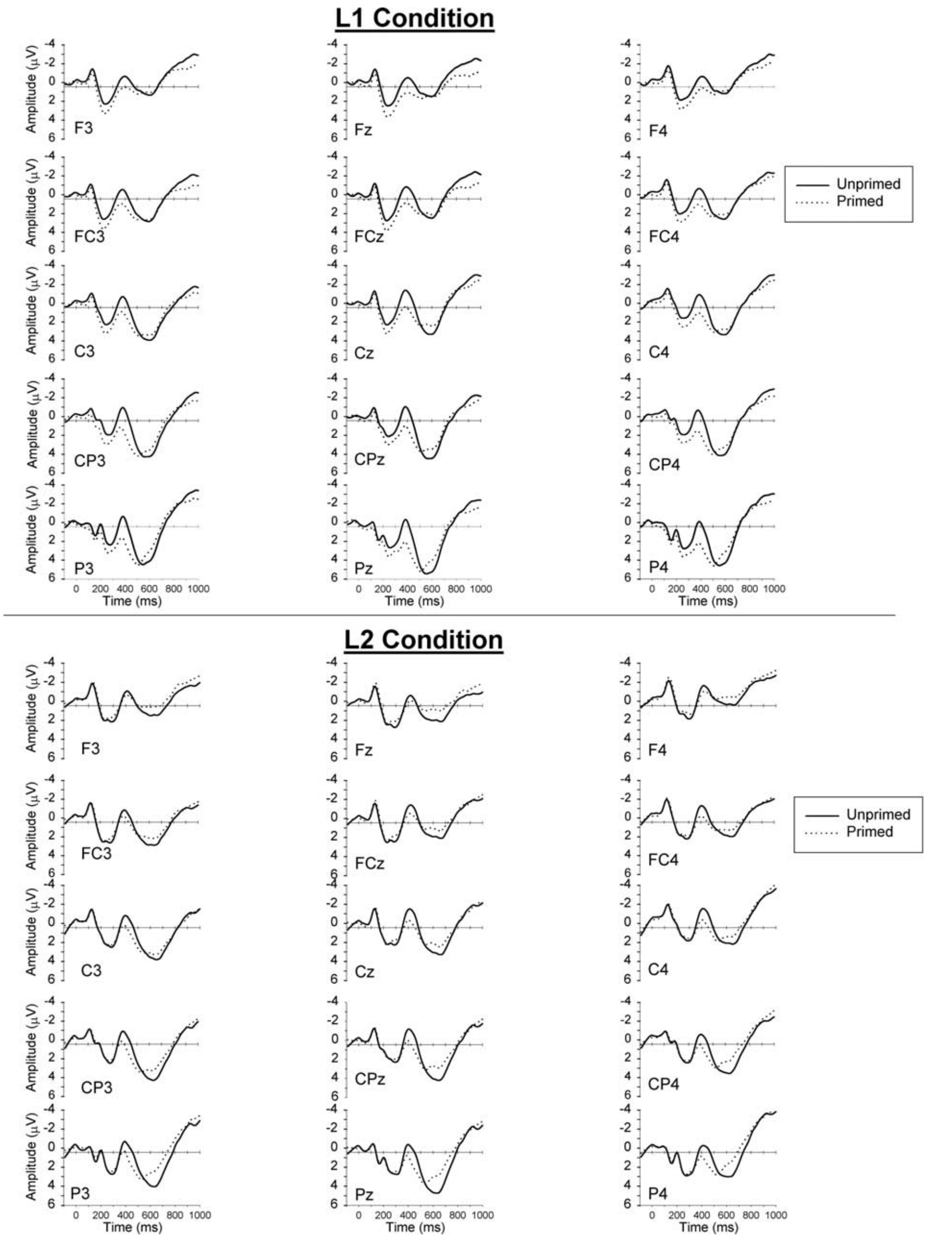


Fig. 2. Grand average ERP waveforms of all participants ($n = 29$) in Experiment 2. Waveforms are depicted as a function of prime trial (unprimed, primed) during the L1 (top panel) and L2 (bottom panel) condition blocks. Shown are waveforms recorded at frontal (left: F3; midline: Fz; right: F4), fronto-central (left: FC3; midline: FCz; right: FC4), central (left: C3; midline: Cz; right: C4), centro-parietal (left: CP3; midline: CPz; right: CP4), and parietal (left: P3; midline: Pz; right: P4) electrode locations. Negative amplitude is plotted upwards. Waveforms were low-pass filtered at 10 Hz for display purposes only.

