GAME-BASED SECOND LANGUAGE VOCABULARY TRAINING: IMPLICATIONS FOR LEARNING OUTCOMES AND BRAIN FUNCTION

by

Kiera G.R. O'Neil

Submitted in partial fulfillment of the requirements for the degree of
Master of Science

at

Dalhousie University
Halifax, Nova Scotia
July 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study sought to investigate the influence of training strategy on adult second language vocabulary learning and the development and integration of second language semantic networks in the brain. English speaking participants completed a 10-day game-based training program ("LANGA"), in which they learned 72 French vocabulary words via two training strategies. Words were either paired directly with their associated meaning ("explicit" strategy) or taught in the form of a three-word sentence ("implicit" strategy). Following training, we observed increased recall and recognition of French vocabulary, as well as a significant N400 effect for mismatched word pairs, although this was limited to explicitly learned words. Images semantically related to learned words produced a reduced N400 compared to mismatched pairs, similar to effects observed native French speakers. Our results attest to the plasticity of adult brains and provide evidence for rapid integration of words into existing semantic networks, in a late learned language.

LIST OF ABBREVIATIONS USED

L2	Second Language
L1	First Language
EEG	Electroencephalography
ERP	Event-Related Potential
BIA+	Bilingual Interactive Activation model+
IPNP	International Picture Naming Project
ANOVA	Analysis of Variance

ACKNOWLEDGEMENTS

First I would like to thank Dr. Aaron Newman, for his support and mentorship, as well as everyone in the NCIL lab who helped out when they could. I would also like to thank my committee members Dr. Helene Deacon and Dr. Lauren Sculthorpe-Petley. Special thanks go to Dr. Cheryl Frenck-Mestre, who not only welcomed me into her beautiful home while I was in France, but who also worked exceptionally hard so that I could get this project finished, and provided so much encouragement and thoughtful guidance along the way. Also thanks to Kastoori Kalaivanan and Aurélie Lagarrigue for their help with running the study, and everyone else at the LPL. Finally, thanks to my friends and family, especially my mother, because I wouldn't be here without her.

CHAPTER 1: INTRODUCTION

1.1 Second Language Acquisition

How to adults acquire second languages? Learning a second language (L2) as an adult can be a significant undertaking, that involves learning entirely new grammatical rules, phonology and vocabulary. Learning vocabulary, which may seem straightforward when compared to learning a new grammatical or phonological system, can pose a challenge due to the number of vocabulary words that need to be mastered to participate in regular conversation; estimates range from 2,000-5,000 words that must be known in order to understand ~95% of conversation (Adolphs & Schmitt, 2003). Although both grammar and vocabulary are critical components of language, in practice effective communication may rely less on the correct application of grammatical rules and more on adequate and effective vocabulary knowledge (Verhoeven & de Jong, 1992). The current study sought to address the issue of vocabulary learning, specifically, under what conditions L2 vocabulary learning is most effective.

Is it better to learn a new word in context, or by linking a new word directly to its meaning? Studies investigating L1 vocabulary learning in children have shown that learning unfamiliar words in context is an effective way to acquire new vocabulary (Nagy et al., 1985). In L2 research, linking a word directly to its meaning is often called 'associative learning' as it involves pairing an L2 word with its first language (L1) equivalent (Prince, 1996). In contrast, learning an L2 word in context can involve reading passages of text that contain novel words, and deducing their meaning based on the passage context, or learning a new language in an immersive environment (although immersion involves direct, associative learning as well). There is a general assumption among language instructors that relying on associative learning might cause learners to link new words too tightly to L1 words, limiting L2 autonomy by forcing learners to relay L2 input and output through their L1 (Prince, 1996). Therefore, many courses and books introduce new material in a contextualized format.

Although a small number of studies have reported little benefit of learning words in context (Prince, 1996; Seibert, 1930), introducing novel L2 content within context as a method to promote learning outcomes has found support in a number of other studies. Pickering (1982), who investigated vocabulary learning in context, found that learning in context was slightly better than learning in a context-independent manner. Webb (2008) again investigated learning in context, by comparing retention of L2 words learned in paragraphs containing either high or low numbers of contextual cues, and found that words encountered with high numbers of contextual cues showed better retention. Kang (1995) compared L2 contextual learning to three other non-context based training strategies and found retention of contextually learned words to be superior to all other methods. Extensive reading, a form of context learning, has been shown to be a highly effective way to acquire vocabulary (Pigada & Schmitt, 2006). Prince (2012) investigated L2 vocabulary learning in context, by comparing words learned within a passage that contained narrative framework vs words learned within a passage that had unrelated sentences. Words learned within a narrative frame showed better recall immediately following training, compared to words learned in unrelated sentences. The research suggests that context learning, in a variety of forms, can be an effective way to learn L2 vocabulary, and may be superior to other non-contextual methods.

The process of acquiring new second language vocabulary in context (typically within a passage of text) is called *incidental vocabulary learning* (Huckin & Coady, 1999). When L2 learners encounter a novel word in context, they can employ several strategies to deal with the unknown word – they may ignore it, consult a dictionary, ask an instructor or attempt to infer its meaning from context, which is called lexical inferencing (Nassaji, 2003). Lexical inferencing is the strategy most widely used by L2 learners, and involves making informed guesses regarding a word's meaning based on current linguistic cues, awareness of context and prior linguistic knowledge (Nassaji, 2003). Successful lexical inferencing depends on several factors, including the nature of the word, the nature of the text that contains the word, the information available in the text, the importance of the word in understanding the text and the degree of effort involved in the task (Nassaji, 2003). *Incidental vocabulary learning* can be enhanced by providing glosses in the form of written translations, pictures, videos or sounds, which accompany the text and explain an unknown word's meaning (Yoshii, 2006). While there is behavioural evidence to suggest contextual L2 vocabulary learning may be superior to associative vocabulary learning, few studies have used both behavioural and neurophysiological measures to longitudinally compare learning via each method.

The study reported in this thesis employed a computerized L2 vocabulary-training program and focused on two training strategies, based on associative learning and context learning. The first we labelled "explicit" training, based on associative learning, in which novel L2 words were paired directly with a visual depiction, and the second is labelled "implicit" training, based on context learning, in which novel L2 words were presented within a three-word sentence, accompanied by a visual depiction of the sentence.

Additionally, on a behavioural level, it is possible that different training strategies may not produce quantifiable differences (McLaughlin, Osterhout, & Kim, 2004). For this reason, learning via each strategy was also assessed using a neurophysiological measure, electroencephalography (EEG). EEG also allowed us to track changes in brain activity that occur as a function of acquiring second language vocabulary. The following sections will explain EEG, and how it can be used to investigate second language processing, acquisition and integration into existing language networks.

1.2 EEG and the N400

EEG, which reflects synchronized post-synaptic activity in cortical pyramidal neurons (Luck, 2014), is a neuroimaging method that works by non-invasively recording the brain's electrophysiological activity, via electrodes placed on the scalp. Event-related-potentials (ERPs) are patterns of activity in the EEG signal time-locked to stimulus presentation. A critical advantage of the ERP technique is that it allows for the measurement of neural processes as they unfold over time, given its high temporal resolution. The ERP waveform can be broken down into components, which are portions with characteristic amplitude, morphology, timing and topography, as well as characteristic sensitivity to certain experimental manipulations (Kutas & Federmeier, 2011). Of critical importance to the current study is an ERP component called the N400, which is thought to reflect the processing of semantic information (for review, see Kutas & Federmeier, 2011).

The N400 is a broadly distributed, parietally maximal, negative potential that occurs roughly 400 ms post-stimulus onset. It was first reported by Kutas & Hillyard

(1980), who showed that the N400 effect was elicited when participants were shown sentences containing a semantically anomalous ending, i.e. *I take my coffee with cream and dog*. The amplitude of the N400 is sensitive to the degree of violation; words with low cloze probability (the percentage of individuals who would supply a particular word given the current sentential context) elicit larger N400 amplitudes than words with high cloze probability (Hillyard & Kutas, 1983). The N400 is not only elicited in sentential contexts; it can also be seen when the relationship between pairs of words is manipulated. For example, in lexical priming tasks, when the target word is related in some aspect to a preceding word (i.e. identically, associatively, semantically, or categorically), the N400 effect is sensitive to the degree of relatedness between the pairs – related items show reduced N400 amplitudes relative to less related items (Kutas & Federmeier, 2011). Generally, N400 amplitude is higher when semantic expectations based on previous context are violated or when difficulties integrating semantic content are encountered, and lower when semantic integration is facilitated through prior context – be it sentential context or activation of related words or concepts. Finally, the N400 is not limited to one input modality and can be observed in both written and auditory language tasks, as well as American Sign Language and even pictures (Kutas & Federmeier, 2011).

With an understanding that the N400 is an index of semantic processing, several studies have sought to identify its neural underpinnings (reviewed in Lau, Phillips, & Poeppel, 2008; Van Petten & Luka, 2006). Intra-cranial recordings taken from epileptic patients prior to surgical intervention have revealed a distributed group of brain areas that appear to be co-active with scalp recorded N400 potentials. Regions include the anterior medial temporal lobe, middle and superior temporal areas, inferior temporal areas and prefrontal areas, in both the left and right hemispheres (Guillem et al., 1996; Kutas & Federmeier, 2011; McCarthy et al., 1995; Nobre & Mccarthy, 1995). These brain areas are known to be part of a widespread network responsible for semantic memory and semantic processing. Generally, neuroimaging data suggest that the N400 does not reflect activity from a single static source, but instead a distributed semantic network that starts with a wave of activity (~250 ms) in the posterior left superior temporal gyrus, and then continues forward and ventrally to the left temporal lobe (~365 ms) and from there (370 – 500 ms) to the right anterior temporal lobe and both frontal lobes (Kutas & Federmeier, 2011).

1.3 The N400 and Bilingualism

A large body of research has investigated the N400 in bilinguals, in order to determine differences and similarities in the processing of first and second languages, and how this may be reflected in brain activity (reviewed in Moreno et al., 2008, Frenck-Mestre, Sneed-German & Foucart, 2014). The research has compared L2 and L1 processing within bilinguals and between bilinguals and monolinguals. Characteristics of the N400, including onset, peak latency and amplitude, as well as factors that may influence these characteristics, such as age of acquisition and proficiency, have been examined using a variety of semantic violation tasks. Generally, the research suggests that the processing of L1s and L2s can differ, as reflected by modulations of N400 characteristics, most notably, onset, peak latency and amplitude.

Several studies have reported a delayed peak latency of the N400 in L2 learners, and this effect can be influenced by proficiency and age of acquisition. Ardal, Donald, Meuter, Muldrew, & Luce (1990) observed a delay in the peak latency of the N400 in bilinguals, with monolinguals showing the shortest onset, followed by bilinguals in their first language and then bilinguals in their second language. A delay in N400 peak latency in bilinguals has also been reported by Weber-Fox & Neville (1996), and both age of acquisition and years of experience showed good predictive value for the N400 latency (Weber-Fox & Neville, 1996). Kutas & Kluender (1994) found variations in the peak latency of the N400 in bilinguals two languages, specifically, semantic violations in the less proficient language elicited an N400 with a delayed peak compared to those seen in their more proficient language. Moreno & Kutas (2005) investigated the influence of proficiency by examining the N400 in a group of English-Spanish bilinguals with varying degrees of proficiency across languages. Semantically incongruous sentence endings produced an N400 that began ~10 ms later and peaked ~27 ms later in their non-dominant language, compared to their dominant language. In line with previous findings, both age of acquisition and proficiency independently accounted for a significant amount of variation in N400 latency. Similar results have been reported by other authors as well (Ojima, Nakata, & Kakigi, 2005).

Variations in mean amplitude have also been observed in L2 learners. Typically, the mean amplitude of the N400 is reduced for L2 semantic violations, compared to L1 semantic violations. In addition to a delay in peak latency of the N400, Kutas & Kluender (1994) also found a reduction in amplitude to semantic violations in bilinguals' L2. Similar variations in N400 peak amplitude in bilinguals have been reported by Proverbio, Čok, & Zani (2002), who found larger N400 effects to semantic errors in subjects' L1 than their L2. Midgley, Holcomb, & Grainger (2009) observed reductions in the N400 amplitude that correlated with proficiency in a group of English speakers who were learning French as adults. Newman et al. (2012) compared native English speakers to native Spanish speakers who had learned English as adults, and found that for late-learners, relative to native speakers, there was a reduction in the N400 amplitude associated with reading sentence-based semantic violations. In both late learners and native speakers, the amplitude of the N400 was correlated to proficiency, however the distribution of the effect differed between groups. Reductions in the peak amplitude of the N400 associated with semantic violations were also observed by Hahne (2001), who compared late learners of German to native speakers.

Finally, N400 onset and peak duration have also been found to differ in bilinguals. In addition to reporting a delay in peak amplitude, Newman et al. (2012) observed a delay in peak onset of the N400 to L2 semantic violations compared to L1 semantic violations. Hahne & Friederici (2001) and a follow up study by Hahne (2001) compared the N400 in native speakers of German and adult learners of German, and found that the N400 elicited in response to semantic violations showed a prolonged peak duration in the L2 group, compared to the L1 group.

To summarize, most studies that have used a semantic violation paradigm to investigate the N400 in bilinguals have reported a delay in peak latency, and reduced amplitude, for L2 violations compared to L1 violations. Generally, age of acquisition and proficiency appear to at least partially account for these effects, although this is not always the case. Overall, this suggests that the time course of semantic processing in L1 and L2 may differ, and there may be increased costs associated with lexical semantic integration in lower proficiency speakers.

1.4 The N400 and Longitudinal Second Language Training

Relatively few studies have used longitudinal training studies to investigate changes in cortical activity over the course of learning a second language. There is a general assumption that learning a second language as an adult is a slow, laborious process (Prince, 2006), but very little is known about the progression of L2 word learning in adults. McLaughlin et al. (2004) used ERPs to determine how much L2 exposure is necessary for learners' brains to register the lexical status and meaning of L2 words. Participants in their study were university students enrolled in a French class, who had no prior exposure to French. Over the course of the 9-month instructional period, they recorded ERPs at three time points – after approximately 14, 63, and 138 hours of instruction. At all three sessions, pseudowords elicited a larger N400 effect than related and unrelated word pairs, and the effect increased across sessions. At session 2, the authors observed the emergence of a word meaning effect, which manifested as a reduction of the N400 for semantically related pairs, compared to semantically unrelated pairs, and the effect increased in amplitude across sessions. At the final session, participants ERP responses were qualitatively indistinguishable from responses seen in native-speakers, showing typical distribution over midline electrodes, and posterior distribution over lateral sites (McLaughlin et al., 2004). Behavioural responses showed only a modest increase in sensitivity to lexical status from session 2 -3.

The results of McLaughlin et al. (2004) are particularly interesting, as they show that adult second language learners' brains can very rapidly (after only ~14 hours of instruction) integrate L2 word information, with word form being initially integrated, followed later by word meaning. This indicates that adult second language learning is not a uniformly slow process – some aspects of linguistic information can be acquired quickly, although word meaning effects were not present until after ~60 hours of instruction. Additionally, their results suggest that ERPs may be able to better reflect learning progress than traditional behavioural measures.

Stein et al. (2006) conducted a similar longitudinal study, although they focused on single word processing and only recorded ERPs prior to language training and then again approximately 5 months later. Native English-speaking students learning German were recruited and their brain activity was recorded while they tried to understand nouns in three languages (English, German and an unrelated, unknown language, Romansh). When ERP results for German words were compared from session 1 to session 2, the authors found changes in scalp topographies, with more positive values over left central electrodes and more negative values over left temporal and occipital electrodes. No differences in scalp topographies were observed across sessions for Romansh words. Overall, these results provide further evidence of plasticity in neuronal networks underlying adult second language learning.

An interesting variation on longitudinal L2 training studies involves teaching a miniature artificial language (a language whose phonology, grammar and vocabulary have been derived for human or human-like communication) to adults. Kersten & Earles (2001) examined the ability of adults to learn a miniature artificial language introduced in either small segments (i.e. individual words) or complex segments (i.e. sentences containing three words). Meaning was conveyed in both cases via accompanying animated events (similar to ASL signs). They found that adult participants acquired the meaning and morphology of words better when participants were first introduced in small segments followed by complex segments consisting of previously introduced words, compared to adults who were only presented with complex segments. This suggests that adults can acquire novel language characteristics in a short amount of time (the training involved learning 72 novel events and typically lasted only 40 minutes) and that initially introducing less complex segments of language may promote better learning outcomes (Kersten & Earles, 2001).

1.5 The N400 and Language Learning Methods

In line with longitudinal studies investigating general L2 learning, studies have used both ERPs and behavioural measures to investigate learning of both L1 and L2 words and artificial languages using different training methods. Generally, these studies suggest that learning novel language content (both vocabulary and grammar) is best when content is introduced in context. In Mestres-Missé et al. (2007) native speakers of Spanish performed a Spanish word reading task where they learned the meaning of novel words based on sentence context. Sentences had varying degrees of contextual constraint; increasing contextual constraint across three sentences ensured participants would gradually decode the meaning of the novel word. This learning condition was compared to two other conditions, one in which the meaning could not be derived from the sentence context, and one in which known Spanish words were presented. An additional ERP experiment was conducted in which words learned in each of the three conditions were presented as pairs – either with their real Spanish counterpart, with a semantically related Spanish word, or with an unrelated word.

After 3 exposures to novel words, the N400 to novel words in a relevant context were indistinguishable from real words. Additionally, when presented as pairs in a priming paradigm, novel words introduced under high contextual constraint facilitated the semantic processing of their real-word counterparts, which manifested as a reduced N400 effect. This is interesting, as it shows that meaning can be rapidly associated with novel words, given that novel words are introduced under relevant context (Mestres-Missé et al. 2007). However, this study only focused on one learning mode, and did not address learning of second language vocabulary.

Chun, Choi, & Kim (2012) used translational learning (linking a word directly to its L1 counterpart) and extensive reading (introducing novel words in text) to compare learning of second language vocabulary. Native Korean speakers learned English vocabulary through one of these methods over 9 weeks. Behavioural measurements and ERPs were recorded prior to initiating training, immediately following training, and then again 5 weeks after training ceased. In the ERP task, matching or mismatching translation pairs were presented, with Korean words as the prime and English words as the target. Vocabulary test results indicated that there was no difference in retention between learning methods immediately following training, however, words learned via the extensive reading method showed better retention at the 5-week follow-up. Similarly, ERP results showed no difference in the N400 effect between training methods immediately post-training, but at the 5-week follow-up, words learned through extensive reading showed a greater N400 effect than translational learning words. This suggests that extensive reading leads to better long-term retention of L2 vocabulary than translational learning, a more direct, explicit method (Chun, Choi, & Kim 2012).

Finally, Morgan-Short, Steinhauer, & Sanz (2012) used an artificial language paradigm to examine whether explicit training (which explicitly taught word meanings and grammatical rules) and implicit training (which involved deducing meaning based on examples in within sentences) can differentially impact both behavioural and ERP measures of syntactic processing. Behaviourally, no difference in performance between training groups was observed at either low (at the start of training) or high proficiency (at the end of training), whereas only the implicitly trained group at high proficiency showed native-like (similar to neurophysiological signals in native speakers) brain responses to syntactic violations. Although this study did not address vocabulary learning directly, it does provide further evidence of rapid changes in brain activity associated with learning a new language, and the conditions under which a language is learned can influence whether learners achieve native-like brain activity, specifically, contextual/implicit L2 training can lead to better learning and more native-like patterns of brain activation.

1.6 Theories of Bilingualism

Once a second language is acquired, how is it represented in the brain? A substantial body of research has investigated how bilinguals handle their languages while listening, reading and speaking. Studies have questioned how the bilingual system is organized on a neural level, whether languages influence each other during processing, how age of acquisition and proficiency play a role in language interaction, how bilinguals control their languages and what effects are associated with language transfer. These studies have led to the development of several models of bilingual processing and organization, that do not all share the same theoretical focus – focus can include language representation, language processing, language learning, language comprehension (both visual and auditory), or language production (van Heuven & Dijkstra, 2010). Of particular interest to the current study is a model that attempts to explain single word processing and whether the bilingual lexicon is integrated or separated by language.

The BIA+ model attempts to explain bilingual word processing via the existence of two sub-systems: an encapsulated language processing or word identification system, with an integrated lexicon for different languages; and a task/decision system (for a review, see van Heuven & Dijkstra, 2010). The two main assumptions of this model are that in bilinguals, the lexicon is integrated across languages, and that word identification is non-selective— automatic co-activation of information in both language systems occurs, regardless of whether activation is contextually appropriate. Orthographic, phonological and semantic representations form an interactive network with different sub-pools (semantics, lexical orthography, lexical phonology, sublexical orthography, sublexical phonology). Non-selective language models like BIA+ predict that cross-linguistic interaction, in which various aspects of individual languages interact and influence linguistic performance, is possible. Cross-linguistic interaction is generally assumed to be slower for L2 to L1, although this may be influenced by the level of L2 proficiency (van Heuven & Dijkstra, 2010).

The BIA+ model is supported by both behavioural and neuroimaging studies. For example, Heuven, Dijkstra, & Grainger (1998) examined the impact of orthographic neighbourhood effects in bilingual word recognition. Orthographic neighbours are any words that differ by a single letter, respecting length and letter position (e.g., *cat* and *bat*). They were interested in how the recognition of target words exclusive to one language can be influenced by the number of orthographic neighbours within the same language and in the other language of bilinguals. They found that increasing the number of cross-linguistic orthographic neighbours slowed response times to target language words in Dutch/English bilinguals. Neighbourhood size has also been shown to modulate the N400 effect in bilinguals. Midgley, Holcomb, Walter, & Grainger (2008) investigated the influence of cross-linguistic neighbours on the N400 effect in a group of French/English bilinguals. For both L1 and L2 words, the number of cross-linguistic neighbours was found to modulate the size the of the N400, with words sharing a large number of cross-linguistic neighbours eliciting a larger N400 effect than words with a small number of cross linguistic neighbours (Midgley et al., 2008). These results can only be accounted for by a non-selective model, which allows for the simultaneous activation of word representations in both languages, as opposed to a selective model, which suggests that the order in which L1 and L2 representations are examined depends on language context (Heuven, Dijkstra, & Grainger, 1998).

Evidence for a non-selective language access model has found further support in studies investigating unconscious translation in bilinguals. For example, in Chinese/English bilinguals, Thierry & Wu (2007) found that when making a semantic relatedness decision for pairs of English words, manipulating the composition of characters in the corresponding Chinese translation had an effect on the amplitude of the N400. A larger N400 effect was observed for English word pairs whose Chinese translations contained repeated characters. Given that the task was conducted entirely in English and participants were unaware of the Chinese character manipulation, this suggests that English words automatically activate their Chinese counterpart.

