

# The Effects of Joint Attention Contexts on Adult Novel Word Learning

by

Krista Elliott

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs  
in partial fulfillment of the requirements for the degree of

Master of Arts

in

Applied Linguistics and Discourse Studies

Carleton University  
Ottawa, Ontario

© 2015, Krista Elliott

## **Abstract**

The effects of joint attention contexts on adult novel word learning

Krista L. Elliott, M.A. Applied Linguistics and Discourse Studies

Carleton University, 2015

Joint attention is influential in infant language acquisition, however less is known of the effects of interactive conditions in adult processing and learning. With adults, digital avatar-to-human interactive tasks have resulted in increased image recognition (Kim & Mundy, 2011) and human-to-human experiments have demonstrated joint attention to significantly impact L2 word learning (Hirotani et al., under revision). The aim of the current study was to validate Hirotani et al.'s (under revision) conclusions using a digital interactive paradigm as opposed to as live, face-to-face design. Nine subjects interacted with a video participant in 3 separate learning blocks consisting of 40 picture-pseudoword pairs in four joint attention contexts (responding, initiating, simultaneous and non). Results indicated no performance difference across blocks or conditions and no interaction effect. Further testing is required to determine whether interactive digital environments can also play an implicit role in the effectiveness of joint attention in adult lexical development.

## **Acknowledgments**

I would like to thank my supervisor, Dr. John Logan, whose understanding, encouragement and willingness to dedicate extra time and work to this progress of this experiment made it achievable.

I would like to extend my gratitude to the following people for contributing to the success of this project: Dr. Natasha Artemeva for her invigorating support and tenacious dedication to the ALDS cohort; my committee members for a fair evaluation of my progress; the team of Language and Brain Lab Research Assistants and lab members, without whom no data could have been collected; and Dr. Masako Hirotsu for granting me use of the equipment in Carleton's Language and Brain Laboratory and for sharing expertise in the design of this project. A special thank you to Kate Carroll for all of her help and support, particularly with project and stimuli design.

Thank you mom, dad, grandmas, brothers, Daniella, and Chris; without the mental and emotional support of my family, friends, and significant other this work would never have materialised.

## Table of Contents

Abstract .....	I
Acknowledgments .....	II
List of Figures .....	IV
List of Tables .....	V
<b>Joint Attention</b> .....	3
<b>Infant and Child Joint Attention Studies</b> .....	6
<b>Joint Attention and Adults</b> .....	8
<b>Significance of Gaze to Learning and Joint Attention</b> .....	10
<b>Purpose of the Present Study</b> .....	12
Method .....	15
<b>Participants</b> .....	15
<b>Apparatuses</b> .....	16
<b>Stimuli</b> .....	16
<b>Procedure</b> .....	22
Results .....	27
Discussion .....	34
<b>Limitations</b> .....	36
<b>Future Directions</b> .....	39
Conclusion .....	43
References .....	45
Appendix A SONA Announcement .....	55
Appendix B-1 Pseudowords .....	56
Appendix B-2 List of words removed from original 80 .....	57
Appendix B-3 Wordlist deletion justifications .....	58
Appendix C-1 Unfamiliar target picture objects .....	59
Appendix C-2 Unfamiliar non-target picture objects .....	60
Appendix C-3 Unfamiliar target picture object ratings Hirotsu et al. (under revision) .....	61
Appendix C-4 Unfamiliar target picture object ratings Hirotsu et al. (under revision) .....	63
Appendix D-1 Sample visual stimuli sequence of word learning sessions .....	65
Appendix D-2 Example audio-visual stimuli used to test learned words in test sessions .....	66
Appendix E Informed Consent Form .....	67
Appendix F Language background questionnaire (based on Wartenburger et al., 2003) .....	72
Appendix G Debriefing Form .....	75
Glossary .....	78

## List of Figures

<i>Figure 1.</i> Simultaneous observation example. This figure illustrates an instance of .....	4
<i>Figure 2.</i> Example illustration of classic joint attention interaction paradigm (similar to the .....	5
<i>Figure 3.</i> Test trial time sequence. This figure displays the time sequence of test trials along .....	23
<i>Figure 4.</i> Mean correct responses per block. The error bars in this figure correspond to 95%.....	28
<i>Figure 5.</i> Mean response times per block. The error bars in this figure correspond to 95%.....	29
<i>Figure 6.</i> Mean number of correct responses per joint attention condition. The error bars in this .....	30
<i>Figure 7.</i> Mean response times per joint attention condition. The error bars in this figure .....	31
<i>Figure 8.</i> Mean correct responses as a function of JA condition and block. The error bars in this .....	32
<i>Figure 9.</i> Mean response times as a function of JA condition and block. The error bars in this .....	33

## **List of Tables**

<i>Table 1.</i> Independent measures ANOVA. This table displays the ANOVA pitch results for the .....	18
<i>Table 2.</i> Independent measure ANOVA familiarity. This table displays the familiarity rating .....	19
<i>Table 3.</i> Independent measure ANOVA familiarity. This table displays the familiarity ratings .....	19
<i>Table 4.</i> Visual angle table. This table summarizes the pertinent display computer information .....	20

## The Effects of Joint Attention Contexts on Adult Novel Word Learning

An increasing number of studies investigating second language acquisition across all ages have begun to cast doubt on the validity of appealing to Lenneberg's Critical Period Hypothesis (1967)—and its softened version, Oyama's Sensitive Period (1976; 1978)—to claim that innate language acquisition capabilities greatly diminish or disappear after puberty (c.f. Long, 1990; Singleton, 2005; Rothman, 2008). Not only can older children acquire language more efficiently and more quickly than younger children, but in certain instances adults also reach particular stages of acquisition earlier than children (Snow & Hoefnagel-Höhle, 1978; Krashen, Long & Scarcella, 1979). Nonetheless, what appears to have stood the test of time is that in normal child populations, setting sociological factors aside, first and second language development follows fairly predictable paths to acquisition (Brown 1973; Ervin-Tripp, 1974), whereas adult language endeavours vary widely in degrees of success (Rothman, 2008).

