Effects of sentence context in L2 natural speech comprehension

Ian FitzPatrick[1,2]
Supervisors: Peter Indefrey[1,2]

[1]*F.C. Donders Centre for Cognitive Neuroimaging, Nijmegen, The Netherlands*
[2]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Electrophysiological studies consistently find N400 effects of semantic incongruity in non-native written language comprehension. Typically these N400 effects are later than N400 effects in native comprehension, suggesting that semantic processing in one's second language (L2) may be delayed compared to one's first language (L1). In this study we were firstly interested in replicating the semantic incongruity effect using natural auditory speech, which poses strong demands on the speed of processing. Secondly, we wished to investigate whether a possible delay in semantic processing might be due to bilinguals accessing lexical items from both their L1 and L2 (a more extensive lexical search). We recorded EEG from 30 Dutch-English bilinguals who listened to English sentences in which the sentence-final word was: (1) semantically fitting, (2) semantically incongruent, (3) initially congruent: semantically incongruent, but sharing initial phonemes with the most probable sentence completion within the L2, (4) semantically incongruent, but sharing initial phonemes with the L1 translation equivalent of the most probable sentence completion. We found an N400 effect in each of the semantically incongruent conditions. This N400 effect was significantly delayed to L2 words that were initially congruent with the sentence context. We found no effect of initial overlap with L1 translation equivalents. Taken together these findings firstly demonstrate that non-native listeners are sensitive to semantic incongruity in natural speech, secondly indicate that semantic integration in non-native listening can start on the basis of word initial phonemes, and finally suggest that during L2 sentence processing listeners do not access the L1 lexicon.

Keywords: Language, Bilingualism, Speech Comprehension, Lexicon, Semantics, Phonology, Event-Related Potentials

Correspondence to: Ian FitzPatrick, F.C. Donders Centre for Cognitive Neuroimaging, P.O. Box 9101, NL-6500 HB, Nijmegen, The Netherlands; e-mail: Ian.FitzPatrick@fcdonders.ru.nl.

1. Introduction

Langage comprehension in a language other than one's native language can be a challenging feat. Interestingly, electrophysiological studies of written language comprehension in bilinguals have consistently revealed a delay in bilingual semantic processing relative to monolinguals (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Hahne, 2001; Weber-Fox, Davis, & Cuadrado, 2003).

Bilinguals have also been shown to perform slower than monolinguals in lexical decision tasks (e.g., Scarborough, Gerard, & Cortese, 1984; Soares & Grosjean, 1984). This observation has been related to bilinguals having to perform a more extensive lexical search than monolinguals (Soares & Grosjean, 1984).

Later studies, also employing lexical decision paradigms, revealed that bilinguals are slower at recognising interlingual homophones (Dijkstra, Grainger, & Van Heuven, 1999) and pseudohomophones (Nas, 1983) than control items. This indicates that lexical items, from both the bilinguals L1 and L2, compete for selection during word recognition.

To date, only a handful of studies have been published that directly investigate non-native auditory sentence comprehension. As far as we know, only one such study has employed natural connected speech (Sanders & Neville, 2003). Therefore, in the present study, we are firstly interested whether we can replicate the finding of delayed semantic processing in bilinguals using natural speech, and secondly, whether this delayed semantic processing, at the sentence level, is due to bilinguals non-selectively accessing lexical candidates from both L1 and L2, thus performing an extensive lexical search.

1.1 Non-selective lexical access

There is converging evidence that word recognition in bilinguals occurs in a language non-selective manner (for an overview see: Dijkstra, 2005). In the domain of auditory word recognition, findings indicate that even though bilinguals have been shown to be able to accurately judge language membership on the basis of very little acoustic information (Grosjean, 1988; Li, 1996), interlingual competition is still very much apparent. In a cross-modal priming, lexical decision paradigm Schulpen, Dijkstra, Schriefers and Hasper (2003) found evidence for priming in Dutch-English bilinguals from both English and Dutch pronunciations of an interlingual homophone (/li:f/ (Dutch)-LEAF, /li:f/ (English)-LEAF) as compared to unrelated control words. This suggests that interlingual homophones activate lexical candidates from both languages, which then compete for selection. More compelling evidence was obtained by Spivey and Marian (1999), who showed that Russian-English bilinguals fixated on interlingual and intralingual competitors more often than on unrelated competitors in an eye-tracking paradigm while listening to spoken Russian; thus indicating the existence of competition both within and between languages in bilingual speech comprehension (Marian & Spivey, 2003a, 2003b; Marian, Spivey, & Hirsch, 2003; Spivey & Marian, 1999). This observation was replicated by Weber and Cutler (2004) in Dutch-native speakers of English listening to spoken English.

Thus it seems that, for bilingual speakers, even in a purely monolingual situation, there is spurious activation from the non-target language.

1.2 Event-related potentials

The most studied electrophysiological effect in language comprehension is a negative deflection in the ERP, that peaks around 400 ms after stimulus onset, termed the N400 (Kutas & Hillyard, 1980). The N400 occurs in response to words which are semantically incongruous within the current language context but can also be observed after presentation of non-words (Holcomb & Neville, 1990) or words with a low cloze probability in a particular sentence (Kutas & Hillyard, 1984). The amplitude of the N400 has been shown to be sensitive to word frequency (Van Petten & Kutas, 1990), with less frequent words eliciting a larger N400. The most commonly held interpretation of the N400 is that it is related to semantic processing of the eliciting word and that it's amplitude indexes the ease of semantic integration (e.g., Brown & Hagoort, 1993; Hagoort & Brown, 2000; Holcomb, 1993). Interestingly, non-native language comprehension has been associated with a delayed peak latency of the N400 effect (Ardal et al., 1990) and N400 component (Weber-Fox & Neville, 1996) and an attenuated N400 effect (Hahne, 2001) in response to semantic incongruity versus congruity, and a larger N400 in response to non-words (Sanders & Neville, 2003). Furthermore, Hahne (2001) observed a more negative N400 component for semantically congruous words in L2 speakers as compared to native speakers.