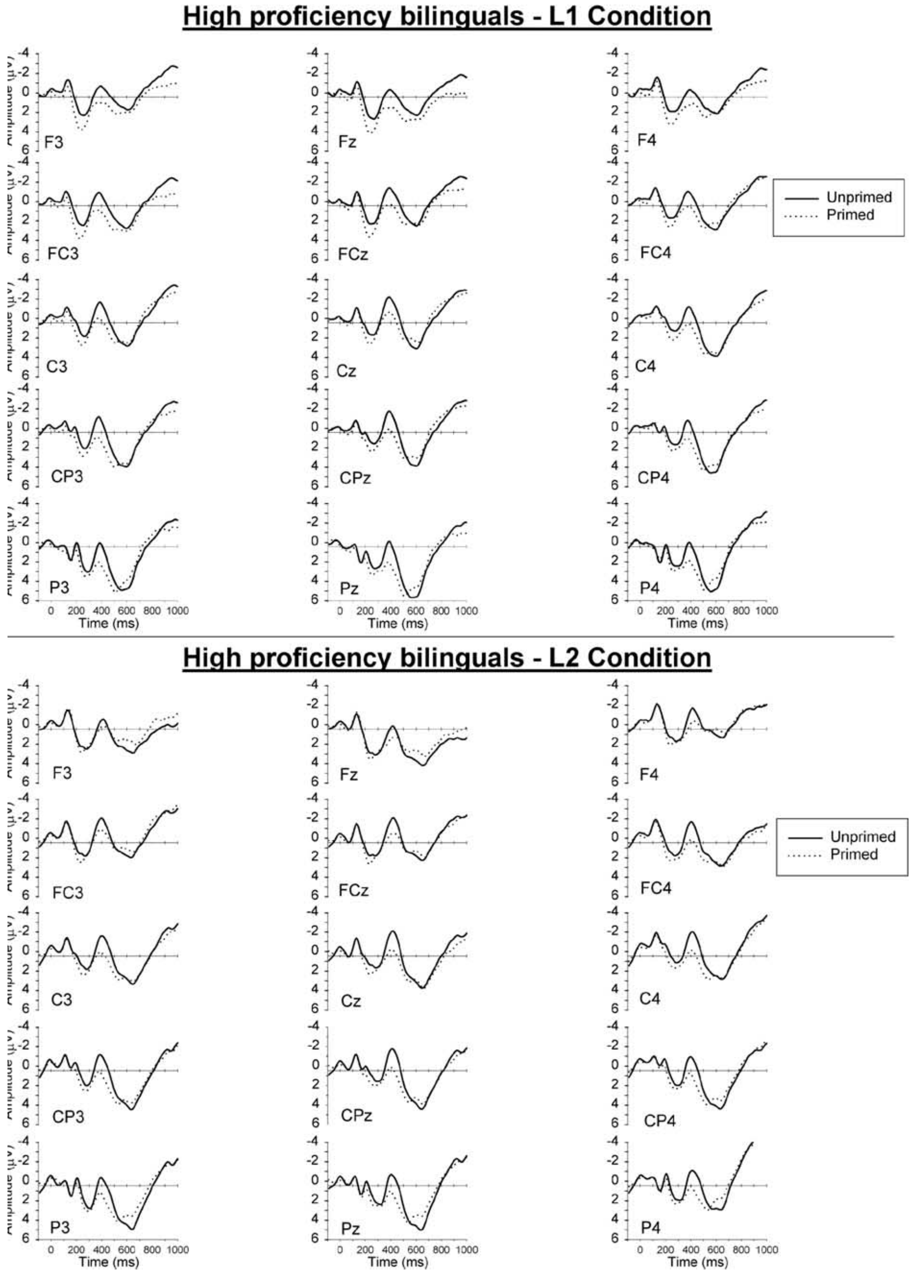
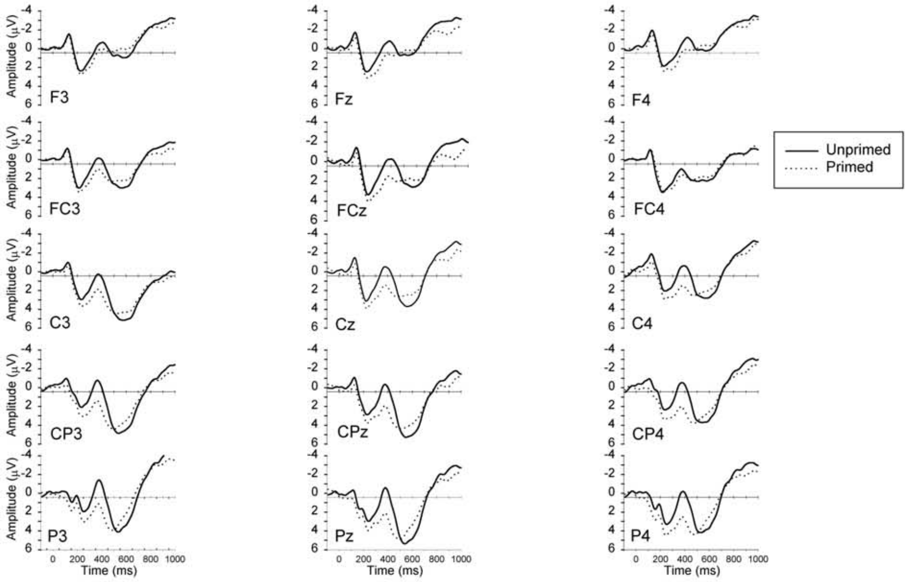