Asserting the Critical Period Hypothesis begs the question: what is meant by the use of the terms *learning* and *language acquisition*? Stephen Krashen (1981) distinguished the two by positing *acquired* knowledge as that which occurs organically and unconsciously through exposure to -and interaction with- a language; and *learned* knowledge as that which occurs through direct instruction (VanPatten & Williams, 2012). Acquisition, then, involves implicit emphasis on meaning and what is being communicated; whereas during instances of learning, explicit focus is placed on recognizing rules and patterns in an effort to understand how language should be used (VanPatten & Williams, 2012). There are many other factors contributing to the differences between how these two terms are used, and one may be hard-pressed to propose definitions that all experts in the various domains of linguistics may agree upon. Although the current experiment blended the conditions dividing the terms as defined by Krashen, for the purpose of clarity, "learning" is used in this thesis to reference to the type of language exposure that the subjects

experienced over the course of their participation in this experiment. More specifically, “learning” encompasses the ability to recognize and clearly distinguish between pseudoword-picture pairs that were displayed in various social interactive conditions.

The role of social cognition in language acquisition incorporates the ability of young children to parse and map words correctly onto environmental targets through directed interaction, and is thought to be crucial to the further development of the child’s linguistic and social capacities (Bruner, 1981; Baldwin & Markman, 1989; Tomasello, 1995). The early ability to follow the directed gestures of a speaker, such as gaze or pointing, has been strongly associated with earlier production of language (Tomasello & Farrar, 1986). However, as crucial this ability is in infancy and childhood, much less is known about how adults continue to use such pragmatic cues to acquire language, whether in their first or subsequent tongues. Schumann’s (1976) taxonomy of factors influencing second-language acquisition highlights a plethora of interwoven, intervening elements that attempt to account for the intricacies of adult learning processes. Although the model strongly emphasizes how individual traits complicate the adult acquisition environment, it also draws particular attention to the social, psychological, cognitive and instructional variables influencing the motivation and ability of an adult to learn a new language. More contemporary research has placed a greater emphasis on analyzing the role of these communal factors, with particular interest being allocated to social cognition and interaction in language acquisition during adulthood. As new words are constantly emerging, and as the growing global community places greater demands on the instrumentality of multilingualism, the implications of uncovering how specific social interactive conditions can affect lexical development in adults are important to furthering theoretical knowledge and applied language teaching practices.

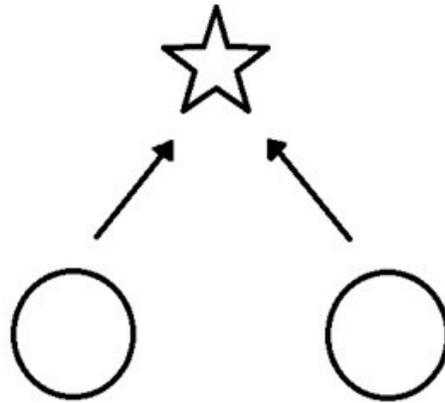
This document describes my thesis project which was designed to explore the relationship between the ability of adults to jointly engage in learning activities and the results of these conditions on lexical recognition. The goal was to explore if, and how, various joint attention conditions could help adults comprehend unfamiliar vocabulary, resulting in the formation of word-object associations. The following paragraphs describe requisite conditions for instances of joint attention, along with previous studies exploring the importance of interpreting these pragmatic cues and the prevalence and significance of this phenomena. The above is all discussed with regards to the applications of the findings to educational environments and practices. The purpose, procedure and results of this experiment are all outlined and addressed with regards to the implications elaborated on in the closing discussion section.

### **Joint Attention**

Social Cognition Theory (Bandura, 1986) applied to language learning postulates, broadly, that language can be acquired through social reinforcement both indirectly, through observation, and directly, through interaction. This capacity for interactive coordination has been previously referred to as Joint Attention (henceforth, JA).

The classical interpretation of JA has grown from Jerome Bruner's use of the term *joint attention* (1975), and Bakeman and Adamson's (1984) work, in which they introduce two types of engagement: *passive* and *coordinated*. Passive engagement involves a single observer directing

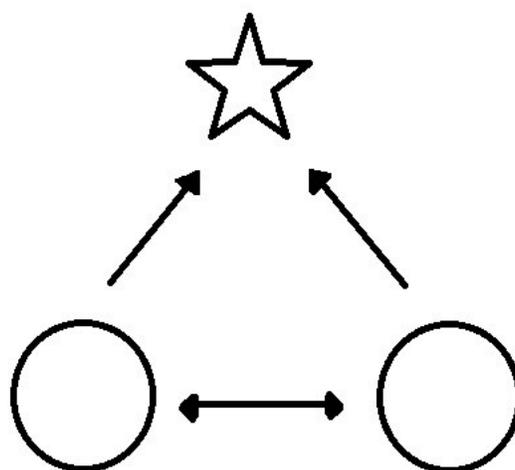
his/her attention to an object, whereas coordinated engagement is described as the three-way interaction between an object and the simultaneous attention of two observers (see Figure 1).



*Figure 1.* Simultaneous observation example. This figure illustrates an instance of simultaneous observation of an object by two observers.

In later work, the type of shared experience necessary to establish JA came into question. If one individual was to examine an airplane flying overhead at one end of Central Park, and another person was observing the airplane from the other end of Central Park, was this to be considered an episode of joint attention? What about people in closer proximity, for example a tour group? If a person at the front of a group views a painting simultaneously as a person in tail of the group, does their coinciding attention exemplify joint attention? This description was later found to warrant elaboration on the inclusion criteria defining a “meaningful” joint attention experience between the observers. Tomasello (2008) expanded this definition of JA by establishing two possible formations of this element termed “common ground”. He named these processes “bottom-up” and “top-down”, framing the former as an instance when an object or event in the environment attracts the attention of both observers, and the latter as when one partner initiates a signal to a referent, resulting in a mutual understanding of a shared goal, and the coordination of perspective between observers. Although a consistent description of JA remains contested amongst researchers,

Carpenter and Liebal (2011) distinguish “true” JA as that which necessarily involves synchronicity of both partners’ knowledge, directing their focus at the same referent simultaneously. This involves bidirectional concentration between interlocutors serving to establish a triadic interaction of attention between them and an object of focus (Carpenter & Liebal, 2011). In short, Carpenter and Liebal insist that simply attending to an object of focus concurrently is not sufficient to establish a JA event, as the experience can remain a very private and individual one. It is the knowledge of the other’s attendance, and the knowledge of sharing the focus on an object, person, event or topic at the same point in time that creates the consciously shared experience which is the requisite for JA (see Figure 2).



