Bilingualism: Language and Cognition

cambridge.org/bil

Bilinguals benefit from semantic context while perceiving speech in noise in both of their languages: Electrophysiological evidence from the N400 ERP

2042

Kristina Coulter[1,2,3], Annie C. Gilbert[3,5], Shanna Kousaie[3,6], Shari Baum[3,5], Vincent L. Gracco[5,7], Denise Klein[3,6], Debra Titone[3,8] and Natalie A. Phillips[1,2,3,4]

[1]Department of Psychology, Concordia University; [2]Centre for Research in Human Development; [3]Centre for Research on Brain, Language and Music; [4]Bloomfield Centre for Studies on Aging; [5]School of Communication Sciences and Disorders, McGill University; [6]Montreal Neurological Institute, McGill University; [7]Haskins Laboratories and [8]Department of Psychology, McGill University

Research Article

Cite this article: Coulter K, Gilbert AC, Kousaie S, Baum S, Gracco VL, Klein D, Titone D, Phillips NA (2021). Bilinguals benefit from semantic context while perceiving speech in noise in both of their languages: Electrophysiological evidence from the N400 ERP. *Bilingualism: Language and Cognition* **24**, 344–357. https://doi.org/10.1017/S1366728920000516

Received: 3 June 2020
Revised: 8 July 2020
Accepted: 17 July 2020
First published online: 5 October 2020

Keywords:
speech perception; semantic context; N400; bilingualism; age of acquisition; working memory

Address for correspondence: Natalie Phillips, Ph.D. Department of Psychology, Concordia University
7141 Sherbrooke Street West, Montreal, Quebec, Canada H4B 1R6
E-mail: natalie.phillips@concordia.ca

Abstract

Although bilinguals benefit from semantic context while perceiving speech-in-noise in their native language (L1), the extent to which bilinguals benefit from semantic context in their second language (L2) is unclear. Here, 57 highly proficient English–French/French–English bilinguals, who varied in L2 age of acquisition, performed a speech-perception-in-noise task in both languages while event-related brain potentials were recorded. Participants listened to and repeated the final word of sentences high or low in semantic constraint, in quiet and with a multi-talker babble mask. Overall, our findings indicate that bilinguals do benefit from semantic context while perceiving speech-in-noise in both their languages. Simultaneous bilinguals showed evidence of processing semantic context similarly to monolinguals. Early sequential bilinguals recruited additional neural resources, suggesting more effective use of semantic context in L2, compared to late bilinguals. Semantic context use was not associated with bilingual language experience or working memory.

Introduction

Speech perception often occurs in suboptimal listening conditions with background noise or multiple talkers. In these conditions, speech perception is more effortful (Pichora-Fuller, Kramer, Eckert, Edwards, Hornsby, Humes, Lemke, Lunner, Matthen, Mackersie, Naylor, Phillips, Richter, Rudner, Sommers, Tremblay & Wingfield, 2016; Zekveld, Heslenfeld, Johnsrude, Versfeld & Kramer, 2014) and relies on the successful use of bottom-up and top-down processes (Bradlow & Alexander, 2007). Cognitive demands are even greater when listening in one's second language (L2) compared to one's first (L1) (e.g., Borghini & Hazan, 2018; Mayo, Florentine & Buus, 1997), likely due to L2 listeners having imperfect language knowledge (Garcia Lecumberri, Cooke & Cutler, 2010). Moreover, higher-order processes (e.g., prosodic and syntactic processing) may be less efficient in one's non-native language (e.g., Akker & Cutler, 2003; Clahsen & Felser, 2006). Thus, the extent to which L2 listeners effectively use top-down processes and higher-order cues (e.g., semantic context) while perceiving speech-in-noise is unclear. The present study examined the use of semantic context during L1 and L2 speech perception in noise in bilinguals through the recording of electroencephalography (EEG) measures. Furthermore, we examined whether semantic context use in L2 speech perception is influenced by L2 knowledge and experience, as well as working memory.

Under optimal listening conditions, auditory word recognition is theorized to be driven by bottom-up perceptual processes and informed by higher-order cognitive processes (e.g., Gaskell & Marslen-Wilson, 1997; Marslen-Wilson & Tyler, 1980). However, background noise distorts the neural representations of speech and impedes the listener's ability to extract the signal from competing background noise (e.g., Parbery-Clark, Marmel, Bair & Kraus, 2011). Speech signals become less intelligible with decreasing signal-to-noise ratios (SNRs) and increasing spectral overlap between the speech signal and the masking noise (e.g., Rogers, Lister, Febo, Besing & Abrams, 2006). In order to disambiguate the target signal from background noise, it is argued that listeners employ cognitive resources to a greater extent compared to listening in quiet (e.g., Mattys, Davis, Bradlow & Scott, 2012; Peelle, 2018; Pichora-Fuller et al., 2016). In line with this, perceiving degraded speech is associated with increased pupil dilation (a correlate of cognitive effort; e.g., Zekveld et al., 2014). Increased listening effort is thought to shift the balance between bottom-up and top-down processes, such that listeners rely more on higher-order processes and cues while perceiving speech-in-noise (Mattys et al., 2012; Peelle, 2018). Consistent with this, working memory capacity may be related to the ability to successfully perceive speech-in-noise in a native language (e.g., Ingvalson, Dhar, Wong & Liu, 2015; Millman & Mattys, 2017). Moreover, researchers have repeatedly shown that listeners benefit from semantic context while perceiving speech-in-noise in their native language (e.g., Boothroyd & Nittrouer, 1988; Cohen & Faulkner, 1983; Kalikow, Stevens & Elliott, 1977a; Miller, Heise & Lichten, 1951).

Most research on speech perception in noise has either been conducted with native listeners or without mention of participants' language backgrounds. However, the number of bilingual individuals is increasing globally, with bi-/multilingualism often being more common than monolingualism (e.g., Statistics Canada, 2017; Ryan, 2013; Eurostat, 2015). It is therefore important to understand the processes involved in speech perception in noise in both native and non-native languages.

Bilinguals are hypothesized to experience speech perception challenges in L2 compared to L1 due to their increased difficulty with aspects of L2 language processing. For example, bilinguals have shown limits in vocabulary (e.g., Bialystok, Craik & Luk, 2008) and less efficient syntactic (e.g., Hwang, Shin & Hartsuiker, 2018) and phonological processing (e.g., Navarra, Sebastián-Gallés & Soto-Faraco, 2005). Research does suggest that speech perception is more challenging and effortful for L2 listeners. For example, Borghini and Hazan (2018) reported greater pupil dilation, suggesting greater cognitive effort, for L2 compared to L1 listeners. Notably, when listening in quiet, bilinguals are able to use semantic information provided by sentences in their L2 to the same extent as in their L1 (Dijkgraaf, Hartsuiker & Duyck, 2017). Thus, compared to L1, L2 listeners may experience processing challenges that make speech perception more effortful but are able to use higher-order cues to successfully perceive speech in quiet.

In contrast, perceiving L2 speech-in-noise is thought to require more effort than in quiet and an even greater involvement of higher-order processes (e.g., processing syntactic and prosodic cues; Garcia Lecumberri et al., 2010). L2 listeners have been repeatedly shown to be more affected by noise (i.e., perform with lower accuracy) compared to L1 listeners (e.g., Cooke, Garcia Lecumberri & Barker, 2008; Garcia Lecumberri & Cooke, 2006; Mayo et al., 1997; Rogers et al., 2006; Shi, 2009, 2010). However, higher-order processes may be less efficient during L2 listening (e.g., Akker & Cutler, 2003; Clahsen & Felser, 2006), further compounding the language processing difficulties mentioned above.

Findings are mixed as to whether bilinguals benefit from semantic context while perceiving L2 speech-in-noise. While some studies failed to observe a benefit of semantic context (Golestani, Rosen & Scott, 2009; Hervais-Adelman, Pefkou & Golestani, 2014), others show a benefit although to a lesser extent than in a native language (Bradlow & Alexander, 2007; Shi, 2014). Other studies highlight the modulating influence of L2 AoA and language proficiency on the ability to capitalize on semantic context while perceiving L2 speech-in-noise. For example, Mayo and colleagues (1997) and Shi (2010) found that simultaneous and early bilinguals benefited more from semantic context than late bilinguals. Similarly, a recent study by our group (Kousaie, Baum, Phillips, Gracco, Titone, Chen, Chai & Klein, 2019) found that simultaneous and early bilinguals, but not late bilinguals, benefited from semantic context while perceiving L2 sentences in noise using an fMRI paradigm. Furthermore, Gor (2014) reported that high-proficiency heritage speakers benefited more from semantic context while perceiving sentences in noise, compared to both low-proficiency heritage speakers and late L2 learners with high or low L2 proficiency. Thus, to better understand semantic context use during L2 speech perception in noise, factors related to language knowledge and experience must be considered.

As noted above, most studies of speech perception in noise have been conducted with monolinguals or native listeners, with relatively few studies on bilinguals. The few studies that have examined bilingual semantic context use during speech perception in noise have mainly used behavioural measures of speech recognition (i.e., accuracy in reporting a target word or phrase) and word-pair stimuli. Few studies have examined the influence of factors relating to bilingual language experience (e.g., AoA), with previous studies often focusing on bilinguals with a specific language background (e.g., late bilinguals with limited L2 proficiency). Therefore, in this study, we examine semantic context use during L1 and L2 speech perception in noise in highly proficient bilinguals with varying ages of L2 acquisition by measuring the N400 event-related brain potential (ERP). We examined whether the amplitude and latency of the N400 is modulated by L2 knowledge and experience, as well as working memory capacity. This EEG study builds on work by Kousaie et al. (2019) who used an fMRI SPIN task with similar stimuli and similar participant groups, with some participants overlapping across both studies. The combined use of these two methods allows us to determine the brain regions implicated in perceiving speech-in-noise with fMRI, while using EEG enables us to further examine the cognitive processes as they unfold over time.