Other electrophysiological evidence of an influence of sentence context on word recognition relates to an early negative effect preceding the N400. This early negative effect has, in some cases, been interpreted as a discernible ERP component referred to as the N200 or *phonological mismatch negativity (PMN)*.[1] In an auditory word recognition task using high cloze probability sentence contexts (Connolly, Stewart, & Phillips, 1990) found an early negative effect in the 270-300 ms range, which they initially termed the N200 and later refer to as the PMN (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992). A PMN was elicited when the initial phonemes of the critical word in a sentence did not match those of the highest cloze probability word.

Van Petten, Coulson, Rubin, Plante and Parks (1999) found an early negative effect to semantically incongruent target words (e.g., "It was a pleasant surprise to find that the car repair bill was only seventeen *scholars*") compared to semantically congruent target words (e.g., "It was a pleasant surprise to find that the car repair bill was only seventeen *dollars*"). This early negative effect preceded the isolation point of the target words and was absent in cases where the target word shared initial phonemes with the semantically congruent sentence completion (e.g., "It was a pleasant surprise to find that the car repair bill was only seventeen *dolphins*"). The authors took this as evidence that semantic integration can start even on incomplete acoustic information.

In a similar study, Van den Brink, Brown and Hagoort (2001) found evidence of an effect of sentence context, possibly even as early as 140 ms after word onset. They contrasted three conditions in an experimental setting in which participants listened to high cloze sentences. In the fully congruent (FC) condition, the critical word, in the sentence final position, was a high cloze probability word. In the fully incongruent (FI) condition, the critical word was semantically incongruent in relation to the sentence. Finally, in the initially congruent (IC) condition, the critical word shared a phonemic onset with the highest cloze probability sentence continuation. Event Related Potentials time-locked to the onset of the critical word revealed an N400 component that was significantly more negative in the FI and IC conditions. This was taken to reflect the difficulty of integrating a semantically incongruent word with the sentence context. More interestingly they also found an early negative effect (N200) in the 150-250 ms time window in each condition, which was significantly more negative in the FI condition. Thus, in the situation where the incoming initial phonemes of the sentence final word matched those of the most probable continuation of the sentence (i.e., in the FC and IC conditions) the early negative effect was attenuated. Additionally, in the aforementioned studies, the N400 in the initially congruent conditions was delayed with respect to its onset (Van den Brink et al., 2001) and peak latency (Connolly & Phillips, 1994; Van Petten et al., 1999).

In a recent study, Diaz and Swaab (in press), investigated the early negative effect in response to violations of expectancy in auditorily presented alliterative word lists (e.g., "chat, champ, chaff, chant, challis, chad, chap, *address/chapter*"), categorical word lists (e.g., "giraffe, sheep, bear, wolf, rabbit, lamb, elephant, *desk/dog*"), and sentence contexts that were either (1) semantically congruent, (2) initially semantically congruent, (3) semantically congruent but of a lower cloze probability, or (4) semantically incongruent. The authors found a dissociation between the early negative effect, which was evident in alliterative word list contexts in response to phonologically incongruent words, and the N400 effect, which arose in the comparison between congruent and incongruent words in categorical lists. In sentence contexts the authors replicated the finding of an early negative effect that was greater for semantically incongruent words than for semantically congruent words, but did not differ between initially congruent words and congruent words. The early negative effect in response to phonological violations in list contexts had a different scalp topography compared to the early negative effect in response to semantic violations in sentence contexts, the latter having a distribution consistent with that of the N400.

The mechanisms underlying the early negative effect are subject to debate. Connolly and Phillips (1994) argue that the PMN reflects the matching of the incoming initial phonemes to an acoustic template, which is generated based on the expectation of the upcoming word. Newman (2003) made an intriguing case for this interpretation. They presented participants with a visual prime word (CLAP) and instructed them to internally delete the initial phoneme (LAP). They then aurally presented the correct deletion (LAP), an incorrect deletion (CAP), a deletion of the initial cluster (AP) or an unrelated word (NOSE). The PMN was comparable across all incorrect conditions, but was found to be attenuated in the correct condition suggesting that the PMN indeed reflects an acoustic comparison process.

In contrast to this phonological account, Van den Brink et al. propose a semantic account for the N200, namely that the N200 cannot be interpreted as a mismatch negativity as the early negative effect is present in all experimental conditions including conditions that are initially or fully congruent (when there is no mismatch). Van den Brink, et al. (2001) suggest that the N200 could reflect a process whereby candidates in the auditory cohort are individually assessed as to their goodness-of-fit within the sentence context (a lexically driven process) and would thus be part of the lexical selection process. There are also suggestions that the early negative effect could be an early onset of the N400 (Diaz & Swaab, in press; Van Petten et al., 1999), based on the fact that a number of studies do not find a differential scalp topography between the early negative effect and the N400 (Connolly & Phillips, 1994; Diaz & Swaab, in press; Van den Brink & Hagoort, 2004) or indeed any evidence pointing to two discernible components in either single subject averages or grand averages (Van Petten et al., 1999).

Clearly, while interpreting sentences in one' s native language, sentence context can influence the recognition of high cloze probability words even on the basis of only their initial phonemes. Given the complexities of non-native language comprehension, it is an open question whether contextual constraints can exert a similar early influence on word recognition in non-native language comprehension as is evident in native language comprehension.

1.3 Objectives

Thus our main question is whether we can obtain electrophysiological evidence for an early influence of sentence context in non-native language comprehension at all.

Secondly, given the converging evidence for language non-selective access to the bilingual lexicon, we are interested whether we could find evidence for activation of L1 lexical items in L2 natural speech comprehension.

Additionally we hope to find evidence that will elucidate the functional nature of the early negative effect and would enable us to further the debate surrounding its interpretation.

To investigate this we presented participants with auditory sentences in their L2, in an ERP paradigm similar to the one employed by Van den Brink et al. (2001). The study incorporated the Fully Congruent (FC) and Fully Incongruent (FI) conditions from the original design. It additionally included a within language overlap condition in which the initial overlap was with the highest cloze probability word (ICL2), and a between language overlap condition in which the initial overlap was with the translation of the highest cloze probability word into the participants' L1 (ICL1) (see Table 1 for examples of stimulus materials).

We hypothesise that if non-native listeners are capable of utilising contextual information at an early stage of the word recognition process we will notice an attenuated early negative effect and/or a delayed N400 component in the ICL2 condition as compared to the FI condition.