Finally, semantic priming has been observed across languages in bilinguals. Martin, Dering, Thomas, & Thierry (2009) asked highly proficient Welsh/English bilinguals to make judgments about the length of Welsh and English words. They were asked to selectively attend to the length of either Welsh or English words presented in series, while their brain activity was recorded with ERPs. Stimuli was presented in pairs, such that consecutive words were either in the same or different language, and semantically related or unrelated. They observed semantic priming within and across languages, irrespective of the language being attended to. This manifested as a reduced N400 effect for semantically related word pairs, in both Welsh and English. Semantic relatedness did not modulate behavioural responses. This provides evidence for the automatic activation of the meaning associated with written words, in both languages of fluent bilinguals.

Overall, there appears to be strong evidence for a bilingual word processing model that suggests a shared lexicon for multiple languages and interaction between languages during processing. While this is interesting, it is not yet known how quickly this integration can occur during second languages acquisition, since studies showing cross-linguistic interactions have typically used fluent bilinguals. Longitudinal second language training studies (i.e. McLaughlin et al., 2004; Stein et al., 2006) have so far focused on the time course of L2 word meaning acquisition, without addressing how quickly newly acquired L2 words can interact with existing semantic networks. The current study will address this question, specifically, can cross-linguistic semantic priming occur after only a short period of training?

1.7 The Current Study

The current study assessed L2 vocabulary learning, with a focus on outcomes associated with different training strategies. In addition, we also assessed whether newly learned L2 words were integrated into existing language networks. French vocabulary words were taught to English speakers via computerized games. Two training strategies were compared, based on associative learning and context learning. For the purposes of this study, we labelled these strategies explicit (based on associative learning) and implicit (based on context learning). These labels were chosen based on previous literature (Morgan-Short et al., 2010 & 2012). We were interested in whether different training strategies promote different learning outcomes, whether newly learned L2 words were integrated into existing semantic networks and if L2 learners could show native-like brain responses after only a short training period.

We used both behavioural and neurophysiological measures to address these questions. We compared participants' results prior to initiating training to their results immediately following the training program. Behavioural assessment included a picture naming task, in which participants were required to name visual depictions of French words, as well as accuracy on a Match/Mismatch task, in which matching or mismatching picture/auditory French word pairs were presented, and participants made a match judgment. In the post-training assessment, two additional conditions were added, in which pairs of words were either semantically related or semantically unrelated. In both cases, the prime was a visual picture of an unlearned word. Neurophysiological assessment included EEG recording while the participants performed the Match/Mismatch task, with a focus on the amplitude of the N400. Participants' behavioural and neurophysiological results from the post-training Match/Mismatch task were compared to a group of native French speakers who also completed the post-training EEG protocol.

Post-training, we expected to see an increase in the total number of words participants were able to name, reflecting an increase in their ability to correctly produce the correct French word associated with a picture. We expected that words taught in the implicit condition would be better-learned than those taught in the explicit condition. Additionally, we expected to see an increase in accuracy on the Match/Mismatch task, reflecting an increase in their ability to correctly identify the correct French word associated with a picture. Again, we expected higher accuracy for words taught via the implicit strategy, compared to the explicit strategy.

Prior to training, participants were not expected to have semantic knowledge of the French words, so mismatches were not expected to elicit an N400 effect. No differences in the ERP waveform were expected for the match and mismatch conditions. Post-training, we expected that participants would have acquired semantic knowledge of the French words. As such, we expected to see an N400 effect for mismatch trials, compared to match trials, and this effect should not vary based on grammatical category (nouns/verbs). We expected to see similar patterns of brain activity for words learned on each day of training. In line with our behavioural expectations, we expected implicit words would be learned better than explicit words. Thus, the N400 elicited for mismatches of implicitly taught words would show greater amplitude, compared to those elicited by mismatches of explicitly taught words.

We expected that newly learned French words would be integrated into semantic networks, so cross-linguistic semantic priming would occur. For the semantically related condition, we expected to see a reduction in accuracy, compared to the semantically unrelated, match and mismatch conditions, in both the learners and the controls. In line with previous research on cross-linguistic semantic priming, we expected that this would result in a reduction of the N400 amplitude for semantically related mismatches, compared to semantically unrelated mismatches, and learned mismatches. We expected that the pattern of ERP results would not differ between learners and controls, specifically that there would be a reduction in N400 amplitude for semantically related trials, compared to mismatch, and semantically unrelated trails, in both groups.

CHAPTER 2: METHODS

Methods were approved by the Dalhousie University Health Sciences Human Research Human Research Ethics Board (protocol #2012-2703) and the Centre National de la Recherche Scientifique (CNRS). This study was conducted at the Laboratoire Parole et Langage, part of the Brain and Language Research Institute (Aix-Marseille Université), in Aix-en-Provence, France.

2.1 Subjects

2.1.1 L2 Learners

Thirteen subjects ("Learners", eight females, five males, Mean age = 25.7) were recruited by word of mouth and by posted advertisements distributed in the community, primarily at Aix-Marseille Université, the International American University (IAU) College and French Language schools. Participants were excluded if they had prior exposure to other Romance languages (Spanish, Portuguese, Italian or Romanian). Five subjects were monolingual native English speakers. Five subjects were bilingual English speakers, with Hindi, Bengali, Tamil, Japanese and Mongolian as additional spoken languages (English age of acquisition range = 0 – 5 years). Three other subjects spoke English as a second language (mean age of instruction in English = 8.2 years), with Turkish and Hungarian (two subjects) as native languages. The average length of prior French instruction among all subjects was 1.5 months (range: 0 - 4 months). Subjects received €100 each for participating.

2.1.2 Native French Speakers

Thirteen native French speakers ("Controls", eleven females, two males, Mean age = 22.1) were recruited via posted advertisements to serve as a comparison group for the post-training EEG task. They participated in a single EEG session only. French subjects received €20 each for participating.

2.2 Stimuli

2.2.1 Training Content

Training content (see appendix Table 1) consisted of 72 French vocabulary words (48 nouns and 24 verbs). Twelve additional French words (8 nouns and 4 verbs) were created for use during initial practice with the training software. Training content selection was based on several criteria. Low frequency words were primarily selected so that content would be unfamiliar, or novel, to participants with very low-level French knowledge (average (log) frequency per million, in French, for training content was 1.1). Words were excluded if their meaning was easily deducible based on their English equivalent (English cognates, for example *l'animal*). Additionally, words were restricted to those with no more than one phoneme not found in English. Content included only regular third person singular and third person plural verbs, and both transitive (e.g., to hit) and intransitive (e.g., to jump, which doesn't take a direct object) verbs were included. Special consideration was placed on selecting words that were easily represented visually, because training relied on associating novel words with pictures demonstrating their meanings. Thus we chose concrete objects and "active" verbs to minimize the likelihood that participants would be unsure what each picture referred to. Verbs had a higher (log) frequency/million than nouns (average (log) freq/million for verbs = 2.14, SD = 0.53, average (log) freq/million for nouns = 1.3, SD = 0.68, $t(71) = 5.2$, $p < .001$). For implicit training, words were organized into three-word sentences containing a subject noun, a verb, and an object noun. Sentences were generated so that they were grammatically correct and conveyed a meaning that was as realistic as possible. Sentence organization was based on groups of 16 simple subject-verb-object sentences that were broken down into 2 sets of 8 sentences, each set containing 4 nouns and 2 verbs. Within each set, 2 nouns served as subjects and 2 served as objects, in all possible combinations with the 2 verbs in that set.

A cartoon line drawing for each word in the training set was created by an artist; all the drawings thus had the same style and were developed through iterative feedback cycles to emphasize clarity.

Spoken versions of each word, produced by a female native French speaker using a standard European French accent, were recorded (stereo, 32-bit, 20 db) in a sound attenuating booth in a single session and edited using Audacity software. Each noun was preceded by its appropriate determiner, either *le* for masculine nouns, or *la* for feminine nouns. Verbs were produced in the first person singular, present tense. As well as individual words, three-word, subject-verb-object sentences were also recorded by the same French native speaker. These were spoken using a deliberately slow, yet natural-sounding rate.

2.2.2 Testing Stimuli

In order to determine if word recognition could generalize to depictions of training content outside of the images used during training, black and white line drawings of each training word were taken from the International Picture Naming Project (IPNP) database. If a training word was not available in the IPNP database, an image was found using Google image search. Images were selected so that the depicted word was immediately obvious, based on feedback from a number of people to help ensure that they clearly represented the correct referent. All images were scaled to 600 x 600 pixels. These images were used in both the Naming Task and the Match/Mismatch EEG Task. Audio recordings of training content used during training were also used for the Naming Task and Match/Mismatch EEG Task.

In the post-training Match/Mismatch EEG task, each learned word was paired with an unlearned, semantically related word, generated based on a standardized set of first verbal associates (Alario & Ferrand, 1999; Ferrand & Alario, 1998). If a learned word was not included in the set of first verbal associates (as was the case with most of the training content), or if the first verbal associate of a training word was another training word, or if the first verbal associate was difficult to depict visually, a semantically related word was suggested and then verified by polling other members of the research team. Words in the two new conditions were matched based on length, bigram frequency, number of neighbours and (log) frequency per million.

2.3 Procedure

2.3.1 Training

Training words were distributed equally across training strategy (36 taught via explicit strategy, 36 taught via implicit strategy). Word assignment to category was balanced across participants, so that two training programs were used (A or B), containing opposite explicit/implicit word lists. Twelve words (8 nouns and 4 verbs) were taught on each day, with explicit and implicit training days alternating. Explicit words were organized into 3 groups of 4 words, consisting of 4 subject nouns, 4 verbs and 4 object nouns. Implicit words were organized into 16 simple subject-verb-object sentences. These 16 sentences were broken down into 2 sets of 8 sentences, each set containing 4 nouns and 2 verbs. Within each set, 2 nouns were presented as subjects and 2 were presented as objects, in all possible combinations with the 2 verbs in that set. The presentation order of the sentences was fixed so that combinations containing the same verb were grouped together.

Training games were part of a development version of LANGA (Copernicus Studios Inc., Halifax, NS, Canada), hosted online and created using the *Flash* and *ActionScript* (Adobe Systems Inc.) software. Training games utilized a commercial speech recognition algorithm (Telisma/OnMobile; Bengalaru, India). Accuracy was based on a threshold that allowed for the distinction between correct and incorrect answers, while also permitting participants who had difficulty pronouncing French phonemes to progress.

Participants were required to complete 10 lab visits (see Figure 2.1). On Day 1, participants completed necessary paperwork and the pre-training behavioural and EEG assessments. On Day 2, participants completed a practice run-through of all training games; this is when the practice stimuli were used. Six days (Days 3-8) of training followed, with 12 new words introduced each day. Day 9 was used for a review of all training content. On Day 10, participants completed the post-training behavioural and EEG assessments. Access to the lab was only allowed on weekdays, which resulted in breaks in training. Participants began the study on either Thursday or Friday, which resulted in 2 breaks, the first between Day 1 and 2, or Day 2 and 3, and the second between Day 6 and 7, or Day 7 and 8.

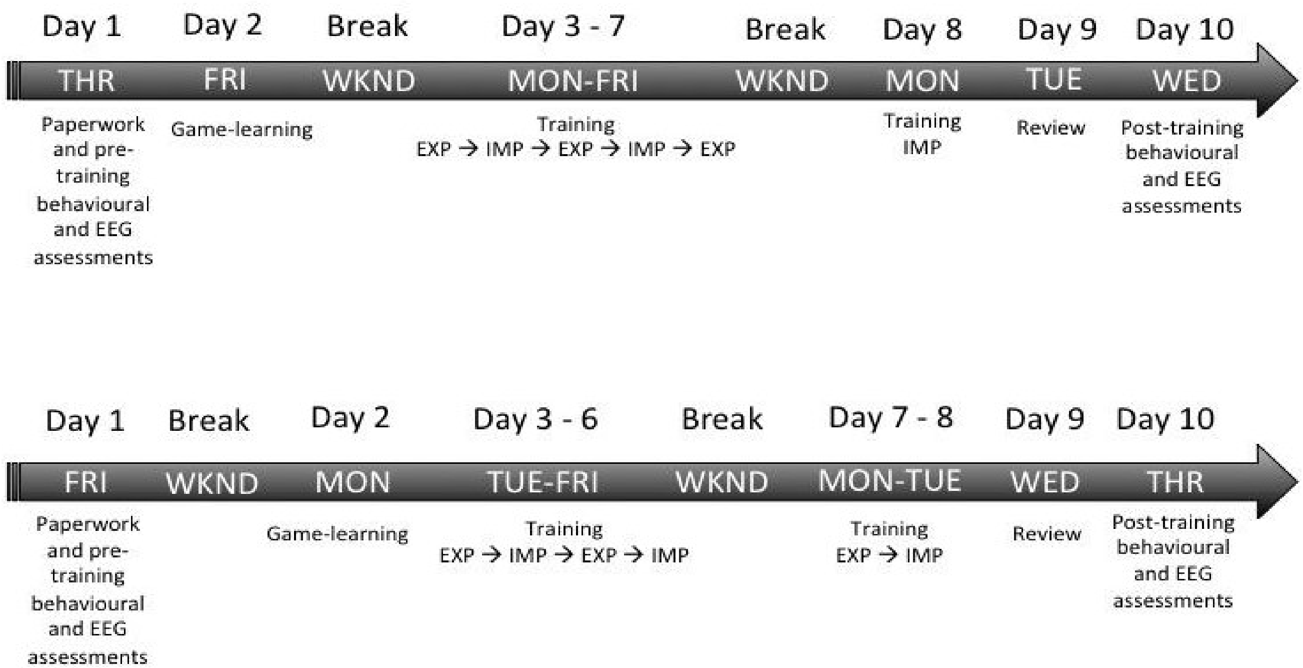


Figure 2.1. Overview of full training schedule.

The game-learning session (~1 hour) and the 6 training sessions (~30 min each) and the review session (~30 min) were completed in the lab under the direct supervision of a member of the research team. Participants were seated in a quiet room and wore ear-bud type headphones while completing the training on a laptop computer.

On each training day, content in both training styles was introduced using the *Learn* game. This game consisted of a set of 4 images, each depicting either an individual French word (explicit training) or a three-word French sentence (implicit training). Participants were familiarized with the games prior to initiating training (on Day 2), but were allowed to watch a brief instructional video explaining the game again at the start of Days 3 and 4, if they wished. Participants were instructed to click on each individual image, listen to the content spoken by a female native French speaker, and then repeat the word or sentence. Accuracy of their repetition was determined by the speech recognition engine. If participants made 3 incorrect answers in a row, the written form of the content appeared below the image. Correct answers for all 4 presented images were required before the participant was allowed to progress to the next game. Feedback was given in the form of both visual and auditory cues, i.e. correct answers elicited a pleasant tone and a green heart, while incorrect answers elicited an unpleasant tone and a broken red heart.

On explicit training days, words were then practiced over 3 subsequent games (Figure 2.2). Within each game, words were presented 4 times. In the *Doubles* game, participants had to correctly identify and name the target word that was presented twice in an array of 5 images. In the *Missing* game, participants had to identify the word within the current training set that was absent from an array of 3 images. In the *Match-3* game, participants were presented with an array of image tiles organized in a matrix. The goal of this game was to arrange 3 tiles depicting the same content word in a row. Participants moved tiles by first clicking on the tile and then naming the associated image. Four correct answers were required for each word before the participant could advance. After completing these 3 games, words in the next training set were introduced via the *Learn* game. At the end of each explicit day, participants reviewed all the words from that day, via the *Shuffle* game, which presented each word individually in random order, which required the participant to name each word as it was presented. Words were again reviewed using the *Learn* game and the *Shuffle* game at the start of the next 2 training days.

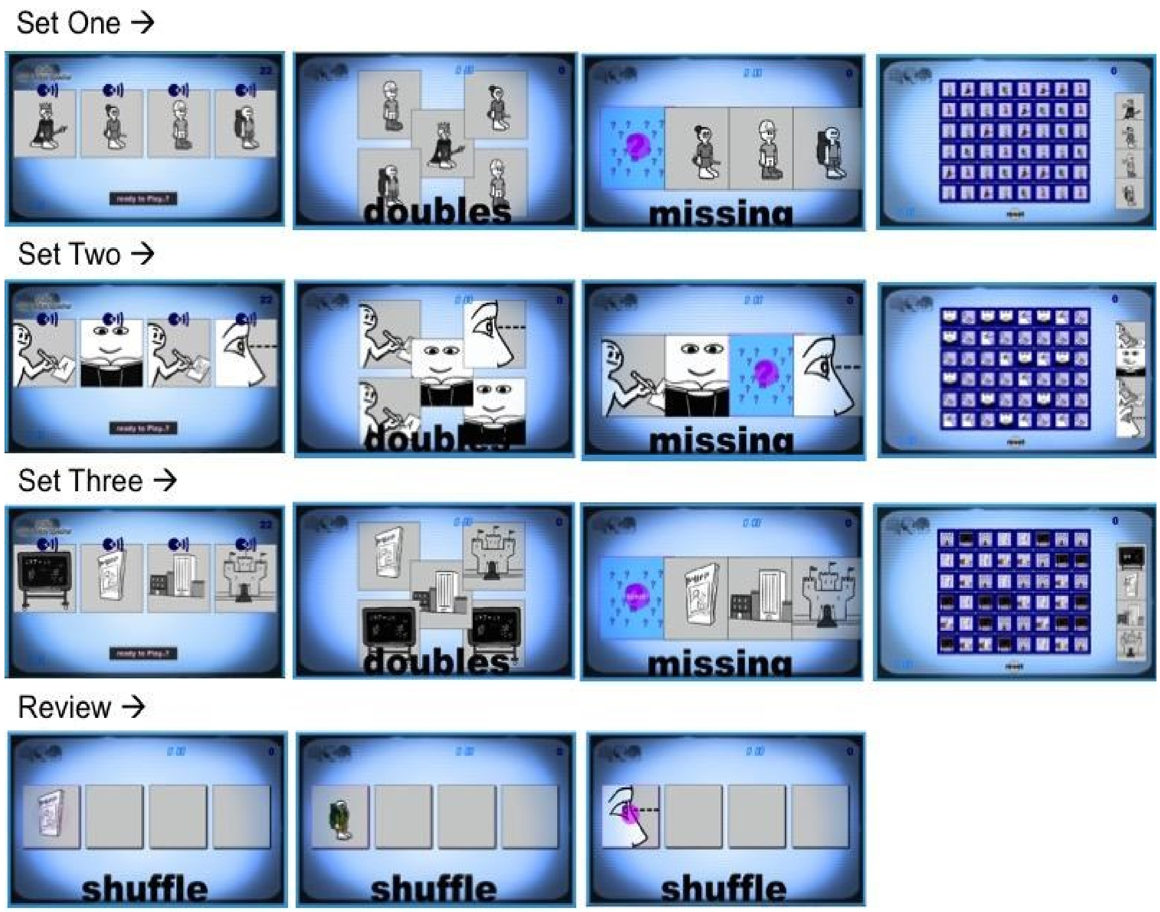


Figure 2.2. Overview of explicit day organization and games.

Words on implicit days were also introduced using the *Learn* game (Figure 2.3). However, rather than the pictures being of individual objects or actions, the pictures showed a depiction of the sentence (a subject performing an action with an object), and the sound file presented in association with each picture comprised three-word subject-verb-object sentences. Correct responses for all 4 sentences presented in the *Learn* game were required before the participant could advance to other games. Content was practiced over 3 subsequent games, with each sentence presented once within a game. The *Fast/Slow* game involved 2 auditory presentations of a sentence accompanied by the corresponding image. The first presentation was spoken at a natural rate for a French native speaker, and the second was spoken at a deliberately slower rate that emphasized each distinct word. Following the slow presentation of the sentence, the participant would repeat the sentence back. Next was the *Slot Machine* game. Three spinning wheels, each representing one component of the sentence, would stop on a word within the training set and the participant would have to combine the 3 components and speak the sentence out loud. This was followed by the *Doubles* game, which was previously described above for the explicit training. After completing all 3 mini-games, new sentences were introduced with the *Learn* game. The second set of mini-games introduced a new verb grouped around the same subject/objects from the first set. The third set introduced new subjects/objects grouped around a new verb. The fourth set introduced a new verb grouped around the subjects/objects from the third set.

At the end of the day, words were reviewed in sentence form using the *Shuffle* game. Words were also reviewed at the start of the next two days using the *Learn* game and the *Shuffle* game. During review, one sentence from each of the four sets was presented, so that all words within a training day were presented once only during review.

Figure 2.3. Overview of implicit day organization and games.

The number of total exposures per word was kept constant across days and learning type. Words were seen 14 times on their learning day (4 times within sets of 3 mini games, plus an additional 2 exposures during the review *Shuffle* game). They were reviewed an additional 6 times over the next 2 days (once during the *Learn* game and twice during the *Shuffle* game, at the start of two subsequent days). Words taught on later days (Day 7 and Day 8) were not reviewed over 2 subsequent days in the same manner as earlier days (Day 7 words were reviewed on Day 8 and during the final full review, and Day 8 words were reviewed on the final review day alone), but exposures were kept constant by increasing the number of exposures for those words on the final review day. The final day of training (Day 9) was used as a full review and each word taught on Days 3-6 was reviewed once during the *Learn* game, *Shuffle* game and *Doubles* game, for a total of 3 exposures; words taught on Day 7 were reviewed twice in each of these games, and words taught on Day 8 were reviewed three times in each of the games. Therefore, words were practiced a total of 23 times throughout the entire training study.

2.3.2 Naming Task (Pre- and Post-Training)

Images corresponding to all 72 training words (not including those used on Day 2 for game practice) were presented on a laptop computer using E-Prime software. Each image was presented on a black background in random order and remained on the screen until participants either named the picture and pressed a key to advance, or skipped the image by immediately pressing a key to advance. Spoken responses were recorded for later scoring.

2.3.3 Match/Mismatch EEG Task

Pre- and post-training, learners performed a picture/auditory word match/mismatch task, presented on a computer with StimPres 3.0 software, while their brain activity was recorded using electroencephalography (EEG). Controls completed the post-training Match/Mismatch EEG Task.