The N400, a negativity peaking approximately 400 milliseconds after a word stimulus (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980), is thought to be related to semantic processing. The N400 may reflect semantic access during language comprehension (e.g., Kuperberg & Jaeger, 2016; Kutas & Federmeier, 2011), with smaller amplitudes reflecting facilitated processing due to pre-activation of upcoming language representations. The N400 may also reflect lexico-semantic integration (e.g., Brown & Hagoort, 1993; Hagoort, Baggio & Willems, 2009), with larger amplitudes reflecting more effortful integration of a word into the preceding context. The N400 is modulated by sentence context, with larger amplitudes evoked by words that are semantically anomalous with respect to the preceding sentence context or occur in sentences low in contextual constraint (e.g., Connolly, Phillips, Stewart & Brake, 1992; Kutas & Hillyard, 1980). Additionally, longer N400 latencies have been observed in response to low-constraint compared to high-constraint sentences (e.g., León-Cabrera, Rodríguez-Fornells & Morís, 2017). N400 studies in bilinguals typically focus on visually presented word stimuli (e.g., Jankowiak & Rataj, 2017; Zirnstein, van Hell & Kroll, 2018). Nevertheless, researchers have found smaller N400 amplitudes and longer N400 latencies in response to auditorily presented semantically incorrect L2 sentences compared to L1 sentences (e.g., Hahne, 2001).

Based on the literature reviewed above, we expect bilinguals to benefit more from semantic context while perceiving L1 speech-in-noise compared to L2, reflected by an earlier and larger N400 effect (greater amplitude difference between high-constraint and low-constraint sentences) in L1 compared to L2. Furthermore, we expect the N400 effect during L2 speech perception to be modulated by L2 language experience. That is, we hypothesize that more years of L2 experience, higher L2 proficiency, and more time spent using L2 compared to L1, will be

https://doi.org/10.1017/S1366728920000516

associated with a greater benefit of semantic context during L2 speech perception, as indicated by a larger and/or earlier N400 effect. Lastly, we anticipate better working memory to be associated with a larger and/or earlier N400 effect while perceiving L2 speech-in-noise.

Methods

Participants

Participants were recruited using advertisements at Concordia University, McGill University, and Université de Montréal and included 19 simultaneous bilinguals, 20 early bilinguals, and 18 late bilinguals. As shown in Table 1, participants were bilingual speakers of English and French, with either as their L1, and had no functional knowledge of a third language. Simultaneous bilinguals had learned both of their languages from birth. Early bilinguals started learning their L2 by age five and late bilinguals after age five. L2 AoA ranged from age 0 to 15 across all participants. Simultaneous bilinguals self-reported which language was their dominant language at the time of testing and this was used as their L1 for our analyses. Participants self-rated their L1 and L2 proficiency in speaking and listening on a scale from 1 ('not at all proficient') to 7 ('native-like proficiency'). On average, participants rated themselves as being moderately to highly proficient in both of their languages with speaking and listening proficiencies ranging from 5 to 7 for L1 and from 4 to 7 for L2. Participants varied in the percentage of their total conversations in which they used each of their languages, with L2 use ranging from 5% to 95%. Importantly, AoA groups did not differ in chronological age (range: 18-36 years old), self-rated L1 and L2 speaking and listening proficiency, or percentage of L1 and L2 language use (see Table 1 for participant demographics). All AoA groups self-reported lower L2 speaking proficiency compared to L1 (all p values < .007) while only the early and late bilinguals self-reported lower L2 listening proficiency compared to L1 (p = .001 and p = .002, respectively). Participants also completed objective measures of L1 and L2 language performance (i.e., letter fluency, category fluency, and sentence repetition) as well as working memory (i.e., forward, backward, and sequencing digit span, and letter number sequencing; see Table 1). These measures are described in more detail below. Notably, simultaneous, early, and late bilinguals did not differ in L1 language performance, L2 language performance, or working memory performance. Performance on the category fluency task did not differ between L1 and L2 for any of the AoA groups (p values > .178). Performance on the letter fluency task was greater for L1 compared to L2 for the early (p = .043) and late bilinguals (p = .001) only. Performance on the sentence repetition task was greater for L1 compared to L2 for the simultaneous bilinguals only (p = .005). All participants were right-handed young adults with normal hearing as assessed by pure-tone average (PTA) thresholds. Participants gave informed consent and were given monetary compensation for participating.

Table 1. Participant Demographics.

	Simultaneous (N = 19)	Early (N = 20)	Late (N = 18)
Age	24.5 (5.43)	24.1 (3.28)	23.9 (4.16)
Number of Females	14	16	12
Number of English L1	10	9	7
L1 Listening Proficiency (/7)	6.8 (.50)	7.0 (.00)	6.9 (.24)
L2 Listening Proficiency (/7)	6.5 (.77)	6.4 (.81)	6.3 (.82
L1 Speaking Proficiency (/7)	6.8 (.54)	6.8 (.55)	6.8 (.51)
L2 Speaking Proficiency (/7)	6.1 (.91)	6.1 (.89)	5.6 (.93)
% L2 Language Use	47.2 (17.19)	41.4 (19.08)	43.7 (29.90)
L1 Letter Fluency	37.5 (10.31)	35.2(10.26)	39.1 (8.76)
L2 Letter Fluency	32.6 (8.96)	30.2 (8.11)	29.0 (8.63)
L1 Category Fluency	20.6 (7.34)	19.4 (5.58)	18.1 (4.38)
L2 Category Fluency	18.1 (5.49)	17.8 (6.03)	15.7 (4.85)
L1 Sentence Repetition	56.6 (8.61)	56.8 (12.34)	58.8 (8.09)
L2 Sentence Repetition	49.3 (13.55)	46.8 (13.44)	43.2 (16.9)
Forward Digit Span	11.0 (2.58)	11.1 (2.47)	11.3 (1.75)
Backward Digit Span	9.7 (2.79)	9.5 (2.86)	9.2 (2.15)
Sequencing Digit Span	8.9 (1.47)	9.0 (2.05)	9.1 (1.68)
Letter Number Sequencing	20.2 (2.18)	20.9 (2.94)	19.0 (2.47)

Note. Standard deviations are presented in parentheses. Simultaneous bilinguals self-reported which of their two languages they felt was their dominant language at the time of testing and this dominant language was used as their L1. Groups did not differ in age, language measures, or working memory performance (all p values > .08).

Materials

SPIN stimuli

The SPIN stimuli used here are the same as those used in the complementary fMRI study by Kousaie et al. (2019); however, the SNR in the noisy listening condition was much lower in the fMRI study (−6 dB) compared to the present ERP study (+1 dB). The SNR for the current study was determined by behavioural performance in the most difficult condition (i.e., low-constraint, L2 sentences in noise) during pilot testing. Our goal was to select an SNR that would be challenging in the easier listening conditions, but not so challenging that it would lead to floor effects in the most difficult condition. Pilot testing revealed that an SNR of +1 dB resulted in a 30% error rate in the most challenging listening condition. This choice is supported by the error rates of our study showing neither floor effects nor ceiling effects, reflected by an error rate of 27% in the most difficult listening condition (i.e., low-constraint, L2 sentences in noise) and 5% in the easiest noise condition (i.e., high-constraint, L1 sentences in noise). A total of 240 sentences were adapted from the Revised Speech Perception in Noise Test (SPIN-R; Kalikow, Stevens & Elliott, 1977b). Semantic context is manipulated in the SPIN-R sentences to yield high- and low-constraint sentences. High-constraint sentences provide rich semantic context, leading to a highly predictable final target word (e.g., "The lion gave an angry **roar**."). In contrast, low-constraint sentences do not provide sufficient semantic context and thus lead to an unpredictable final target word (e.g., "He is thinking about the **roar**."). Sentences from the SPIN-R test are English sentences with five to eight words and six to eight syllables. All terminal words are monosyllabic nouns with mid-range word frequency counts (5 to 150 per million words; Lorge & Thorndike, 1952). Sixty high-constraint and 60 low-constraint sentences were selected from the original SPIN-R list and matched on number of words (high-constraint: M = 5.5, SD = .81; low-constraint: M = 4.9, SD = .79) and number of syllables (high-constraint: M = 6.5, SD = .70; low-constraint: M = 6.6, SD = .70).

https://doi.org/10.1017/S1366728920000516

An additional 120 SPIN-R sentences (60 high-constraint and 60 low-constraint) were selected from the original set and adapted to French. To match high- and low-constraint French sentences on sentence length, French sentences were not all direct translations of English SPIN-R sentences. For example, "The bread was made from whole wheat" was adapted to "Le pain brun est fait de blé". High- and low-constraint French sentences were matched on number of words (high-constraint: $M = 5.8$, $SD = 1.01$; low-constraint: $M = 5.0$, $SD = 1.15$) and number of syllables (high-constraint: $M = 7.7$, $SD = 1.04$; low-constraint: $M = 7.3$, $SD = 1.21$). Target terminal French words were approximately 45% monosyllabic and 55% disyllabic. English and French terminal words were matched on spoken frequency (English: $M = 20.5$, $SD = 27.50$; French: $M = 24.4$, $SD = 28.90$), phonological neighbourhood density (English: $M = 15.4$, $SD = 9.22$; French: $M = 16.4$, $SD = 7.38$), imageability (English: $M = 539.5$, $SD = 65.77$; French: $M = 563.0$, $SD = 48.44$), and familiarity (English: $M = 524.5$, $SD = 51.36$; French: $M = 517.4$, $SD = 55.09$) using the MRC Psycholinguistic Database (Coltheart, 1981), Lexique 3 (New, 2006; New, Pallier, Ferrand & Matos, 2001), and the Corpus of Contemporary American English (Davies, 2008).[1]

All sentences were recorded by a female, simultaneous bilingual speaker of English and French in a sound-attenuated booth using an Olympus recorder with a 44.1 kHz sample-rate and 32-bit resolution. Sentence stimuli were presented in both a quiet and a noise condition. The background noise used for both English and French listening conditions consisted of an English multi-talker babble adapted from Bilger, Nuetzel, Rabinowitz, and Rzeczkowski (1984). The original eight-talker babble, which was low-pass filtered at 7500 Hz, was overlaid three times with a slight temporal jitter to create a babble mask that was less variable in its intensity fluctuations. Although it is possible that using an English-only babble may have influenced our findings, an acoustic analysis revealed that it did not likely provide any informational content. In addition, the low-pass filtering and the overlay of multiple speakers made it difficult to subjectively tell which language the babble was drawn from.