*Figure 2.* Example illustration of classic joint attention interaction paradigm (similar to the diagram in Carpenter & Liebal, 2011)

In the interest of precisely examining the distinct mechanisms underlying social-cognition and the JA phenomenon, recent studies (Mundy et al., 2007; Mundy & Jarrold, 2010; Kim & Mundy, 2012; Hirontani, Shimada, Okazaki, Tanabe, & Sadato, under revision) have further elaborated Tomasello’s (2008) and Carpenter and Liebal’s (2011) definitions and divided JA into separate context dependent classifications that are used more precisely in adult language

experiments. Similar to Tomasello's proposed "bottom-up" JA process, Simultaneous Joint Attention (henceforth, SJA) in recent literature has been used in reference to when two –or more- interlocutors attend to the same object in synchronicity. The attention in a SJA instance is elicited through an intrinsic attractiveness of the referent object, and not through an explicit communicative gesture on the part of either –or any- of the interlocutors. Initiated Joint Attention (henceforth, IJA) is used to specifically label when one interlocutor draws the other's –or another's- attention to an object of interest. The intention is signalled through the use of a communicative gesture (whether a head-nod, a point, a saccade or "gaze shift", or some other observable change in body orientation) to draw the attention of the other to the desired object of focus. Responding to Joint Attention (henceforth, RJA) is used to code when one interlocutor observes and reacts to another's attention to an object, whether intentionally directed through gaze or gesture, or not. Hence where one interlocutor engages the other(s) through IJA, the other(s) demonstrate RJA in acknowledging the transmission and attending to the object of focus. For the purposes of this study, the term Non-Joint Attention (henceforth, NJA) will be employed to signify what some researchers refer to as "solo-attention"; these are contexts void of a communicated pragmatic cue or with an absence of shared attention to an object of interest.

### **Infant and Child Joint Attention Studies**

Developments in children's representational theory-of-mind (i.e., the ability to understand that other people have thoughts that are different from one's own) presuppose JA as integral to the expansion of cognition (Tomasello, 1995; Verga & Kotz, 2013). Infant studies have stressed a strong underlying motivation for the sharing of attention and interest in objects with others to create simultaneous mutual knowledge (Carpenter & Liebal, 2011). This mutual interest occurs through verbal and non-verbal cues, such as gaze shifts and fixations, head motions, and pointing,

directed most often by an adult to the infant or child until the infant becomes capable of reciprocating and initiating (Carpenter, Nagell, & Tomasello, 1998). Research has shown that by the age of 9 months infants begin to become capable of identifying these communicative intentions. The ability to fully partake in JA, as defined above, emerges somewhere around approximately ten months to one year of age (Carpenter & Liebal, 2011).

It has been proposed that the capacity to properly interpret and disambiguate these deictic episodes of engagement is crucial in identifying labels or objects, and stands as a precursor to vocabulary development and social awareness (Baldwin, 1991; Tomasello & Ferrar, 1986; Bruinsma, Koegel & Koegel, 2004; Brooks & Meltzoff, 2005). Many studies with younger children have shown positive learning outcomes in social contexts involving JA tasks in comparison to NJA tasks (Baldwin, 1993; Baldwin, 1995). For examples, Brooks and Meltzoff (2005) found a strong correlation between the ability of 10-11 month-old infants who identified the significance of following the head-turns of adults with their eyes open, compared to an eyes closed condition, and who made vocalizations, and heightened language scores at 18 months. Baldwin, Markman, Bill, Desjardins, Irwin and Tidball (1996) showed that 18-20 month-old infants mapped words to novel objects when the speaker was present and directing attention to the object, but failed to learn the word-object pairs when the speaker was not physically or visually attending to the object in question even when the label was clearly noticed. These results have been supported by more recent findings, such as that of Hirotani, Stets, Striano, and Friederici (2009), which showed that 18-21 month-old infants learned new words significantly more efficiently in JA tasks, where shared attention was directed and paired with positive voice tones, in comparison to words learned in NJA tasks, where focus was not shared and was paired with neutral tones. These results are corroborated by the correlation between the severely impaired

ability of children with autism to identify speaker intentions, along with the late realization of JA patterns, and a later lack of linguistic communicative proficiency (Bruinsma et al., 2004).

### **Joint Attention and Adults**

Social interactions, particularly those involving positive reactions and gestures, have been shown to have a strong effect on child language learning. However, much remains unknown about how these variables influence adult learning. Studies in conversational analysis indicate that the same supra-segmental gestures used to direct infant's attention are also used in adult-to-adult interactions to clarify inferred beliefs, emotions, and intentions. When interacting with people of all ages, adults have a tendency to align their actions, reactions, and linguistic devices (i.e., lexicon, syntax, and prosody) to accommodate context dependent and linguistic knowledge. This is done in order to facilitate comprehension in the complex shared, dialogical process (Pickering & Garrod, 2004; 2006). Without such adjustments, conversation quickly disintegrates into isolated, or "misaligned" communicative actions which stray from achieving 'joint knowledge' and forsake the implicit shared mental representations (i.e., 'common ground') of the interaction (Pickering & Garrod, 2004; 2006).

Although pertinent to all communicative exchanges, these embodied interactions have particular implications in contact between native and non-native language speakers as well (Schegloff, Koshik, Jacoby, & Olsher, 2002). In an eye-tracking experiment where participants interacted with a digital avatar partner, Kim and Mundy (2012) found that pictures learned in IJA conditions were recognized significantly more than those learned in RJA conditions. The authors attempted to explain the results by stating that in comparison with RJA, IJA relies more heavily on self-referenced processing, personal choice, and the plethora of associated neural networks (e.g., proprioception and spatial interpretation, identity representations, reward and motivation

systems). Extending this work, Hirotsu et al. (under revision) remodelled the eye-tracking paradigm to have pairs of monolingual Japanese participants interact in live directed JA conditions, together learning a series of English words. Their results demonstrated that although adult second language word learning was triggered by all four JA conditions, overall SJA conditions clearly produced the most accurate responses after four and five test sessions. Although these findings are promising, more research focused on these multilingual interactions is required in order to determine the extent of the role that other carefully controlled, deliberate, and automatic JA actions can play in facilitating understanding as well as the learning of new languages.