Secondly, if the early negative effect and N400 component are similarly attenuated and/or delayed in the ICL1 condition in comparison to the FI condition, we will have compelling evidence that lexical candidates from the participants' L1 are active even in a monolingual L2 sentence context.

This finding would be appealing, in terms of our third research question, as it has the potential of distinguishing between the conflicting interpretations of the early negative effect. Given that the highest cloze probability item and its direct translation equivalent greatly if not completely overlap in terms of their semantic features, the absence of an early negative effect in the ICL1

Condition	Sentence	Target
FC	To get water from a well, you need a *bucket.*	bucket
FI	The window cleaner carried a sponge, a ladder and a *pie.*	bucket
ICL1	The cleaning lady used soapy water from a large iron *embassy.*	bucket (Dutch: emmer)
ICL2	You need to put something under that leaky roof; do you have a *bubble?*	bucket

Table 1. Examples of stimulus materials. Underlining illustrates phonemic overlap.

condition compared to the FC condition would clearly suggest that the initial overlap activates lexical candidates from the participants' L1. As it is highly unlikely that participants would explicitly expect an L1 word in an L2 sentence context, it would indicate that the early negative effect could not simply reflect the acoustic matching of incoming phonemes to a template, but must rather be interpreted as an effect of semantic processing.

2. Methods

2.1 Participants

Thirty right-handed Dutch-English bilinguals participated in the experiment, 26 of which were included in the final analysis (7 men; mean age 23.7 years). The participants' English proficiency was assessed using the average z-scores on 50 grammaticality judgement items of the Oxford Placement Test (Allan, 1992) (mean score: 43.65, SD = 2.68; maximum score: 50) and a non-speeded lexical decision test (60 items), created by Meara (1996) and later adapted by Lemhöfer, Dijkstra and Michel (2004) (mean score: 44.42, SD = 5.78; maximum score: 60). Participants were either paid a small fee or they received study credits. None of the participants had any neurological impairment. All participants gave their written informed consent.

Dutch	English
aː	ɑː
ɑ	ʌ
ɛ, ɘ	ɛ, æ[2]
ɪ	ɪ
i	iː
ɔ	ɒ
oː	ɔː, əʊ
u, ʏ	uː
@	@
ɔu̯	aʊ

Table 2. Phoneme correspondences between Dutch and English, displayed using the International Phonetic Alphabet (IPA) (International Phonetic Association, 1999), used to define phonemic overlap.

2.2 Materials

Participants listened to English sentences that belonged to one of four conditions. In the FC condition sentences ended in a high cloze probability word, e.g., "When you buy duty free goods, you don't have to pay *tax*". In the FI condition sentences ended in a semantically incongruent word, e.g., "When you buy duty free goods, you don't have to pay *contact*". In the ICL2 condition the sentence final word shared initial phonemic overlap with the highest cloze probability word, e.g., "When you buy duty free goods, you don't have to pay *tadpole*" (initial overlap with *tax*). In the ICL1 condition the sentence final word shared initial phonemic overlap with the direct translation of the highest cloze probability word in the participant's L1, e.g., "When you buy duty free goods, you don't have to pay *baboon*" (initial overlap with "*belasting*" where *belasting* is Dutch for *tax*). We defined a number of correspondences between Dutch and English vowels and diphthongs (see Table 2) which we considered to be sufficiently similar to constitute an overlap. In each case the extent of the overlap was the vowel and initial consonant or consonant cluster.

We created 152 sentence frames in 38 quadruplets. The target words for each quadruplet consisted of one high cloze probability word (FC condition) and three semantically incongruent words (FI, ICL1 and ICL2 conditions). We created two stimulus lists, each with a different pairing of target word and sentence frame. Each sentence frame occurred only once per stimulus list. Seventy-six semantically congruent filler sentences were created and added to both lists to balance the number of sentences that were incongruent and congruent. Half of the participants were presented with stimuli from the first list, and half were presented with stimuli from the second list.

To give us a clear marker of critical word onset for time-locking the EEG, all critical words were chosen from English nouns that had either a plosive onset or vowel onset with a glottal stop. The distribution of critical words with a voiced plosive, unvoiced plosive, and vowel onset was kept constant over conditions. Critical words were controlled across conditions with respect to the number of phonemes and word frequency (see Table 3). Word frequencies were taken from the CELEX English lemma database (Baayen, Piepenbrock, & van Rijn, 1993). None of the critical words were cognates or homophones between English and Dutch.

The experimental sentences, fillers, and practice items were spoken by a female English native speaker with normal intonation and at a normal speaking rate. The materials were digitally recorded

Condition	Frequency (SD)	Phonemes (SD)
FC	3.34 (0.95)	5.29 (2.30)
FI	3.061 (1.03)	5.17 (1.90)
ICL1	2.94 (1.19)	4.90 (1.55)
ICL2	2.89 (1.29)	5.60 (1.90)

Table 3. Mean log frequency per million and mean number of phonemes for the: FC, FI, ICL1 and ICL2 conditions. Standard deviations are given in parentheses.

in a sound attenuating booth and digitised at a rate of 14.4 kHz. Sound files were later equalised to eliminate any differences in sound level.

2.3 Procedure

Participants were exclusively addressed in English by an English native speaker, both preceding and during the experiment, in order to make certain they were in a monolingual L2 language mode (Grosjean, 1982). Participants were placed in a sound attenuating booth and were instructed to listen attentively to the sentences, which were played over two loudspeakers at a distance of roughly 1.5 m, and to try to understand them. The sound level was kept constant over participants. To ensure that participants remained focussed on the task, they were prompted to make an animacy decision regarding the previous sentence, at random intervals during the experiment. They could respond by means of a button box. These data were not analysed. Each trial began with a 300 ms warning tone, followed by 1200 ms of silence, then a spoken sentence. The next trial began 4100 ms after the sentence offset. To ensure that participants did not blink during and shortly after presentation of the sentence, 1000 ms prior to the beginning of the sentence a fixation point was displayed. Participants were instructed not to blink while the fixation point was on the screen. The fixation point remained until 1600 ms after the offset of the spoken sentence. Participants had a practice session with five sentences to familiarize themselves with the experimental setting.