Sentence stimuli were presented in eight experimental conditions (four experimental conditions per language). For example, the English conditions were: high-constraint sentences in quiet; low-constraint sentences in quiet; high-constraint sentences in noise; low-constraint sentences in noise. Each target word was heard in each condition within each language, but two lists were created so that each word was heard only twice in each list. Thus, within each list, each target word was heard once in a high- and once in low-constraint sentence, as well as once in noise and once in quiet; (e.g., the terminal word "**spoon**" was heard in the high-constraint quiet and the low-constraint noise conditions in List 1 and was heard in the low-constraint quiet and high-constraint noise conditions in List 2). Each list consisted of eight experimental blocks. Lists were blocked by listening condition (quiet and noise) and language (English and French), both of which were counterbalanced within each list. Low-constraint and high-constraint sentences were pseudo-randomly intermixed within each block such that they were presented equally often within each block and no more than three similarly constrained sentences appeared in sequence. Each participant heard only one list and lists were counterbalanced across participants.

[1]The difference in the number of monosyllabic versus disyllabic terminal words between the English and French stimuli may have influenced our findings. This difference was unavoidable given the nature of French vocabulary and our desire to match English and French terminal words on the variables previously mentioned. We conducted two ANOVAs to assess this potential confound. Overall, participants were 2% more accurate in French compared to English (a small but reliable difference). However, participants were more accurate (5%) for monosyllabic compared to disyllabic French terminal words, an effect that was greater for the early bilingual group. While this effect did not interact with semantic constraint, it did interact with listening condition such that the monosyllabic versus disyllabic effect was exaggerated in noise. While it is possible that number of syllables may have influenced our ERP results, we had too few trials to test this.

Language proficiency measures

Participants completed letter and category verbal fluency tasks, and a sentence repetition task in each language. These tasks were selected because they are objective, commonly-used language measures that have been shown to distinguish between first- and second-language performance (e.g., Sandoval, Gollan, Ferreira & Salmon, 2010).

In the fluency tasks, participants were asked to say as many words as possible in one minute (excluding proper nouns, numbers, and words differing only in suffix) that began with a given letter of the alphabet (*F*, *A*, *S* for English; *P*, *F*, *L* for French) or drawn from a semantic category (*animals* and *fruit* for English and French, respectively). The number of words produced for all three letters within each language were summed to give a single score for each language. The number of words generated for each category was counted to give a score for category fluency for each language.

In the sentence repetition subtest of the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wiig & Secord, 2003), the experimenter read single sentences aloud and the participant was asked to repeat each sentence immediately after hearing it. Each trial received a score out of three, with a score of three indicating zero repetition errors and a score of zero representing four or more repetition errors. This task was performed in each language, with 24 sentences per language. Scores for each trial were summed to give a total score out of 72 for each language.

Working memory tests

The Digit Span and Letter Number Sequencing subtests of the Wechsler Adult Intelligence Scale Fourth Edition (Wechsler, 2008) were used as measures of working memory and were administered in the participants' native language.

For the Digit Span subtest, participants were asked to repeat a series of spoken numbers in order (Forward Digit Span), in reverse order (Backward Digit Span), or in sequential order from smallest to largest (Sequencing). The span length increased by one digit after every two trials and the task ended after two incorrect responses per level. One point was given for each correct trial and summed to give a total score for each variant of the task.

In the Letter Number Sequencing task, participants were read a series of letters and numbers and asked to repeat the numbers in sequential order, followed by the letters in alphabetical order. The span length increased by one unit after every three trials and the task ended after three incorrect responses in a level. One point was given for each correct trial and the points were summed to give a total score.

Procedure

All participants completed two testing sessions on two different days. In the first session, participants completed the language proficiency tests, working memory tests, a PTA hearing threshold test, and a language background questionnaire in which they self-reported detailed information regarding their L1 and L2 proficiency, AoA, and

https://doi.org/10.1017/S1366728920000516

patterns of language use. In the second session, participants performed the SPIN task while EEG was continuously recorded. During this session, participants were seated in a sound attenuated booth in front of a computer monitor. Participants first completed a practice block of the SPIN task in both languages. Practice trials consisted of 41 sentences (22 English, 19 French; approximately half high-constraint and half low-constraint). Five sentences in each language were presented in quiet and the rest in noise. Participants then completed one list (i.e., 240 sentences; described above) of the experimental SPIN task. Sentences were binaurally presented through EARLINK tube ear inserts (Neuroscan, El Paso, TX, USA) using Inquisit 4.0 (Millisecond Software, Washington). Thus, each participant listened to high- and low-constraint sentences in English and French, in quiet and in noise (SNR +1 dB). During sentence presentation, a fixation cross was presented on the computer screen. Participants were prompted to repeat the final word of the preceding sentence 1000 ms after the end of the sentence (i.e., when "Final Word?" appeared on the computer screen). Responses were manually scored as correct or incorrect by the experimenter. Scoring was lenient in that responses were accepted as correct if the participant made a pluralization error that was semantically and syntactically correct within the context of the sentence. Responses were also accepted if, in addition to reporting the correct word, participants included the appropriate determiner in the French sentences. Only trials scored as correct were included in the EEG analyses. A total of 1410 trials (i.e., 10.3% of trials) were excluded from the EEG analyses due to incorrect responses.

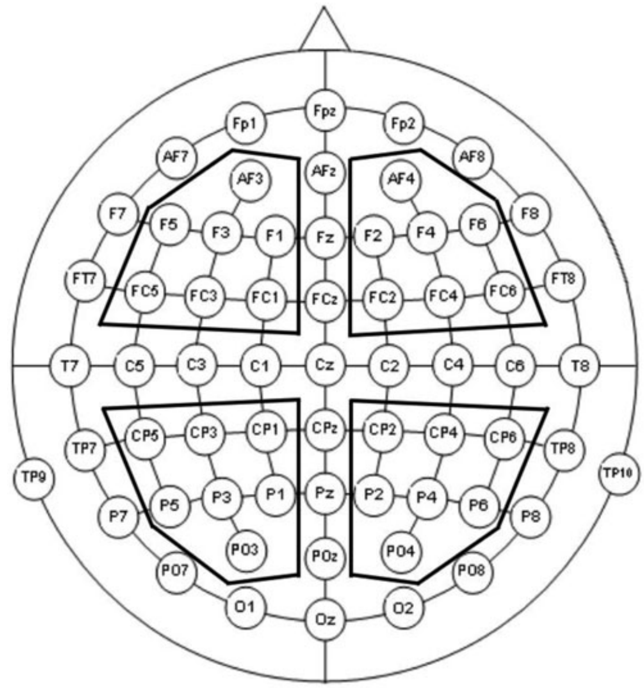

Fig. 1. Electrode clusters used for the topographical analysis of the N400 effect. Left anterior: AF3, F5, F3, F1, FC5, FC3, FC1. Right anterior: AF4, F2, F4, F6, FC2, FC4, FC6. Left posterior: CP5, CP3, CP1, P5, P3, P1, PO3. Right posterior: CP2, CP4, CP6, P2, P4, P6, PO4.

EEG data acquisition and analysis

EEG was recorded using a 64-electrode nylon cap and an ActiveTwo system (Biosemi, Amsterdam, NL) with a sampling rate of 2048 Hz and a bandwidth of .01 to 100 Hz. Additional electrodes were placed above and below the left eye and on the left and right canthi to monitor for horizontal and vertical eye movements. All electrodes were referenced to electrodes placed on the left and right earlobes.

EEG data was processed using BrainVision Analyzer 2.0.3 (Brain Products, Gilching, DE). The data were screened manually to remove exceptionally large artifacts and sections of data between experimental blocks. A low-pass filter of 100 Hz and a high-pass filter of 0.01 Hz were applied, as well as a DC drift correction. Artifacts from vertical and horizontal eye movements were removed using the Ocular Correction Independent Components Analysis. The EEG was then segmented into 1100 ms segments beginning 100 ms before terminal word onset and ending 1000 ms after terminal word onset. These segmentations were performed for each of the eight experimental conditions separately. Semi-automatic artifact rejection was conducted for all segments within each condition by removing segments where: a) the absolute difference between two adjacent data points exceeded 50 microvolts, b) the difference between the maximum and minimum amplitude within a segment exceeded 200 microvolts, or c) the activity fell below 0.5 microvolts. Based on these criteria, a total of 412 segments were removed (3.5%). No systematic differences were found in the number of segments removed as a function of condition. Segments were averaged per condition for each participant. The N400 peak was identified and scored in the averaged waveform for each condition at four midline electrode sites: Fz, FCz, Cz and CPz. Based on previous studies examining the N400 using auditorily presented sentences (e.g., Holcomb & Neville, 1991), the N400 was operationally defined as the most negative peak between 250 and 600 ms, that was temporally consistent across the four midline electrode sites, following terminal word onset. The topographical distribution of the N400 was characterized by examining the mean amplitudes within 300 to 600 ms following terminal word onset at four lateral site electrode regions of interest (ROIs; Figure 1).

Statistical analyses

Statistical analyses were performed using IBM SPSS 23.0 (IBM Corp., 2015). The Greenhouse-Geisser non-sphericity correction was applied to all analyses and the Bonferroni correction was applied to adjust for multiple comparisons.

A mixed factorial ANOVA was conducted to examine potential differences in behavioural accuracy between groups on the SPIN task. Within-subjects factors included language (L1, L2), listening condition (quiet, noise), and context (high-constraint, low-constraint). The between-subjects factor was AoA group (simultaneous, early and late bilinguals).

Two families of ANOVAs were done to analyze the N400; one for midline electrodes (Fz, FCz, Cz, and CPz) and one for lateral site ROIs (Figure 1). First, to examine any differences in the N400 peak amplitude and latency between groups and conditions, a 3 (group) X 2 (language) X 2 (listening condition) X 2 (context) X 4 (electrode) mixed factorial ANOVA was conducted. For each participant, and within each language and listening condition, the high-constraint waveforms were then subtracted from the low-constraint waveforms to better examine the N400 effect (i.e., the context effect). The N400 effect was identified in these subtracted waveforms following the same operational definition mentioned above and analysed in a 3 (group) X 2 (language) X 2 (listening condition) X 4 (electrode) mixed factorial ANOVA.

https://doi.org/10.1017/S1366728920000516

Table 2. Percentage of variance accounted for by each variable for each principal component.