In order to shed light on the ontogenetic factors of human interaction, many contemporary investigations of JA in adults have explored the neurological basis of this social-cognitive phenomenon. Saito et al. (2010) combined MRI and eye-tracking technology to test conditions similar to RJA, IJA, and NJA (as described above) by instructing participants to share eye contact with their partner until one of the two images of a ball changed colour from red to blue, upon which time they were to shift their gaze to the red ball. In discussing the synchronized activation of the right inferior frontal gyrus –a brain region associated with unconscious imitations of facial gestures, self-evaluation, and self-other distinction– in the RJA and IJA conditions, Saito et al. argued that partnered JA interactions may lead to the internalization of the intentional, consciously shared experience. They suggested that neural activation in these interactive settings serves to improve interpretation of communicative gestures and enhance collaborative competence through the creation of a shared psychological state or mental representation (Saito et al., 2010) Similarly, a comparison of recorded interactions to live JA interactions in the eye-tracking and fMRI experiment of Redcay et al. (2010) showed greater stimulation in several brain regions previously associated with social cognition (i.e., the right temporoparietal junction and the right superior

temporal sulcus), suggesting that these areas are crucial to social interactive communication and joint attention in particular. Interestingly, Redcay et al. also elaborated on the activation of a nexus of neural reward centres (such as the ventral striatum, the anterior cingulate cortex, and the amygdala) during live JA interactions, providing evidence that an underlying motivation and benefit to these shared experiences is a sense of reward and enjoyment. In contrast, Redcay et al. (2013) found that fMRI data from adults with autism spectrum disorder do not reveal distinct activation in these regions during JA events, but instead exhibit a greater disparity in stimulation during RJA conditions in comparison to typical subjects, in addition to an atypical increase in activation to these areas during solo-attention (NJA) conditions. This evidence potentially accounts for the continued difficulty this population encounters in perceiving and interpreting conventional social interactive cues in comparison with typical populations.

Taken together, the findings from JA experiments suggest that parallel attention and instances of shared knowledge likely heighten motivation as well as one's sensitivity to referential learning, thereby creating the right neural environment to enhance adults' word learning. However further research is necessary to validate these conclusions.

### **Significance of Gaze to Learning and Joint Attention**

In Western society the eyes are commonly said to be the “window to the soul”. Between the seeing-abled, eye contact is often used as a non-verbal communicative gesture serving each of the following five functions of Patterson's Sequential Model (1982): providing information, regulating interaction, expressing intimacy, exercising social control, and facilitating service or task goals. Gaze tracking abilities emerges early in infancy (Scaife & Bruner, 1975). In experiments designed to explore early social behaviour, Bloom (1975) demonstrated how simply making eye contact with infants encouraged an increase in bouts of vocalizations. Between adults,

shared gaze has been shown to provide comfort in enhancing ease of communication during social interactions (Kleinke, 1986). Egbert (1996) observed that German speakers often use the very specific verbal cue “bitte” (a German equivalent to the English word “pardon”) to imply a desire to establish shared gaze, in order to recast or repair a misunderstanding in a conversation. In situations where mutual gaze was not held (e.g., when the speakers had their backs turned to each other, or when they were in separate rooms), or was not possible (e.g., in telephone conversations) the use of this particular signal increases substantially, indicating that shared gaze holds a significant place in repairing disjointed conversations and facilitating cooperative exchanges.

Shared gaze has also been shown to have positive impacts on retention and recall with both children and adults in educational settings. Otteson and Otteson (1980) reported that children were better able to recall stories told by a teacher who shared gaze with the students than one that did not. Likewise, Fry and Smith (1975) reported that university students instructed by an experimenter who made eye contact with them during the instruction period obtained higher results on a digit coding task than those who were instructed without shared gaze. Breed and Colaiuta (1974) found similar results in classroom settings. Further work with university students has shown that shared gaze between pupil and professor elicits more classroom interaction and participation (Caproni, Levine, O'Neal, McDonald, & Garwood, 1977). This is important when considering classroom size in university settings. Perhaps students may benefit more from smaller classrooms and lecture halls, or lowered available registration numbers (in favour of increased sections of each class in order to provide everyone who desires to take a course the opportunity to do so), allowing for more active, interpersonal, non-verbal engagement (i.e., shared eye contact, visible nodding, etc.) between the instructor and each student?

Whether analyzed separately or in conjunction with facial expressions, posture, and the disposition of the body, oculomotor patterns are generally understood as a reliable means of monitoring interactants' engagement, intentions, and comprehension. This is evidenced by Saito et al.'s (2010) brief elaboration on the between-subjects synchronized activation of the right inferior frontal gyrus in partnered JA interactions, in which they write, "[e]ye contact implies the sharing of various psychological states" (p. 9). The statement highlights that eye-contact, more than other non-haptic communicative gestures, stimulates a deeper sense of interpersonal connection in addition to a heightened sense of inter-subjectivity. As such, the eyes have been crucial to previous studies in joint attention.

### **Purpose of the Present Study**

In the present study participants progressed through alternating learning and testing phases in a word-picture learning task in which JA was manipulated. The study had several goals. By adulthood, a typical person should be well practiced in the conscious and subconscious patterns of JA usage associated with facilitating communication. Although insight into the unconstrained potential for JA to assist language acquisition has been provided (Saito et al., 2010; Redcay et al., 2010; Hirotani et al., under revision), further research is required to determine precisely what features of JA affect adult word learning and whether infants are specially attuned to social learning in ways that differ from adults.

Although infant-adult and child-adult interactions inherently hinge on a hierarchical relationship, adult-to-adult interactions occupy spaces of explicit authority imbalances (e.g., employee-employer and student-professor exchanges) as well as perceived neutral relationships (e.g., coevals, colleagues, companions). Kendall and Schussler's (2012) observations of adult learner perspectives on these relationship dynamics in an educational setting offer insight into

affective factors that may alter learning efficacy. Compiled answers from 110 undergraduate students' opinions on instructor types across disciplines portrayed professors as being entirely more knowledgeable, confident, removed, and authoritative compared to their less distinguished counterparts, teacher assistants. Conversely, graduate teaching assistants were perceived as less competent and less poised, but more approachable and engaging in their instruction styles. With this in mind, the current study was initially designed to be a two-part experiment in which the first group of participants were to be told that their video partner was a teacher assistant and the second group of participants were to be told that their video partner was a professor or a language instructor. Through this division, the primary novel aim of this study was to examine whether initial perceived ideas and implicit attitudes of the JA partner would lead to a noticeable difference in language learning with the expectation that differentiation between instructor types would be found, however it was uncertain which would be more favourable to learning.