After the EEG recording the participants completed a word translation test on the critical items to verify that they were known and a cloze test on all the experimental sentence frames to check whether participants expected the sentence continuation that we had envisaged.

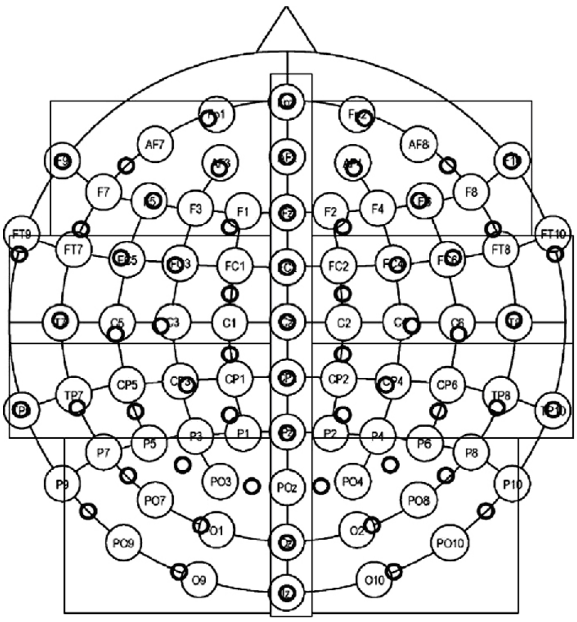

Figure 1. Electrode setup. Radial projection of electrode positions of the 10%-system (black circles) relative to the 10-20 system. Boxes denote regions for statistical analyses. Apostrophes denote not standard electrode locations.

2.4 EEG Recording

The EEG was recorded continuously from 64 sintered Ag/AgCl electrodes, each referred to an electrode on the nose of the participant. The electrodes were mounted in an equidistant elastic cap (http://www.easycap.de) and placed according to the extended 10% system (see Figure 1. for the electrode distribution). The EEG and EOG recordings were amplified with a BrainAmp DC amplifier (Brain Products, München, Germany) using a high-cutoff of 200 Hz, a time constant of 10s (0.016 Hz), and a sampling rate of 500 Hz. Impedances were kept below 5 kΩ. Trials with eye blinks or deflections exceeding 100 µV were rejected.

3. Results

3.1 ERP analysis

Using the results from the post tests, EEG analysis was restricted to those trials for which participants chose the same continuation as we had envisaged (ICL1, ICL2 and FI conditions) or for which they had correctly translated the critical word

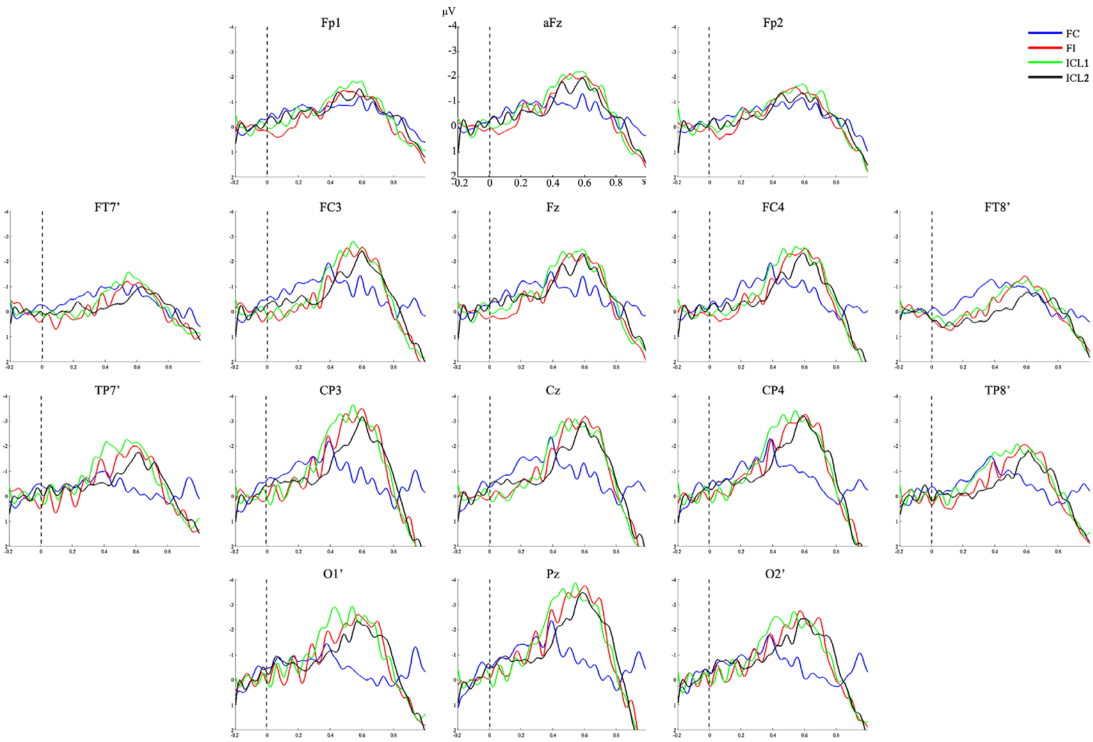

Figure 2. Grand average waveforms. FC (blue), FI (red), ICL1 (green), and ICL2 (black) on 16 scalp electrodes.

(FC condition).

Data from four participants were not analysed. Two participants were excluded due to excessive alpha. Data from one participant were incomplete due to a technical malfunction. One other participant was left out due to failure to complete the post-tests.

The data were analysed using the FieldTrip (http://www.ru.nl/fcdonders/fieldtrip) toolbox for Matlab (http://www.mathworks.com). EEG data were time-locked to critical word onset. Average waveforms were calculated for each participant using a 200 ms pre-stimulus baseline. Grand average waveforms were calculated by averaging the individual average waveforms. Statistical analysis was performed on the grand averaged data, in the latency ranges 150-300 ms and 300-800 ms, using an omnibus analysis of variance (ANOVA) with condition (4 levels) and site (9 levels; see Figure 1) as within subject factors. The latency ranges were chosen based on the previous literature and visual inspection of the grand average waveforms. All *p* values are reported after Greenhouse-Geisser correction (Greenhouse & Geisser, 1959). Contrasts between pairs of conditions were tested using a randomization approach that corrects for multiple comparisons (Maris, 2004). We now include a brief overview of this method.