	PC1 L1/L2 Proficiency	PC2 Working Memory	PC3 Years of Experience
L2 Years of Experience	0.26	1.27	77.23
Letter Fluency	21.76	5.58	0.03
Category Fluency	24.21	3.98	1.78
Sentence repetition	24.57	7.43	0.22
Percent L2 Use	6.10	4.52	4.68
Forward Digit Span	3.31	28.62	1.30
Backward Digit Span	4.77	27.39	0.55
Sequencing Digit Span	6.45	8.66	1.82
Letter Number Sequencing	8.56	12.56	12.38

Second, mean amplitudes within 300 to 600 ms of the subtracted waveforms were analyzed from four lateral site ROI electrode clusters (left anterior, right anterior, left posterior and right posterior; Figure 1) using a 3 (group) X 2 (language) X 2 (listening condition) X 4 (ROI) mixed factorial ANOVA to better characterize any topographical differences between the groups.

Lastly, a principal components analysis (PCA) and Pearson correlations were run to examine the association between L2 semantic context use (i.e., the N400 effect amplitude and behavioural accuracy) and individual differences in language experience and working memory. The PCA was run using the "prcomp" function in R version 3.5.1 (R Core Team, 2018) and included the variables: years of L2 experience, percent L2 use, L2 minus L1 (L2−L1) letter fluency, L2−L1 category fluency, L2−L1 sentence repetition, forward, backward and sequencing digit spans, and letter number sequencing. Number of L2 years of experience was calculated for each participant by subtracting their L2 AoA from their age at testing. L1 scores on the fluency and sentence repetition tasks were subtracted from L2 scores to give measures of relative proficiency balance between the languages. Scores close to zero reflect balanced L1 and L2 proficiency. Larger positive scores reflect greater L2 proficiency relative to L1 and larger negative scores reflect greater L1 proficiency relative to L2. Given that L2 experience can influence L1 proficiency, it is arguably more meaningful to assess a bilingual's language proficiency by taking into account both their L1 and L2 instead of studying either language in isolation. Thus, L1 scores were subtracted from L2 to account for both L1 and L2 proficiency. Variables were centered at zero and scaled to have unit variance using "center" and "scale" arguments in the "prcomp" function. The first three principal components (PCs) accounted for 66.9% of the total variance and were found to reflect language proficiency, working memory performance, and L2 years of experience, respectively (see Table 2 for contributions of variables to each principal component). Pearson correlations were then conducted ("cor.test" function) between each of the PCs and the N400 effect amplitude in L2 quiet and noise conditions, as well as between the three PCs and behavioural accuracy in L2 noise conditions.

Results

Behavioural accuracy

All groups were more accurate on high-constraint sentences ($M = 95.2\%$, $SE = .51$) compared to low-constraint sentences ($M = 84.2\%$, $SE = 1.18$; $F(1, 54) = 150.06$, $p < .001$, $\eta^2 = .141$). Similarly, all groups were more accurate in quiet compared to noise (quiet: $M = 96.3\%$, $SE = .45$; noise: $M = 83.1\%$, $SE = 1.45$; $F(1, 54) = 81.18$, $p < .001$, $\eta^2 = .207$), and this difference was exaggerated for low- compared to high-constraint sentences, as well as for L2 compared to L1 sentences (Figure 2; $F(1, 54) = 67.55$, $p < .001$, $\eta^2 = .063$; $F(1, 54) = 5.10$, $p = .028$, $\eta^2 = .001$; respectively). A main effect of language was observed ($F(1, 54) = 13.33$, $p = .001$, $\eta^2 = .010$), which interacted with group ($F(1, 54) = 3.69$, $p = .031$, $\eta^2 = .005$) such that performance was overall less accurate in L2 compared to L1 for the late bilinguals only. See Table 3 for means and standard errors.

N400

General observations

The waveforms depicting the amplitude and latency effects of the N400 are shown in Figure 3. An N400, larger for the low- compared to high-constraint conditions, is evident beginning around 200 ms and peaking between approximately 300 and 400 ms.

Amplitude

A main effect of constraint was observed: amplitudes were more negative for low- compared to high-constraint sentences (LC: $M = -6.7$ µV, $SE = .41$; HC: $M = -3.4$ µV, $SE = .31$; $F(1, 54) = 145.30$, $p < .001$, $\eta^2 = .140$). Amplitudes were more negative in quiet compared to noise (quiet: $M = -5.5$ µV, $SE = .40$; noise: $M = -4.6$ µV, $SE = .35$; $F(1, 54) = 6.61$, $p = .013$, $\eta^2 = .009$). There was no effect of language ($F(1, 54) = 3.37$, $p = .072$, $\eta^2 = .004$). However, there was a language by group interaction such that amplitudes were more negative in L2 compared to L1 for the early bilinguals only ($F(1, 54) = 4.40$, $p = .017$, $\eta^2 = .011$). A main effect of electrode indicated that peak amplitudes on the unsubtracted waveforms became increasingly more negative from CPz to Cz to FCz and Fz, with no difference between Fz and FCz, nor between Fz and Cz ($F(3, 162) = 18.40$, $p < .001$, $\eta^2 = .011$). See Table 4 for means and standard errors.

Latency

The N400 peak latency was longer in response to low- compared to high-constraint sentences (LC: $M = 400.7$ ms, $SE = 8.44$; HC: $M = 332.1$ ms, $SE = 7.97$; $F(1, 54) = 77.13$, $p < .001$, $\eta^2 = .117$), as well as in quiet compared to noise (quiet: $M = 379.9$ ms, $SE = 8.58$; noise: $M = 352.8$ ms, $SE = 7.56$; $F(1, 54) = 13.85$, $p < .001$,

https://doi.org/10.1017/S1366728920000516

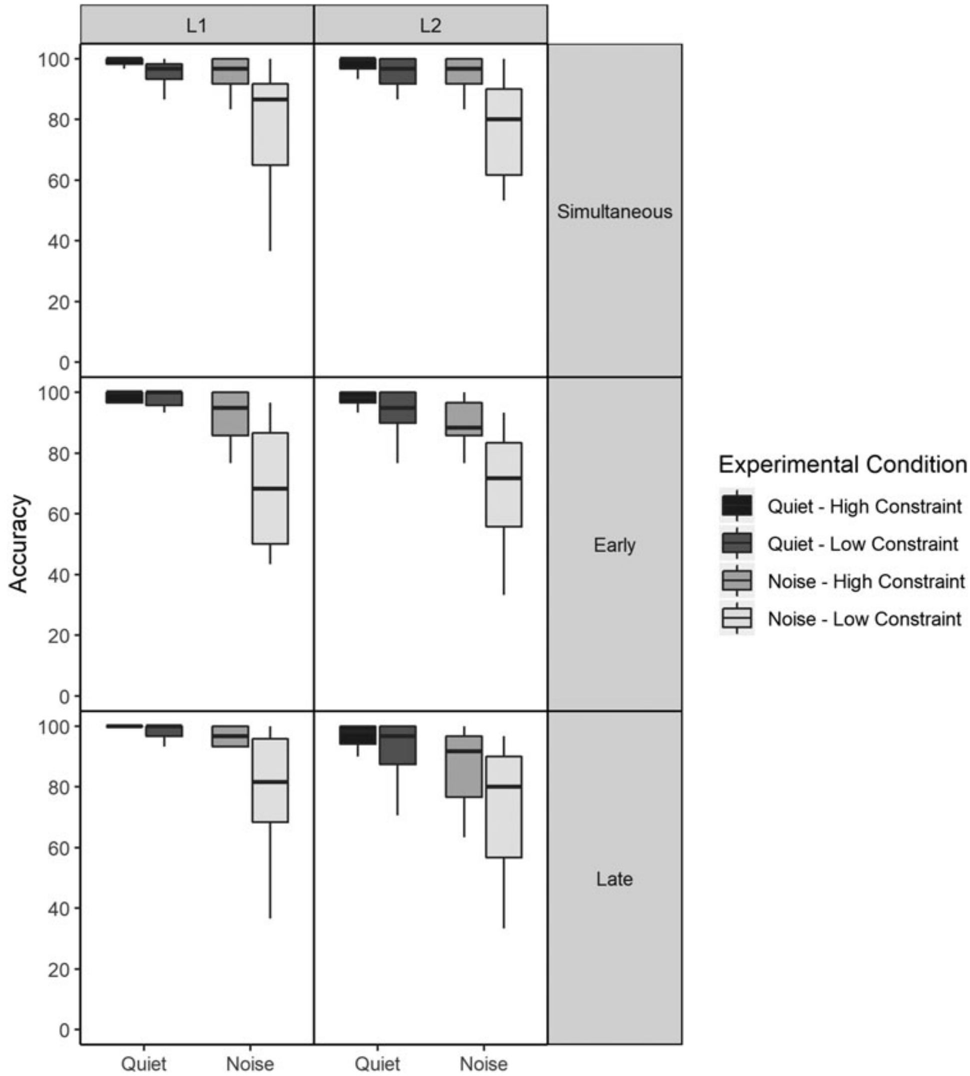

Fig. 2. Behavioural accuracy in repeating sentence terminal words for high and low constraint sentences in quiet and noise, in L1 and L2, for all three AoA groups. Accuracy is shown in percent.

Table 3. Behavioural Accuracy.

	Quiet	Noise	Simultaneous	Early	Late
HC	98.1 (.34)	92.2 (.90)	96.7 (.90)	93.9 (.85)	94.8 (.92)
LC	94.5 (.73)	74.0 (2.17)	86.3 (2.06)	81.0 (1.96)	85.4 (2.12)
L1	97.2 (.52)	85.0 (1.49)	91.8 (1.52)	88.4 (1.44)	93.0 (1.56)
L2	95.4 (.58)	81.2 (1.61)	91.2 (1.57)	86.5 (1.50)	87.3 (1.62)

Note. Mean accuracies are reported in percentages. Standard error is indicated in parentheses. HC: high constraint. LC: low constraint

$\eta^2 = .018$; respectively). No difference was observed between L1 and L2 ($F(1, 54) = .901$, $p = .347$, $\eta^2 = .002$).

N400 effect

General observations

The N400 effect waveforms and topographical distributions are shown in Figure 4 with the N400 peaking at approximately 500 ms.