The present study also aimed to revisit and replicate Hirotani et al.'s (under revision) results in an attempt to learn more about the reliability of the differences between all four JA trial types (i.e., SJA, IJA, RJA, NJA) in adult language learning. The secondary purpose of the study was to analyze eye-tracking and behavioral response data to further examine how several concomitant JA parameters (i.e., duration of shared gaze, duration of gaze directed at a target object, direction and timing of saccades, environment, and number of trial exposures) may impede or support adult lexical development, retention, and recognition of unfamiliar words.

Given the observations of Kim and Mundy (2012) and Hirotani et al. (under revision) that adults, like infants, learned new words more efficiently in socially-driven JA tasks, I predicted that the adult participants of this study would demonstrate a heightened acquisition of word-object pairs that were presented in JA conditions. Similar to Hirotani et al.'s findings, the results were

furthermore expected to demonstrate an advantage for SJA conditions over all other conditions. Based on Kim and Mundy's (2012) results, I expected that IJA conditions would produce superior learning effects in comparison to RJA conditions. In their second study, Kim and Mundy showed that prolonged opening gaze times in the RJA condition did not create a significant deficit in target object gaze times, removing a possible explanation for the differentiation between IJA and RJA results. However, I speculated that the results from the current study could be contingent on whether the early gaze shift of RJA trials made the sharing of gaze direction and timing more explicit than in the IJA trials, where poorer peripheral vision or deep concentration could preclude observances of shared attention. Nevertheless, I predicted that SJA, IJA, and RJA conditions would yield better learning results than the NJA condition.

Analyzing these factors in conjunction with creating a pseudo-foreign language learning environment, the present study was also designed with the peripheral intention of uncovering how the enhanced mastery of word learning by adults in various social interactional contexts could contribute to the academic dialogue surrounding language teaching methods. Important particularly to the domain of second language teaching, results from the present study could offer potential clues as to how to restructure instructional methods to incorporate more pragmatic, gestural communication (i.e., pointing, evident gaze cueing, and other body language) and more explicit coaching in the development and interpretation of pragmatic cues (i.e., explicit focus on sociolinguistic, cognitive and cultural implications and meanings of gestures, and the teaching of how and when these non-verbal components are used). If language learning cohorts are shown to benefit from shared experience between peers or coevals, after demonstration of pragmatic communication, instructors could integrate more JA opportunities for conscious, shared engagement into the language learner's classroom experience.

In an age that is increasingly reliant on digital teaching and learning methods, this research could provide grounds to reflect on why it may be important to synthesize more interpersonal teaching approaches into less socially isolated digital learning spaces. It could raise questions about the efficiency of teaching through televisions and force the analysis of how much pragmatic cues are successfully communicated and interpreted via digital media. This work could lead to the restructuring of teaching styles over digital devices, perhaps calling for more theatrical or explicit actions, or an increase in verbal directions when using video-on-demand or televised lectures. Finally, should JA be shown to contribute significantly in a positive way to learning, then perhaps instructors should rethink the recent push advocating for the increased usage of digital spaces such as digital classrooms and the removal of the physical presence of both teachers and peers from the learning experience.

## **Method**

### **Participants**

Nine participants (7 female, 2 male) were recruited for the experiment through the Carleton University SONA recruitment system (see Appendix A). Prior to commencing the experiment, all participants reported being right-handed, dominantly Canadian English speakers, and having good or corrected vision (with soft contact lenses only). None of the participants reported colorblindness, speech or hearing impairments, or previous diagnosis of any neurological or psychiatric disorder<sup>1</sup>. Each participant was given course credit for their volunteer work in this experiment. The participants ranged in age between 19 and 24 years ( $M = 23.22$ ).

---

<sup>1</sup> Some of the requirements, (i.e., right-handedness) were not pertinent to the eyetracking data, but were crucial to the EEG phase of the experiment that is not reported on in this thesis.

## **Apparatuses**

The experiment was divided into two sections: a learning phase and a testing phase.

The learning phase experiment was assembled using EyeLink Experiment Builder software, version 1.10.1 (SR Research, Inc.) and was displayed using EyeLink 1000 hardware in addition to a host personal computer (PC) using Windows XP. Visual stimuli were presented on a 260 mm x 365 mm monitor screen with a resolution of 1024 x 768 pixels per inch (PPI) and played through speakers set to medium volume.

The testing phase of the experiment was programmed using Presentation software (Neurobehavioural Systems, Inc.) version 16.3 and displayed on a host PC using Windows XP. Visual stimuli were presented using a 325 mm x 520 mm monitor screen set to a resolution of 1024 x 768 PPI and auditory stimuli were played through speakers set to medium volume. Responses were measured using a Cedrus RB-730 desktop response pad (Cedrus Corporation). Participants responded using the leftmost button to signal “yes” and rightmost button to signal “no” on the response pad during testing phases.

## **Stimuli**

Three types of stimuli –one auditory and two visual- were used in this experiment. The learning sessions employed all three, (i.e., video, sound and bordered pictures) and the test session employed two (i.e., sound and non-bordered pictures).

Auditory stimuli consisted of 40 mono-syllabic English pseudowords taken from the ARC database (Rastel, Harrington, & Coltheart, 2002) (see Appendix B-1 for wordlist). The pseudowords represented phonotactically valid but non-existent English words. They were formed with the following restrictions: three phonemes ordered in consonant-vowel-consonant formation, maximum number of phonological neighbours set to 45, minimum set to 28 (with a mean of 32.35).

Initially the number of phonological neighbours was set at 26, but this was expanded to 45 in order to assemble pseudowords that were phonologically different from each other –as opposed to only orthographically different- in addition to pseudowords that were phonologically dissimilar from all existent English words, as opposed to only orthographically non-existent but phonologically existent. The original list was composed of 80 pseudowords, but was narrowed to 40 due to time constraints on the duration of the learning phase and the heightened difficulty of mapping and learning 80 picture-word pairs (for a list of pseudowords removed and justifications see Appendix B-2 and Appendix B-3 respectively). All pseudowords were presented as common nouns, thus occupying the same lexical category over the course of the experiment.