The cluster randomization approach investigates the hypothesis that the observed ERP data is independent of the condition in which it is observed.

In a first step dependent samples *t*-statistics are computed for each sensor x time pair over a specified latency range (here between 300 and 800 ms after critical word onset for the N400). Those *t*-statistics that exceed a pre-defined threshold, that is based on the parametric *t*-distribution ($\alpha=0.05$, one tailed for: condition X < fully congruent), are passed to a cluster finding algorithm that creates clusters of sensor × time pairs which show the same effect that are adjacent either in time or with respect to their location on the scalp. For each cluster the combined *t*-statistic over the cluster is computed. This is called the cluster-level statistic.

Secondly, the distribution of the cluster-level statistic under the null hypothesis is computed by randomly permuting the order of the conditions over all subjects. A Monte Carlo approximation of the distribution is calculated from 500 randomizations. To correct for multiple comparisons the maximum cluster-statistic is used in the reference distribution.

The *p*-value for a given cluster is the proportion of random draws with a cluster-level statistic that exceeds the observed maximum cluster-level statistic. This *p*-value can be compared with a critical alpha level which corresponds to the

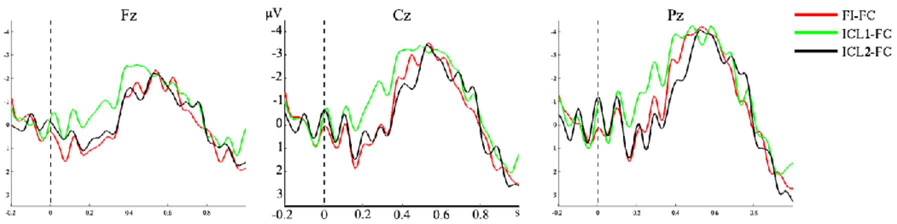

Figure 3. Difference waveforms. FI-FC (red), ICL1-FC (green), and ICL2-FC (black) on 3 midline electrodes.

probability of falsely identifying at least one cluster-level statistic as being significant.

Cluster randomization was performed on the following pairs of conditions: FI versus FC, ICL1 versus FC, ICL2 versus FC, ICL1 versus FI, and ICL2 versus FI; using the same latency ranges as the ANOVA (150-300 ms, 300-800 ms).

To determine the peak latency of the N400, for the three semantically incongruent conditions, we applied a low-pass filter at 7 Hz to the individual averages. The peak of the N400 component was defined as the minimum of the filtered individual averages, in the 300-800 ms latency range. A repeated measures ANOVA was performed on the peak latency of the N400 in the three incongruent conditions with condition (3 levels) as within subject factor. We tested two a priori pair-wise comparisons: ICL1 versus FI, and ICL2 versus FI (see Table 4).

Visual quantification of onset latencies was complicated due to variability of individual averages. We therefore estimated the onset of the N400 effect by performing a repeated measures ANOVA on successive 100-ms latency bins (c.f., Van Petten et al., 1999) with the following a priori pair-wise comparisons: FI versus FC, ICL1 versus FC, and ICL2 versus FC (see Table 5).

Finally, we performed a median split on the participants based on their pooled scores on both

	Peak latency		
Source	***df***	***F***	***p<***
Condition	2,50	5.455	0.01
ICL1 vs. FI	1,25	0.341	n.s.
ICL2 vs. FI	1,25	9.378	0.01

Table 4. Repeated measures ANOVA of the peak latency of the N400 in the semantically incongruent conditions.

proficiency tests. We performed an additional omnibus ANOVA, in the 300-800 ms latency range, with condition (4 levels) and site (9 levels) as within subject factors, and proficiency as between subject factor.

3.2 Grand averages

Figure 2 shows the grand average of each condition on 16 scalp electrodes. The waveforms for the three incongruent conditions (FI, ICL1, and ICL2) show an increased negativity in the 300-800 ms latency range relative to the fully congruent condition. This negativity is most pronounced on the centro-parietal electrodes.

Figure 3 shows the difference waveforms of the incongruent conditions minus the fully congruent condition on three midline electrodes. Figure 4 shows the topographical distribution of potentials in each condition as well as the topography of the significant differences between the incongruent and congruent conditions.

3.3 Omnibus ANOVA

In the 150-300 ms latency range, the ANOVA yielded no significant effect of condition ($F(3, 75) = 2.193$, pGG = n.s., $\varepsilon = 0.899$) and no interaction with site ($F(24, 600) = 1.233$, pGG = n.s., $\varepsilon = 0.241$).

In the 300-800 ms latency range, the ANOVA yielded a significant main effect of condition ($F(3,75) = 5.613$, $p\text{GG} < 0.01$, $\varepsilon = 0.866$) and a significant interaction with site ($F(24,600) = 5.508$, $p\text{GG} < 0.001$, $\varepsilon = 0.239$).

3.4 Fully Incongruent

Relative to the FC condition, there was a significant negative cluster starting at 414 ms after critical word onset ($p < 0.001$, cluster size = 5783 data points) and lasting until 708 ms. This negativity first diverged from the FC condition in the 400-500 ms latency bin (Table 5) and its peak was estimated at 461 ms.