Fig. 3. Grand average ERP waveforms of bilinguals with high proficiency in their L2 ($n = 15$). All other details are as per Fig. 2. The N400 amplitude priming effect was significant during the 300–500 ms interval in L1 (top panel) and during the 350–500 ms interval in L2 (bottom panel).

Low proficiency bilinguals - L1 Condition



Low proficiency bilinguals - L2 Condition

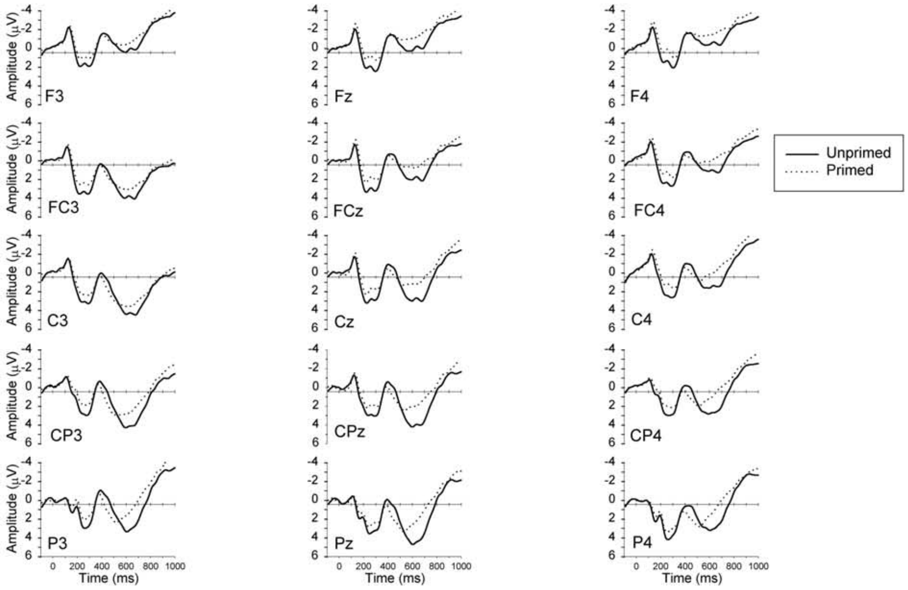


Fig. 4. Grand average ERP waveforms of bilinguals with low proficiency in their L2 ($n = 14$). All other details are as per Fig. 2. These low proficient participants showed a significant N400 amplitude priming effect (i.e. primed target amplitude was significantly less negative than unprimed target amplitude) in the 300–450 ms time window in their L1 (top panel). In contrast, there was no significant effect of priming in L2 (bottom panel).