The pseudowords were recorded using the software Audacity version 2.0.6 in a single session by the same female Canadian English speaker. Through minor compression and elongation modifications using Audacity, the duration of the pseudowords was fixed at 800 ms in duration. The perceived naturalness of the pseudowords was unaffected in the process. The pseudowords were next distributed into four separate groups of ten. Number and type of onset (i.e., first consonant or consonant cluster of a syllable), nucleus (i.e., the middle portion of a syllable), and coda (i.e., the final portion of a syllable), in addition to vowel types and tokens (i.e., numbers) were balanced and distributed as evenly as possible throughout the groups. Statistics show that none of the lists differed significantly in pitch,  $F(3, 36) = 1.34, p = .277$  (see Table 1).

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	324.494	3	108.165	1.340	.277
Within Groups	2906.945	36	80.748		
Total	3231.439	39			

*Table 1.* Independent measures ANOVA. This table displays the ANOVA pitch results for the word groups.

During the learning phase, the pseudowords were presented simultaneously with 80 uncoloured pictures of non-existent objects taken from Kroll and Potter (1984). Although the pictures themselves were colorless, duplicates with red, blue, and yellow borders were created through the use of the software Paint (Microsoft) for the learning phase of the experiment in order to guide participant gaze during various joint attention conditions (see Appendix C-1 for a complete list of target pictures and Appendix C-2 for a complete list of non-target pictures). Thus a total of 320 bordered and non-bordered picture stimuli were employed throughout the entirety of the experiment. The non-objects had previously been rated for similarity to real objects by Hirotani et al. (under revision) and were thus divided into eight groups of ten (i.e., 4 target groups and 4 non-target groups) based on these rating results (see Appendix C-3 for target object similarity ratings and group assignments and Appendix C-4 for non-target similarity ratings and group assignments). The target picture groups were controlled stringently for maximum rating uniformity and showed no significant difference,  $F(3, 36) = .628, p = .601$ , whereas the non-target picture groups were assembled using the remaining pictures and consequently showed greater variation in ratings,  $F(3, 36) = 2.56, p = .070$  (see Table 2 for target groups and Table 3 for non-target groups results). As the participant's attention was expected to be focused most intensely on the target pictures, it was deemed more necessary to ensure a strict uniformity between these four

groups than the non-target groups. Calculation of the visual angle and size of the display monitors resulted in learning phase pictures to be displayed at 250 pixels x 250 pixels and the test phase pictures at 370 px x 370 px (see Table 4 for summary). The invisible eye interest areas situated on the “learning partner” videos measured 75.6 px x 37.8 px (approximately 1.5 cm x 1 cm), while the target interest areas measured 250 px x 250 px.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2.058	3	.686	.628	.601
Within Groups	39.306	36	1.092		
Total	41.364	39			

*Table 2.* Independent measure ANOVA familiarity. This table displays the familiarity rating results for target-object groups

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	7.850	3	2.617	2.560	.070
Within Groups	36.794	36	1.022		
Total	44.644	39			

*Table 3.* Independent measure ANOVA familiarity. This table displays the familiarity ratings results for the non-target object groups.

The second group of visual stimuli consisted of 160 recorded 5200 ms clips of the researcher situated in the eye-tracker while participating in learning sessions. The videos were collected using a Logitech webcam and were cut from 4 same-day recordings of the researcher performing each of the four lists once with the aim of creating videos with appropriate left and right gaze and saccade timing as cued by the task design. The recorded face was framed in windows measuring 260 px x 480 px, with a centimetre portion of the eye-tracker visible for effect and believability. Facial reactions were kept neutral or moderately positive in order to avoid a confounding effect driven by affective bias or positive reinforcement. The clips were placed in the centre of the screen between the target and non-target objects (which altered sides, with 20 trials showing targets on the left and non-target on the right, and 20 trials showing target pictures on the right with non-target pictures on the left, in randomized orders), where they were used to establish the joint attention conditions with the participant (see Appendix D-1 and D-2 for task design examples). These videos were the second versions created in order to guarantee that joint attention was properly simulated with more natural timing parameter. In the first video creation process the “video volunteer participant” was simply following the task design without having established joint attention.

	Screen size (Height x Width)	Screen resolution (Pixels Per Inch)	Participant eye distance
Learning Phase	260mm x 365mm	1024x768	475mm
Test Phase	325mm x 520mm	1024x 768	1000mm

*Table 4.* Visual angle table. This table summarizes the pertinent display computer information for the separate learning and testing chambers.

Each condition required 40 video files; 20 of which were designated to left gaze shifts and 20 of which were designated to right gaze shifts. Once the mini clips were cut from the mother videos, the researcher and her team performed an initial filtering process to eliminate unusable videos based on the following criteria: wrong or mistaken gaze direction, misleading preemptive saccades, misaligned video start and trial onset timing (i.e., absent flash or no closing shared eye gaze), prolonged closures of the eyes and noticeably distracting fatigue effects (i.e., drooping of the eyelids). The successful preliminary selections were next subjected to a larger norming study, in which 10 participants were asked to judge the naturalness of the videos on a scale of 1 to 7 by attending to the facial expressions, movements and overall appearance of the subject. Responses were averaged across trials and participants. Trials scoring lower than one standard deviation from the average of its group were eliminated from the selection pool. The final video files were sorted into condition type and saccade direction before randomized orders were generated. The randomized videos were sectioned into groups of ten per condition per list, with five videos gazing left and five gazing right. The remaining videos were discarded.

The NJA and SJA conditions required the videos to share gaze for similar durations (approximately 1250 ms to a maximum of 1500 ms) before making a saccade to either the left or right. The IJA conditions required the videos to share gaze for a noticeably extended duration (approximately 1700 ms to 2000 ms) because the JA was meant to be “initiated” by the real participant, while the RJA condition required videos of shorter duration (approximately 1000 ms to 1200 ms). In the NJA and SJA conditions the gaze shift occurred simultaneously with the appearance of the borders (indicated by a very subtle noticeable flash on the video participant’s face). In the IJA condition the gaze shift occurred slightly later than the flash, while in the RJA condition the video shift occurred slightly earlier. Finally, the SJA, IJA, and NJA videos were

matched to pictures with red and blue borders, the RJA condition was matched to pictures with yellow border in order to force the participant to rely on the video to indicate the correct picture.