Over each EEG recording session, each word was presented twice, once paired with the picture reflecting its meaning (match) and once with a different, semantically unrelated picture corresponding to one of the other words in the training set (mismatch). This resulted in a total of 144 trials, including 72 match and 72 mismatch. Mismatch pairs were created by dividing the word list by training type (explicit or implicit), learning day, gender and grammatical category. Words were paired with other words of the same training type, learned on a different day (i.e. Day 1 explicit words were paired with either Day 3 or Day 5 explicit words). Nouns were paired with nouns of the same gender, and verbs were paired with verbs. Mismatch pairs of the same two words were avoided – i.e. if an image of *le cheval* was paired with an auditory presentation of *le livre*, an image of *le livre* would not be paired with an auditory presentation of *le cheval*. Given that words were presented twice during the task (once as a match and once as a mismatch), presentation order was pseudo-randomized so that 2 consecutive trials would not contain the same word and presentation order of match/mismatch was balanced across participants. Four separate pseudo-randomized lists were created for each training program (A or B, containing reversed explicit/implicit training lists). Each participant saw only 1 list. Two additional mismatch conditions (a semantically related condition and a semantically unrelated condition) were included in the post-training Match/Mismatch EEG task; these trials were randomly intermixed with the match and mismatch trials used in the pre-training session (please refer to Table 1 in the appendix for a full list of semantically related and unrelated pairs, and appendix Figure 1 for a complete set of visual stimuli used during pre- and post- training assessments). In the semantically related condition, the pair was semantically related, with the picture representing an unlearned word (not taught during training) semantically related to a learned word, which was presented as the auditory target (for example, a picture of a saddle, followed by the learned word *le cheval*). In the semantically unrelated condition, the pair was unrelated, with the picture representing an unlearned word, which was unrelated to the learned word presented as the auditory target (for example, a picture of a fork, followed by the learned word *le cheval*). The post-training task had 72 learned match trials, 72 learned mismatch trials, 72 semantically related trials and 72 semantically unrelated trials, for a total of 288 trials. Pre-training, each learned word was presented twice as a picture (once as a match and once as a mismatch) and twice as an auditory target (once as a match and once as a mismatch). Post-training, each learned word was presented twice as a picture (once as a match and once as a mismatch) and four times as an auditory target (once as a match, once as a mismatch, and once in the semantically related condition and once in the semantically unrelated condition).

During each trial, an image was presented followed 1 s later by an auditory word. The image remained on the screen during the auditory presentation. A response prompt, consisting of a screen displaying "Match?" followed 500 ms after auditory stimulus offset. Participants indicated their responses by pressing buttons corresponding to "Yes" or "No" on a button box. The hand used (left or right) to press each button was counterbalanced across participants. The response prompt remained on screen until the participant responded. A "Blink!" screen, displayed for 1 second, followed 200 ms after their response. This was used to encourage the participant to only blink between trials because blinks can create artifacts. Participants completed three practice trials at the start of the task and were allowed three short breaks (at the 25%, 50% and 75% completion points).

2.4 EEG Recording and Pre-processing

Scalp potentials were continuously recorded from 21 tin electrodes attached to an elasticized cap (Electrocap International). Locations were standard International 10-20 scalp sites (Jasper, 1958) over frontal, temporal, central, posterior temporal, parietal and occipital areas of the left and right hemispheres (FP1/2, F7/8, F3/4, C3/4, T5/6, P3/4, O1/2), as well as over midline (Fz, Cz, Pz). In addition, electrodes were placed centrally between homologous anterior and central sites (FC5/6), central and parietal sites (CP5/6). Horizontal eye movements were monitored by means of an electrode placed at the outer canthus of the right eye while blinks and vertical eye movements were monitored via an electrode beneath the left eye. Electrodes were placed over both the left and right mastoids; all electrodes were referenced to the left mastoid during the recording and then re-referenced to the average of the mastoid electrodes for analyses. The EEG was amplified with a bandpass of 0.1–40 Hz (3 dB cutoff) by means of an SAI Bioamp 32 channel model and was digitised online at 200 Hz. EEG were later lowpass filtered with a cutoff of 30 Hz. The electrode impedance threshold value was set to 3 kΩ for scalp electrodes and 10 kΩ for face electrodes. Epochs began 100 ms prior to auditory word onset and continued 1100 ms thereafter. Amplitude was normalized using calibration pulses (50 μV). Average ERPs were formed off-line from trials free of muscular and/or ocular artefact and amplifier blocking (rejection was performed by a computerised routine, total rejection = Learners Pre-training: Match: 3.8%, Mismatch: 4.7%; Learners Post-training: Match: 6.9%, Mismatch: 8%, Semantically Related: 7.5%, Semantically

Unrelated: 7.1%; French Controls: Match: 9%, Mismatch: 9%, Semantically Related: 9%, Semantically Unrelated: 10%).

2.5 Behavioural Data Analysis

For the Naming Task, participants were given a correct answer (1 mark) if they responded with the full word (including correct gender article) and a partial correct answer (0.5 mark) if they responded with the correct word stem but incorrect gender article or a partially correct word stem. Naming results were compared across session, training strategy and grammatical category (noun or verb). Due to a computer error, post-training responses for the Naming Task were lost for 6 participants. Changes from pre- to post-training for the 7 subjects with complete data were analyzed using a repeated measures ANOVA with factors session (pre/post) and grammatical category (noun/verb). Finally, a paired samples t-test was used to compare post-training naming results for each training method.

Accuracy for the Match/Mismatch EEG task was determined by recording the number of correct responses in each condition (pre-training match/mismatch and post-training match/mismatch/semantically related/semantically unrelated) and converting those numbers to a percentage value. Pre-training results in each condition were compared using a paired samples t-test, and we also performed a *d'* analysis to determine if participants had a bias towards a specific answer. A repeated measures ANOVA with factors session (pre/post), condition (match/mismatch), and grammatical category (noun/verb) was used to compare performance across sessions. A separate ANOVA with factors condition (match/mismatch/semantically related/semantically unrelated) and group (learners/controls)) was used to compare the learners' performance on the post-training task to a group of native speakers. Finally, paired samples *t*-tests were used to compare performance within each condition (match/mismatch) across training strategies.

2.6 EEG Data Analysis

The ERP data were quantified by determining mean voltage amplitude during the 300-500 ms time window, which roughly corresponds to the time window associated with the N400, the ERP component of interest. Data were analyzed using linear mixed effects (LME) modeling, a type of general linear model. LME modeling was accomplished via the *lmer()* function in the *lmer4* library (Bates, Maechler & Bolker, 2009) in *R* version 2.10 (R Core Team, 2013). LME models can include both fixed and random effects, and offer advantages over standard repeated measures ANOVA, including more extensive modeling of random effects, the ability to account for unbalanced data and the capacity to deal with missing data and non-sphericity (Baayen et al., 2008; Bagiella et al., 2000; Gelman & Hill, 2006). This method of analysis has previously been used in our lab to analyze ERP data (Newman et al., 2012) and has been used by other authors as well (Pritchett et al., 2010; Wierda, van Rijn, Taatgen, & Martens, 2010).

Analyses were performed at 11 electrode sites (Fz, Cz, Pz, FC1/FC2, CP1/CP2, C3/C4, P3/P4). In our analyses, fixed effects predictors included condition (match/mismatch/semantically related/semantically unrelated), session (pre/post), grammatical category (noun/verb), group (learners/controls), training strategy (explicit/implicit), training day (1-6), and electrode location. Subject variability was used as a random effects predictor. To determine the best mixed-effects model for each dependent measure, progressively simpler models were compared to more complex models using log-likelihood ratio testing, which removes factors and interactions that do not explain significant amounts of variability in the data. The 'best' model was that which was the simplest, in that it accounted for the most variability using the fewest number of factors and interactions. Analyses were performed on a pre-selected array of electrodes associated with the N400. The N400 is typically maximal over centro-parietal cites, so it was possible that the magnitude of the effect would vary across electrodes (as indicated by a significant main effect of electrode). However, in the case of the presence of a significant main effect of electrode, but the absence of a significant interaction between electrode and another fixed effect, Cz was chosen as a representative electrode for follow-up analyses investigating other interactions. *F* tests were used to compare main effects and interactions and *t* tests (Bonferroni corrected) were used for specific follow-up comparisons. Bonferroni correction involved multiplying the uncorrected *p*-values by the number of comparisons made, and comparing the new, corrected *p*-value to an alpha level of .05.

In order to compare N400 onset between the learners and controls, we also computed difference waves at electrode Cz (match-mismatch, and semantically related-semantically unrelated) and performed a fractional area latency analysis. This routine provides an estimate of onset latency by computing the area under the component of interest, and then determining the time at which cumulative area under the curve reaches a pre-specified percentage of this total area (in this case, 15%). N400 onset between groups was compared using independent samples *t*-tests.

CHAPTER 3: RESULTS

3.1 Behavioural Results

3.1.1 Naming Task Results

Pre-training, participants were able to name an average of 5 (SD = 7.02) words out of the set of 72. Among the 7 subjects for whom we obtained post-training naming data, we observed a significant increase in the total number of words named post-training (M_{post}= 49.8 words, SD=13.4, M_{post}-M_{pre} = 42.5), $t(6) = 9.99$, $p < .001$, $d = -4.1$. Additionally, they were able to name more words learned through explicit training (M = 26.3 out of 36, SD = 6.15) than through implicit training (M = 23.6 out of 36, SD = 7.41), $t(6) = 2.87$, $p = .028$, $d = 0.22$.

Given the difference in number of nouns and verbs taught, total number of nouns and verbs named pre- and post-training were converted to percent accuracy. A repeated measures ANOVA was used to compare changes in naming accuracy for grammatical categories, across sessions. We observed a significant main effect of session, $F(1, 6) = 89.0$, $p < .001$, $\eta^2 = 0.93$, indicating an increase in accuracy across sessions and a significant main effect of grammatical category, $F(1,6) = 9.7$, $p = 0.02$, $\eta^2 = 0.62$; the session × grammatical category interaction was not significant, $F(1, 6) = 0.55$, $p = .48$, $\eta^2 = 0.08$. Follow up t-tests (Bonferroni corrected for 4 comparisons) revealed that pre-training, the accuracy for nouns (M = 9.8%, SD = 12.4) and verbs (M = 2.9%, SD = 3.9) did not differ, $t(6) = 2.05$, $p = .34$, $d = 0.75$, and post-training, the accuracy for nouns (M= 72.6%, SD = 17.7) and verbs (M = 62.5%, SD = 21.6) also did not differ, $t(6) = 2.7$, *p* = .12, *d* = 0.51. There was a significant increase in accuracy across sessions for both nouns, $t(6) = 10.83$, $p < .001$, $d = -4.3$ and verbs, $t(6) = 7.67$, $p < .001$, $d = -3.8$.

3.1.2 Match/Mismatch EEG Task

Total number of correct answers in each condition was converted to percent accuracy. Pre-training, participants had higher mismatch accuracy (M = 84.4%, SD = 16.0) than match accuracy (M = 63.9%, SD = 13.5), $t(12) = 5.804$, $p < .001$, $d = 1.3$. Given that the task was a forced-choice paradigm (participants had to answer either yes or no, even if they were unsure), a *d'* analysis was conducted. The results indicate that participants had a slight bias to answer "No" compared to "Yes" ($d' = 1.3$, $C = 0.3$).

A repeated measures ANOVA was used to assess whether participants' accuracy scores in each condition increased following training, and if changes in accuracy varied by grammatical category, as seen in Figure 3.1. We found significant main effects of session (pre/post), $F(1, 12) = 69.65$, $p < .001$, $\eta^2 = 0.85$, indicating an increase in accuracy in across sessions, and condition (match/mismatch), $F(1,12) = 34.8$, $p < .001$, $\eta^2 = 0.74$, indicating differences in accuracy between conditions. The main effect of grammatical category, $F(1, 12) = 3.77$, $p = .07$, $\eta^2 = 0.23$, was not significant. The only significant interaction observed was between condition and session, $F(1, 12) = 54.22$, $p < .001$, $\eta^2 = 0.81$.

Follow up post hoc *t*-tests (Bonferroni corrected for 5 comparisons) revealed that there was a significant increase in accuracy in both the match (pre-training: M = 63.8%, SD = 13.50, post-training: M = 96.04%, SD = 3.48), $t(12) = 8.54$, $p < .001$, $d = 3.2$, and mismatch condition (pre-training: M = 84.4%, SD = 16.04, post-training: M = 99.03%, SD = 1.04), $t(12) = 3.72$, $p = .01$, $d = 1.2$, although the increase was larger for the match condition (M_{diff}= 32.1%, SD = 13.5) than it was for the mismatch condition (M_{diff}= 14.6%, SD = 16.1), $t(12) = 5.12$, $p < .001$, $d = 1.1$. Finally, the post-training accuracy was higher in the mismatch condition, than it was in the match condition, $t(25) = 6.86$, $p < .001$, $d = 1.1$.

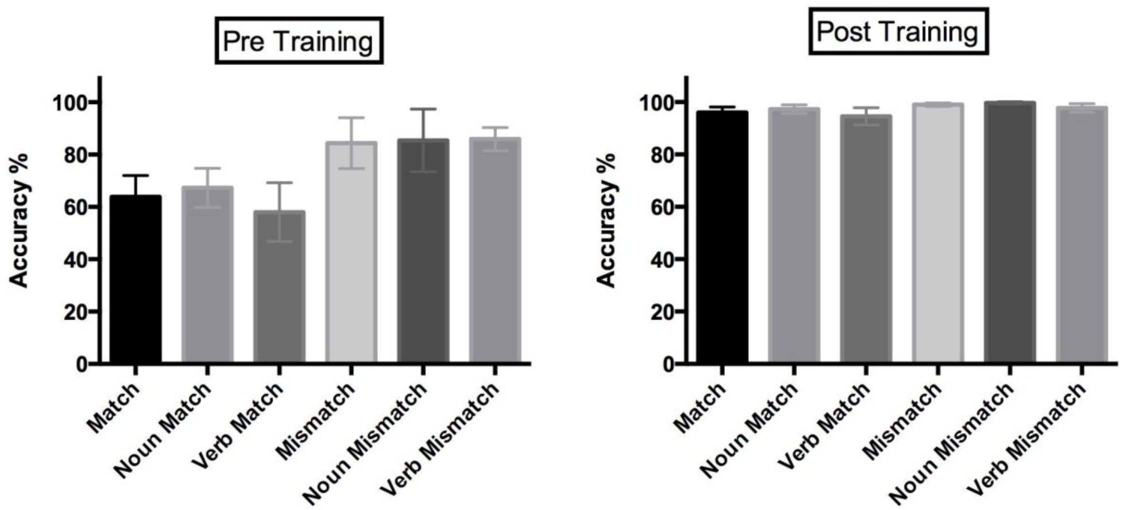


Figure 3.1. Pre- and Post-training accuracy for Match/Mismatch EEG task. Error bars indicate 95% confidence intervals.

The post-training Match/Mismatch EEG task incorporated two new conditions, a semantically related condition and a semantically unrelated condition. As described in the Methods, this task was also administered to a group of native French speakers. A repeated measures ANOVA was used to investigate differences in accuracy across all four conditions and whether accuracy scores varied between the learners and native French speakers, as seen in Figure 3.2.

The assumption of sphericity was violated (due to heterogeneity of variance between the groups), so the Greenhouse-Geisser correction was used. A significant main effect of condition was found, $F(1.2, 29.6) = 175.74$, $p < .001$, $\eta^2 = 0.88$, but the interaction between condition and group was not significant, $F(1.2, 29.6) = 2.11$, $p = .10$, $\eta^2 = 0.08$, nor was there a main effect of group, $F(1.2, 29.6) = 0.54$, $p = 0.8$, $\eta^2 = .002$. . Follow-up post hoc t-tests (Bonferroni-corrected for 16 comparisons) revealed that for learners, accuracy was lower in the semantically related condition compared to each of the other conditions (semantically related vs match, $t(12) = 9.5$, $p < .001$, $d = -3.6$; semantically related vs mismatch, $t(12) = 11.56$, $p < .001$, $d = -4.4$; semantically related vs unrelated, $t(12) = 11.90$, $p < 0.001$, $d = -3.9$). This pattern of results was also observed in the controls, with significantly lower accuracy in the semantically related condition compared to all other conditions (semantically related vs match, $t(12) = 9.7$, $p < .001$, $d = -3.7$; semantically related vs mismatch, $t(12) = 9.9$, $p < .001$, $d = -3.4$; semantically related vs unrelated, $t(12) = 5.90$, $p < 0.001$, $d = -2.4$). For both learners and controls, the accuracy in the match condition was lower than the accuracy in the mismatch condition (learners match vs mismatch $t(12) = 3.2$, $p = 0.007$, $d = -1.1$; controls match vs mismatch $t(12) = 8.6$, $p = 0.007$, $d = -3.5$). All other comparisons were not significant (ps > .05).

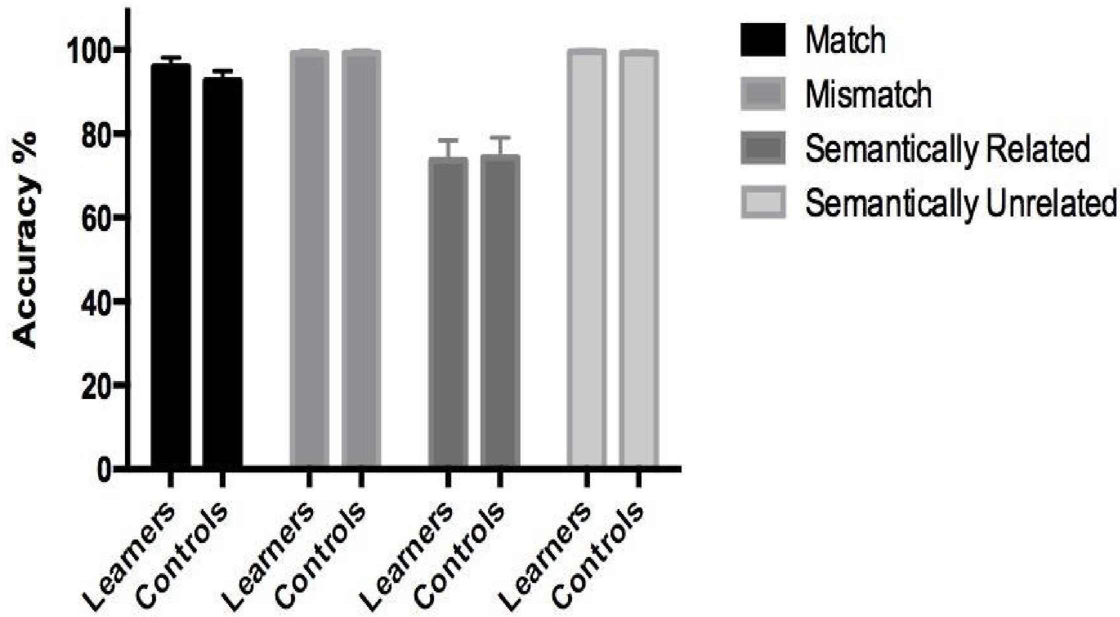


Figure 3.2. Post-training condition-by-group analysis. Error bars represent 95% confidence intervals.

To investigate whether post-training accuracy varied as a function of training strategy, paired samples *t*-tests were used to compare accuracy between training strategies within each condition. This data is shown in Figure 3.3. For the match condition, there was no difference in accuracy between explicit words (M = 96.6%, SD = 3.6) and implicit words (M = 95.5%, SD = 4.17), $t(12) = 1.1$, $p = .29$, $d = 0.2$. For the mismatch condition, accuracy for explicit words (M = 99.8%, SD = 0.77) was higher than for implicit words (M = 98.3%, SD = 2.13), $t(12) = 2.21$, $p = .04$, $d = 0.94$. Although this latter difference was statistically significant, we note that it was only on the order of 1.5%, and that post-training Match/Mismatch accuracy was extremely high in all conditions.

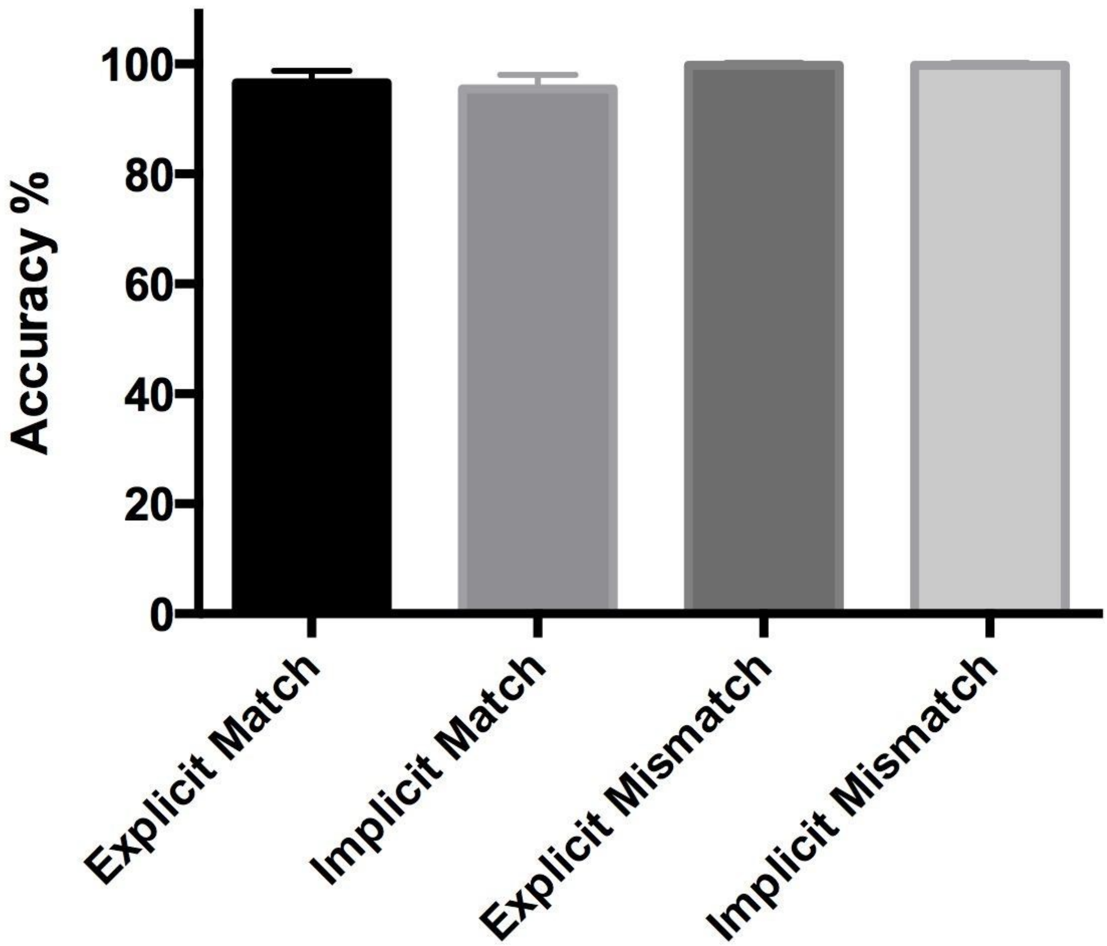


Figure 3.3. Effects of training strategy. Error bars represent 95% confidence intervals.

3.2 EEG Results

3.2.1 Pre to Post Comparison

We first compared learner's pre- and post-training data, with the aim to determine if there was an emergence of the N400 on mismatch trials post-training, and if the effect varied by grammatical category. Visual inspection of the pre-training grand average, across both nouns and verbs suggested the presence of a larger negativity for the mismatch than match condition, occurring approximately 400 ms post stimulus onset, as seen in Figure 3.4 (top) — consistent with an N400 effect. This result also appeared to be present when the data for nouns (Figure 3.5, top) and verbs (Figure 3.6, top) were inspected separately. In the post-training data, the larger N400 effect for mismatch compared to match trials appeared to increase relative to pre-training in the grand average (Figure 3.4, bottom), and also when data from nouns and verbs were examined separately — although the increase appeared larger for nouns (Figure 3.5, bottom), compared to verbs (Figure 3.6, bottom).