($F(3,81) = 3,01$, MSE $= 54.81$, $p = 0.05$, $\varepsilon = 0.677$). In light of the fact that Language and Prime factors differed as a function of Proficiency Group, we then conducted repeated measures ANOVAs (within factors: Language, Prime, Coronal region, Laterality region, Time interval) separately for each proficiency group.

Fig. 3 illustrates the primed and unprimed grand average waveforms of the High L2 Proficient subjects during the L1 and L2 condition blocks. The ANOVA yielded a main effect of Prime ($F(1,14) = 33.73$, MSE $= 41.09$, $p < 0.001$), Time ($F(3,42) = 8.45$, MSE $= 187.79$, $p = 0.003$, $\varepsilon = 0.56$), and a significant Language X Prime X Coronal region X Time interaction ($F(12,168) = 4.30$, MSE $= 0.73$, $p = 0.005$, $\varepsilon = 0.27$). The latter reflected the fact that, in L1, these participants exhibited a widespread N400 priming effect (Unprimed more negative in amplitude than Primed) that was significant across the entire 300–500 ms epoch from frontal to centro-parietal regions (350–500 ms over the parietal region). In contrast, significant priming effects in L2 were 50 ms later, beginning at 350 ms interval and lasting to 500 ms, over central to parietal regions. In other words, significant priming effects were observed in both L1 and L2, confirming that the electrophysiological measures reflected aspects of semantic processing in this paradigm, but the N400 effect occurred later and had a more centro-parietal distribution in L2.

Fig. 4 illustrates the primed and unprimed grand average waveforms of the Low L2 Proficient subjects during the L1 and L2 condition blocks. The ANOVA yielded a main effect of Time ($F(3,39) = 8.35$, MSE $= 96.08$, $p = 0.001$, $\varepsilon = 0.72$), and a significant Language X Prime X Time interaction ($F(3,39) = 4.64$, MSE $= 18.26$, $p < 0.03$, $\varepsilon = 0.56$). The latter revealed a significant priming effect in L1 during the 300–450 ms post-stimulus interval; however, there were no significant effects of Prime in L2 during any time interval. A Prime X Coronal region interaction, $F(4,52) = 7.68$, MSE $= 14.73$, $p < 0.02$, $\varepsilon = 0.29$) indicated that the N400 priming effect, when it was present, was maximal over central to parietal scalp regions. To summarize, the Low L2 Proficient subjects showed a significant N400 priming effect in their L1 but not in their L2.

Coefficient of Variation analysis. For each averaged waveform per participant (L1 primed, L1 unprimed, L2 primed, L2 unprimed), we computed the CV of the N400 waveform. The logic of this computation was the following. Each participant's averaged ERP waveform was the average of the EEG epochs associated with correct trials (for the primed and unprimed conditions, the maximum possible was 60 trials for each). The SD of each waveform represents the variability of those trials. The entire 1100 ms duration of the averaged waveform was composed of 110 digital values, each one representing the mean amplitude of the waveform sampled every 10 ms (since our digitization rate was 100 Hz). Thus, the period comprising the N400 peak (300–500 ms) was represented by 20 digital values, each one representing the mean amplitude at that point in time (e.g. at 360 ms) and its corresponding SD. The N400 amplitude CV calculated for any one of those time points (using the mean and the SD) is conceptually similar to the RT CV, the first representing the variability of the amplitude of the EEG response evoked on the eligible trials and the other representing the variability of the RT elicited on the same trials. However, there were two ways in which our computation of the N400 CV differed from the RT CV. First, since the N400 waveform may not be precisely measured at any one single time point (e.g. its activity is not limited to the 360 ms time period), we chose to base our ERP CV measure on the average of the 5 digital values (and the average of their SDs) comprising each of four 50 ms time intervals (i.e. 300–350, 350–400, 400–450, 450–500 ms). The second difference was that the peak amplitude of the most positive trough prior to the N400 (between 200 and 300 ms) was scored and subtracted from each averaged amplitude window to control for inter-subject differences in baseline amplitudes. We then computed the CV (i.e. the average of the 5 SDs divided by the averaged mean amplitude which had been trough-to-peak corrected) for each time interval. The 50 ms interval in which the N400 amplitude effect was maximal in each language was determined for each subject through visual inspection and was chosen for analysis. For all participants, these were either the 350–400 or 400–450 ms intervals. Only data from electrode sites P3, Pz, and P4 were selected for analysis, as the N400 effect was most robust at these locations. Despite these computational differences between the behavioural and electrophysiological measures of CV, the N400 CV represents the variability of trials contributing to the waveform (measured over a given 50 ms interval) and remains conceptually similar to the behavioural CV since both were derived from the same trials.