In order to guarantee maximum uniformity between all four stimuli groups, the target picture-pseudoword-non-target picture pairings in the learning blocks were unaltered across all lists (i.e., p01 and s01 were coupled consistently in every list). However, the conditions –and corresponding matched videos– were rotated through the picture-pseudoword pairs. For example: list one matched the NJA condition to picture and pseudoword items 1-10, RJA to 11-20, IJA to 21-30 and SJA to 31-40, whereas list two matched RJA to items 1-10, IJA to 11-20, SJA to 21-30 and NJA to 31-40 and so forth.

## **Procedure**

After signing acknowledgement and agreement of the informed consent (see Appendix E to view the form), participants were given a brief oral explanation of their expected role throughout the learning and testing sessions of the experiment. The directives were fully reiterated in two separate slideshow presentations which also provided visual examples of task instructions to better accommodate all individuals. Following the instruction slides, participants were given practice sessions for both the learning and testing phases, the former first and the latter second. Each practice block consisted of four trials designed to replicate the experiment trials. In the interest of eliminating strong reactions to the first few non-object-pseudoword pairs upon presentation in the actual experiment, the practice sessions substituted spare non-object and pseudoword recordings that were previously unused in the experiment design. This method also served to avoid confounding the learning results with pre-exposure. Once participants were comfortable with their tasks the experimental sequence began.

Participants first proceeded to a pretest in which they were asked to set the Cedrus response pad on their lap and to indicate either “yes” or “no” to whether they thought the audio recording was congruent with the single, borderless image that appeared on the screen. As participants had not yet been exposed to a learning session, they were asked to make guesses by pushing the left and right buttons at random, allowing for this block to serve as a baseline with which to compare proceeding test results. The test sessions were held in a sound attenuated chamber that housed the display monitor on a desk, with speakers placed evenly to each side. Subjects sat in a chair placed approximately 100 cm from the monitor in order to maximize the visual field and perceptual comfort of the individual.

Each test trial progressed in the following manner: a fixation cross was shown for 500 ms, followed by a 2000 ms presentation of a single, unbordered non-object picture in conjunction with an 800 ms auditory recording of a pseudoword that was either consistent (congruent) or inconsistent (incongruent) to the learning phase picture-sound pairs, after which another fixation cross to center participant focus was displayed for 1500 ms. Following the last fixation cross, the participants were prompted to judge the validity of the picture-sound pairs through the appearance of a question mark on the display screen, which timed-out after 3000 ms. Finally, a blank screen appeared during the inter-trial interval for approximately 500 ms at which time the cycle restarted. The entire test phase consisted of 4 separate blocks of 40 word-picture pairs each, with each trial lasting up to 8 seconds.



*Figure 3.* Test trial time sequence. This figure displays the time sequence of test trials along with representations of what the participant saw at successive checkpoints.

Participants were informed that the fixation cross was the signal that a new trial was beginning and that they were expected to pay close attention to the picture and sound. They were asked to fixate on the cross and the following image while trying their best to hold their blinks until the question mark appeared on screen. This was done in order to guarantee their full exposure to the picture presented in each trial. Subjects were also instructed that in order for their responses to be recorded, they must respond only while the question mark was being displayed on the screen. Finally, they were asked to try to remain relaxed but aware, and to respond according to their intuition or first impression (i.e., to not overthink or overanalyze the coupled stimuli). If they were unsure of the match, participants were instructed to simply guess and not to worry should they feel an error was made.

Upon completing the first test block the participants were led to a nearby room containing the eye-tracking equipment. The learning phase of the study involved a picture-sound association learning and memory task designed to record the exact locations and durations of eye gaze and saccades. During this task participants sat in a chamber containing the stimuli display computer equipped with speakers and an EyeLink1000. Participants were asked to face the display computer, align themselves with the EyeLink 1000 and pull the chair as close to the table as possible before comfortably resting their chin and forehead on the eye-tracker's designated rest pads. Once positioned they were asked to place their hands on their legs beneath the table and to focus on the monitor ahead. The display screen was approximately 47.5 cm away from the participant's face, with the speakers placed at an equal distance away from the participant.

Each learning session began with camera setup, consisting of adjusting the camera's resolution to best suit the participant's eye. This was followed by calibration, validation and drift

check in order to establish the most accurate tracking of the participant's right eye. The camera setup period varied between 4 minutes to 15 minutes in duration, depending on whether the participant was required to remove eye makeup or not.

Each learning trial followed a consistent pattern and was approximately 6 seconds in length. First, a fixation cross would appear on the screen for 500 ms. This was followed immediately by the presentation of the "learning partner" video gazing forward for times ranging between 1000 ms to 2000 ms (depending on the joint attention condition being fulfilled) simultaneously with the appearance of the non-bordered pictures (located parallel with the video, one on the left and one on the right) for 1200 ms. Next, the coloured borders emerged for 1000 ms concurrently with the 800 ms audio recording. To establish the various joint attention conditions, the video "learning partner's" controlled gaze shift occurred either slightly before, in conjunction with, or slightly after the appearance of the colour borders. The sequence was finalized with the instantaneous disappearance of the borders in addition to the closing shared eye gaze, in which the "learning partner's" gaze once again shifted to the centre and looked forward for 1000 ms to 2000 ms. After this sequence, a blank screen would briefly flash in order to signal the commencement of a new trial.

All 40 trials were presented in each learning phase, however the orders were randomized differently in each block using the Randomization option in Experiment Builder (SR Research, Inc.). Similar to the design of Bayliss, Paul, Cannon and Tipper (2006), participants were asked to focus on the fixation cross until the appearance of the video, with which they were instructed to establish and maintain eye contact. Eye contact was trained to be held until the borders appeared on screen, after which participants were to gaze at the target object in order to form the congruent picture-word association. The instructions dictated that if the borders surrounding the pictures

were blue and red, the picture with the red border was the target object being named by the recording and the picture with the blue border, along with the gaze patterns of the video, were to be disregarded. When the borders appeared yellow, the participants were told to follow the direction of gaze cued by the video as it would indicate the correct object being named by the sound. The participants were strongly encouraged to return their eye gaze to share contact with the video participant once the borders disappeared. Aside from gaze fixation durations and saccade directions, no response from the participant was required for this portion of the experiment.