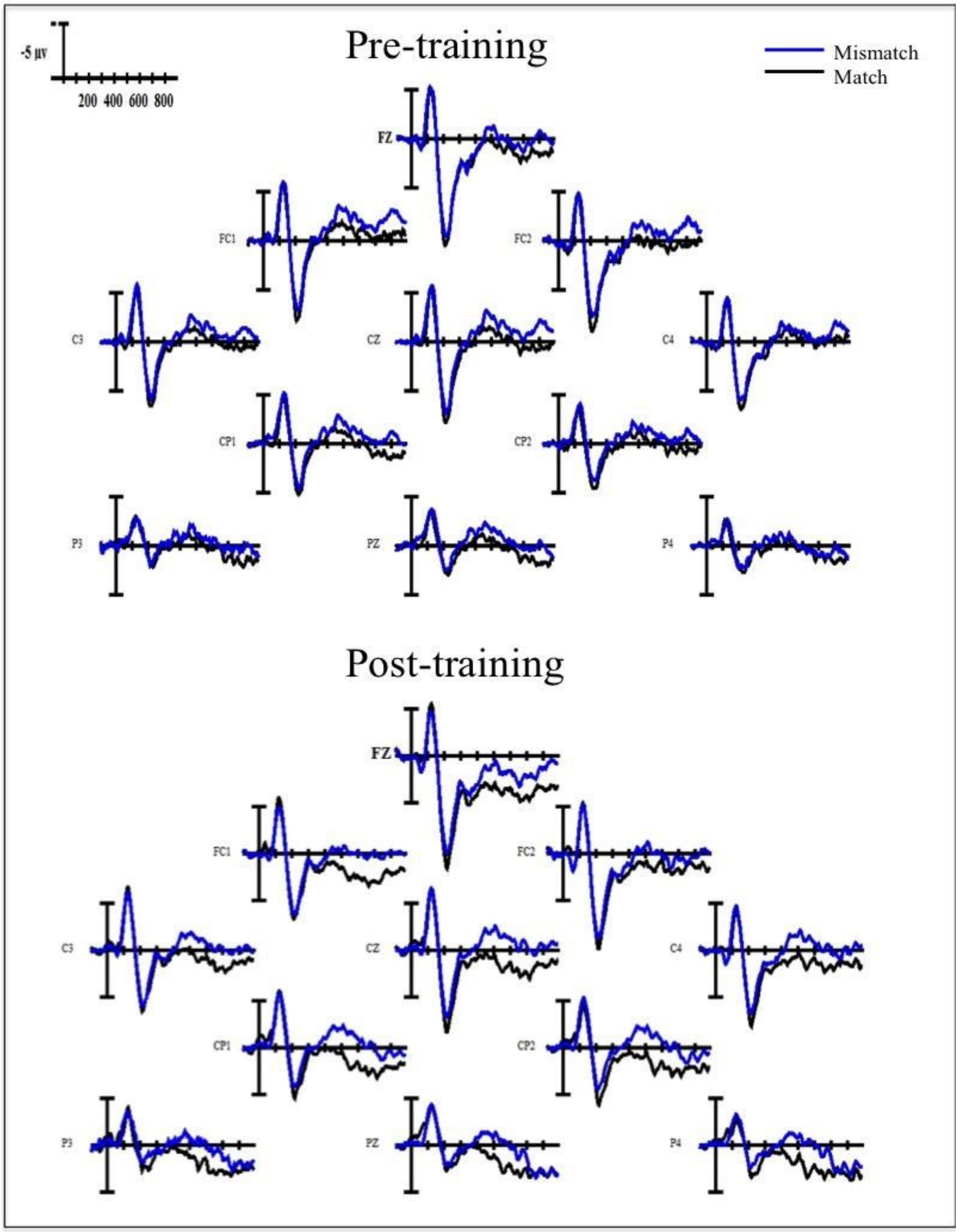


Figure 3.4. Overall pre-training and post-training grand average ERP waveforms at 11 electrode sites.

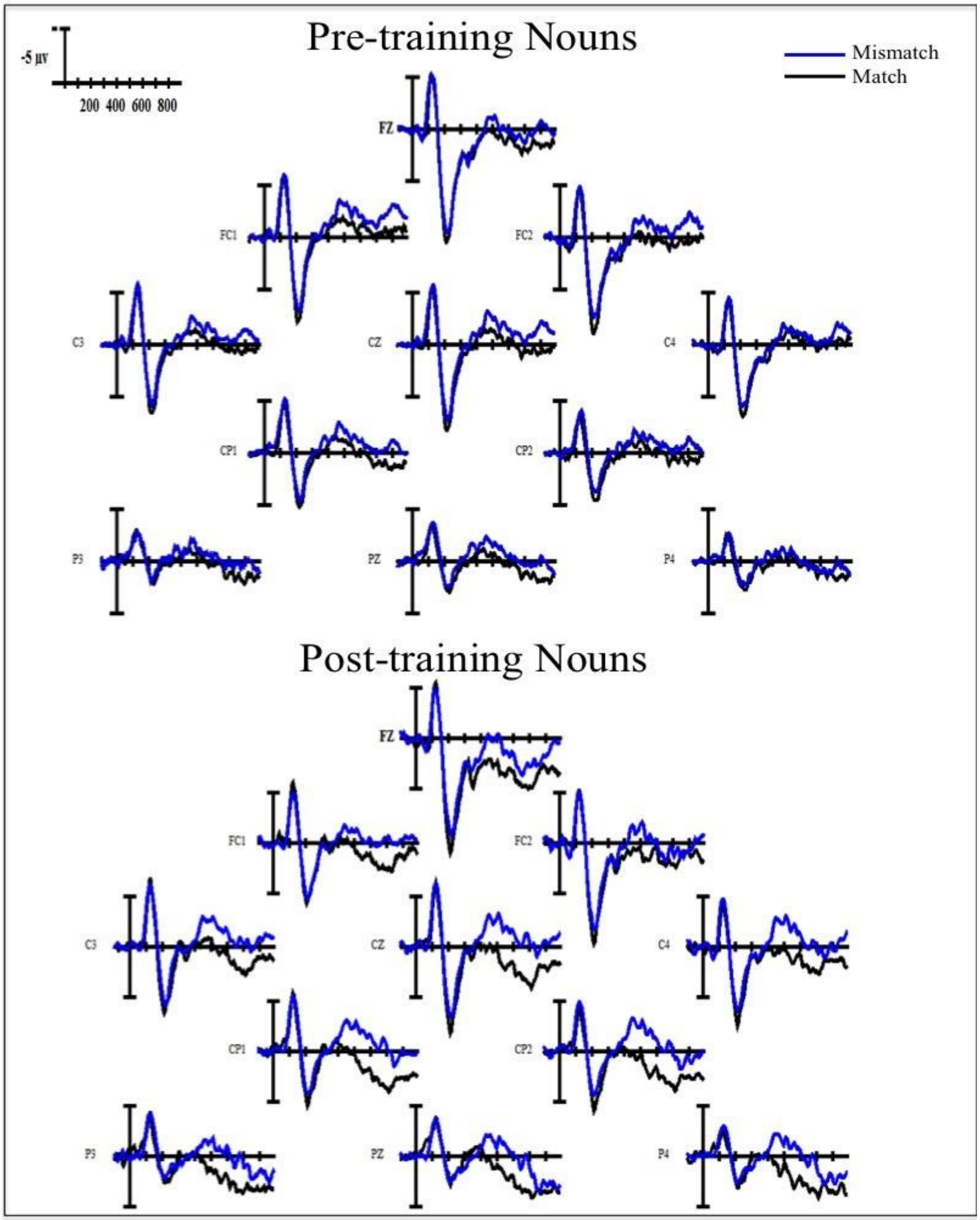


Figure 3.5. Pre-training and post-training grand average ERP waveforms at 11 electrode sites, for nouns.

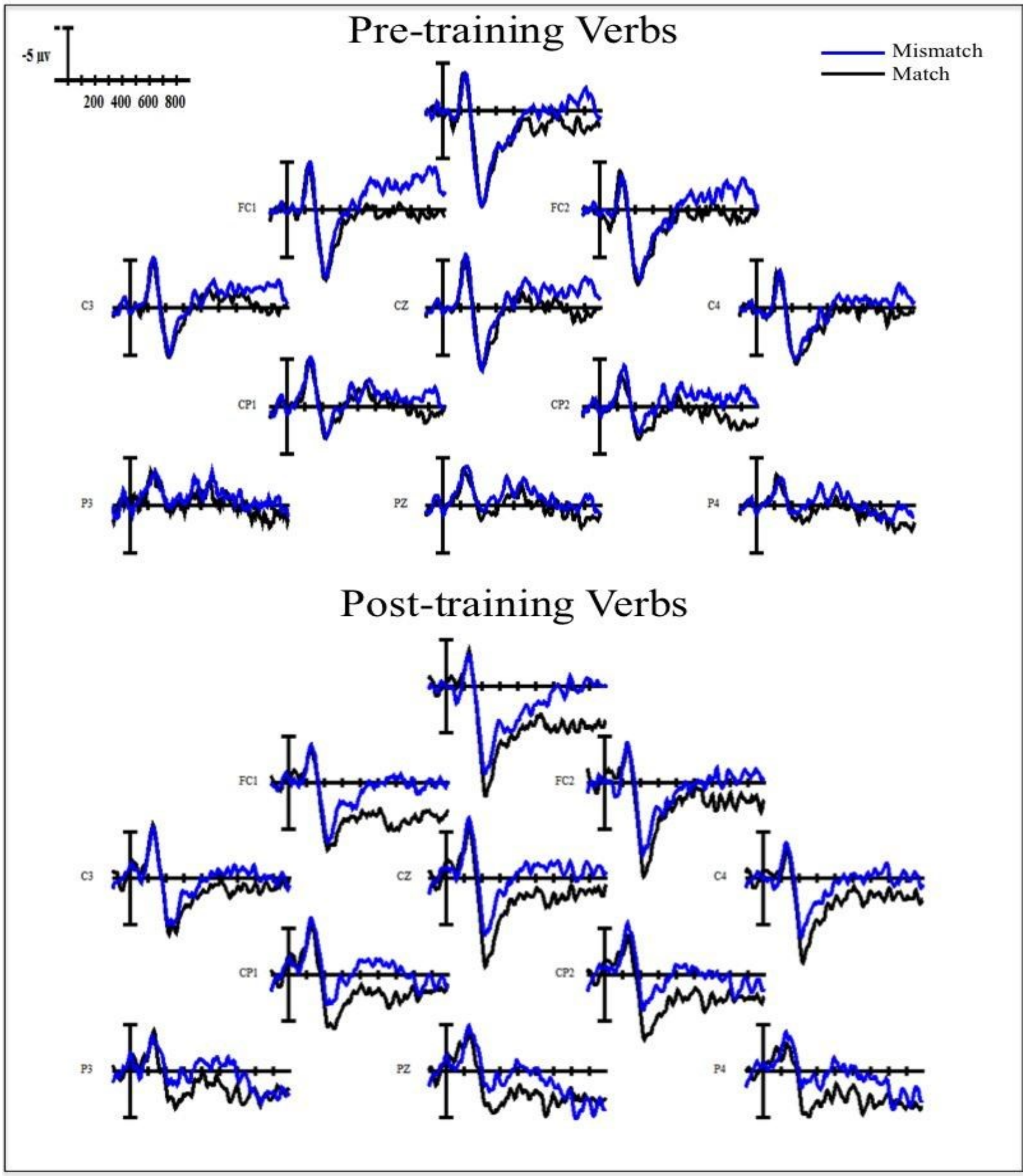


Figure 3.6. Pre-training and post-training grand average ERP waveforms at 11 electrode sites, for verbs.

To analyze the amplitude of the N400, we performed LME modeling on the mean amplitude from 300-500, using session (pre/post), condition (match/mismatch), grammatical category (noun/verb), and electrode as fixed effects and subject as a random effect (outliers removed: n = 21, 1.8%). The optimal model was found to include the factors electrode, session, condition and grammatical category, as well as the interactions session × condition, session × grammatical category, condition × grammatical category, and session × condition × grammatical category (Table 3.1A).

The presence of a significant three-way interaction (session × condition × grammatical category) suggested that the change in the N400 effect seen from pre- to post-training may be dependent on grammatical category (see Figure 3.7). To further investigate this, we conducted follow-up post hoc *t*-tests (Bonferroni corrected for 6 comparisons) at electrode Cz (where the N400 amplitude was maximal) comparing the N400 effect during the pre- and post-training sessions, separately for nouns and verbs (see Table 3.1B for *t*- and *p*-values). Pre-training, the ERP amplitude of the match and the mismatch conditions did not differ for nouns ($M_{mismatch} - M_{match} = -0.1$ μV) but for verbs the amplitude of the mismatch condition was found to be significantly more negative than the amplitude of the match condition ($M_{mismatch} - M_{match} = -1.6$ μV). Post-training, the mismatch amplitudes were significantly more negative than match for both nouns ($M_{mismatch} - M_{match} = -0.9$ μV) and verbs ($M_{mismatch} - M_{match} = -1.1$ μV). The size of the N400 did not change significantly from pre- to post- training for nouns ($M_{diff} = -0.7$ μV), or for verbs ($M_{diff} = 0.5$ μV). To summarize, although there was a significant N400 effect for mismatch trials post-training for nouns, the change in the N400 effect from pre- to post-training was not significant. The N400 effect was present on mismatch trials pre-training for verbs, and effect was maintained in the post-training data.

Table 3.1. Pre- and post-training comparison. Results of linear mixed effects analysis for N400 amplitude in the 300–500 msec time window (denominator lower-bound df = 1092; upper-bound df = 1105) and post hoc tests. *P*-values were computed using lower-bound degrees of freedom. * = significant based on alpha level of .05

Coefficient	*df*	*SumSq*	*MeanSq*	*F*	*p (lower)*
A. Linear Mixed Effects Results					
Electrode	10	504.08	50.4	8.05	< .001*
Session (pre/post)	1	62.97	62.97	10.06	.001*
Condition (match/mismatch)	1	259.52	259.52	41.46	< .001*
Grammatical class (noun/verb)	1	47.78	47.78	7.63	.005*
Session x Condition	1	0.4	0.4	0.06	.79
Session x Gram	1	72.88	72.88	11.64	< .001*
Gram x Condition	1	46.65	46.65	7.45	< .001*
Session x Condition x Gram	1	31.82	31.82	5.08	.02*
Post Hoc Comparisons				*t*	*p (lower)*
B. Session × Condition × Gram					
Pre noun match vs mismatch				0.60	.54
Pre verb match vs mismatch				5.60	< .001*
Post noun match vs mismatch				3.10	.001*
Post verb match vs mismatch				3.55	< .001*
Pre noun mismatch-match vs post noun mismatch-match				1.79	.42
Pre verb mismatch-match vs post verb mismatch-match				1.39	0.96

Post hoc probability values were Bonferroni corrected for 6 comparisons.

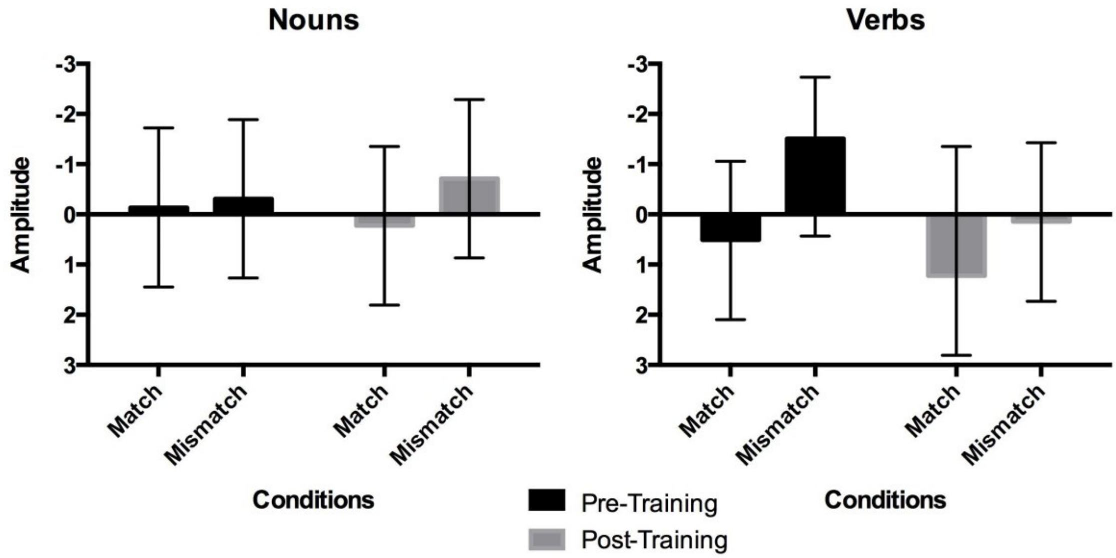


Figure 3.7. ERP amplitude (μV) during 300-500 ms time window at electrode Cz. Pre-training and post-training data compared, for both nouns (left) and verbs (right). Error bars indicate 95% confidence intervals.

3.2.2 Effects of Training Strategy

We assessed whether training strategy had an impact on the amplitude of the post-training mismatch N400 effect. Post-training, the overall grand average indicated that for both explicit and implicit words, the N400 effect was greater for mismatch than match trials post-training, however the effect appears larger for the explicit words (Figure 3.8, top), compared to the implicit words (Figure 3.8, bottom).

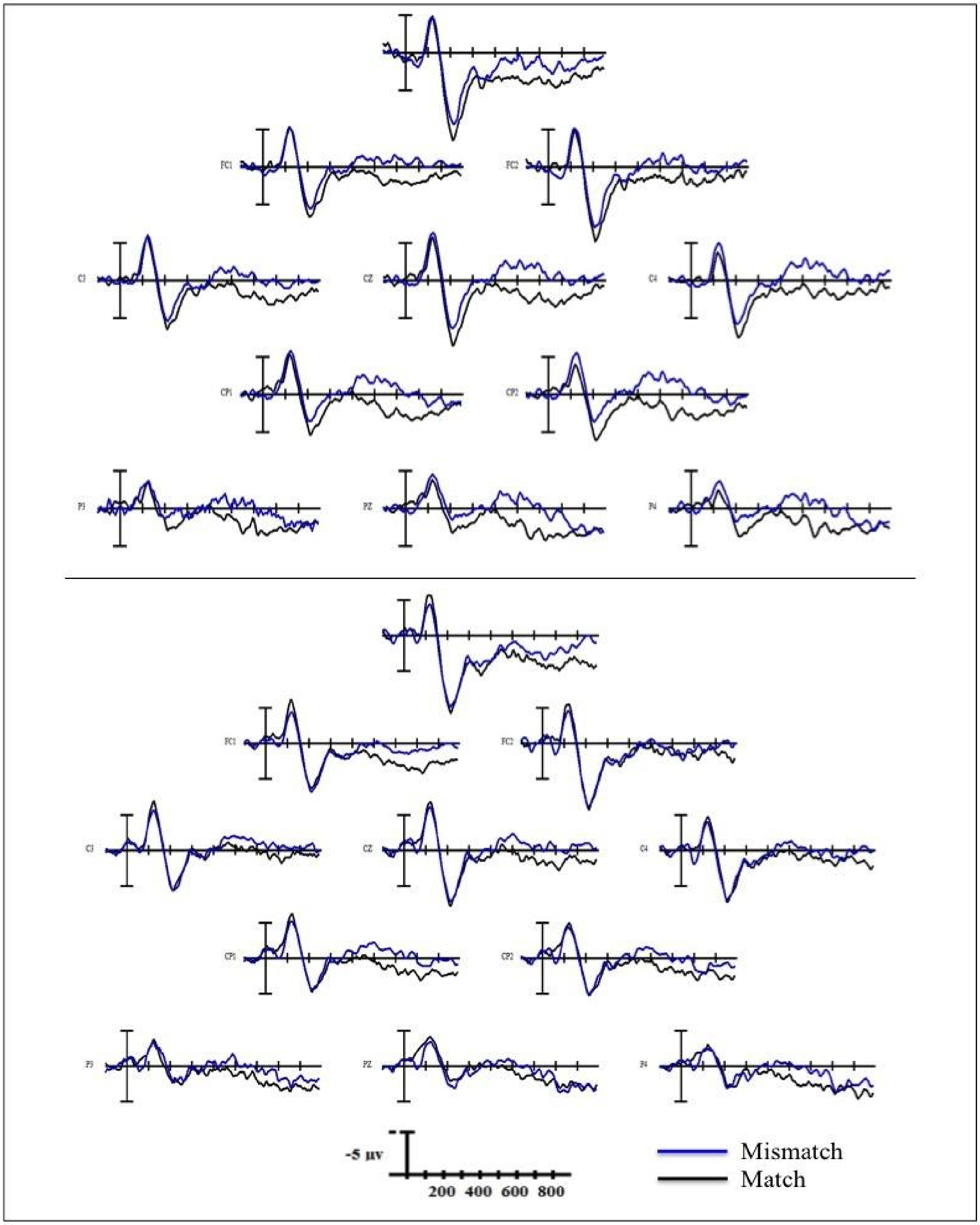


Figure 3.8. Post-training ERP amplitudes for words taught using explicit (top) and implicit (bottom) training strategies.

LME modeling was conducted with session (pre/post), condition (match/mismatch), training strategy (explicit/implicit), and electrode as fixed effects, and subject as a random effect (outliers removed: n = 12, 2.09%). It should be noted that pre- and post-training, the explicit and implicit training strategies do not represent different groups of words, as word assignment to category was balanced across participants. The optimal model contained the factors session, condition, training strategy, electrode and the interactions session × condition, session × training strategy, condition × training strategy and the three-way interaction session × training strategy × condition (Table 3.2A).

Follow up post hoc comparisons investigating the three-way interaction between session, training strategy and condition (Bonferroni corrected for 6 comparisons) at electrode Cz (see Table 3.2B for *t*- and *p*-values) indicated that pre-training, for explicit words, the mismatch and match condition did not differ ($M_{mismatch} - M_{match} = -0.6$ μV), and for implicit words, the mismatch condition was more negative than the match condition ($M_{mismatch} - M_{match} = -0.95$ μV). Post-training, for explicitly trained words, the mismatch condition was more negative than the match condition ($M_{mismatch} - M_{match} = -1.6$ μV). For implicitly trained words, the mismatch and match conditions did not differ ($M_{mismatch} - M_{match} = -0.4$ μV). The change in the difference between match and mismatch from pre- to post-training was significant for explicitly-training words ($M_{diff} = -1.12$ μV), but was not significant for implicitly trained words ($M_{diff}= 0.42$ μV) To sum up, we found a significant N400 effect for mismatch trials post-training, but the effect was only significant for words taught via the explicit training strategy and the change from pre- to post-training was only significant for explicitly trained-words (Figure 3.9).