We were especially interested in whether the N400 waveform CV would reflect processing efficiency or automaticity in a way similar to that found for the behavioural CVs. To examine this, the N400 amplitude CVs, computed in the manner described above, were subjected to a mixed factors ANOVA, with the within factors Language, Priming, and Site, and the between factor Proficiency Group. There were no significant main effects for Language, Site, or Proficiency Group. However, there was a main effect of Priming ($F(1,27) = 21.18$, MSE = 8.16, $p < 0.001$) which revealed that the CV of the N400 amplitude elicited on primed trials was significantly greater ($M = 3.56$, SE = 0.35) than on unprimed trials ($M = 2.15$, SE = 0.14; mean primed-unprimed difference = 1.41, SE = 0.306), paralleling the pattern obtained in the behavioural analysis reported earlier. To illustrate these effects, Fig. 5 shows the mean N400 amplitude CVs and the behavioural CVs for primed and unprimed conditions, collapsed across language and proficiency groups.

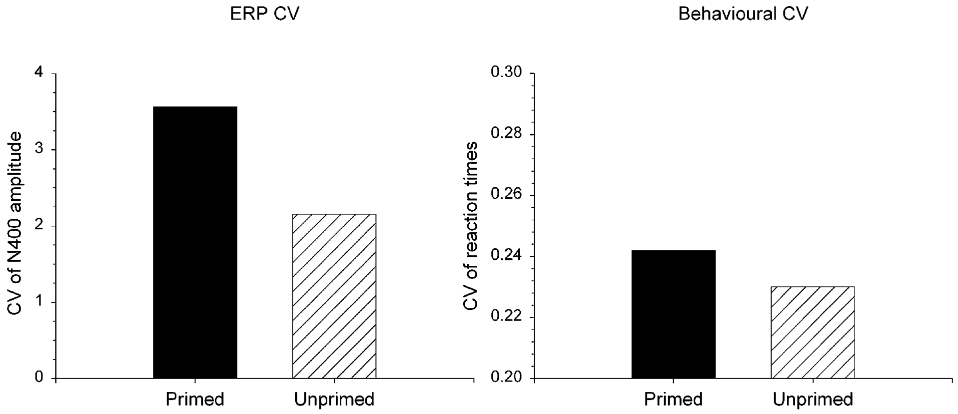


Fig. 5. Intra-individual mean CVs of N400 amplitudes and behavioural response times for primed and unprimed conditions, collapsed across first and second language conditions and proficiency groups ($n = 29$). The standard error of the difference between primed and unprimed conditions was 0.306 and 0.0056 for the N400 and behavioural measures, respectively.

7. Discussion

We will first review the consistency between the behavioural results of Experiments 1 and 2 and will then discuss the behavioural and electrophysiological results of Experiment 2. The behavioural results of the two experiments were very consistent. Significant RT priming effects were obtained, indicating that our task was successful in inducing semantic processing. In both experiments, participants performed more quickly in L1 than in L2. The fact that CVs were lower in L1 than in L2 indicated that the faster responding was due to more automatized and less variable performance (reflecting greater processing efficiency) in L1. This interpretation is also supported by the fact that, in both studies, the L2-specific CV of the highly proficient subjects was smaller than that of the low proficient subjects. Finally, baseline CV and RT measures were correlated, which again strengthens the argument that more efficient and less variable processing in L2 (indexed by the L2-specific CV) underlies general L2 proficiency (indexed by faster L2 RTs). Thus, the behavioural results of these experiments revealed two important findings. First, general L2 proficiency (defined as the residualized L2 RT after partialling out the L1 RT) was significantly related to the degree to which participants performed automatically in their L2, indicating that faster performance was related to more efficient processing. Second, the more L2-proficient bilinguals performed more quickly and more automatically during the L2 block than did the less L2-proficient group.