Amplitude

The N400 effect was largest in amplitude at CPz and Cz and then decreased from FCz to Fz ($F(3, 162) = 13.15$, $p < .001$, $\eta^2 = .008$, $\varepsilon = .715$). There was no reliable amplitude difference between the groups ($F(2, 54) = 2.94$ $p = .061$, $\eta^2 = .029$). An electrode by group interaction indicated a group difference in the topographical distribution of the N400 effect ($F(3, 162) = 4.52$, $p = .002$, $\eta^2 = .006$, $\varepsilon = .715$; see Table 5 for means and standard errors). The N400 effect was larger (i.e., more negative) at anterior electrode sites

https://doi.org/10.1017/S1366728920000516

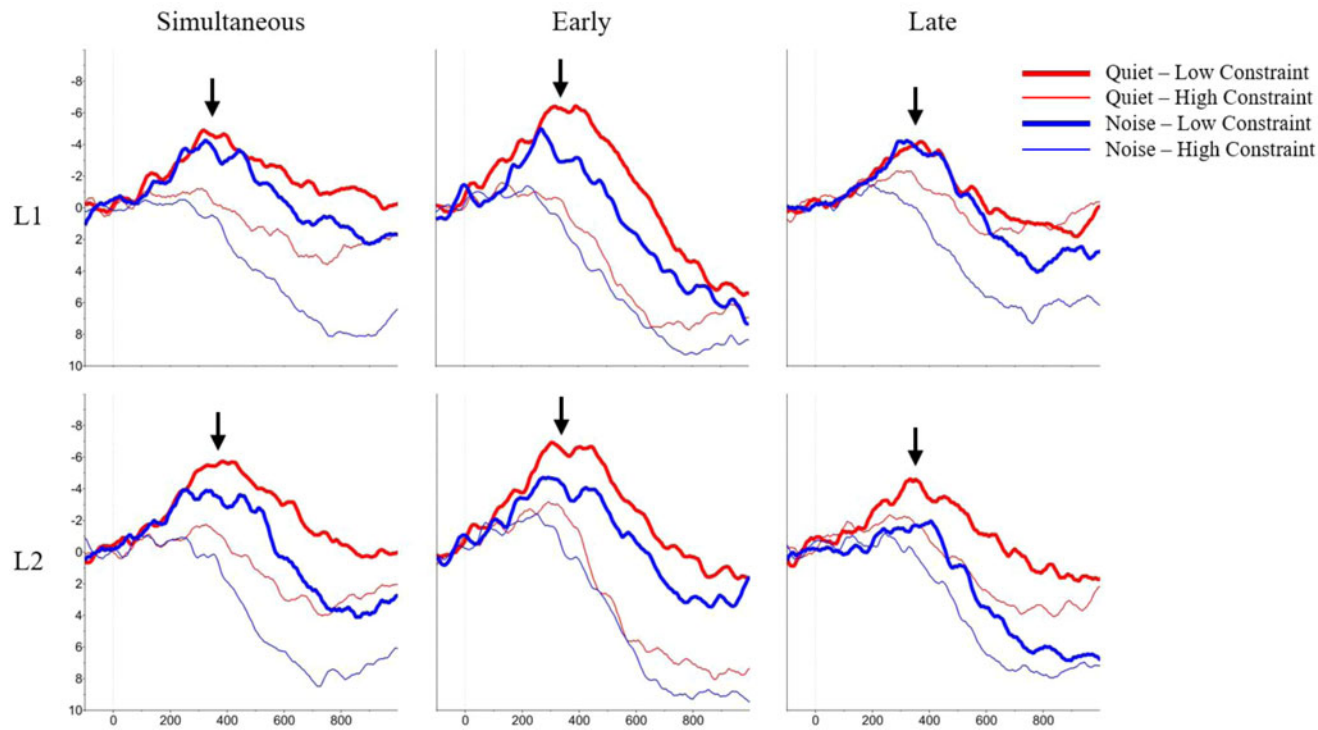

Fig. 3. High and low constraint grand average waveforms in quiet and noise, L1 and L2 for all three AoA groups. Waveforms are displayed at electrode site Cz, which is broadly representative of the effects. Arrows indicate the approximate N400 peaks. An effect of context is displayed, with larger amplitudes and longer latencies for low-constraint compared to high-constraint sentences. Larger amplitudes and longer latencies were also observed in the quiet compared to noise conditions. Moreover, more negative amplitudes were observed in L2 compared to L1 for early bilinguals only.

Table 4. Peak amplitudes of the unsubtracted waveforms.

	Simultaneous	Early	Late
L1	−4.7 (.57)	−4.8 (.55)	−4.8 (.59)
L2	−5.0 (.70)	−6.6 (.69)	−4.3 (.72)
Fz	−5.1 (.67)	−6.4 (.65)	−4.6 (.69)
FCz	−5.5 (.62)	−6.3 (.60)	−4.9 (.63)
Cz	−4.9 (.56)	−5.4 (.54)	−4.6 (.57)
CPz	−4.0 (.58)	−4.9 (.57)	−4.2 (.60)

Note. Mean amplitudes are reported in microvolts. Standard errors are reported in parentheses.

(Fz, FCz) for early compared to late bilinguals and tended to be more negative (Fz) in early compared to simultaneous bilinguals ($p = .058$). The amplitude of the N400 effect was not modulated by listening condition or language ($F(1, 54) = .03$ $p = .857$, $\eta^2 = .000$; $F(1, 54) = .45$ $p = .504$, $\eta^2 = .002$; respectively).

Latency

The N400 effect was later in quiet compared to noise (quiet: $M =$ 443.5 ms, $SE = 10.78$; noise: $M = 416.5$ ms, $SE = 10.11$; $F(1, 54) =$ 4.46, $p = .039$, $\eta^2 = .020$), but was not modulated by AoA group, nor whether participants were listening in L1 or L2 ($F(2, 54) = .27$, $p = .764$, $\eta^2 = .004$; $F(1, 54) = .09$, $p = .766$, $\eta^2 = .000$; respectively).

Topography by regions of interest

Mean amplitudes for the left and right posterior electrode clusters were more negative compared to anterior electrode clusters ($F(3, 162) = 8.255$, $p < .001$, $\eta^2 = .008$). A trend was observed for a difference between the groups in mean amplitudes across the four electrode clusters ($F(6, 162) = 2.330$, $p = .051$, $\eta^2 = .004$; see Table 5). Planned simple effects comparisons revealed that only the simultaneous bilinguals showed more negative mean amplitudes for posterior compared to anterior electrode sites.

Correlational analyses

No reliable association was found between any of the three principal components and the amplitude of the N400 effect during L2 speech perception (Table 6). Exploratory correlations were computed between the three principal components and the N400 effect during L1 speech perception. Following the False Discovery Rate (FDR) correction for multiple correlations, using the Benjamini-Hochberg procedure, no correlations were statistically significant. Another exploratory analysis was run to determine whether there was any association between the three principal components and the midline topographical differences reported above (i.e., greater anterior negativity for the early compared to the simultaneous and late bilinguals). For each participant, the amplitude of the N400 effect at CPz was subtracted from that at Fz (amplitudes were collapsed across listening and language conditions). Amplitude differences closer to zero reflect more evenly distributed topographical negativities along the midline (i.e., reflecting individuals with negative amplitudes at anterior electrode sites). In contrast, larger negative values would reflect more typical centro-parietal distributions, as seen in the simultaneous bilinguals, for example. No reliable associations were observed between any of the three principal components and the anterior−posterior topography of the N400 effect.

https://doi.org/10.1017/S1366728920000516

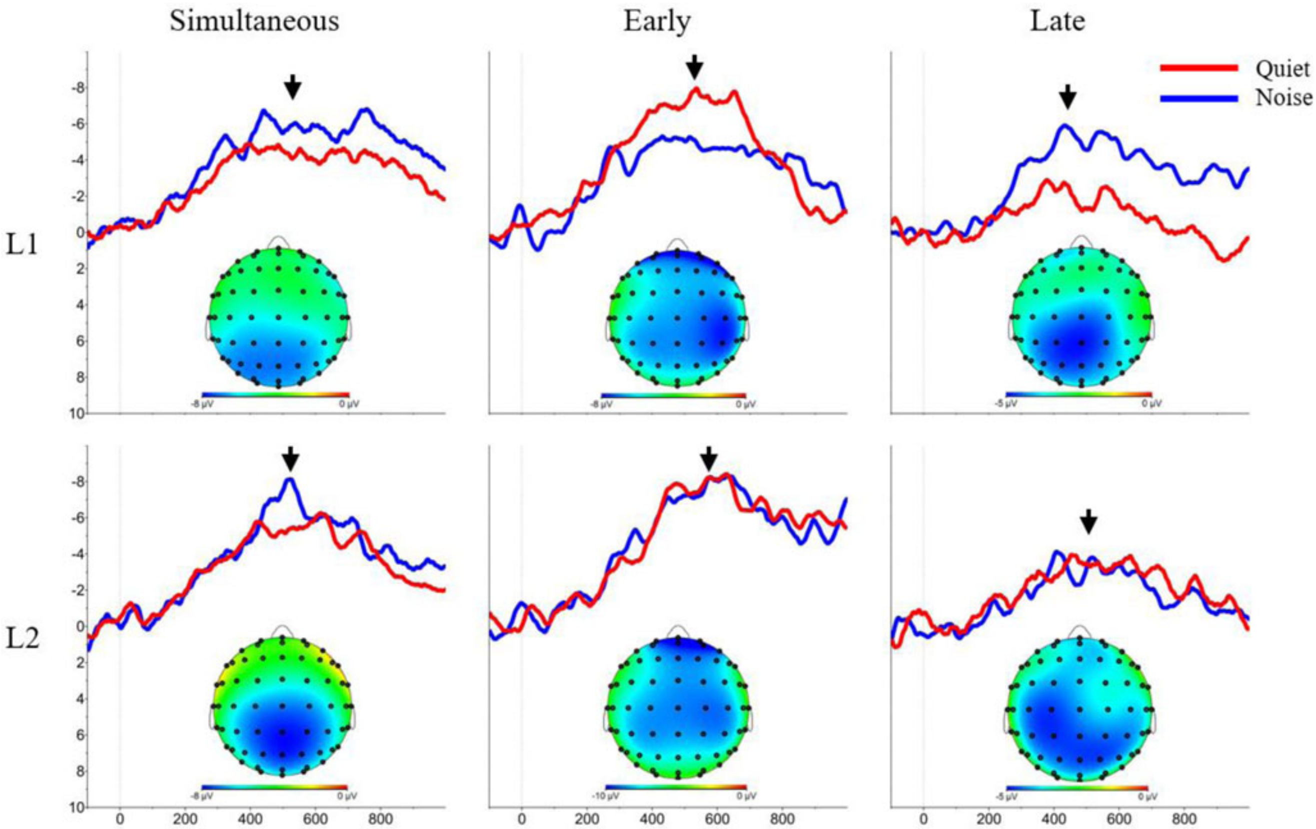

Fig. 4. Topographical distributions and grand average waveforms of the N400 effect at Cz (difference waveforms = low constraint minus high-constraint waveforms). Arrows indicate the approximate N400 effect peaks. The scalp plots depict the topographical distributions of the N400 effect (i.e., of the difference waveforms) in L1 and L2, collapsed across listening conditions, for all three groups. The N400 effect is more frontally distributed in early bilinguals compared to simultaneous and late bilinguals.