Table 3.2. Comparison of training strategies. Results of linear mixed effects analysis for N400 Amplitude in the 300–500 msec time window (denominator lower-bound df =1012 ; upper-bound df = 1103) and post hoc tests. *P*-values were computed using lower-bound degrees of freedom. * = significant based on alpha level of .05.

Coefficient	*df*	*SumSq*	*MeanSq*	*F*	*p (lower)*
A. Linear Mixed Effects Results					
Electrode	10	507.05	50.70	17.96	.29
Session	1	3.02	3.02	1.07	.30
Condition	1	31.15	31.15	11.04	< .001*
Training Strategy	1	3.08	3.08	1.09	.29
Training Strategy x Condition	1	12.88	12.88	4.96	.03*
Session x Condition	1	9.33	9.33	3.29	.06
Session x Training Strategy	1	40.60	40.60	14.38	.0002*
Session x Condition x Training Strategy	1	42.06	42.06	14.90	.0001*

Post Hoc Comparisons	*t*	*p (lower)*
B. Session x Condition x Training		
Strategy	1.81	.42
Pre Explicit match vs mismatch	2.83	.02*
Pre Implicit match vs mismatch	7.21	< .001*
Post Explicit match vs mismatch	2.13	.12
Post Implicit match vs mismatch	3.90	.0006*
Explicit pre mismatch-match vs		
Explicit post mismatch-match	1.49	.78
Implicit pre mismatch-match vs		
Implicit post mismatch-match		

Post hoc probability values were Bonferroni corrected for 6 comparisons.

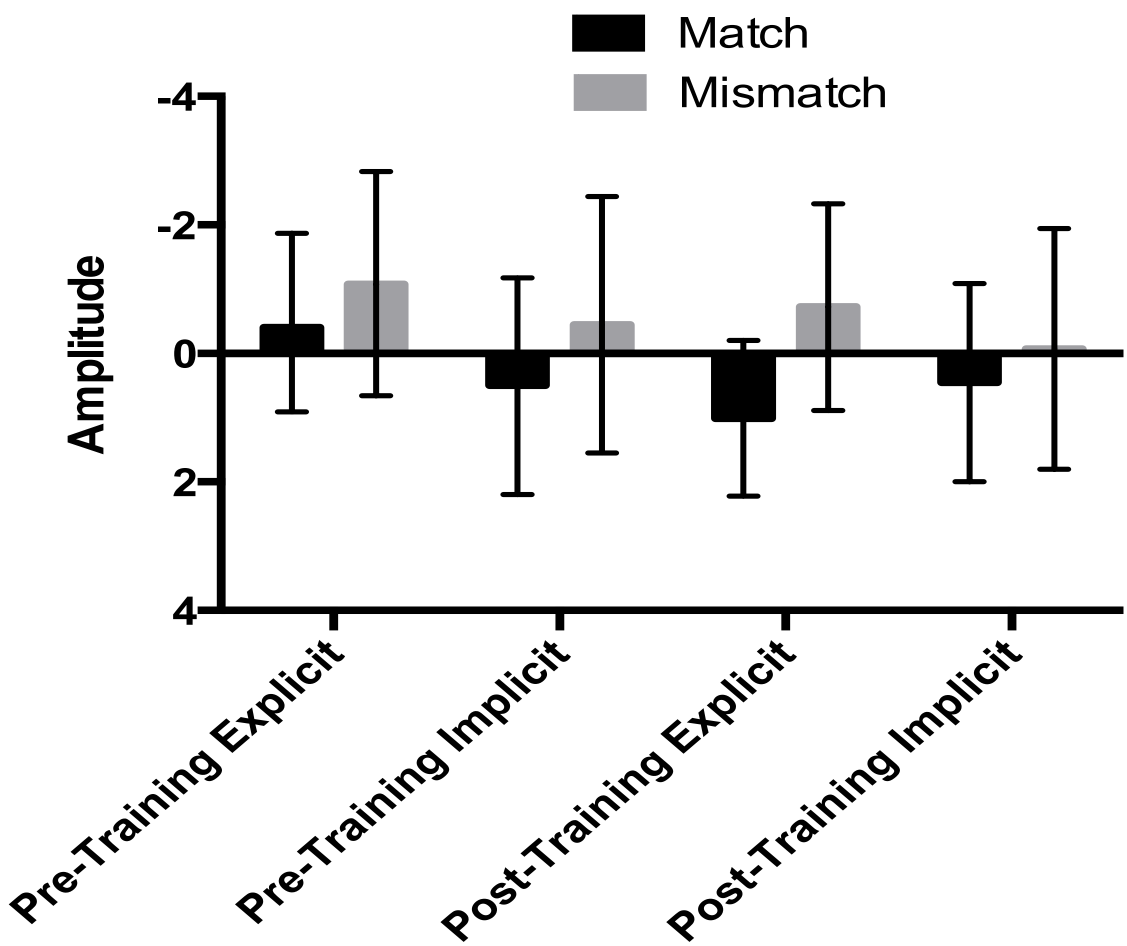


Figure 3.9. Pre- and post-training ERP amplitudes (µV) at electrode Cz comparing training strategy. Error bars indicate 95% confidence intervals.

3.2.3 Post-Training Group Comparison

Next we investigated whether manipulating the semantic relationship between mismatch pairs had an influence on the post-training N400 effect, and if any modulation of the N400 effect was consistent between learners and controls. Visual inspection of grand averages suggested that in both learners and controls, the semantically related condition showed reduced negativity compared to the mismatch and semantically unrelated conditions, as seen in Figure 3.10.

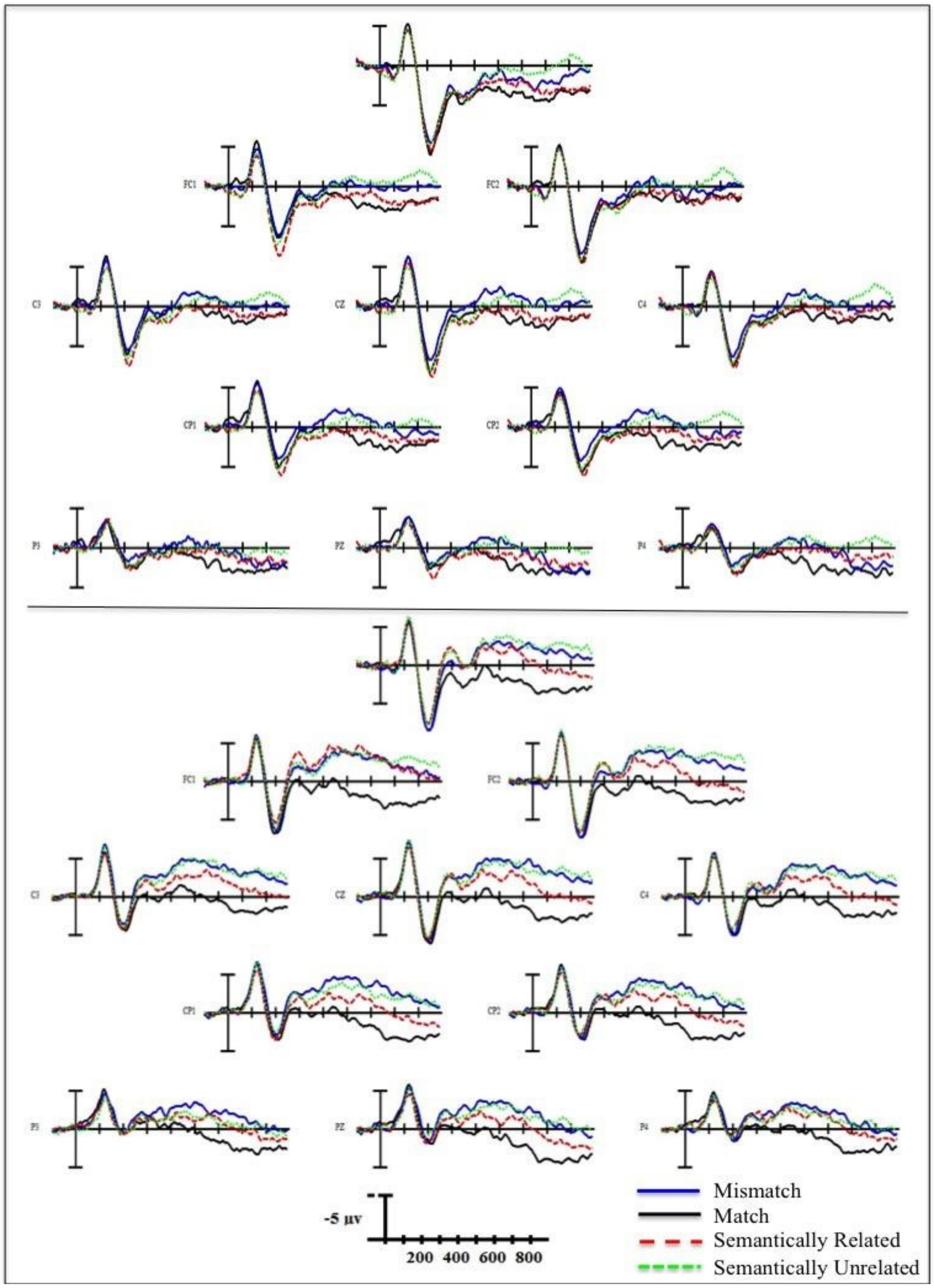


Figure 3.10. Post-training ERP waveforms of all conditions, for both learners (top) and controls (bottom).

We performed LME modeling on the 300-500 ms mean amplitude values with condition (match/mismatch/semantically related/semantically unrelated), group (learners/controls), and electrode as fixed effects and subject as a random effect (outliers removed: n = 22, 1.9%). The optimal model included the factors group, condition, electrode and the interaction group × condition (Table 3.3B).

Follow up post hoc *t*-tests (Bonferroni corrected for 16 comparisons) investigating the group × condition interaction at electrode Cz (see Table 3.3B for *t*- and *p*-values) revealed that for the learners, both the amplitude of the mismatch condition and the amplitude of the semantically unrelated condition were more negative than the amplitude of the match condition ($M_{mismatch}$ - M_{match} = -1.0 µV, M_{semrel} – M_{match} = -0.5 µV). The match and semantically related conditions did not differ (M_{semrel} – M_{match} = 0.3 µV). The mismatch condition was more negative than the semantically related condition ($M_{mismatch}$ – M_{semrel} = -1.3 µV), but did not differ from the semantically unrelated condition ($M_{mismatch}$ – $M_{semunrel}$ = -0.5 µV). Finally, the semantically unrelated condition was more negative than the semantically related condition ($M_{semunrel}$ – M_{semrel} = -0.8 µV). To summarize, the N400 was larger for both mismatch and semantically unrelated trials than for either match or semantically related trials; further, mismatch and semantically unrelated responses did not differ from each other, and similarly, match and semantically related responses did not differ from each other (see Figure 3.11).

For the native French speaker control group, the mismatch condition, the semantically unrelated condition and the semantically related condition were all more negative than the match condition ($M_{mismatch}$ – M_{match} = -2.7 µV; $M_{semunrel}$ – M_{match} = -2.5 µV; M_{semrel} – M_{match} = -1.9 µV). The mismatch condition was more negative than the semantically related condition ($M_{mismatch}$ – M_{semrel} = -0.7 µV), but did not differ from the semantically unrelated condition ($M_{mismatch}$ – $M_{semunrel}$= -0.2 µV). Finally, the semantically related condition did not differ from the semantically unrelated condition ($M_{semunrel} - M_{semrel} = -0.6\ \mu V$). Overall, the pattern of results was similar to those of the learners, showing larger N400 effects for both the mismatch and semantically unrelated conditions than the match condition. However, whereas for the learners no difference between the match and semantically related conditions were observed, for the controls these conditions did differ with a larger N400 effect for related than for matching words.

Table 3.3. Comparison of learners and controls. Results of linear mixed effects analysis for N400 Amplitude in the 300–500 msec time window (Denominator Lower-bound df = 1008; Upper-bound df = 1034) and post hoc tests. *P*-values were computed using lower-bound degrees of freedom. * = significant based on alpha level of .05.

Coefficient	*df*	*SumSq*	*MeanSq*	*F*	*p (lower)*
A. Linear Mixed Effects Results					
Electrode	10	492.05	49.20	13.53	< .001*
Condition	3	586.05	195.35	53.73	< .001*
Group	1	14.80	14.80	4.07	.04*
Group x Condition	3	224.78	74.92	20.60	< .001*
Post Hoc Comparisons				*t*	*p (lower)*
B. Group x Condition					
Controls					
match vs mismatch				11.94	< .001*
match vs semantically related				8.77	< .001*
match vs semantically unrelated				11.26	< .001*
mismatch vs semantically related				3.17	.01*
mismatch vs semantically unrelated				0.69	.99

Post Hoc Comparisons	*t*	*p (lower)*
Controls (continued)		
semantically related vs unrelated	2.48	.16
Learners		
match vs mismatch	4.51	< .001*
match vs semantically related	1.33	.99
match vs semantically unrelated	2.26	.02*
mismatch vs semantically related	5.90	< .001*
mismatch vs semantically unrelated	2.24	.24
semantically related vs unrelated	3.61	.003*
Controls vs Learners		
C – match vs L- match	0.55	0.99
C – mismatch vs L – mismatch	2.17	0.47
C – semantically related vs L – semantically unrelated	2.75	0.09
C – semantically unrelated vs L – semantically unrelated	2.50	0.19

Post hoc probability values were Bonferroni corrected for 16 comparisons. C = Control, L = Learners

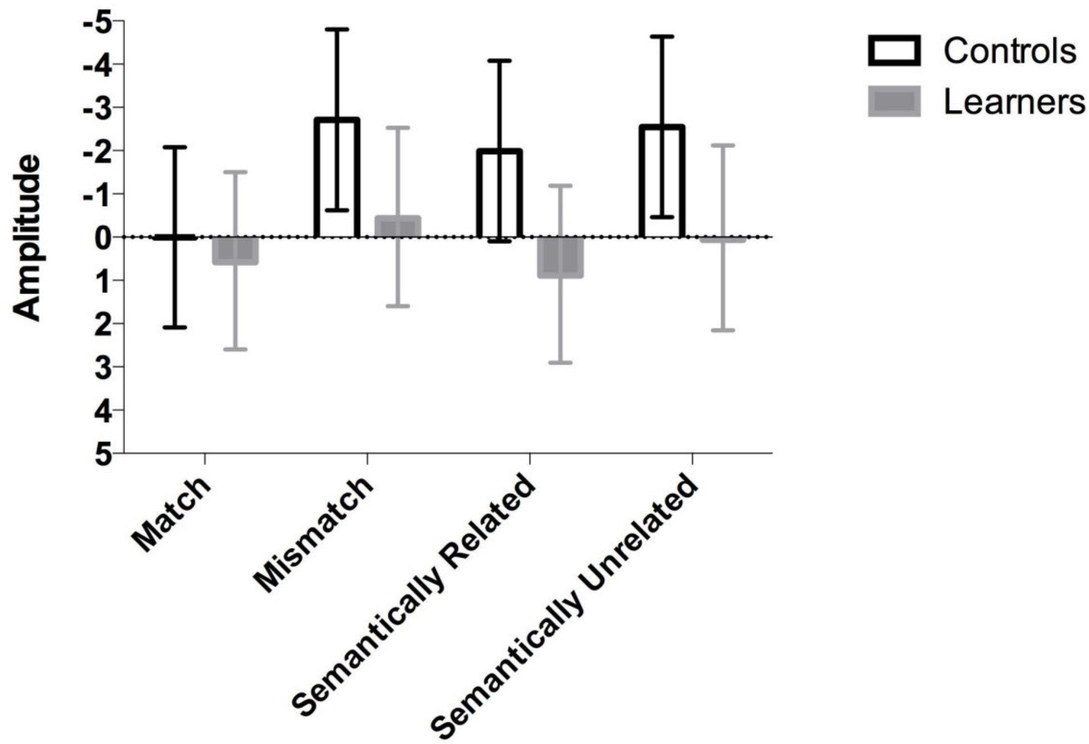


Figure 3.11. ERP amplitude (μV) during 300-500 ms time window at electrode Cz comparing all post-training conditions, in both learners and controls. Error bars indicate 95% confidence intervals.

Consistent with previous studies comparing the N400 effect between native and late learners, visual inspection of the present data suggested that the onset of the N400 effect may have differed between the leaners and controls. We thus computed difference waves at electrode Cz (match-mismatch, and semantically related-semantically unrelated) and performed a fractional area latency analysis. This routine provides an estimate of onset latency by computing the area under the component of interest, and then determining the time at which cumulative area under the curve reaches a pre-specified percentage of this total area (in this case, 15%). However, no significant differences in onset latency between were observed for the match-mismatch N400 (controls = 355 ms, SD = 40.4, learners = 366 ms, SD = 47.6), $t(24) = 0.68$, $p = 0.49$, or the semantically related-semantically unrelated N400 (controls = 381.5 ms, SD = 52.41, learners = 348.1 ms, SD = 32.63), $t(24) = 1.95$, $p = 0.06$.

3.2.4 Effect of Training Strategy on Semantic Integration

Next we investigated whether training strategy had an impact on semantic integration of new vocabulary into existing semantic networks (please see Figure 3.12). LME modeling of the mean amplitude during the 300-500 ms with condition, training strategy and electrode as fixed effects and subject as a random effect (outliers removed = N=18, 1.5%), resulted in a model with training strategy, condition, electrode and the interaction condition × training strategy (see Table 3.4A).

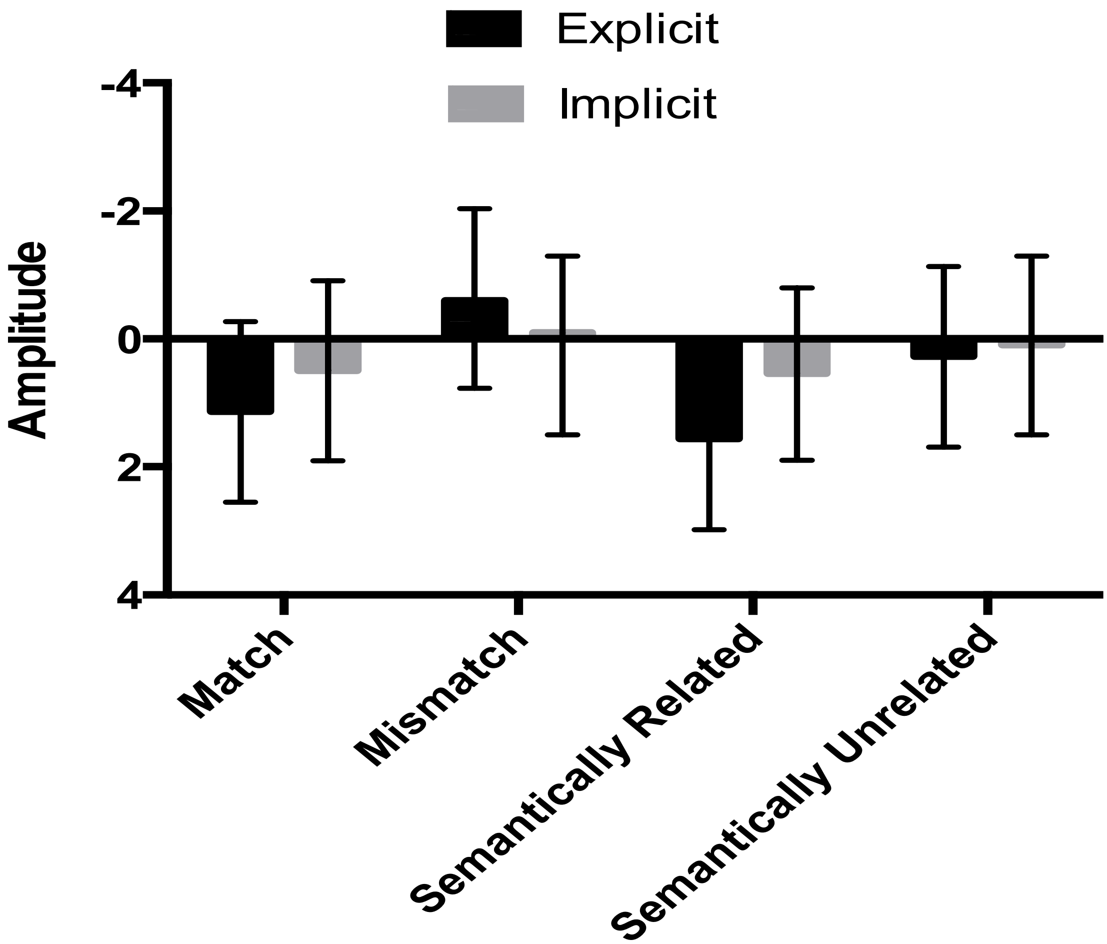


Figure 3.12. Mean ERP amplitude at electrode Cz during the 300-500 ms time window, comparing the effect of training strategy on semantic integration. Error bars indicate 95% confidence intervals.

Follow-up comparisons (Bonferroni corrected for 12 comparisons) revealed that for explicit words, in addition to the difference between the mismatch and match conditions previously reported (see section 3.2.2), the semantically related condition was reduced in negativity compared to the mismatch condition ($M_{mismatch}$-M_{semrel} = 0.43 μV) and the semantically unrelated condition ($M_{semunrel}$-M_{semrel} = -1.29 μV), but did not differ from the match condition (M_{semrel} – M_{match} = -0.43 μV). The mismatch condition was more negative than the semantically unrelated condition ($M_{mismatch}$ – $M_{semunrel}$ = 0.91 μV).

For implicitly learned words, none of the conditions differed from each other ($M_{mismatch} - M_{match}$ = -0.39 μV, $M_{mismatch} - M_{semrel}$ = 0.43 μV, $M_{mismatch} - M_{semunrel}$ = -0.01, $M_{semrel} - M_{match}$ = 0.04 μV, $M_{semrel} - M_{semunrel}$ = -0.45, $M_{semunrel} - M_{match}$ = 0.40 μV). Please see Table 3.4B for *t* and *p* values.

Table 3.4. Effects of training strategy on semantic integration. Results of linear mixed effects analysis for N400 amplitude in the 300–500 msec time window (denominator lower-bound df =1082; upper-bound df = 1108) and post hoc tests. *P*-values were computed using lower-bound degrees of freedom. * = significant compared to alpha level of .05

Coefficient	*df*	*SumSq*	*MeanSq*	*F*	*p (lower)*
A. Linear Mixed Effects Results					
Electrode	10	460.11	46.01	9.78	< .001*
Condition	3	312.48	104.16	22.16	< .001*
Training Strategy	1	24.62	24.62	5.23	.02*
Condition x Training Strategy	3	123.84	41.28	8.78	< .001*

Post Hoc Comparisons	*t*	*p (lower)*
B. Condition x Training Strategy		
Explicit		
Match vs mismatch	6.9	<.001*
Match vs semantically related	1.67	.99
Match vs semantically unrelated	3.31	.01*
Mismatch vs semantically related	8.60	<.001*
Mismatch vs semantically unrelated	3.52	.004*
Semantically related vs unrelated	4.90	<.001*

Post Hoc Comparisons	*t*	*p (lower)*
Implicit		
Match vs mismatch	1.50	.99
Match vs semantically related	0.17	.99
Match vs semantically unrelated	1.52	.99
Mismatch vs semantically related	1.90	.96
Mismatch vs semantically unrelated	.05	.99
Semantically related vs unrelated	1.74	.96

Post hoc probability values were Bonferroni corrected for 12 comparisons.