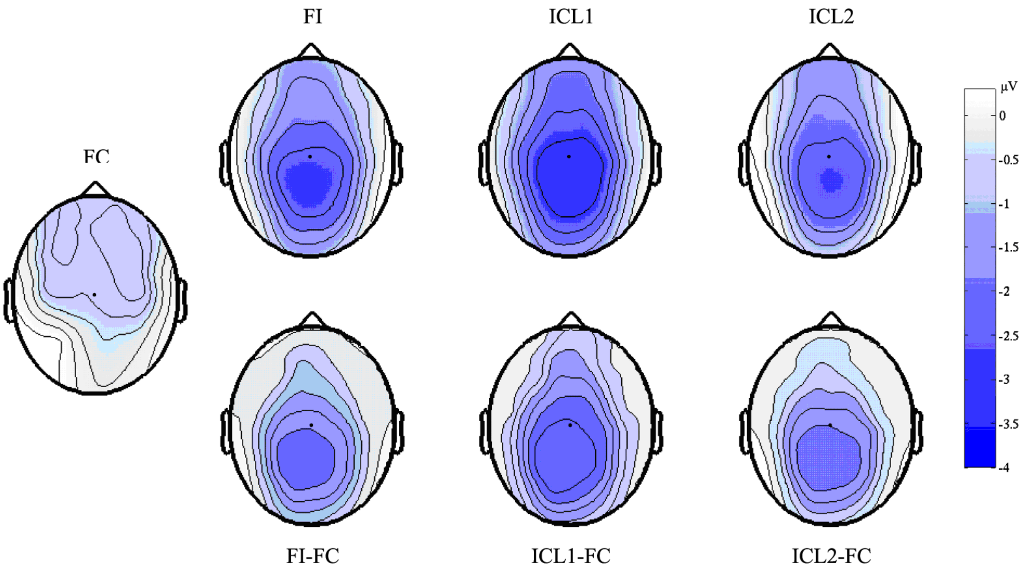

Figure 4. Topographic plots. (top) Topographic distribution of potentials between 300 and 800 ms for the FC (far left), FI (left), ICL1 (centre), and ICL2 (right) conditions. Blue denotes a negative value, white denotes a positive value. (bottom) Cluster randomization results for FI-FC(left), ICL1-FC(middle), and ICL2-FC(right) averaged between 300 and 800 ms and masked with the significant cluster.

3.5 Initially Congruent with the L1

Relative to the FC condition, there was a significant negative cluster starting at 358 ms ($p < 0.01$, cluster size = 7887 data points) and lasting until 712 ms. No significant clusters were found in the comparison of the ICL1 condition with the FI condition. The negativity first diverged from the FC condition in the 400-500 ms latency bin (Table 5) and the peak latency in the 300-800 ms time window was not significantly different from the corresponding negativity in the FI condition (see Table 4).

3.6 Initially Congruent with the L2

Relative to the FC condition, there was a negative cluster starting at 476 ms (**p** < 0.001, cluster size = 5738 data points) and lasting until 776 ms. No significant clusters were found in the comparison of the ICL2 condition with the FI condition. The negativity first diverged from the FC condition in the 500-600 ms latency bin (Table 5) and the peak latency in the 300-800 ms time window was significantly delayed compared to the corresponding negativity in the FI condition (see Table 4). The average delay was estimated at 67 ms.

3.7 Proficiency

Figure 5 shows the global field potentials for each condition for both high proficient and low proficient participants. Although low proficient participants seemed to have attenuated negativities in the 300-800 ms latency range in comparison with high proficient participants, no significant interactions were found between proficiency and condition ($F(3,72) = 1.304$, pGG = n.s., = 0.850) proficiency and site ($F(8,192) = 0.914$, pGG = n.s., $\varepsilon = 0.325$), and there was no significant three-way interaction between condition, site, and proficiency ($F(24,576) = 1.288$, pGG = n.s., $\varepsilon = 0.240$).

4. Discussion

The present study investigated the effect of sentence context and both intra- and interlingual initial phonemic overlap, on L2 natural speech comprehension. English-Dutch bilinguals listened to sentences in English that ended in a word that was: semantically congruent (FC condition), semantically incongruent (FI condition), semantically incongruent but initially overlapping with the most probable sentence completion (ICL2 condition), or semantically incongruent but initially overlapping with the L1 translation equivalent of the most probable sentence completion.

4.1 Influence of sentence context in non-native listening

We observed a significant N400 effect in each of the semantically incongruent conditions compared

latency bin (ms)	FI vs. FC		ICL1 vs. FC		ICL2 vs. FC	
	F(1,25)	*p*<	*F*(1,25)	*p*<	*F*(1,25)	*p*<
0-100	1.982	n.s.	0.001	n.s.	0.03	n.s.
100-200	2.416	n.s.	0.226	n.s.	1.184	n.s.
200-300	0.393	n.s.	1.844	n.s.	0.394	n.s.
300-400	0.242	n.s.	2.668	n.s.	0	n.s.
400-500	6.417	0.05	9.05	0.05	1.944	n.s.
500-600	15.104	0.01	15.054	0.01	14.582	0.01
600-700	13.864	0.01	17.155	0.001	15.556	0.01
700-800	2.609	n.s.	5.461	0.05	9.363	0.01

Table 5. Time course of the semantic incongruity effect. Latency bins are measured from critical word onset.

to the fully congruent condition. As far as we know, the only studies of natural speech processing in monolingual English speakers to report peak latency measures of the N400 to semantically incongruent sentence final words are Connolly and Phillips (1994) and Federmeier, McLennan, De Ochoa and Kutas (2002). Whereas the N400 in their semantic incongruity conditions peaked around 420 ms and 406 ms respectively, we found the average peak latency of the N400 in the Fully Incongruent condition to be approximately 460 ms. Although our study did not include a monolingual control condition, which would allow for a more direct comparison of N400 latencies in native and non-native listening, we note that this apparent delay is consistent with earlier findings of delayed N400s in non-native written language comprehension (Ardal et al., 1990; Hahne, 2001; Weber-Fox & Neville, 1996).

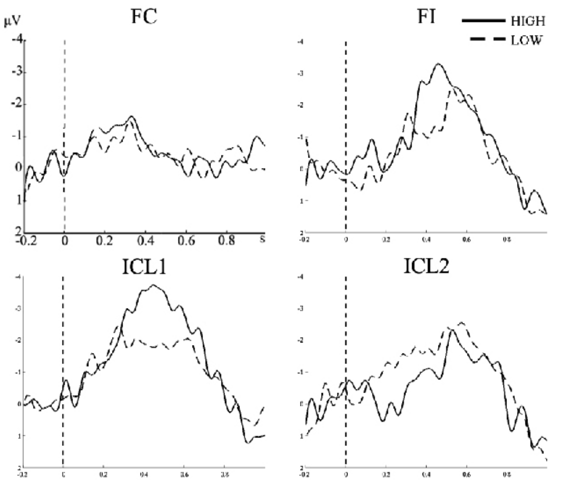

Figure 5. Proficiency. Global field potentials, for each condition, for high proficient participants (solid line) and low proficient participants (dashed line)

Also in line with earlier findings (e.g., Hahne, 2001; Weber-Fox et al., 2003; e.g., Weber-Fox & Neville, 1996), we observed a non-significant trend for the N400 to be attenuated in low proficient participants. We consider it likely that the high- and low proficient groups in our study differed insufficiently in terms of their language proficiency to observe a significant differential ERP effect. However this may also be attributable to a low signal-to-noise ratio by virtue of having too few participants per group.