Correlations were run between the three principal components and behavioural accuracy in the two L2 noise conditions (Table 6). Following FDR correction, a statistically significant correlation was found between PC1 (relative language proficiency) and behavioural accuracy for high-constraint, L2 sentences in noise, such that greater accuracy was associated with greater L2 proficiency relative to L1 proficiency. No correlations were found between behavioural accuracy in L2 noise conditions and PC2 or PC3.

Discussion

The aim of this study was to examine whether bilinguals benefit from semantic context while perceiving speech in their L2 to the same extent as in their L1 and whether the benefit from semantic context was related to L2 language experience and/or working memory. We studied English–French bilinguals who had varying ages of L2 acquisition, but were highly proficient, regular users of both of their languages. Our behavioural and electrophysiological evidence indicates that bilinguals can benefit from semantic context in both of their languages. Moreover, semantic context use did not vary with language experience or working memory. In the following subsections, we discuss the effects of semantic context, listening condition, language, and individual differences on bilingual speech perception in noise. Although these effects will be discussed in separate subsections, some factors interacted with each other and these interactions will also be discussed.

Before interpreting our findings, it is important to note that there is still no strong consensus concerning the functional significance of the N400. There are two commonly held views in the literature. The first is that the N400 amplitude reflects semantic access during language comprehension (e.g., Kuperberg & Jaeger, 2016; Kutas & Federmeier, 2011). This view hypothesizes that higher-level information and previously encountered content allows one to predict or pre-activate potential upcoming representations, facilitating the processing of new input. Smaller N400 amplitudes are then taken to reflect this facilitated processing. In contrast, the N400 has also been viewed as reflecting lexico-semantic integration (e.g., Brown & Hagoort, 1993; Hagoort, Baggio & Willems, 2009). Within this view, larger N400 amplitudes are taken to reflect more effortful integration of a word into the preceding context. Lau, Phillips, and Poeppel (2008) note that the functional significance of the N400 is difficult to untangle because factors that facilitate lexical/semantic access may also facilitate integration and argue that neuroanatomical models of semantic processing based on fMRI and MEG studies support the semantic access view of the N400. We also note that the literature is mostly based on studies using visually presented language stimuli. Thus, it is not clear how strongly either of these views are supported by experimental work with spoken language. Despite this, it seems that the literature currently

https://doi.org/10.1017/S1366728920000516

Table 5. N400 Effect Amplitudes.

	Simultaneous	Early	Late
L1:			
Fz	−7.7 (.83)	−9.0 (.81)	−6.4 (.85)
FCz	−8.4 (.79)	−9.3 (.77)	−6.7 (.81)
Cz	−9.0 (.84)	−9.5 (.82)	−7.1 (.86)
CPz	−9.6 (.92)	−9.2 (.90)	−8.3 (.94)
L2:			
Fz	−6.2 (.87)	−9.2 (.85)	−6.4 (.90)
FCz	−8.0 (.90)	−8.8 (.87)	−6.7 (.92)
Cz	−8.8 (.91)	−8.9 (.89)	−6.9 (.94)
CPz	−9.3 (.93)	−9.0 (.91)	−7.0 (.96)
Regions of Interest:			
Left anterior	−3.6 (.75)	−4.3 (.73)	−2.6 (.77)
Right anterior	−3.7 (.77)	−5.1 (.75)	−2.5 (.79)
Left posterior	−5.4 (.67)	−5.0 (.67)	−2.9 (.70)
Right posterior	−5.4 (.73)	−5.3 (.71)	−3.0 (.75)

Note. Amplitudes are reported in microvolts. Standard errors are presented in parentheses.

Table 6. Pearson correlation coefficients.

	PC1 L1/L2 Proficiency	PC2 Working Memory	PC3 Years of Experience
N400 Effect Amplitude:			
L1 Quiet	.08 (p = .54)	−.06 (p = .68)	.20 (p = .13)
L1 Noise	.06 (p = .64)	−.12 (p = .39)	−.04 (p = .76)
L2 Quiet	−.07 (p = .61)	.07 (p = .60)	−.01 (p = .93)
L2 Noise	−.24 (p = .07)	.01 (p = .93)	.19 (p = .15)
N400 Effect Topography:			
Fz−CPz Amplitude	.01 (p = .95)	−.02 (p = .88)	−.11 (p = .41)
Behavioural Accuracy			
L2 Noise:			
High Constraint	.50 (p < .001)	−.25 (p = .06)	−.20 (p = .14)
Low Constraint	−.07 (p = .61)	−.19 (p = .16)	−.14 (p = .30)

Note. The p-values shown are uncorrected for multiple correlations.

shows stronger support for the semantic access view and we will interpret our findings accordingly.

Semantic context

All participants were more accurate in repeating terminal words for high- compared to low-constraint sentences, replicating the robust effect of semantic context during speech perception (e.g., Boothroyd & Nittrouer, 1988; Bradlow & Alexander, 2007; Miller et al., 1951; Shi, 2014). The context benefit in our study was greater when listening in noise compared to quiet. This suggests that participants relied more on semantic cues when the speech signal was degraded.

Our ERP results also revealed an effect of contextual constraint for all participant groups such that low-constraint sentences elicited larger (i.e., more negative) amplitudes compared to high-constraint sentences. This suggests that the terminal words of low-constraint sentences were more effortful to process compared to high-constraint sentences (e.g., Connolly et al., 1992; Hagoort & Brown, 2000; Kutas & Hillyard, 1980). Thus, the semantic context of high-constraint sentences facilitated semantic access and processing of terminal words. The N400 was also delayed following low- compared to high-constraint sentences for all three groups, suggesting that semantic access took longer for low-constraint sentences.

Listening condition

Overall, participants were more accurate in quiet compared to noise, an effect that interacted with both semantic context and language. These interactions are discussed in the ***Semantic context*** and ***L1 vs L2*** subsections.

Amplitudes of the unsubtracted waveforms were more negative in quiet compared to noise. Given that this difference was not observed on the subtracted waveforms (i.e., the difference between high-constraint and low-constraint sentence waveforms), it does not reflect a meaningful difference in the N400 effect *per se* or the effect of semantic context on processing speech in quiet and noise. This finding is, however, consistent with previous research showing larger N400 responses for more vs less intelligible speech (Obleser & Kotz, 2011) and for sentences presented in isolation vs with competing speech (Carey, Mercure, Pizzioli & Aydelott, 2014). Such studies suggest that poor signal quality can disrupt semantic processes during comprehension and is reflected in smaller average N400 amplitude in noise compared to quiet. It is also possible that the noise condition may elicit trial-by-trial N400 latency jitter due to variability in the masking effect of the babble across trials. This would lead to greater variability in the N400 latency in noise compared to quiet, resulting in overall reduced amplitudes in the averaged waveforms in the noise condition.

In the current study, the N400 peak was earlier in noise compared to quiet for both the subtracted and unsubtracted waveforms. This contrasts with previous research that has found shorter N400 latencies for quiet compared to noise (e.g., Aydelott, Dick & Mills, 2006; Connolly et al., 1992). However, there are important methodological differences between these previous studies and our study that may explain this inconsistency. For example, Aydelott and colleagues (2006) auditorily presented congruent and incongruent sentences that were acoustically intact or degraded by low-pass filtering. In contrast, the current study used auditorily presented high- and low-constraint sentences that were masked by a multi-talker babble noise. Importantly, Connolly and colleagues (1992) used similar stimuli to the current study (i.e., auditorily presented high- and low-constraint sentences and a multi-talker babble mask). However, as with the other studies, Connolly et al. examined monolingual participants. Given that bilinguals typically show slower lexical access than monolinguals (e.g., Ivanova & Costa, 2008; Shook, Goldrick, Engstler & Marian, 2015), the difference between our findings and those of Connolly et al. may be due to a difference in the populations tested. Additionally, although the exact SNR used by Connolly and colleagues was unreported, it is possible that our SNR of +1 dB was too favourable to delay the N400 in the noise conditions. However, this would be

https://doi.org/10.1017/S1366728920000516

surprising given that our SNR was sufficient to produce a behavioural effect. Visual inspection of Figure 3 suggests that our latency effect may be due to an earlier N400 peak for high-constraint sentences in noise compared to quiet. This would be consistent with the idea that semantic context facilitates speech perception in noise. However, post-hoc analyses revealed that the comparison between N400 peak latency for high-constraint sentences in noise compared to quiet was not statistically reliable. In the absence of a monolingual group, it is difficult to determine whether the difference between our latency effect and that previously seen in the literature is due to a difference in the populations tested or the nature of the stimuli used. Future studies could address this by attempting to replicate our latency effect.

Given that our experimental trials were blocked by listening condition, our latency effect in quiet compared to noise could be a result of the different task demands elicited by our listening conditions. Listening in noise is more difficult and cognitively taxing than in quiet. Participants may have been more actively engaged in the task during the noise blocks in order to compensate for the increased task demands. In contrast, participants may have perceived the quiet conditions as being easier, leading to more passive engagement during the quiet blocks. More active engagement during noisy blocks may have led to the shorter N400 latencies in noise compared to quiet.

L1 vs L2

Although a main effect revealed greater accuracy for L1 compared to L2, an interaction indicated that this was only the case for late bilinguals (see ***Age of acquisition*** subsection). The effect of context on behavioural accuracy did not interact with language, suggesting that participants benefited from semantic context to the same extent in both of their languages. Consistent with previous research (e.g., Mayo et al., 1997; Rogers et al., 2006), our participants were more affected by noise in their L2 compared to their L1. This was evidenced by a greater difference in error rate between quiet and noise listening conditions in L2 compared to L1.