3.2.5 Post-Training Learning Across Days

Finally, we investigated whether the mismatch N400 effect differed across learning day. Day learned for this analysis corresponded to the day on which the auditory target word was learned. When broken down by day, the post-training results at electrode Cz appeared to show that words learned on Days 1-3 showed a mismatch N400 effect, while words learned on Days 4-6 showed a reduced or delayed effect, as shown in Figure 3.13.

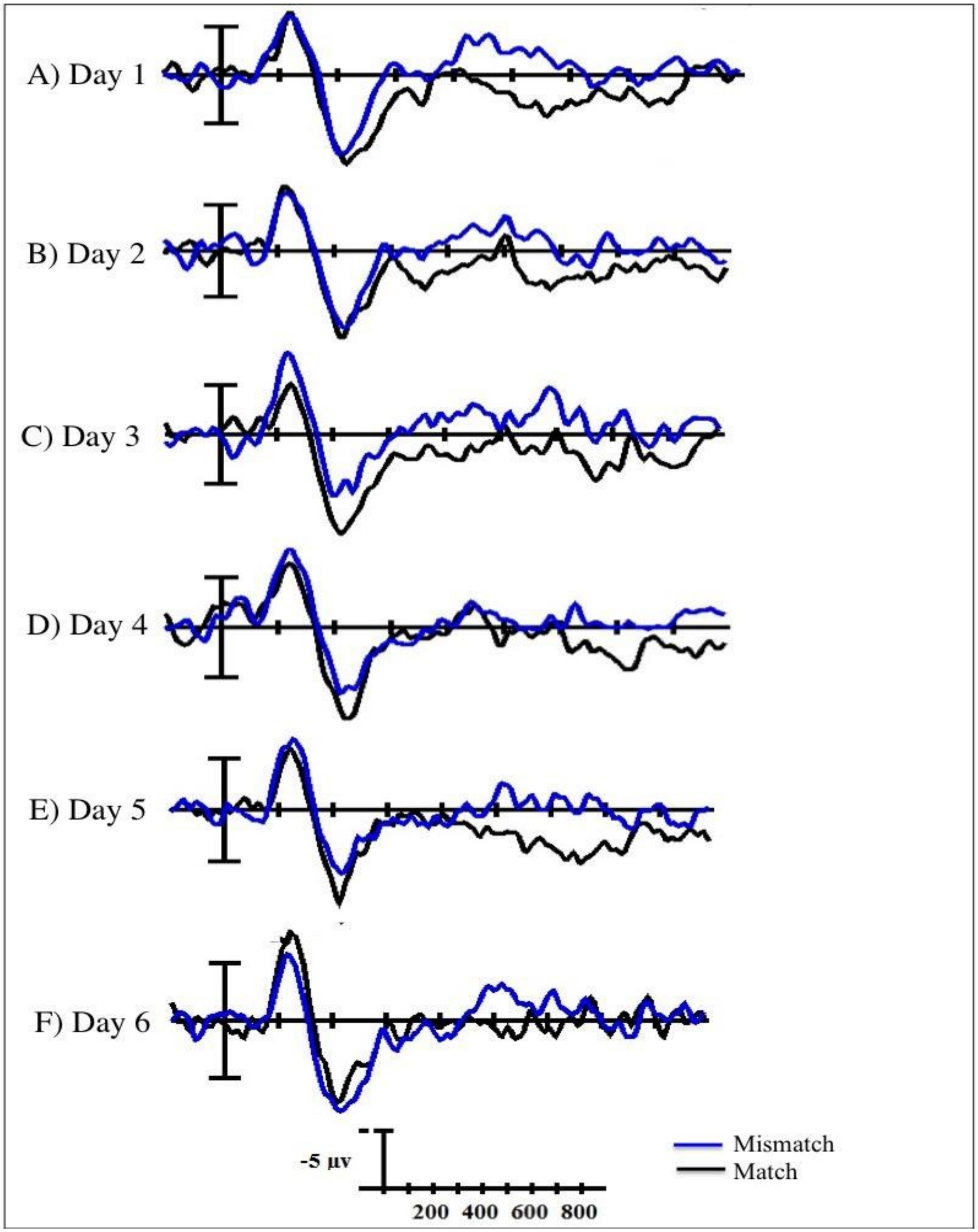


Figure 3.13. Post-training ERP amplitudes comparing the N400 at electrode Cz, across days. A) Day 1 B) Day 2 C) Day 3 D) Day 4 E) Day 5 F) Day 6

LME modeling of mean amplitude values during the 300-500 ms time window, run with condition, day and electrode as fixed effects and subject as a random effect (outliers removed: n = 40, 2.33%), indicated the optimal model contained the factors condition, day, electrode and the interaction day × condition (Table 3.5A).

Follow up post hoc comparisons (Bonferroni corrected for 6 comparisons) investigating the day × condition interaction were conducted at electrode Cz (see Table 3.5B for *t*- and *p*-values). We observed a significant difference between conditions, with mismatch showing increased negativity compared to match, for words learned on Day 1 ($M_{mismatch}$ - M_{match} = -2.7 μV), Day 2 ($M_{mismatch}$ - M_{match} = -1.5 μV) and Day 3 ($M_{mismatch}$ - M_{match} = -2.2 μV), but not for words learned on Day 4 ($M_{mismatch}$ - M_{match} = 0.6 μV), Day 5 ($M_{mismatch}$ - M_{match} = -0.2 μV) or Day 6 ($M_{mismatch}$ - M_{match} = 0.3 μV). This indicates that words learned on Days 1-3 exhibited a significant mismatch N400 effect within the 300-500 ms time window post-training, whereas words learned on Days 4-6 did not (Figure 3.14).

Table 3.5. Comparison across days. Results of linear mixed effects analysis for N400 amplitude in the 300–500 msec time window (denominator lower-bound df =1631; upper-bound df = 1644) and post hoc tests. *P*-values were computed using lower-bound degrees of freedom. * = significant compared to alpha level of .05

Coefficient	*df*	*SumSq*	*MeanSq*	*F*	*p (lower)*
A. Linear Mixed Effects					
Electrode	10	752.56	75.25	7.52	< .001*
Condition	1	400.17	400.17	40.17	< .001*
Day	5	189.47	37.89	3.78	< .001*
Day x Condition	1	632.10	126.42	12.64	< .001*
Post Hoc Comparisons				*t*	*p (lower)*
B. Learn x Condition					
1 – match vs 1 - mismatch				7.00	< .001*
2 – match vs 2 - mismatch				4.11	< .001*
3 – match vs 3 - mismatch				5.80	< .001*
4 – match vs 4 – mismatch				1.50	.79
5 – match vs 5 – mismatch				0.68	.99
6 – match vs 6 - mismatch				0.71	.99

Post hoc probability values were Bonferroni corrected for 6 comparisons. 1-6 indicates learning day

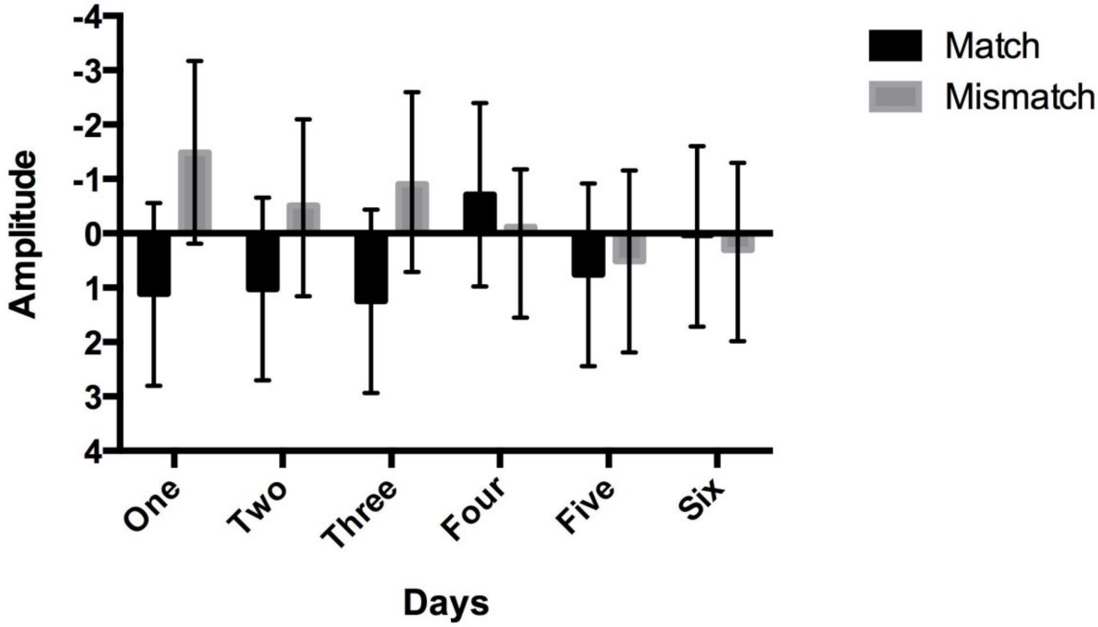


Figure 3.14. ERP amplitudes (μV) during the 300-500 ms time window at electrode Cz, comparing conditions across learning day. Error bars indicate 95% confidence intervals.

CHAPTER 4: DISCUSSION

4.1 Interpretation of Results

4.1.1 Behavioural Results

The primary goal of this research was to investigate whether a relatively short period of L2 training could lead to vocabulary learning and associated neurological changes as indexed by the N400 and well as how different training strategies can impact L2 vocabulary learning. Additionally, we were interested in whether newly learned L2 vocabulary words were integrated into existing L1 semantic networks in the brain. French vocabulary was taught to English speakers via two strategies over approximately 10 hours of instruction. Behavioural and EEG recordings taken prior to initiating training and immediately following training were compared. Results from the learners' post-training assessments were compared to those of native French speakers. We expected that words learned via the implicit strategy, which required participants to deduce a word's meaning based on context, would be learned better than those taught via the explicit strategy. We also expected that newly learned French words would be integrated into existing L1 semantic networks, such that cross-linguistic semantic priming would occur following training, and the learners results would be similar to those of a native French speaker. Overall, we did observe both behavioural and neurophysiological evidence of French vocabulary learning. Both training methods proved to be an effective way to learn L2 vocabulary — thus our hypotheses regarding learning outcomes associated with different training strategies were not supported. However, we did find evidence to support our hypothesis that newly learned L2 words can be rapidly integrated into L1 networks: L2 learners' N400 responses showed similar modulation by semantic relatedness as did those of native speakers, after only ~3 hours of L2 training.

Our results indicated that the training was effective, in that there was a significant increase in the number of words participants were able to name. Contrary to expectations, learners showed better retention of explicitly-learned words, compared to those learned implicitly. However, because this is only a subset of the data, these results may not be representative of a wider range of individuals, and should be further investigated in future studies. Focus in this Discussion will therefore be placed on the results of the Match/Mismatch EEG task.

Prior to initiating training, participants had unexpectedly high accuracy on the Match/Mismatch EEG task ($M_{mismatch}$ = 84.4%, SD = 16.0, M_{match} = 63.9%, SD = 13.5), especially considering they were only able to name an average of 5 of the 72 training words. However, it is possible that although participants could not recall the correct French word, they could recognize the correct word when it was presented as a picture/auditory pair; even in the absence of certainty, participants may have made accurate guesses based on familiarity. Additionally, mismatch accuracy was higher than match accuracy, although this may be partially explained by a slight bias to answer 'No', as opposed to 'Yes'. Given that the task was a forced choice paradigm, and that participants' French knowledge was relatively minimal, when forced to make a guess about the relationship between a pair of words, they may have been more inclined to decide they did not match. Overall, while we did expect that the pre-training accuracy scores for the Match/Mismatch EEG task would be close to chance (because the task was forced choice), the overall pattern of results could be due to pre-existing exposure to French.

After completing training, overall accuracy (collapsing across training strategies and grammatical category) in both the match and mismatch conditions was near ceiling, with significant increases compared to pre-training accuracy in both the match and mismatch conditions. This demonstrates that participants were successful in learning new L2 vocabulary in a short amount of time. Grammatical category did not impact learning, with no differences in accuracy between nouns and verbs in either the pre- or post-training data. With respect to training strategy, in the match condition, training strategy had no effect on accuracy scores, whereas in the mismatch condition, accuracy was higher for explicit words, compared to implicit words. While this is contrary to expectations, it should be noted that, in both cases, accuracy was close to 100%, which indicates that both training strategies were highly effective. Given that accuracy for both training strategies was so high, no concrete conclusions can be drawn about the relative effectiveness of each training strategy, based on behavioural data alone.

Finally, we compared all conditions in the post-training Match/Mismatch EEG task (match, mismatch, semantically related, semantically unrelated), between the two experimental groups (learners and controls). No effect of group was observed, indicating that L2 learners did not differ from native French speakers in terms of their ability to judge the relationship between picture/auditory French word pairs, which indicates that after only 3.5 hours of instruction and 23 exposures to each word, L2 learners were performing comparably to native speakers. As expected, for both groups, there was significantly lower accuracy for the semantically related condition, compared to all other conditions. Lower accuracy as evidence of semantic priming may at first appear counterintuitive. When other types of tasks are used, such as lexical decision tasks, semantic relatedness between word pairs has a facilitatory effect on accuracy and reaction times, increasing and decreasing these measures, respectively (Holcomb & Neville, 1990). However, in the case of lexical decision tasks, the relatedness between the word pairs facilitates the correct answer, which is typically a positive response ('yes'). In contrast, when the correct response is a negative ('no'), overlap between word pairs can inhibit, or interfere with, the correct answer. For semantically related pairs, while they are related, they are not a matching pair, so the correct response is 'No'. Therefore, a reduction in accuracy for semantically related pairs indicates that newly learned L2 vocabulary words were integrated into existing semantic networks, and the influence cross-linguistic semantic priming on behavioural measures is similar to that seen as a result of L1 semantic priming in native speakers.

4.1.2 EEG Results

4.1.2.1 Pre- to Post-training Comparison

When we compared participants' cortical responses to mismatched word pairs from pre- to post-training, we observed a significant N400 component pre-training for verbs, but not for nouns, which could be explained by the higher (log) frequency/million of verbs, compared to nouns. Post-training, the N400 for both nouns and verbs was significant. However, in neither case did we observe a significant increase in the mismatch N400 effect from pre- to post-training. However, examination of the ERP waveforms for nouns in Figure 3.4 reveals that the mismatch waveform was somewhat more negative than the match waveform even pre-training. This effect may not have reached statistical significance, but may account for the lack of a significant change in N400 mismatch effect size for nouns from pre- to post- training. The pre-training N400 effects parallel the behavioural results on this task — although there was a significant increase in participants' performance from pre- to post-training, pre-training scores were above chance indicating pre-existing familiarity with the words. We observed a pre-training N400 effect that, while it did not reach significance, does indicate that the N400 is sensitive to familiarity, and not just recall. This is in line with McLaughlin et al. (2004) who observed an N400 effect in a lexical decision task, prior to accurate behavioural performance. The presence of an overall mismatch N400 effect post-training indicates that brain activity associated with normal/native language lexical-semantic access was elicited by newly-learned words, indicating that even the relatively short amount of training led to the establishment of typical patterns of brain activity associated with accessing word meanings

4.1.2.2 Training Strategy Comparison

A central focus of the current study was to determine if different training strategies impacted learning outcomes. Pre-training, we did observe a significant N400 effect for implicit words, although this is most likely an anomaly, as counterbalancing across participants ensured all words were taught both explicitly and implicitly. As previously mentioned, behavioural results indicated that both strategies promoted learning outcomes. With regards to the neurophysiological data, we predicted that learning words in context would lead to better retention, which would result in larger post-training N400 amplitudes to implicit mismatches, compared to explicit mismatches. However, this was not was what observed. Explicitly taught words elicited a significant

N400 effect, whereas implicitly taught words did not — although a small N400 effect was evident in the EPR waveforms. In addition, only explicitly trained words showed semantic integration, as indicated by the reduced N400 effect for semantically related explicit word pairs, compared to mismatches. It is important to emphasize that implicit words were actually learned very well, as indicated by the behavioural results — thus the lack of N400 to implicitly taught words occurred in spite of subjects' very high behavioural accuracy. It's possible that the N400 may reflect ease of access or the costs associated with accessing a words meaning and not a binary 'learned or not learned' outcome.

Discrepancies between behavioural and neurophysiological outcomes measures have been previously reported (McLaughlin et al., 2004; Morgan-Short, Sanz, & Steinhauer, 2010; Morgan-Short et al., 2012). In McLaughlin et al., (2004) adult L2 learners' brain activity reflected distinctions in L2 words and pseudowords after a short training program, whereas their behavioural responses did not, suggesting that behavioural measures may not completely reflect early acquisition of certain aspects of a new language. Also relevant to the current study are results obtained by Morgan-Short, Sanz & Steinhauer (2010) and Morgan-Short et al., (2012) who directly compared L2 grammar learning via two training programs. In both instances, behavioural outcome measures did not reflect differences in training strategy efficacy, whereas ERP measures indicated that implicit teaching strategies promoted more native-like patterns of brain activity.

In line with previous findings, in this study we found ERP differences between words trained implicitly and explicitly, in the absence of behavioural differences.

However, in contrast to Morgan-Short and colleagues (2010, 2012) we found more native-like ERP effects for the explicitly than implicitly trained words. It should be noted however, that in both studies reported by Morgan-Short and colleagues (2010, 2012), emphasis was placed on grammar training, as opposed to vocabulary training, and so their results may not apply to vocabulary training specifically. This discrepancy between expectations and results, as well as results reported by previous authors (Chun et al., 2012; Morgan-Short et al., 2010; Morgan-Short et al., 2012) could be explained by several factors related to the nature of the LANGA games used during training.

First, automated speech recognition was used to evaluate responses during the games. For explicit words, a single word response was required. However, for implicit words, which were taught in the form of a three-word sentence, the participant had to accurately produce the entire three-word sentence as a response. If the participant made an error on the first or second word of the sentence, s/he would receive immediate negative feedback, effectively halting their response before they were finished. This would mean that for those subjects that consistently made errors, the second and third word of sentences may have been practiced less. In addition, participants had to make a complete response within a certain amount of time. Producing a correct three-word sentence at normal speech speed may have proved difficult, especially for learners struggling with correct pronunciation. Focus may have been placed on producing a 'correct sounding' sentence, as opposed to focusing on sentence content and meaning. Finally, during training, implicit words were only ever seen within a sentence. However, during the pre- and post-assessments, words were presented individually. The ability to recall an item can depend on the similarity between the conditions under which the item was learned and the conditions under which the attempted retrieval occurs (Tulving & Thomson, 1973). The change in context, from viewing words in sentential format, to viewing words individually, may have impacted their ability to recall the correct word associated with the picture. These factors may not have had an influence on behavioural outcome measures, as both training strategies proved effective, however they could partially account for the ERP results obtained.

It is also possible that the results obtained are not due to limitations of the LANGA games. Perhaps the added complexity associated with processing syntax (albeit very simple syntax) may prevent or impede vocabulary learning. At early stages of learning, for very low proficiency learners, providing a direct connection to meaning may be more beneficial. Support for this explanation can be found in the literature. Kersten & Earles (2001) found that adults learned an artificial miniature language better when they were first introduced to one word at a time, than when they were immediately introduced to full sentences. Prince (1996) compared translational learning (linking a word directly to its L1 translation) and context learning and found translation learning to be superior, in terms of recall. In order to concretely ascertain which method provides better learning outcomes, the problems with the training games would first need to be addressed.

4.1.2.3 Semantic Integration

We were also interested in whether cross-linguistic semantic priming could be reflected in brain responses, after only a short training period. We found an effect of semantic relatedness, such that the amplitude of the N400 was reduced for targets preceded by semantically related primes — and this effect was similar in L2 learners and French native speakers. In line with previous studies on cross-linguistic semantic priming

(Chen & Ng, 1989; Frenck-Mestre et al., 2014; Kerkhofs et al., 2006; Law et al., 2005; Martin et al., 2009; Phillips et al., 2004), this indicates that newly learned French words interact with existing semantic concepts during processing, and integration occurs rapidly (after ~3.5 hours of instruction). The fact that the brain responses were comparable to those of native French speakers indicates that cross-linguistic interaction during semantic processing can be similar to within L1 interaction. This provides support for models of bilingual processing that suggest multiple languages exist in a shared lexicon and language activation is non-selective (Dijkstra & Heuven, 1998)

Previous literature on bilingual language processing suggests that there are qualitative differences in the ERP signals produced by semantic violations in first and later-learned languages, specifically that L2 violations are associated with an N400 that shows reduced amplitude, delayed onset and delayed peak latency, compared to L1 violations (Ardal et al., 1990; Hahne & Friederici, 2001; Hahne, 2001; 1994; Midgley et al., 2009; Moreno et al., 2008; Moreno & Kutas, 2005; Newman et al., 2012; Ojima et al., 2005; Proverbio et al., 2002; Weber-Fox & Neville, 1996). In this study, we did not observe significant differences in the overall N400 amplitude of match/mismatch pairs between learners and controls, and N400 onset. While we did observe discrepancies between groups in the overall pattern of results, our results indicate that after only a short period of training, L2 learners can achieve native-like brain activity associated with semantic processing.

4.1.2.4 The Time Course of Learning

Finally, we compared the N400 mismatch effect across learning days. The results of this analysis were surprising, as it indicated words taught on Days 1-3 were learned in such a way as to elicit a significant mismatch N400 effect post-training, while words taught on Days 4-6 were not. This may be because later-learned words (on Day 5 and 6) were not reviewed in the same manner as previous words, because of the training schedule. Instead of reviewing words over two subsequent days following introduction, additional practice was given for those words on the final review day. It is possible that reviewing words in smaller doses, over multiple days, promotes better learning, than a large amount of exposure all at once. Pimsleur (1967) found that there is a specific pattern of repetitions that leads to better learning outcomes, which he called *gradual interval recall*. Generally, repetitions that are spaced out and remind the learner of the language material before they've forgotten it promotes better retention. Another, non-exclusive possible explanation comes from theories of memory consolidation, which suggest that new memories consolidate slowly over time (McGaugh, 2000). Long-term memory and long-lasting memory (months – lifetime) rely on the interaction of brain systems to reorganize and stabilize distributed connections. It's possible that later learned words may not have entered into long-term memory, whereas earlier-learned words did, either simply due to the number of days between training and testing, or with the help of practice spaced over more different sessions. Note that the total amount of practice was the same for all words; items taught on Days 5 and 6 simply had that amount of practice concentrated in fewer sessions.