Turning to the electrophysiology and behavioural results of Experiment 2, we observed that N400 priming effects differed according to language and proficiency status.

Participants with high proficiency in L2 showed significant N400 and RT priming effects in both their L1 and in their L2. The low L2 proficient subjects manifested significant RT priming effects in L1 and L2 and a significant ERP priming effect in L1; however, they did not manifest a significant ERP priming effect in L2. How is it, then, that in these participants' response times were facilitated by related primes in L2 while the N400 effect was absent? One possibility is that the ERP and RT measures reflected different stages or aspects of cognitive processing. There is controversy concerning the extent to which the N400 is sensitive to automatic priming processes (Deacon et al., 2000; Kiefer, 2002; Kutas & Hillyard, 1989), or later post-lexical/integrative processes (Brown & Hagoort, 1993; Chwilla et al., 1995; Chwilla, Kolk, & Mulder, 2000; Holcomb, 1993). The same is true for RT-based measures, which have been shown to be influenced both by automatic spreading of activation in semantic memory and by attentional and/or expectancy-driven effects (Favreau & Segalowitz, 1983; Neely, 1991). One possible explanation for the dissociation in L2 between the N400 and RT in low proficient subjects is that the N400 effect is determined largely by post-lexical integration processes while RT priming effects are determined, at least in part, by automatic spread of activation which operates earlier, at the time of lexical access. It is possible that the task of judging each word as living or non-living in L2 for low proficient subjects was too taxing (as evidenced by their prolonged RTs) to maintain the preceding word in working memory and establish a higher order semantic context, thus leading to a difficulty in post-lexical integration as reflected by the N400. In the absence of such a context, each successive word, regardless of whether it was primed or unprimed, would elicit an N400 response due to a failure in post-lexical integration. Inspection of the grand average waveforms supports this notion.

The L2 primed waveform of the low proficient subjects is characterised by a robust N400 similar to the unprimed waveform, and shows no evidence of a reduction in amplitude as seen in the primed waveforms in L1 for these subjects. In contrast, automatic activation spreading from the prime to the critical word would be unaffected by the difficulty of the semantic judgement task in L2 and, thus, could facilitate response times.

The ERP data in Experiment 2 revealed another effect in addition to the general findings paralleling the RT data. In the high proficiency participants, the N400 effect began approximately 50 ms later in L2 than in L1, a finding consistent with previous results reported for sentence anomalies (Ardal et al., 1990; Weber-Fox & Neville, 1996, in their late L2 exposed bilinguals). This finding accords with the idea that processing is slower in L2 and could reflect, for example, the need to access L2 words via the L1 lexicon, as would be predicted by the revised hierarchical model of the bilingual mental lexicon, proposed by Kroll and Stewart (1994). However, this 50 ms delay in the N400 response in L2 versus L1 was not paralleled by a similarly large delay in RT in the high proficient bilinguals where, collapsed over primed and unprimed trials the difference between L1 and L2 was only approximately 11 ms. This again suggests that the N400 and RT priming measures are indexing very different stages of lexical and semantic processing.

We believe this study is the first to examine intra-subject variability (operationalized here as the CV) around the N400 waveform as a measure of semantic processing. We found a smaller CV for unprimed than for primed ERP waveforms, which is consistent with the findings for the behavioural CV data. These results can be understood as follows. In the unprimed condition, target words and the preceding stimulus bore a low associative relationship to one another. In this case participants should have received little processing benefit from the preceding word and the impact of that word, if any, should have been relatively uniformly low, contributing little to processing variability. In contrast, in the primed condition the target and the preceding word bore a strong semantic relationship to one another. Moreover, this relationship varied in strength from one case to the next and also varied in type. For example, sometimes the prime and target were both lexically and semantically associated, as in the case when a trial with the stimulus SOCK was followed by a trial with the stimulus SHOE (the words SOCK and SHOE frequently co-occur in the language and they are drawn from the same semantic category). Sometimes prime and target were only lexically associated, as in the case of successive trials with KITCHEN followed by SINK (these words frequently co-occur in the language but they are drawn from different semantic categories). These sources of variation in the strength and nature of prime-target associations could be expected to contribute to processing variability. Work by Kotz (2001) indicates that ERPs and RTs from associative and categorically based priming can vary differentially in bilinguals. Thus, one could expect more processing variability—and hence greater CVs—in the primed than in the unprimed condition in the present experiment.