In contrast to studies of monolingual auditory language comprehension, we did not observe an early negative effect in non-native listeners. This discrepancy could arise from differential processing requirements between native and non-native listening. While semantic integration in native listening may start on the basis of very little acoustic information (Van Petten et al., 1999), activation of the appropriate L2 lexical candidates might require the availability of a larger portion of the speech signal. For instance, using an auditory gating task whereby Dutch and English words were progressively revealed in increments of 40 ms, Schulpen et al. (2003) report that Dutch-English bilinguals are later and less proficient at identifying English words than they are at identifying Dutch words. For English words, they also consider multiple candidates even after the correct target can be uniquely identified on the basis of the acoustic information. Thus, in the latency range of the early negative effect, the available acoustic/phonological information may be insufficient for the non-native listener to be able to ascertain the presence of a semantic incongruity. Delayed semantic processing in non-native listening could potentially result in the absence of the early negative effect.

Alternatively, one might argue, following suggestions from Van Petten (1999) and Diaz (in press) that we can reconcile our findings with observations of an early negative effect by assuming that initial incongruity with a sentence context elicits an early N400. Logically, when there is no phonological and/or semantic incongruity on the basis of the initial phonemes of the target word, the onset of the N400 would be delayed (until such an incongruity arises). This delayed onset would show up as an absence of negativity in the latency range preceding the main N400 component. In our study the delayed N400 in the initially congruent with the L2 condition is evidenced by the delay in the peak latency of the N400.

4.2 Effects of initial phonological overlap with an L2 target word

We hypothesised that if non-native listeners are capable of utilising contextual information at an early stage of the word recognition process we would observe an attenuated early negative effect and/or a delayed N400 component in the initially congruent with the L2 condition as compared to the fully incongruent condition. While we did not find evidence for a negative ERP component preceding the N400 in any of our conditions (c.f., Van den Brink et al., 2001; Van den Brink, Brown, & Hagoort, 2006; Van den Brink & Hagoort, 2004) initial phonemic overlap with the most probable sentence continuation delayed the peak latency of the N400 by nearly 70 ms compared to the semantically fully incongruent condition. Furthermore, the initially congruent with the L2 condition began to diverge from the fully congruent condition in the 500-600 ms latency range, whereas the fully incongruent condition already diverged from the fully congruent condition in the 400-500 ms latency range.

In addition to the PMN, Connolly and Phillips (1994) also report a delayed peak latency of the N400 to items with initial phonological overlap with the correct target item. They propose that this observation reflects a delay in the semantic process underlying the N400. In conditions with initial phonological overlap, the incongruity is only apparent after the initial phonemes of the critical word. Thus the electrophysiological effects of the semantic mismatch would be delayed until the incongruity is detected. We interpret this finding as a confirmation of our hypothesis as non-native listeners treat the initial segment of critical words in the initially congruent with the L2 condition as if they were congruent with the sentence context.

4.3 Activation of L1 candidates

Our second hypothesis related to the question of language non-selectivity. If the early negative effect and N400 component were attenuated and/or delayed in the initially congruent with the L1 condition in comparison to the fully incongruent condition, we could interpret this as evidence for language non-selective activation of L1 lexical candidates. We found no differential effect between in the initially congruent with the L1 and fully incongruent conditions in the early latency range. Nor did we find a delay in the peak latency of the N400 in the initially congruent with the L1 condition as was the case in the initially congruent with the L2 condition. Also, the initially congruent with the L1 condition diverged from the fully congruent condition in the 400-500 ms latency range, as was the case in the fully incongruent condition. Although this is a negative finding it does suggest to us that non-native listeners do not treat initial overlap with the translation of the most likely sentence continuation as though it were initially congruent with the sentence context. This, in turn, would imply that word candidates from the participants' L1 are neither explicitly expected nor available for semantic processing, in non-native processing of natural speech.

This finding contrasts with studies that observe non-selective access to the bilingual lexicon. This may be due to the fact that many such studies employed single word paradigms (Paulmann, Elston-Guettler, Gunter, & Kotz, 2006; Schulpen et al., 2003) rather than investigating lexical access in the presence of a sentence context. A sentence context may be sufficient to restrict the lexical search process of a non-native listener to items from only the relevant language.

Eye-tracking paradigms such as those employed by Marian and Spivey (Marian & Spivey, 2003a, 2003b; Spivey & Marian, 1999) generally do embed target words in a sentence context. Notably though, these studies typically employ invariant sentence contexts, such as: "Pick up the *stamp*." (Spivey & Marian, 1999). This calls into question whether such experiments can be considered to reflect true sentence processing. Furthermore, participants in eye-tracking studies, by necessity, have a visual representation of the target objects before them, which would enable them to retrieve word-form and semantic information about the targets before they are referred to in the experimental sentences.

4.4 Semantics or phonology?

Our third research question related to the possibility of an absence of an early negative effect in the initially congruent with the L1 condition compared to the fully congruent condition. Such a finding would have allowed us to confidently state that the early negativity does not reflect a phonological mismatch but rather an effect of the initial semantic mismatch, as it is unlikely that participants would create a phonological expectation of an L1 word in an L2 context. As we did not find a negativity preceding the N400, these data do not allow us to rule out the phonological account. Nonetheless, given the indistinguishable scalp distributions between the N400 and the early negativity in monolingual speech comprehension (Connolly & Phillips, 1994; Diaz & Swaab, in press; Van den Brink & Hagoort, 2004; Van Petten et al., 1999), we do consider the semantic account to be more parsimonious.

5. Summary and Conclusions

Firstly, while a number of N400 studies have shown that non-natives are sensitive to semantic incongruity in written language processing, our study shows that non-native listeners are also sensitive to semantic incongruity in natural speech.

Secondly, consistent with previous findings, the peak latency of the N400 component seems to be longer in non-native listeners compared to latencies observed in monolingual natural speech comprehension.