The effect we observed across days raises an interesting question regarding the interaction in this study between time and training strategies – training strategies alternated days, such that Days 4 and 6 were both Implicit days. Recall that although post-training performance on the naming and match/mismatch tasks did not differ between implicitly- and explicitly-trained words, N400 mismatch effects were only significant post-training for explicitly-trained words. Thus since Days 4-6 had a higher number of implicit training sessions, the lack of N400 mismatch effects for words learned on these days may be an artifact of their being learned implicitly. On the other hand, the opposite direction of causality is also possible – that the apparent difference in N400 mismatch effect size between explicitly- and implicitly-trained words is an artifact of the day on which they were learned. In order to properly assess whether the differences observed in N400 amplitude are due to the nature of the learning strategies, or due to problems with memory consolidation, training strategies should be counterbalanced across training days in future studies. This could also be also addressed by using a between-subject design, with separate subject groups receiving different training strategies.

Finally, it is important to note that analyses were performed only an ERP amplitudes within the 300-500 ms time window. As previous literature has demonstrated (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Weber-Fox & Neville, 1996; Kutas & Kluender, 1994; Moreno & Kutas, 2005; Ojima, Nakata, & Kakigi, 2005; Newman et al. 2012), the N400 can show delayed peak onset in bilinguals. Visual inspection of the waveforms indicates that the N400 effect may be delayed for Days 4-6 (especially for Day 5), so further analysis should investigate later time periods, to determine if the effect is indeed present, but simply shows delayed onset.

4.2 Implications and Future Directions

Our results indicate that adult second language learners can rapidly acquire second language vocabulary, after only a short period of training, and newly acquired L2 vocabulary is integrated into existing semantic networks. This study provides further evidence of a bilingual word processing model that suggests a shared lexicon and non-selective language access. Additionally, we have shown that adult second language learning is not a uniformly slow, laborious process; rather, semantic knowledge can be acquired and integrated within a short period of time. Although we did not find evidence to support our hypothesis that words taught in sentence contexts (implicitly) would be learned better than words taught via paired association (explicitly), further research is required to clarify the effects of each type of training, and separate these from the effects of amount of consolidation time and number of review sessions.

In order to address the methodological issues that may have influenced the findings of this study, several steps should be taken. First, it is difficult to determine if the results concerning training strategy are due to problems related to the LANGA training games, or due to a lack of memory consolidation of later learned implicit words. Problems with speech recognition and feedback would need to be fixed, such that participants are permitted to make a complete response before receiving feedback, at a speed that allows them to focus on content over pronunciation. In addition, training strategies would need to be properly counterbalanced across days or a between subject design could be used. Once these issues are addressed, future studies could not only attempt to determine if learning words in context promotes greater retention, but other areas of language acquisition could be evaluated, such as grammar and pronunciation.

For example, although the games incorporate speech recognition and feedback is given on pronunciation, we did not measure changes or improvements in pronunciation directly. Although speech recognition is an integral part of the LANGA protocol, its effectiveness is not currently known. It is also not known if providing feedback on pronunciation has any effect on vocabulary learning. A potentially interesting line of inquiry could investigate improvements in second language pronunciation, in addition to vocabulary learning. The speech recognition software compares participant's utterances to that of a native speaker and responses need to be above a certain threshold in order to register as correct. Simply recording participant's accuracy for each utterance would be a straightforward way to track progress. Additionally, feedback currently only informs participants if their utterance was correct or incorrect, but they are not given suggestions on how to improve. Providing direction on exactly which aspect of the word was mispronounced could be greatly beneficial. Lastly, it would be interesting to determine whether pronunciation training can impact vocabulary learning. Dual-coding theory (Clarke & Paivio, 1991) suggests that the ability to code a stimulus two different ways (both visual and auditory) increases the chances of remembering that item, compared to when a stimulus is represented in only one modality. Enhancing pronunciation may lead to a stronger auditory representation of the word, which could impact vocabulary learning overall. Additionally, participants may be hesitant to speak in their second language due to concerns with their accent. Perhaps promoting proper pronunciation encourages second language use, which could improve long-term outcomes.

Future studies could also investigate whether learning words in context can influence grammatical learning. While the specific purpose of teaching words in the form of three-word sentences was to evaluate vocabulary learning, it also introduces basic grammatical structure (subject-verb-object). Perhaps words learned in context are more readily associated with grammatical structure and are therefore more sensitive to syntax violations. The P600, an ERP component sensitive to grammatical and syntax violations, could be used to evaluate potential differences in cortical responses to syntax violations in sentences containing explicit and implicit words. This would be especially interesting if in actuality, vocabulary learning is best when words are taught via paired association. Perhaps costs to early vocabulary learning are offset by benefits related to grammatical development. If the long-term goal of second language learning is the ability to communicate effectively, it may be better to forgo basic semantic knowledge in favour of other types of language development.

Finally, it should be noted that the participants in this study had varying levels of or prior language experience – some were monolingual while others were fluent bilinguals. Bilingualism has been shown to impact third language acquisition, specifically, bilinguals can obtain higher levels of proficiency in a third language than monolinguals acquiring a second language (Cenoz, 2003). Third language users could be considered "expert" language learners, compared to monolingual second language learners. They may have developed efficient learning/processing strategies that provide an advantage over "novice" second language learners (Cenoz, 2003). In the case of this study, this may have contributed to variability in participant's overall final proficiency. Further investigations into the impact of prior language experience on learning a new language could be conducted, by comparing the final behavioural and EEG outcome measures of bilingual and monolingual participants.

In conclusion, the present study provided evidence of fast second language vocabulary learning in adults. It attests to the neuroplasticity of language networks in the adult brain, given that changes in brain activity associated with semantic processing were observed after only a short period of training. We further reported evidence of rapid integration of L2 semantic knowledge, such that newly learned L2 words can interact with existing semantic networks during word processing. Brain responses of second language learners are similar to those of native speakers when seeing items that they did not know the L2 word for, but that were semantically related to a known L2 word. This again attests to the neuroplasticity of the adult brain. Overall, these results suggest that game-based vocabulary training is an effective means to teach second languages to adults and second language learning is not a uniformly slow and laborious process.

REFERENCES

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied linguistics*, *24*(4), 425–438.

Alario, F., & Ferrand, L. (1999). A set of 400 pictures standardized for French: norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition . *Behavioural Research Methods, Instruments and Computers*, *31*, 531–552.

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, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bagiella, E., Sloan, R. P., & Heitjan, D. F. (2000). Mixed-effects models in psychophysiology. *Psychophysiology*, *37*(01), 13–20.

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*(1), 71-87.

Chen, H.-C., & Ng, M.-L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, *17*(4), 454–462.

Chun, E., Choi, S., & Kim, J. (2012). The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study. *Neuroscience Letters*, *521*(2), 125–129.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational psychology review*, *3*(3), 149-210.

Dijkstra, T., & Heuven, W. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), (p. 384). Psychology Press, 2013.

Ferrand, L., & Alario, F. (1998). Normes d'associations verbales pour 366 noms d'objects concrets. *L'Anée Psychologique*, *98*, 659–70.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226-248.

Frenck−Mestre, C., Sneed-German, E., & Foucart, A. (2014). Qualitative differences in native and non-native semantic processing as revealed by ERPs In R.R. Heredia & J. Altarriba (Eds.) Foundations of Bilingual Memory. Springer: New York, pp. 237-256.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press: Cambridge, U.K.

Guillem, F., N'Kaoua, B., & Rougier, A. (1996). Differential involvement of the human temporal lobe structures in short-and long-term memory processes assessed by intracranial ERPs. *Psychophysiology*, *33*, 720–730.

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

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4*(2), 123–141.

Heuven, W. J. B., Dijkstra, T, & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Lanugage, 39*, 458–483.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual review of psychology*, *34*(33), 33–61.

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.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language. *Studies in second language acquisition*, *21*(02), 181–193.

Kang, S.-H. (1995). The effects of a context-embedded approach to second-language vocabulary learning. *System*, *23*(1), 43–55.

Kerkhofs, R., Dijkstra, T., Chwilla, D.J., & de Bruijn, E. (2006). Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects. *Brain Research*, *1086*, 170–183.

Kersten, A., & Earles, J. (2001). Less Really Is More for Adults Learning a Miniature Artificial Language. *Journal of Memory and Language*, *44*(2), 250-273.

Kutas, M., & Federmeier, K. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205.

Kutas, M., & Kluender, R. (1994). What is who violating? A reconsideration of linguistic violations in light of event-related brain potentials (pp. 183–210). Birkhäuser Boston.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature reviews. Neuroscience*, *9*(12), 920–33.

Law, S.-P., Yeung, O., Wong, W., & Chiu, K. (2005). Processing of semantic radicals in writing Chinese characters: Data from a Chinese dysgraphic patient. *Cognitive Neuropsychology*, *22*(7), 885–903.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique (Cognitive Neuroscience)*. MIT Press.

Martin, C.D., Dering, B., Thomas, E.M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the "active"and the "non-attended"language of early bilinguals. *NeuroImage*, *47*, 326–333.

McCarthy, G., Nobre, A.C., & Bentin, S. (1995). Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. *The Journal of Neuroscience, 15*(2), 1080–1089.

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248–251.

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

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Midgley, K.J,, Holcomb, P.J., & Grainger, J. (2009). Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics, 22*, 281–200.

Midgley, K.J., Holcomb, P.J., Walter, J.B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research*, *1246*, 123–135.

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research, 22*, 205–220.

Moreno, E., Rodríguez-Fornells, A., & Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics*, *21*(6), 477-508.

Morgan-Short, K., Sanz, C., & Steinhauer, K. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60*(1), 153-193.

Morgan-Short, K, Steinhauer, K, & Sanz, C. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience, 24*(4), 933-947.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*(2), 233-253.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *Tesol Quarterly*, *37*(4), 645-670.

Newman, A.J., Tremblay, A., Nichols, E.S., Neville, H.J., & Ullman, M.T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, *24*(5), 1205–1223.

Nobre, A.C., & Mccarthy, G. (1995). Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. *The Journal of Neuroscience, 5*(2), 1090–1098.

Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Cognitive Neuroscience, 17*(8), 1212–1228.

Phillips, N.A., Segalowitz, N., & O'Brien, I. (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.

Pickering, M. (1982). Context-free and context-dependent vocabulary learning: An experiment. *System*, *10*(1), 79–83.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1–28.

Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51*(2), 73-75.

Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, *80*(4), 478–493.

Pritchett, S., Zilberg, E., Xu, Z., Myles, P., Brown, I., & Burton, D. (2010). Peak and averaged bicoherence for different EEG patterns during general anaesthesia. *BioMedical Engineering OnLine*, *9*(76), 1-21.

Proverbio, A.M., Čok, B., & Zani, A. (2002). Electrophysiological measures of language processing in bilinguals, *14*(7), 994–1017.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Seibert, L.C. (1930). An experiment on the relative efficiency of studying French vocabulary in associated pairs versus studying French vocabulary in context. *Journal of Educational Psychology*, *21*(4), 297–314.

Stein, M., Dierks, T., Brandeis, D., Wirth, M., Strik, W., & Koenig, T. (2006). Plasticity in the adult language system: A longitudinal electrophysiological study on second language learning. *NeuroImage*, *33*(2), 774–783.

Thierry, G., & Wu, Y.J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *8*(5), 352–375.

Van Heuven, W., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews*, *64*, 104–122.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and language*, *97*(3), 279–93.

Verhoeven, L., & de Jong, J. (1992). *The construct of language proficiency: applications of psychological models to language assessment*. John Benjamins Publishing Company.

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language, 20*(2), 232–245.

Weber-Fox, C., & Neville, H. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Cognitive Neuroscience*, *8*(3), 231-256.

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2010). Distracting the mind improves performance: An ERP Study. *PLoS ONE*, *5*(11).

Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning and Technology, 10*(3), 85–101.

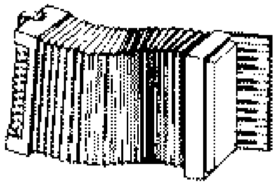
APPENDIX

Table 5. Stimulus lists, including (at the bottom of each column) natural log bigram frequency, number of neighbours, and (log) frequency per million (at bottom). Words in the semantically related and semantically unrelated conditions were presented as pictures. The English translation is provided for clarity. For the semantically related list, * indicates words that were generated based on a set of published first verbal associates.

English	French Trained Word	Semantically Related	Semantically Unrelated
apple	la pomme (9.1, 8, 1.7)	pear*	duck
backpacker	le routard (9.7, 1, 0)	hitchhiker	skateboard
badger	le blaireau (9.7, 0, 0.3)	ferret	cactus
bakeress	la boulangère (10.3, 0, 1.0)	bread	paint
blackboard	le tableau (9.5, 0, 2.0)	chalk*	pizza
boar	le sanglier (10.3, 0, 0.6)	pig	dig
book	le livre (9.3, 7, 2.3)	shelf	laugh
building	le bâtiment (10.3, 0, 1.6)	scaffolding	light switch
builder	le maçon (9.2, 2, 0.8)	hardhat	hammock
buys	achete (10.0, 1, 1.9)	money	whale
carries	porte (9.9, 3, 2.9)	baby	wheel
cart	le chariot (9.8, 1, 1.0)	donkey	square
castle	le château (9.3, 0, 1.7)	drawbridge	lighthouse
catches	attrape (10.2, 0, 1.4)	net	jar
chair	la chaise (9.7, 2, 1.8)	table*	mixer
cleans	nettoie (9.8, 0, 1.2)	mop	bee
climbs	grimpe (9.9, 2, 1.4)	mountain*	elephant
combs	peigne (9.7, 1, 1.2)	hair*	tree
deer	la biche (8.9, 10, 0.9)	moose	bench
dozes	dort (9.0, 0, 2.1)	pillow	puzzle
draws	dessine (10.2, 1, 1.7)	sketchbook	toothbrush
dwarf	le nain (9.2, 9, 1.0)	troll	acorn
eat	mange (10.0, 6, 2.2)	food*	boat
fairy	la fée (6.7, 1, 0.9)	wand	lamp
falls	tombe (9.6, 1, 2.4)	trip	bird
farmer	le fermier (10.2, 0, 1.0)	barn	clap
fence	la barrière (9.8, 1, 1.2)	gate	bell
fireman	le pompier (9.9, 1, 1.0)	hose	bath
fish	le poisson (9.91, 3, 1.7)	snorkeling	rhinoceros
frog	la grenouille (10.3, 0, 0.9)	lily pad	ladybug
ghost	le fantôme (10.0, 0, 1.4)	halloween	accordion
goat	la chèvre (9.2, 0, 1.1)	hoof	snail
groundhog	la marmotte (9.7, 2, 0)	burrow	bottle
hides	cache (9.8, 6, 2.1)	search	flower
hit	frappe (9.8, 0, 2.1)	Punching bag	grasshopper
horse	le cheval (9.1, 1, 2.1)	saddle	hammer

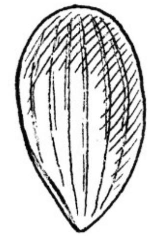
house	la maison (9.8, 2, 2.5)	family*	mirror
hunter	le chasseur (9.9, 1, 1.4)	gun	row
jockey	le cavalier (10.1, 0, 1.4)	boot	bone
jumps on	saute sur (9.9, 6, 1.9)	bed	cut
king	le roi (8.4, 10, 2.0)	queen*	smoke
knocks over	renverse (10.5, 0, 1.5)	wrecking ball	ironing board
lemon	le citron (9.8, 3, 0.9)	orange	ladder
looks at	regarde (10.0, 2, 2.7)	glasses	tractor
loves	aime (9.8, 6, 2.7)	wedding	whistle
lumberjack	le bûcheron (10.1, 0, 0.6)	axe	ski
magazine	la revue (9.2, 2, 1.5)	newspaper	lightning
mat	le tapis (9.4, 5, 1.6)	door	crab
ox	le bœuf (8.4, 0, 1.3)	yoke	knot
peanut	la cacahuète (9.4, 0, 0.2)	almond	guitar
priest	le prêtre (9.3, 0, 1.5)	church	vacuum
pulls	tire (10.1, 7, 2.4)	tug-of-war	typewriter
pumpkin	la citrouille (10.2, 0, 0.1)	gourd	ladle
rabbit	le lapin (9.3, 8, 1.2)	carrot*	turtle
reads	lit (9.4, 18, 2.6)	ebook	igloo
rides	monte (10.4, 4, 2.4)	bicycle	dentist
rock	la pierre (10.0, 1, 2.4)	boulder	leopard
scares	effraie (9.7, 0, 1.4)	mask	harp
seagull	la mouette (9.7, 4, 0.9)	ocean	arrow
sheep	le mouton (9.7, 2, 1.3)	wool*	vase
skunk	le putois (9.3, 1, 0)	smelly	banana
sleigh	le traineau (9.5, 0, 0.4)	santa	berry
store	le magasin (9.6, 0, 1.6)	cashier	rainbow
street	la rue (8.3, 9, 2.5)	crosswalk	corkscrew
takes	prend (10.0, 0, 3.0)	give	leaf
teacher	la maîtresse (9.9, 0, 1.6)	desk	iron
thief	le voleur (9.6, 2, 1.2)	handcuffs	asparagus
throws	jette (10.1, 5, 2.0)	football	envelope
truck	le camion (9.7, 1, 1.5)	car*	sun
walks	marche (9.9, 1, 2.6)	cane	bear
woman	la femme (9.1, 3, 2.7)	man*	hat
writes	écrit (9.6, 0, 2.5)	letter	finger
	(9.0, 5.3, 1.5)	(9.1, 4.5, 1.1)	(9.2, 4.5, 1.1)