8. General discussion

This study investigated variability in behavioural responses and variability in electrophysiological responses in bilinguals performing a semantic priming task to determine whether such variability may underlie individual differences in L2 proficiency. Experiment 1 confirmed that the task yielded an expected semantic priming effect, and it demonstrated that the L2-specific CV measure of automatic processing was significantly related to L2 proficiency. The results of Experiment 2 supported these basic conclusions and provided additional data regarding ERPs generated during this task. Experiment 2 also demonstrated that priming and the participant's level of L2 proficiency were reflected in the obtained patterns of N400 latencies and amplitudes. Moreover, the results of Experiment 2 demonstrated that *variability* in N400 amplitudes, as measured in this study by the CV, reflected sensitivity to semantic priming in a manner that paralleled the behavioural measures. This last result suggests that intra-individual processing efficiency, of the kind that can be associated with automatic processing, can be observed in patterns of brain activity.

There are, of course, some limitations to the present study. First, it would have been advantageous to include a group of monolingual speakers to be able to evaluate the effect of bilingualism per se on N400 amplitude, timing, distribution, and variability. Few ERP studies have done this, with the exception of that of Ardal et al. (1990), which found delays in N400 latency for L1 processing in bilinguals. Second, repeating critical words within our lists allowed us to compare the effect of an associated and unassociated prime on the same critical words, thereby avoiding the need to compare different words. However, this design was less than ideal given the finding that repeating a word can reduce the N400 effect (Besson, Kutas, & Van Petten, 1992; Karayanidis, Andrews, Ward, & McConaghy, 1991; Rugg, 1990; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991). However, all repetitions were counterbalanced across primed and unprimed trials, and order of language blocks alternated across participants, so it is unlikely to have accounted for the effects observed here.

In closing, the present results may open up a new dimension of research into the nature of language processing in the bilingual, by extending investigations that were previously confined to behavioural measures to include measures of brain activity. One particular ERP measure examined here—the N400 CV—may provide an important tool that could be used in future research on L2 skill acquisition. For example, it may be possible to conduct longitudinal ERP studies of gains in processing efficiency as a function of maturation and specific learning experiences—instructional methods, learning context, and the amount and nature of language practice. Beyond this, it may be possible to study changes in processing efficiency and automaticity in other skill domains too, thereby enriching our general understanding of how processing during the acquisition and performance of complex cognitive skills becomes increasingly automatic as a function of training and practice.

Acknowledgements

This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada (to NP and NS) and by a grant from the Canadian Institutes for Health Research (to NP). We would like to thank L. Cameli, S. Gagnon, L. Ingenito, M. Lagüe-Beauvais, and J. Mercier for their assistance in the ERP study. The principal authors can be reached by e-mail by writing to Natalie Phillips at Natalie.Phillips@concordia.ca or Norman Segalowitz at segalow@vax2.concordia.ca.

References

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*, 187–205.

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*, 132–149.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*, 34–44.

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*(2), 155–168.

Chwilla, D. J., Brown, C. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology*, *32*, 274–285.

Chwilla, D. J., Kolk, H. H., & Mulder, G. (2000). Mediated priming in the lexical decision task: evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*, 314–341.

Coles, M. G. H., & Rugg, M. D. (1995). Event-related brain potentials: an introduction. In M. D. Rugg, & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1–39). Oxford, UK: Oxford University Press.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33A*, 497–505. http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm.

Connolly, J. F., & Phillips, N. A. (1994). Event-related components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256–266.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, *94*, 276–287.

Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, *9*, 137–146.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 121–125). Cambridge, MA: Cambridge University Press.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Memory and Cognition*, *11*, 565–574.

Frenck-Mestre, C., & Prince, P. (1997). Second language autonomy. *Journal of Memory and Language*, *37*, 481–501.

Greenhouse, S. W., & Geisser, S. (1959). On the methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

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.

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: implications for the role of the N400 in language processing. *Psychophysiology*, *30*, 47–61.

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, *4*, 169–178.

Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*(1), 32–51.

Karayanidis, F., Andrews, S., Ward, P. B., & McConaghy, N. (1991). Effects of inter-item lag on word repetition: an event-related potential study. *Psychophysiology*, *28*, 307–318.

Kiefer, M. (2002). The N400 is modulated by unconsciously perceived masked words: further evidence for an automatic spreading activation account of N400 priming effects. *Cognitive Brain Research*, *13*, 27–39.

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*(2), 143–154.

Kroll, J. F. & de Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: mapping form to meaning in two languages. In A. M. B. de Groot, & J. F. Kroll (Eds.), *Tutorials in bilingualism* (pp. 169–199). Mahwah, NJ: Erlbaum.

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.

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. L. Nicol, & T. Langendoen (Eds.), *Language processing in bilinguals* (pp. 49–71). Cambridge, MA: Blackwell.

Kutas, M., & Hillyard, S. A. (1980). Reaction senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.

Kutas, M., & Hillyard, S. A. (1989). An electrophysiological probe of incidental semantic association. *Journal of Cognitive Neuroscience*, *1*, 38–49.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology*. Greenwich, CT: JAI Press.

Lambert, W. E. (1955). Measurement of the linguistic dominance of bilinguals. *Journal of Abnormal and Social Psychology*, *50*, 197–200.

Magiste, E. (1986). Selected issues in second and third language learning. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 97–122). *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives*, Hillsdale, NJ: Lawrence Erlbaum.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): an electrophysiological study of code switching. *Brain and Language*, *80*, 188–207.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://w3.usf.edu/FreeAssociation/

Polich, J., & Herbst, K. L. (2000). P300 as a clinical assay: rationale, evaluation, and findings. *International Journal of Psychophysiology*, *38*, 3–19.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition*, *18*, 367–379.

Segalowitz, N. (1986). Skilled reading in the second language. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives* (pp. 3–19). Hillsdale, NJ: Lawrence Erlbaum.

Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382–408). Oxford: Blackwell.

Segalowitz, N., & Freed, B. F. (in press). Context, contact and cognition in oral fluency acquisition: learning spanish in "At Home" and "Study Abroad" contexts. *Studies in Second Language Acquisition*.

Segalowitz, N., & Hulstijn, J. (2003). Automaticity in bilingualism and second language learning. In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches*. Oxford, UK: Oxford University Press, in press.

Segalowitz, N., Poulsen, C., & Segalowitz, S. (1999). RT coefficient of variation is differentially sensitive to executive control involvement in an attention switching task. *Brain and Cognition*, *38*(1), 255–258.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: evidence from second language word recognition. *Applied Psycholinguistics*, *14*(3), 369–385.

Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, *19*, 53–67.

Segalowitz, N., Watson, V., & Segalowitz, S. (1995). Vocabulary skill: single case assessment of automaticity of word recognition in a timed lexical decision task. *Second Language Research*, *11*(2), 121–136.

Tzelgov, J., Henik, A., Sneg, R., & Baruch, O. (1996). Unintentional word reading via the phonological route: the Stroop effect with cross-script homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 336–349.

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.

Vasos, H (1983). *Semantic processing in bilinguals*. Unpublished PhD dissertation, Concordia University, Montreal, Quebec.

Vigneau, F. (1998). Automaticité et automatisation du traitement de l'information. In F. P. Buchel, J. L. Paour, Y. Courbois, & W. Scharnhorst (Eds.), *Attention, mémoire, apprentissage: Études sur le retard mental, Lucerne* (pp. 43–52).

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.

Woltz, D. J. (1999). Individual differences in priming: the roles of implicit facilitation from prior processing. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 135–159). Washington, DC: American Psychological Association.