Thirdly, the delayed peak latency of the N400, in the condition with initial phonological overlap with the most probable continuation, indicates that semantic integration in non-native natural speech comprehension can start on the basis of the word-initial phonemes.

Finally, we find no evidence that L1 lexical items are active during non-native natural speech comprehension.

Footnotes

1. This does not refer either to the N2 component (often referred to as the N200) observed in go/no-go paradigms (e.g., Rodriguez-Fornells et al., 2005; Schmitt, Schiltz, Zaake, Kutas, & Munte, 2001) or the N200 component in studies of orthographic processing (e.g., Elston-Guettler & Friederici, 2005; Kramer & Donchin, 1987; Niznikiewicz & Squires, 1996).

2. It is well established that Dutch natives are insensitive to the phonemic contrast between /ɛ/ and /æ/. Thus they would not by able to discriminate minimal pairs such as *cattle* and *kettle* (c.f., Weber & Cutler, 2004).

Acknowledgements

We thank Ton Dijkstra, Daniëlle van den Brink and Robert Oostenveld and three anonymous reviewers for their insightful comments. We also express our gratitude to Michel Bex, Vasiliki Folia, Jana Hanulova, Lilla Magyari, Stephan Miedl, Anouk Peijnenborgh, Willemijn Schot, and Kirsten Weber for assisting with the EEG electrode application.

References

Allan, D. (1992). *Oxford Placement Test*. Oxford: Oxford University Press.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., & Luce, M. (1990). Brain Responses to Semantic Incongruity in Bilinguals. *Brain and Language, 39*(2), 187-205.

Baayen, H., Piepenbrock, R., & van Rijn, H. (1993). The CELEX lexical database [CD-ROM]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400 - Evidence from Masked Priming. *Journal of Cognitive Neuroscience, 5*(1), 34-44.

Connolly, J. F., & Phillips, N. A. (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience, 6*(3), 256-266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-Related Potential Sensitivity to Acoustic and Semantic Properties of Terminal Words in Sentences. *Brain and Language, 43*(1), 1-18.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The Effects of Processing Requirements on Neurophysiological Responses to Spoken Sentences. *Brain and Language, 39*(2), 302-318.

Diaz, M. T., & Swaab, T. Y. (in press). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*.

Dijkstra, T. (2005). Bilingual Visual Word Recognition and Lexical Access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179-201). New York: Oxford University Press.

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*(4), 496-518.

Elston-Guettler, K. E., & Friederici, A. D. (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language, 52*(2), 256-283.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133-146.

Greenhouse, S. W., & Geisser, S. (1959). On Methods in the Analysis of Profile Data. *Psychometrika, 24*(2), 95-112.

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.

Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language an Cognitive Processes, 3*, 233-274.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia, 38*(11), 1518-1530.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30*(3), 251-266.

Holcomb, P. J. (1993). Semantic Priming and Stimulus Degradation - Implications for the Role of the N400 in Language Processing. *Psychophysiology, 30*(1), 47-61.

Holcomb, P. J., & Neville, H. J. (1990). Auditory and Visual Semantic Priming in Lexical Decision - a Comparison Using Event-Related Brain Potentials. *Language and Cognitive Processes, 5*(4), 281-312.

Kramer, A. F., & Donchin, E. (1987). Brain Potentials as Indexes of Orthographic and Phonological Interaction during Word Matching. *Journal of Experimental Psychology-Learning Memory and Cognition, 13*(1), 76-86.

Kutas, M., & Hillyard, S. A. (1980). Reading Senseless Sentences - Brain Potentials Reflect Semantic Incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain Potentials during Reading Reflect Word Expectancy and Semantic Association. *Nature, 307*(5947), 161-163.

Lemhöfer, K., Dijkstra, T., & Michel, M. C. (2004). Three languages, one ECHO: Cognate efects in trilingual word recognition. *Language an Cognitive Processes, 19*(5), 585-611.

Li, P. (1996). Spoken Word Recognition of Code-Switched Words by Chinese–English Bilinguals. *Journal of Memory and Language, 35*(6), 757-774.

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173-193.

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97-115.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and Language, 86*(1), 70-82.

Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology, 41*(1), 142-151.

Meara, P. M. (1996). *English vocabulary tests: 10k*. Swansea, UK: Center for Applied Language Studies.

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 and Verbal Behavior, 22*(5), 526-534.

Newman, R. L., Connolly, J. F., Service, E., & Mcivor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology, 40*(4), 640-647.

Niznikiewicz, M., & Squires, N. K. (1996). Phonological processing and the role of strategy in silent reading: Behavioral and electrophysiological evidence. *Brain and Language, 52*(2), 342-364.

Paulmann, S., Elston-Guettler, K. E., Gunter, T. C., & Kotz, S. A. (2006). Is bilingual lexical access influenced by language context? *Neuroreport, 17*(7), 727-731.

Rodriguez-Fornells, A., van der Lugt, A., Rotte, M., Britti, B., Heinze, H. J., & Munte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience, 17*(3), 422-433.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research, 15*(3), 214-227.

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of Lexical Access in Bilingual Word Recognition. *Journal of Verbal Learning and Verbal Behavior, 23*(1), 84-99.

Schmitt, B. M., Schiltz, K., Zaake, W., Kutas, M., & Munte, T. F. (2001). An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience, 13*(4), 510-522.

Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology-Human Perception and Performance, 29*(6), 1155-1178.

Soares, C., & Grosjean, F. (1984). Bilinguals in a Monolingual and a Bilingual Speech Mode - the Effect on Lexical Access. *Memory & Cognition, 12*(4), 380-386.

Spivey, M., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281-284.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*(7), 967-985.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2006). The cascaded nature of lexical selection and integration in auditory sentence processing. *Journal Of Experimental Psychology-Learning Memory And Cognition, 32*(2), 364-372.

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*(6), 1068-1084.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(2), 394-417.

Van Petten, C., & Kutas, M. (1990). Interactions between Sentence Context and Word-Frequency in Event-Related Brain Potentials. *Memory & Cognition, 18*(4), 380-393.

Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language, 85*(2), 231-244.

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*(3), 231-256.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1-25.