The unsubtracted waveform amplitudes were overall more negative for L2 compared to L1 sentences. Although this finding does not reflect a difference in the N400 effect *per se*, it may reflect more effortful semantic access and processing in L2 overall, compared to L1.

The N400 context effect (i.e., the difference in amplitude between the waveforms elicited by the high- and low-constraint sentences) did not differ in amplitude or latency between L1 and L2 sentences. Consistent with our behavioural results, this suggests that all groups benefited from semantic context to the same extent in both of their languages.

Previous studies have found delayed N400 latencies in L2 compared to L1 (e.g., Phillips, Segalowitz, O'Brien & Yamasaki, 2004). However, no effect of language was observed in N400 latency of the subtracted or unsubtracted waveforms, suggesting that our participants did not differ in processing speed between L1 and L2. This could be due to the high L2 proficiency of our participants. It could also be due to a difference in task demands. For example, in the study by Phillips and colleagues (2004), participants made an animacy judgment in response to visually presented word pairs in both of their languages. In contrast, the task in our current study involved perceiving L1 and L2 sentences in noise and is arguably more challenging than the animacy judgment task. This added effort may have resulted in a delayed N400 overall, washing out the latency effect previously reported in the literature. Consistent with this, the N400 latency in response to L1 sentences was about 60 ms later in the current study compared to that reported by Phillips et al. (2004).

Age of acquisition

As mentioned above, simultaneous and early bilinguals performed with similar accuracy in both of their languages. In contrast, late bilinguals were overall less accurate in their L2 compared to their L1, although they were still highly accurate. This is to be expected given that late bilinguals have likely had less experience with their L2 compared to simultaneous and early bilinguals. However, this effect did not interact with contextual constraint, suggesting all groups benefited from having semantic context in both of their languages to the same extent. By contrast, Kousaie and colleagues (2019) observed that only simultaneous and early bilinguals benefited from semantic context in their L2, despite similar experimental methodology and some overlap in participants across our two studies. Notably, the paradigm used by Kousaie and colleagues involved a much lower SNR (−6 dB) compared to our study (+1 dB). Therefore, late bilinguals may be able to benefit from semantic context when listening in higher SNR conditions, but may not in more difficult listening conditions.

The topographical distribution of the N400 effect differed between groups. Examining left and right anterior and posterior regions revealed that the N400 effect is distributed more posteriorly for simultaneous bilinguals. This more posterior topographical distribution is typical of the N400 effect seen in native monolingual listeners (e.g., Connolly, Stewart & Phillips, 1990; Van den Brink, Brown & Hagoort, 2006). In contrast, the early and late bilinguals in our study showed a more evenly distributed N400 effect across the four lateralized electrode ROIs. These topographical differences between our groups have two implications. First, the more posterior distribution of the simultaneous bilinguals suggests that they may be processing semantic context similarly to monolinguals. Second, the more distributed topography seen in the early and late bilinguals indicates that these two groups may be recruiting additional neural resources to support successful task performance compared to the simultaneous bilinguals. Moreover, the early bilinguals showed a stronger negativity at anterior midline electrode sites compared to the simultaneous and late bilinguals. This further suggests that the early bilinguals may be recruiting additional neural resources while processing low-constraint sentences compared to both simultaneous and late bilinguals. This is consistent with our finding that early bilinguals performed similarly in both their languages whereas late bilinguals performed worse in L2 compared to L1. Thus, recruitment of additional neural resources may be successfully supporting the maintained performance by the early bilinguals in their L2.

Due to the low spatial resolution of ERP measurements, we cannot comment on the neural sources underlying these topographical differences. However, the more distributed topography of the early and late bilinguals, as well as the stronger anterior negativity of the early bilinguals, may be consistent with literature implicating the left inferior frontal cortex in speech processing and semantics (for reviews see Lau et al., 2008; Peelle, 2018, 2019). Increased activity in left inferior frontal cortex, as indexed by the BOLD response, has been associated with semantic processing of auditorily presented sentences (e.g., Cardillo, Aydelott, Matthews & Devlin, 2004). Some researchers have

https://doi.org/10.1017/S1366728920000516

proposed that the inferior frontal cortex mediates top-down, controlled semantic retrieval and selection of lexical representations (Lau et al., 2008). Others have proposed that the inferior frontal cortex is involved in lexico-semantic integration (e.g., Hagoort, 2013). Therefore, it is possible that the topographical differences reported above reflect a greater recruitment of inferior frontal cortex in early and late bilinguals.

Individual differences

Language experience

Previous studies have found that L2 AoA (e.g., Mayo et al., 1997; Shi, 2009) and L2 proficiency (e.g., Bradlow & Alexander, 2007; Gor, 2014) moderate the benefit of semantic context, measured using behavioural accuracy, while perceiving L2 speech-in-noise. Consistently, we found that speech perception of L2 sentences in noise was associated with participants' relative balance of L1 and L2 proficiency. Specifically, greater accuracy in perceiving L2 sentence terminal words in noise was associated with greater L2 proficiency relative to L1. In contrast, we found that the amplitude of the N400 effect was not associated with individual differences in participants' relative L1 and L2 proficiency. Notably, the N400 in our study is based only on trials successfully perceived, whereas behavioural accuracy necessarily reflects both success and failure of speech perception. Thus, individual differences in L2 language experience may be more strongly associated with the success (or failure) of perceiving words but may not reliably moderate semantic processing (as indexed by the N400).

Years of L2 experience were not associated with behavioural accuracy or the N400 effect in the current study; however, the simultaneous, early, and late bilinguals did differ in their behavioural accuracy of L1 and L2 sentences, with only the late bilinguals performing more poorly in L2 versus L1. Thus, our behavioural data are consistent with previous behavioural studies that have found a relationship between AoA and semantic context use while perceiving L2 speech-in-noise (e.g., Mayo et al., 1997; Shi, 2009). It therefore appears that L2 AoA modulates behavioural performance during speech perception in noise but not the N400 effect. It is possible that this effect with behaviour was not observed in our correlational analyses because we correlated behaviour with years of experience instead of AoA. Given that one-third of our sample consists of simultaneous bilinguals who have an L2 AoA of 0 years, raw AoA data would not have provided a sufficient data distribution for conducting meaningful correlational analyses. Thus, L2 years of experience was used as a rough proxy of AoA. However, the mapping between these two variables may not be perfect because years of experience is necessarily confounded by age. It is also important to note that we did not use a pure measure of years of experience. Instead, we used a principal component (i.e., PC3). Although L2 years of experience makes up the majority of the loading on PC3, this component still has contributions, albeit smaller, from other variables.

Working memory

The amplitude of the N400 effect in our study was not related to individual differences in working memory, despite previous studies showing an association between working memory performance and behavioural accuracy on L1 SPIN tasks (e.g., Ingvalson et al., 2015; Millman & Mattys, 2017). However, our finding is consistent with some studies reporting no effect of working memory on behavioural performance during non-native SPIN tasks (e.g., Kilman, Zekveld, Hällgren & Rönnberg, 2014; Schmidtke, 2016). Notably, working memory may play a more important role during speech perception in older adult populations and populations with hearing impairments compared to normal-hearing, young adults (Füllgrabe & Rosen, 2016).

Limitations

First, the correlations in the current study were run using composite variables from a PCA and not the raw individual difference variables. A PCA was done in this study for multiple reasons. First, it allowed us to more simply test our hypotheses and increase the statistical power of our tests by reducing both the number of variables examined and the number of tests conducted. Second, it is, at the present time, impossible to accurately capture the bilingual language experience with a single test due to its inherent complexity. The PCA allowed us to examine the language experience of our participants more efficiently by providing a smaller number of variables that reflect the underlying constructs common to the multiple facets of language experience assessed. Despite these benefits, some variance is necessarily lost in computing composite variables. The principal components used in our analyses accounted for about 67% of the variance in the raw individual difference variables. This must be considered when interpreting our correlational analyses.

Second, it is possible that differences between the English and French stimuli may have influenced our findings, including the previously noted difference in the number of syllables between English and French terminal words. As noted, this difference had a small and inconsistent effect on behavioural performance and may have contributed to some variability in our data.

Despite these limitations, it is clear from our findings that proficient bilinguals do benefit from semantic context while perceiving speech-in-noise in both of their languages. This contrasts with previous studies reporting that bilinguals do not benefit from semantic context in their L2 (Golestani et al., 2009; Hervais-Adelman et al., 2014). Importantly, these previous studies used small sample sizes of moderately proficient, late bilinguals. Their SPIN task involved semantically-related and unrelated word pair stimuli. In contrast, our study examined a large sample of highly proficient bilinguals with a range of AoAs using a more ecologically valid SPIN task with sentence stimuli. Consequently, our findings more strongly suggest that bilinguals do use semantic context while perceiving speech-in-noise in both of their languages.

Conclusion

To our knowledge, this is one of the first studies to examine the electrophysiology of semantic context use at the sentence level during speech perception in noise in bilinguals. It is also the first study, to our knowledge, to examine the association between an electrophysiological measure of sentence context use during bilingual speech perception in noise and individual differences in bilingual language experience and working memory. Based on the behavioural and electrophysiological evidence presented above, proficient bilinguals benefit from semantic context while perceiving speech in both of their languages. However, although they do still benefit from semantic context in their L2, bilinguals who learn their L2 at a later age (after age 6 in this study) may be more limited in their use of semantic context to facilitate semantic processing while perceiving L2 speech compared to bilinguals who learn their L2 earlier in life (before age 5). The more effective

https://doi.org/10.1017/S1366728920000516

use of semantic context by bilinguals who acquired their L2 early in life appears to be supported by recruiting additional neural resources as compared to late bilinguals. Our findings also suggest that bilinguals who learn their two languages from birth may process semantic context during speech perception similarly to monolinguals. Moreover, the most reliable individual difference variable influencing our findings was participants' age of second language acquisition, which indicated differences between simultaneous, early, and late bilinguals in the recruitment of neural resources.

Acknowledgements. This work was supported by a Master's Award from the Fonds de Recherche du Québec – Santé and a Graduate Student Stipend from the Centre for Research on Brain, Language and Music to KC, a grant from The Natural Sciences and Engineering Research Council of Canada (NSERC; grant # 203751) to NP, a team grant from the Fonds de Recherche du Québec – Société et Culture (FRQSC, grant #171276), and the Blema and Arnold Steinberg Family Foundation.