Accordion

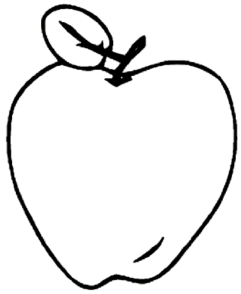


Acorn

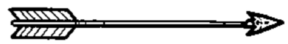
Almond



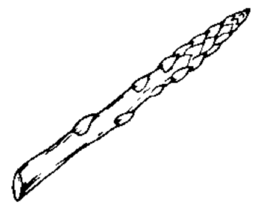
Apple



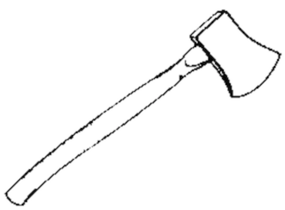
Arrow



Asparagus



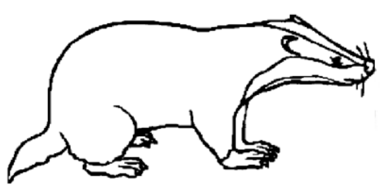
Axe



Baby

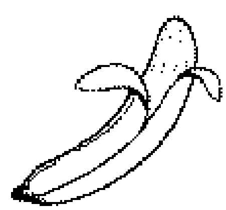
Backpacker

Badger

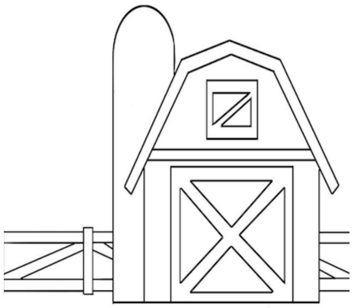


Bakeress

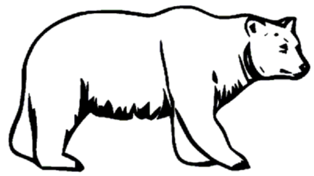
Banana



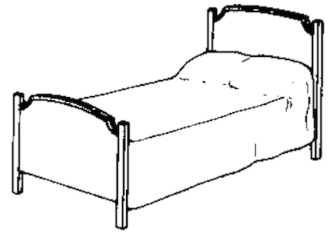
Barn



Bear

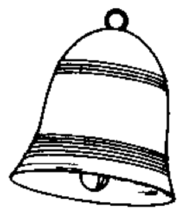


Bed

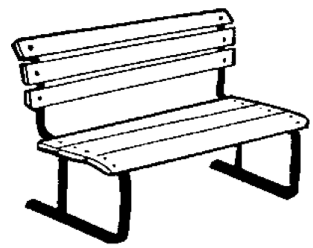


Bee

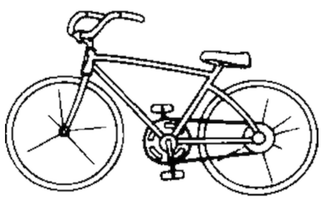
Bell



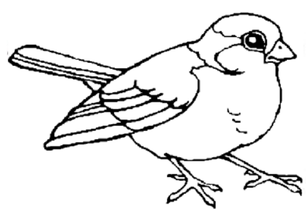
Bench



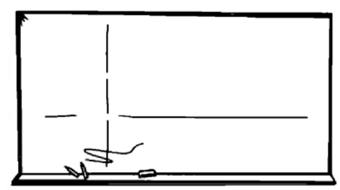
Bicycle



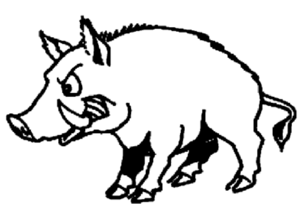
Bird



Blackboard

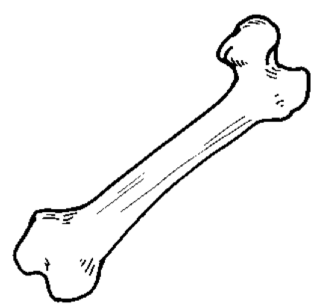


Boar

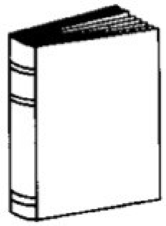


Boat

Bone

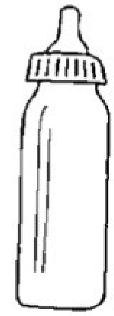


Book

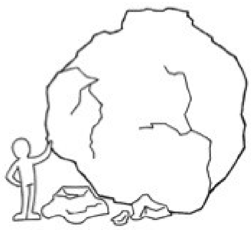


Boots

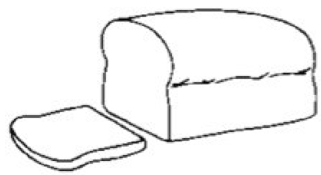
Bottle



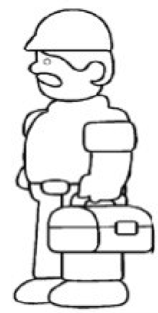
Boulder



Bread

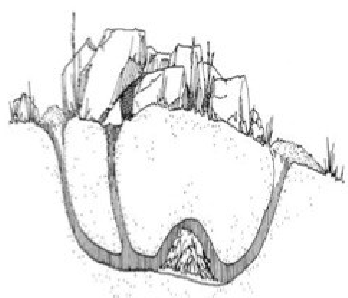


Builder



Building

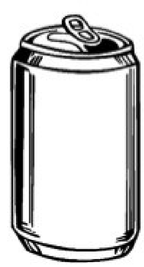
Burrow



Buys

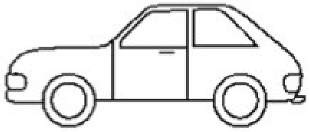
Cactus

Can

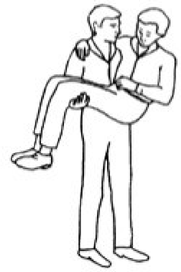


Cane

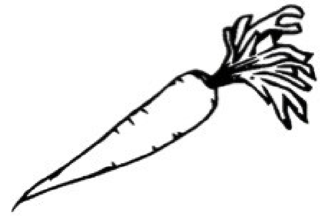
Car



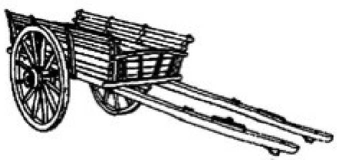
Carry



Carrot

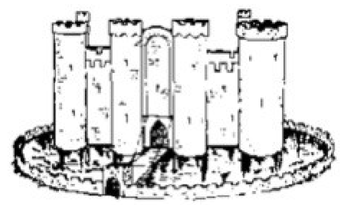


Cart



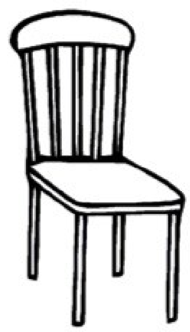
Cashier

Castle



Catch

Chair



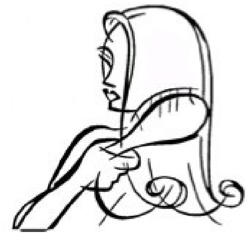
Chalk

Church

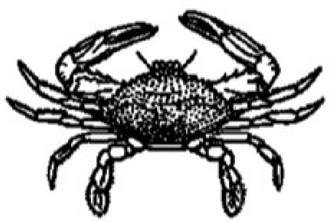
Claps

Cleans

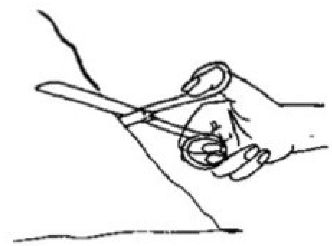
Climbs

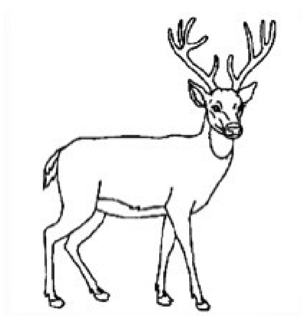

Combs


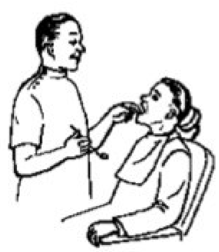
Corkscrew


Crab


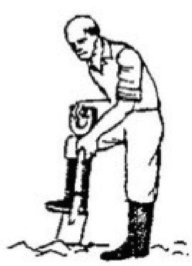
Crosswalk

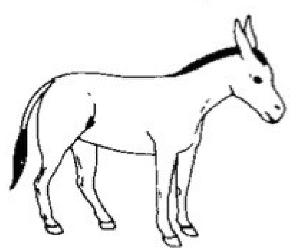

Cut


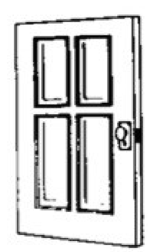
Deer


Dentist


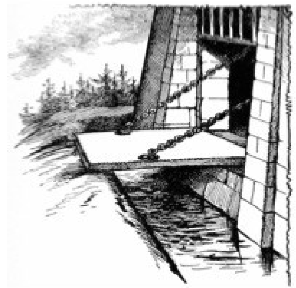
Desk


Dig


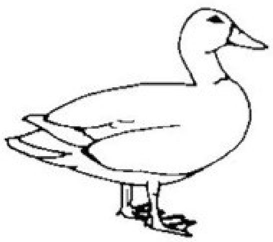
Donkey


Door


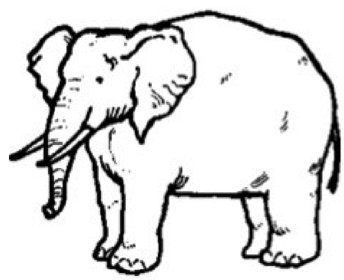
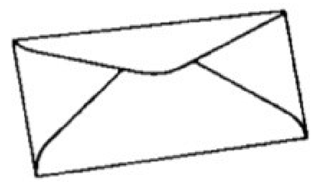
Dozes Drawbridge Draws


Duck Dwarf Eats

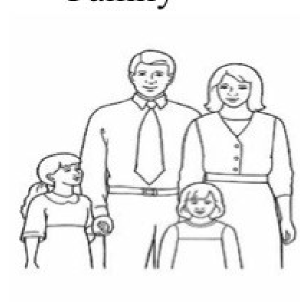


Ebook Elephant Envelope

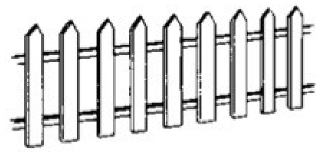
Fairy Falls Family



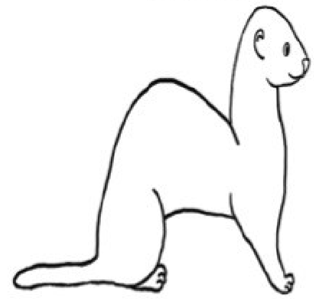


Farmer

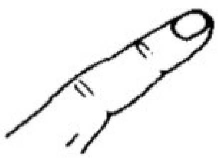
Fence



Ferret

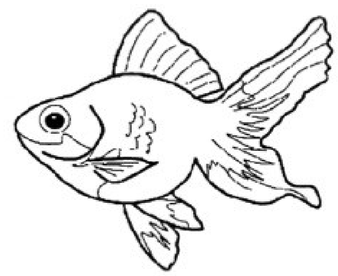


Finger

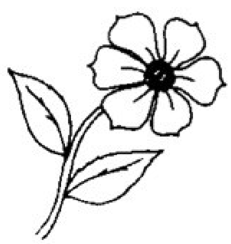


Fireman

Fish

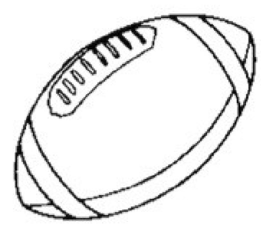


Flower

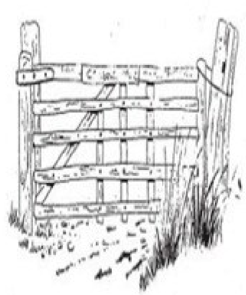


Food

Football



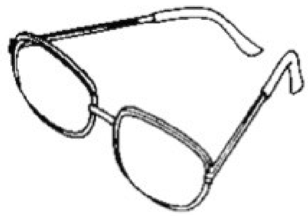
Gate



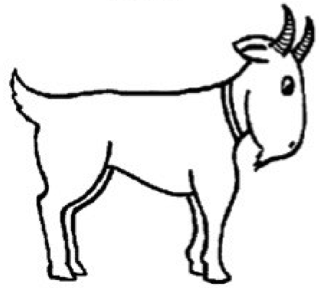
Ghost

Give

Glasses

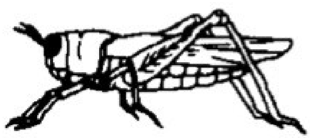


Goat



Gourd

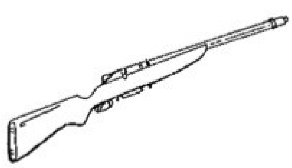
Grasshopper



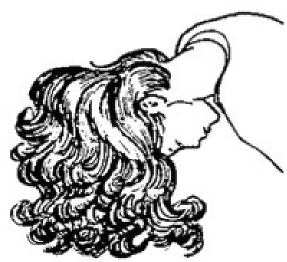
Groundhog

Guitar

Gun



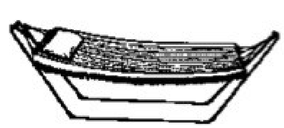
Hair



Halloween

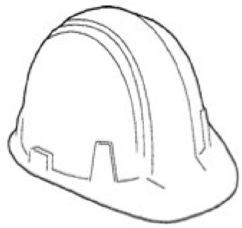
Hammer

Hammock

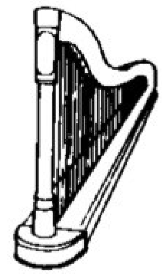


Handcuffs

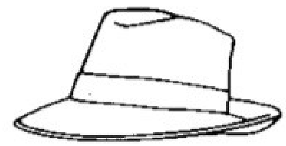
Hardhat



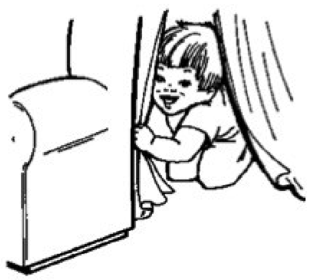
Harp



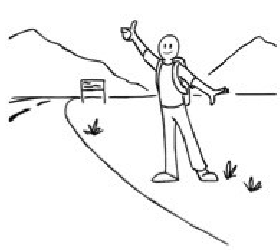
Hat



Hides

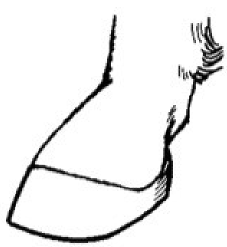


Hitchhiker

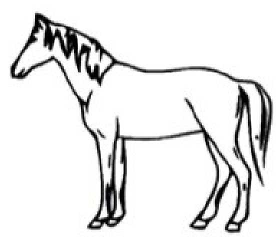


Hit

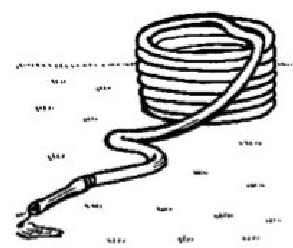
Hoof



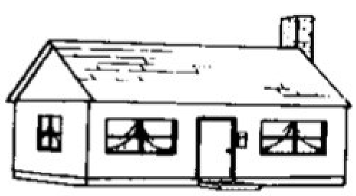
Horse



Hose

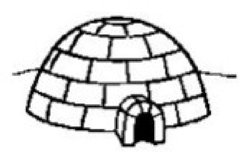


House

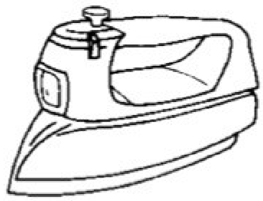


Hunter

Igloo



Iron

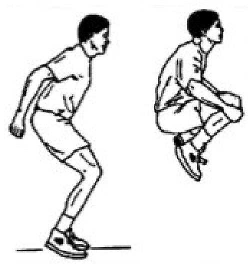


Ironing Board

Jar

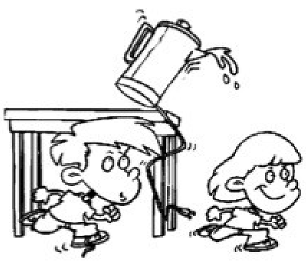
Jockey

Jump

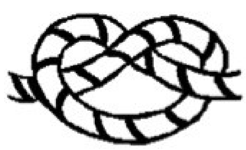


King

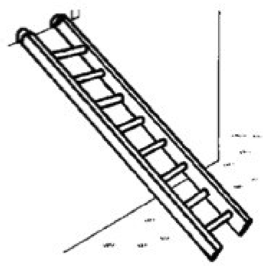
Knock Over



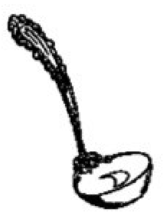
Knot



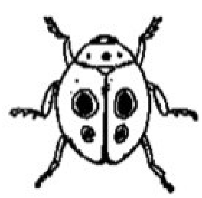
Ladder



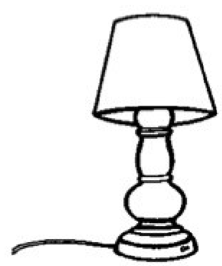
Ladle



Ladybug

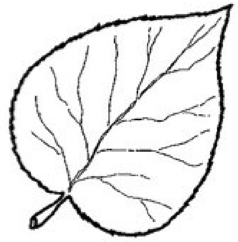


Lamp

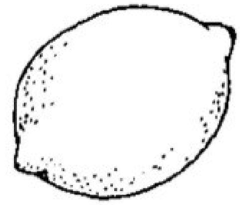


Laugh

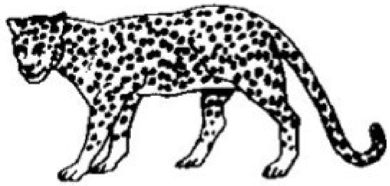
Leaf



Lemon



Leopard

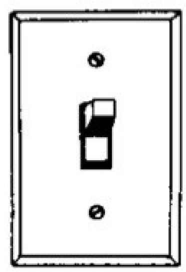


Letter

Lighthouse

Lightning

Light Switch



Lily pad

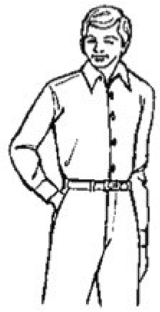
Looks

Loves

Lumberjack

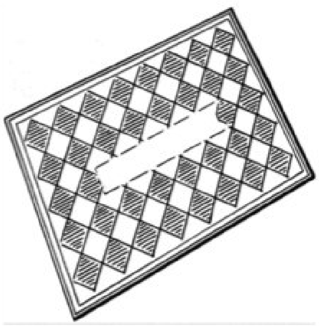
Magazine

Man

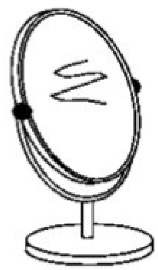


Mask

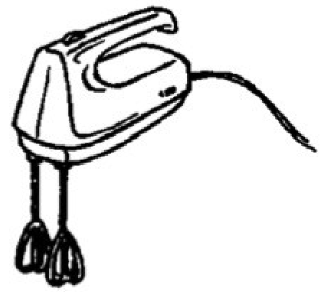
Mat



Mirror

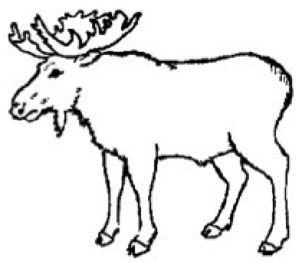


Mixer

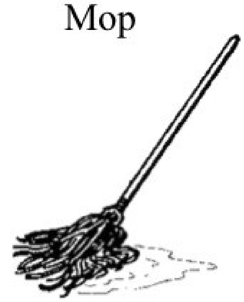


Money

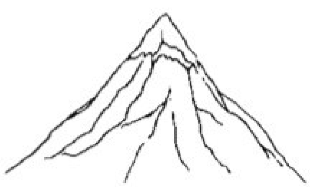
Moose



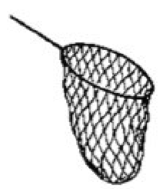
Mop



Mountain



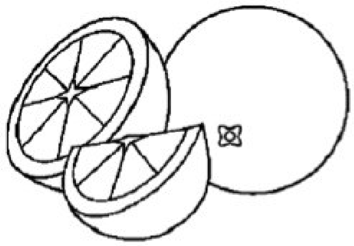
Net



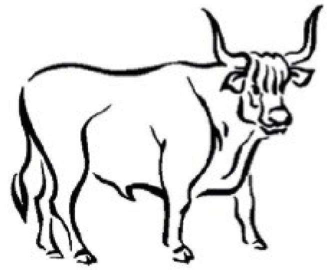
Newspaper

Ocean

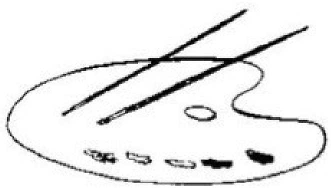
Orange



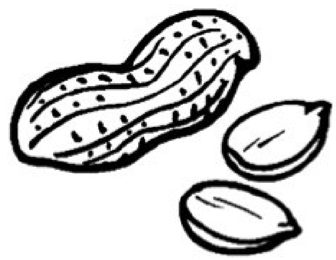
Ox



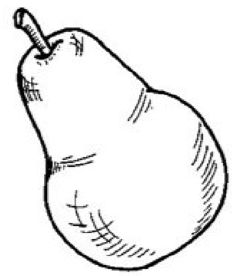
Paint



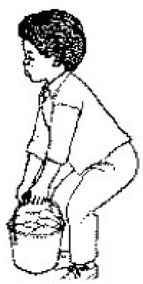
Peanut



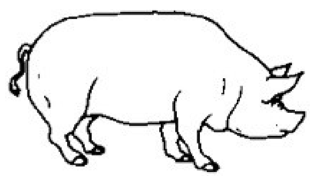
Pear



Picks Up



Pig



Pillow

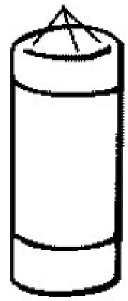
Pizza

Priest

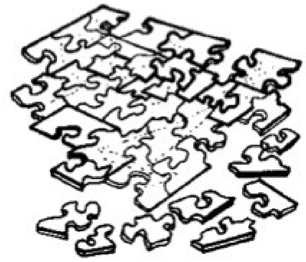
Pulls

Pumpkin

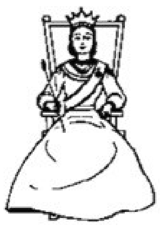
Punching Bag



Puzzle

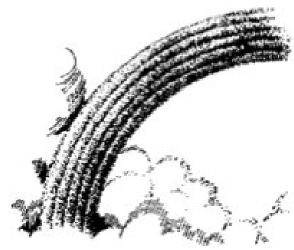


Queen



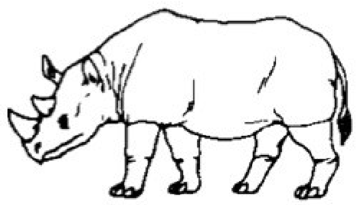
Rabbit

Rainbow



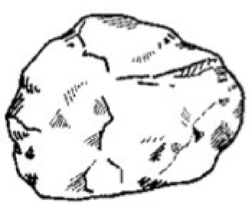
Reads

Rhino

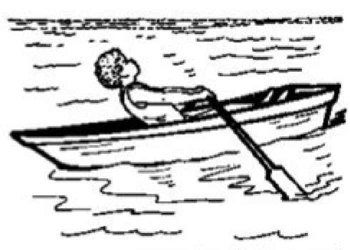


Ride

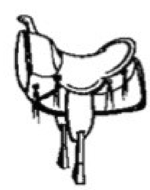
Rock



Row

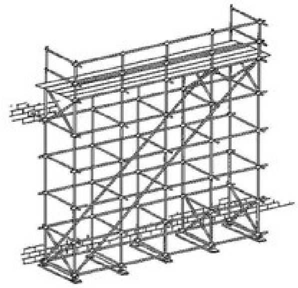


Saddle



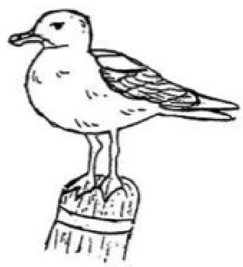
Santa

Scaffold



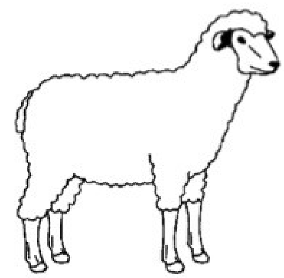
Scares

Seagull



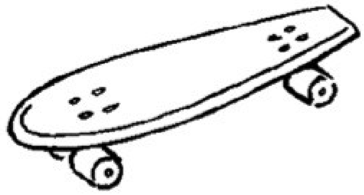
Search

Sheep

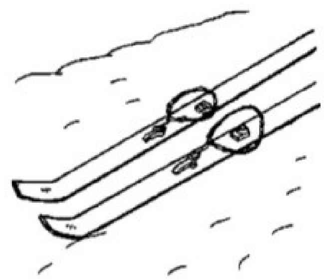


Shelf

Skateboard

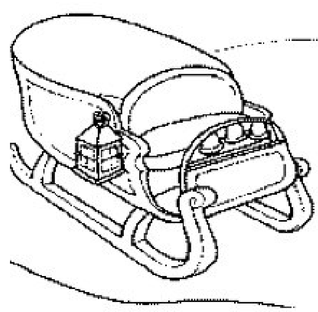


Skis



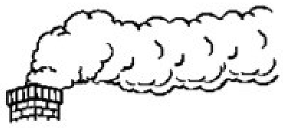
Skunk

Sleigh

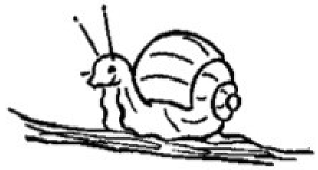


Smell

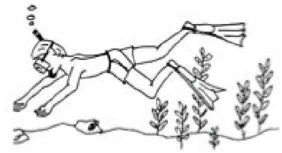
Smoke



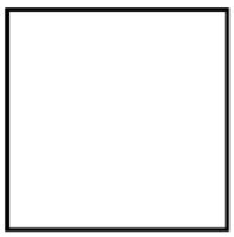
Snail



Snorkeling

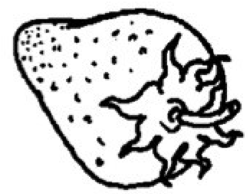


Square



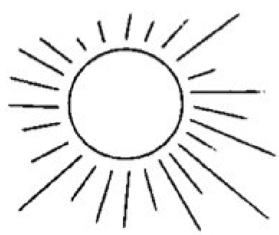
Store

Strawberry

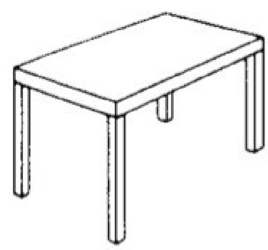


Street

Sun



Table



Takes

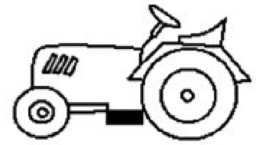
Teacher

Thief

Throws

Toothbrush

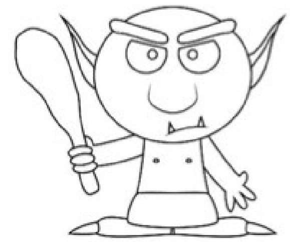
Tractor



Tree

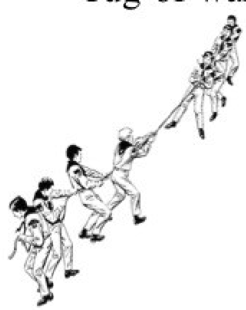
Trip

Troll

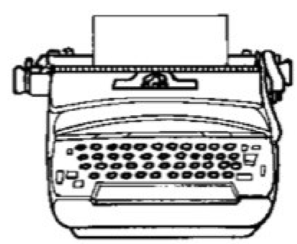


Truck

Tug-of-war

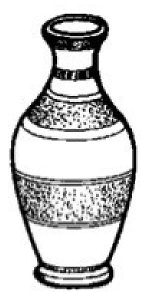


Typewriter



Vacuum

Vase



Walks

Figure 1. Visual stimuli used in pre- and post-training behavioural and EEG tasks.