References

Akker E and Cutler A (2003) Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition* **6**(2), 81–96.
Aydelott J, Dick F and Mills DL (2006) Effects of acoustic distortion and semantic context on event-related potentials to spoken words. *Psychophysiology* **43**(5), 454–464.
Bialystok E, Craik FIM and Luk G (2008) Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics* **21**(6), 522–538.
Bilger RC, Nuetzel JM, Rabinowitz WM and Rzeczkowski C (1984) Standardization of a test of speech perception in noise. *Journal of Speech, Language, and Hearing Research* **27**(1), 32–48.
Boothroyd A and Nittrouer S (1988) Mathematical treatment of context effects in phoneme and word recognition. *The Journal of the Acoustical Society of America* **84**(1), 101–114.
Borghini G and Hazan V (2018) Listening effort during sentence processing is increased for non-native listeners: a pupillometry study. *Frontiers in Neuroscience* **12**, 152.
Bradlow AR and Alexander JA (2007) Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America* **121**(4), 2339–2349.
Brown C and Hagoort P (1993) The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience* **5**(1), 34–44.
Cardillo ER, Aydelott J, Matthews PM and Devlin JT (2004) Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *Journal of Cognitive Neuroscience* **16**(9), 1552–1561. https://doi.org/10.1162/0898929042568523
Carey D, Mercure E, Pizzioli F and Aydelott J (2014) Auditory semantic processing in dichotic listening: Effects of competing speech, ear of presentation, and sentential bias on N400s to spoken words in context. *Neuropsychologia* **65**, 102–112. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2014.10.016
Clahsen H and Felser C (2006) How native-like is non-native language processing? *Trends in Cognitive Sciences* **10**(12), 564–570.
Cohen G and Faulkner D (1983) Word recognition: Age differences in contextual facilitation effects. *British Journal of Psychology* **74**(2), 239–251.
Coltheart M (1981) The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology* **33**(4), 497–505.
Connolly JF, Phillips NA, Stewart SH and Brake WG (1992) Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language* **43**(1), 1–18.
Connolly JF, Stewart SH and Phillips NA (1990) The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language* **39**(2), 302–318.
Cooke M, Garcia Lecumberri ML and Barker J (2008) The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America* **123**(1), 414–427.
Davies M (2008) *The corpus of contemporary American English*. BYE, Brigham Young University.
Dijkgraaf A, Hartsuiker RJ and Duyck W (2017) Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition* **20**(5), 917–930.
Füllgrabe C and Rosen S (2016) On the (un) importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Frontiers in Psychology* **7**, 1268.
Gaskell MG and Marslen-Wilson WD (1997) Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* **12**(5–6), 613–656.
Golestani N, Rosen S and Scott SK (2009) Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism: Language and Cognition* **12**(3), 385–392.
Gor K (2014) Raspberry, not a car: context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Frontiers in Psychology* **5**, 1449.
Hagoort P (2013) MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2013.00416
Hagoort P, Baggio G and Willems RM (2009) Semantic unification. In *The cognitive neurosciences, 4th ed.* (pp. 819–836). MIT press.
Hagoort P and Brown CM (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.
Hervais-Adelman A, Pefkou M and Golestani N (2014) Bilingual speech-in-noise: Neural bases of semantic context use in the native language. *Brain and Language* **132**, 1–6.
Holcomb PJ and Neville HJ (1991) Natural speech processing: An analysis using event-related brain potentials. *Psychobiology* **19**(4), 286–300.
Hwang H, Shin J and Hartsuiker RJ (2018) Late Bilinguals Share Syntax Unsparingly Between L1 and L2: Evidence From Crosslinguistically Similar and Different Constructions. *Language Learning* **68**(1), 177–205.
Ingvalson EM, Dhar S, Wong PCM and Liu H (2015) Working memory training to improve speech perception in noise across languages. *The Journal of the Acoustical Society of America* **137**(6), 3477–3486.
IBM Corp (2015) IBM SPSS statistics, version 23.0. IBM Corp Armonk, NY.
Ivanova I and Costa A (2008) Does bilingualism hamper lexical access in speech production? *Acta Psychologica* **127**(2), 277–288.
Jankowiak K and Rataj K (2017) The N400 as a window into lexico-semantic processing in bilingualism. *Poznan Studies in Contemporary Linguistics* **53**(1), 119–156.
Kalikow DN, Stevens KN and Elliott LL (1977a) Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America* **61**(5), 1337–1351.
Kalikow D, Stevens K and Elliott J (1977b) The speech perception in noise (SPIN) test. *The Journal of the Acoustical Society of America* **61**, 1337–1351.
Kilman L, Zekveld A, Hällgren M and Rönnberg J (2014) The influence of non-native language proficiency on speech perception performance. *Frontiers in Psychology* **5**, 651.
Kousaie S, Baum S, Phillips NA, Gracco V, Titone D, Chen J.-K., Chai XJ and Klein D (2019) Language learning experience and mastering the challenges of perceiving speech in noise. *Brain and Language* **196**, 104645. https://doi.org/10.1016/J.BANDL.2019.104645
Kuperberg GR and Jaeger TF (2016) What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* **31**(1), 32–59.
Kutas M and Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology* **62**, 621–647.
Kutas M and Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* **207**(4427), 203–205.
Lau EF, Phillips C and Poeppel D (2008) A cortical network for semantics: (de) constructing the N400. *Nature Reviews Neuroscience* **9**(12), 920.

Lecumberri MLG and Cooke M (2006) Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America* **119**(4), 2445–2454.

Lecumberri MLG, Cooke M and Cutler A (2010) Non-native speech perception in adverse conditions: A review. *Speech Communication* **52**(11–12), 864–886.

León-Cabrera P, Rodríguez-Fornells A and Morís J (2017) Electrophysiological correlates of semantic anticipation during speech comprehension. *Neuropsychologia* **99**, 326–334.

Lorge I and Thorndike EL (1952) *The teacher's word book of 30 000 words*. New York: Columbia University Press.

Marslen-Wilson W and Tyler LK (1980) The temporal structure of spoken language understanding. *Cognition* **8**(1), 1–71.

Mattys SL, Davis MH, Bradlow AR and Scott SK (2012) Speech recognition in adverse conditions: A review. *Language and Cognitive Processes* **27**(7–8), 953–978.

Mayo LH, Florentine M and Buus S (1997) Age of second-language acquisition and perception of speech in noise. *Journal of Speech, Language, and Hearing Research* **40**(3), 686–693.

Miller GA, Heise GA and Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology* **41**(5), 329.

Millman RE and Mattys SL (2017) Auditory Verbal Working Memory as a Predictor of Speech Perception in Modulated Maskers in Listeners With Normal Hearing. *Journal of Speech, Language, and Hearing Research* **60**(5), 1236–1245.

Navarra J, Sebastián-Gallés N and Soto-Faraco S (2005) The perception of second language sounds in early bilinguals: new evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance* **31**(5), 912.

New B (2006) Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles* (*TALN* 2006).

New B, Pallier C, Ferrand L and Matos R (2001) Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary fren[ch]: LEXIQUE™. *L'année Psychologique* **101**(3), 447–462.

Obleser J and Kotz SA (2011) Multiple brain signatures of integration in the comprehension of degraded speech. *Neuroimage* **55**(2), 713–723.

Parbery-Clark A, Marmel F, Bair J and Kraus N (2011) What subcortical–cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience* **33**(3), 549–557.

Peelle JE (2018) Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear and Hearing* **39**(2), 204.

Peelle JE (2019) The neural basis for auditory and audiovisual speech perception.

Phillips NA, Segalowitz N, O'Brien I and Yamasaki N (2004) Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics* **17**(2–3), 237–262.

Pichora-Fuller MK, Kramer SE, Eckert MA, Edwards B, Hornsby BWY, Humes LE, Lemke U, Lunner T, Matthen M, Mackersie CL, Naylor G, Phillips N, Richter M, Rudner M, Sommers M, Tremblay KL and Wingfield A (2016) Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, 37. Retrieved from https://journals.lww.com/ear-hearing/Fulltext/2016/07001/Hearing_Impairment_and_Cognitive_Energy___The.2.aspx

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rogers CL, Lister JJ, Febo DM, Besing JM and Abrams HB (2006) Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing. *Applied Psycholinguistics* **27**(3), 465–485.

Ryan C (2013) Language Use in the United States: 2011. American Community Survey Reports, *US Census Bureau*.

Sandoval TC, Gollan TH, Ferreira VS and Salmon DP (2010) What causes the bilingual disadvantage in verbal fluency? The dual-task analogy. *Bilingualism: Language and Cognition* **13**(2), 231–252.

Schmidtke J (2016) The bilingual disadvantage in speech understanding in noise is likely a frequency effect related to reduced language exposure. *Frontiers in Psychology* **7**, 678.

Semel E, Wiig EH and Secord WA (2003) *Clinical evaluation of language fundamentals, (CELF-4).(5-8 age) The Psychological Corporation*. A Harcourt Assessment. Inc.: San Antonio, USA.

Shi L.-F. (2009) Normal-hearing English-as-a-second-language listeners' recognition of English words in competing signals. *International Journal of Audiology* **48**(5), 260–270.

Shi L.-F. (2010) Perception of acoustically degraded sentences in bilingual listeners who differ in age of English acquisition. *Journal of Speech, Language, and Hearing Research* **53**(4), 821–835.

Shi L.-F. (2014) Measuring effectiveness of semantic cues in degraded English sentences in non-native listeners. *International Journal of Audiology* **53**(1), 30–39.

Shook A, Goldrick M, Engstler C and Marian V (2015) Bilinguals show weaker lexical access during spoken sentence comprehension. *Journal of Psycholinguistic Research* **44**(6), 789–802.

Van den Brink D, Brown CM and 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.

Wechsler D (2008) *Wechsler Adult Intelligence Scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: The Psychological Corporation.

Zekveld AA, Heslenfeld DJ, Johnsrude IS, Versfeld NJ and Kramer SE (2014) The eye as a window to the listening brain: neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage* **101**, 76–86.

Zirnstein M, van Hell JG and Kroll JF (2018) Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition* **176**, 87–106.