In total the participants completed three learning phases, which were alternated between four test blocks and incorporated short breaks between each set. The learning block trials were presented in an automatic unique randomized order, under the constraints of avoiding any joint attention condition or target picture location (either left or right) to be presented more than three times consecutively. There were four different test lists to match the four different learning lists, which were circulated through the participants in an equal systematic order. In each test list, 50% of the picture-sound pairs were congruent and 50% were incongruent. Each picture-sound pair was presented in a congruent condition twice and in an incongruent position twice, meaning that in two of the four lists the picture was paired with a different sound than was learned in the learning sessions. The participants were never presented the same picture-sound incongruent pairs in the test sessions.

Upon completion of the main tasks the participants were asked to fill out a language background questionnaire outlining their language history and some demographic information (see Appendix E for the Language Background Questionnaire). The experiment lasted approximately three hours. As it was jointly comprised of an EEG component not included in this report, the

learning and test session alterations composed approximately an hour and a half of the three hour timespan for which the participants were credited.

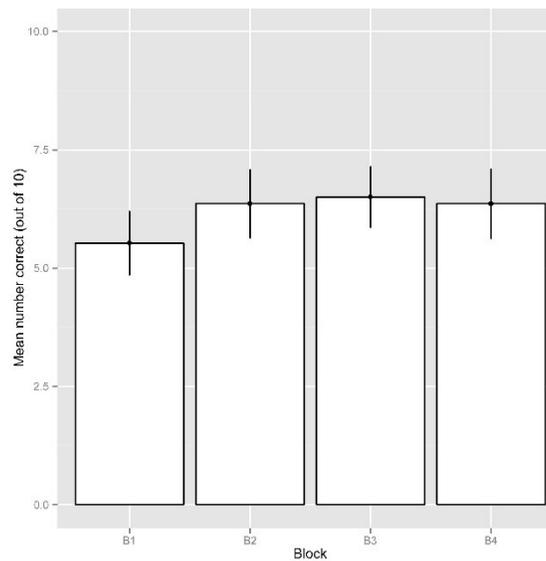
## **Results**

Test trial results for each condition of each learning Block were obtained by calculating the mean gaze duration on the interest areas set to the “learning partner’s” eyes and the target objects. Mean opening shared gaze-time across participants was 497 ms, indicating that participants spent between slightly less than half to less than a quarter of the allotted 1000 ms to 2000 ms of shared eye gaze time focusing on the video participant’s eyes. This could perhaps point to a design flaw, as shared gaze for a minimum of 1000 ms was determined as the precondition in establishing “common ground” or the knowledge of shared directed gaze. In contrast, mean gaze-time at target objects across participants was 2181 ms, indicating that on average nearly the entire 2200 ms assigned to direct focus at the target-object was used for its intended purpose. Due to data retrieval difficulties, data from subject 8 were removed from the learning sessions, hence averages were calculated on the remaining eight participants.

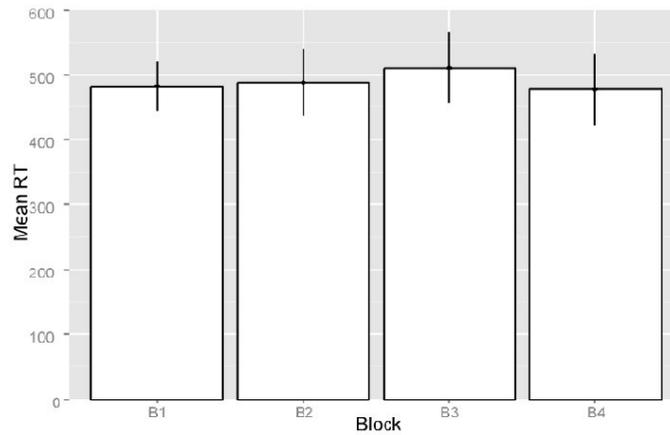
Mean percent correct values and mean response times (in ms) were calculated for each participant as a function of Block and joint attention condition. As a two alternative forced choice task was used for the test sessions, chance performance was 50%. Results were analyzed using a repeated measures analysis of variance (ANOVA), with four levels for the joint attention variable and four levels for the block variable.

Average performance across participants in Block 1 was 55%, near chance level. The results increased to 63% in Block 2 and 65% in Block 3. However, by Block 4 the average performance in percent correct totalled 63%, resulting in an overall 8% increase in percent correct

when comparing Block 1 results to Block 4 results. Overall no evidence for significant learning was observed,  $F(3, 24) = 0.844, p = .48$ . Figure 3 shows the mean accuracy across the four training blocks, with the error bars indicating 95% confidence intervals. The lack of difference is corroborated by the average response time (RT) results across participants: 484 ms for Block 1, 491 ms for Block 2, 504 ms for Block 3, and 480 ms for Block 4. Increased accuracy over learning blocks would typically be accompanied by faster RTs. Here, the RTs are relatively constant across blocks,  $F(3, 24) = .294, p = .83$ . Figure 4 shows the mean response times across the four training blocks, with the error bars indicating 95% confidence intervals.

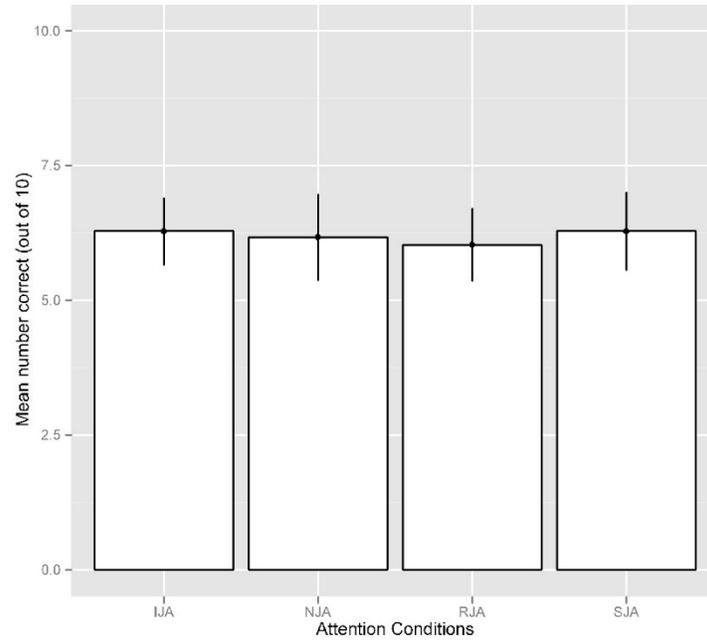


*Figure 4.* Mean correct responses per block. The error bars in this figure correspond to 95% confidence intervals.

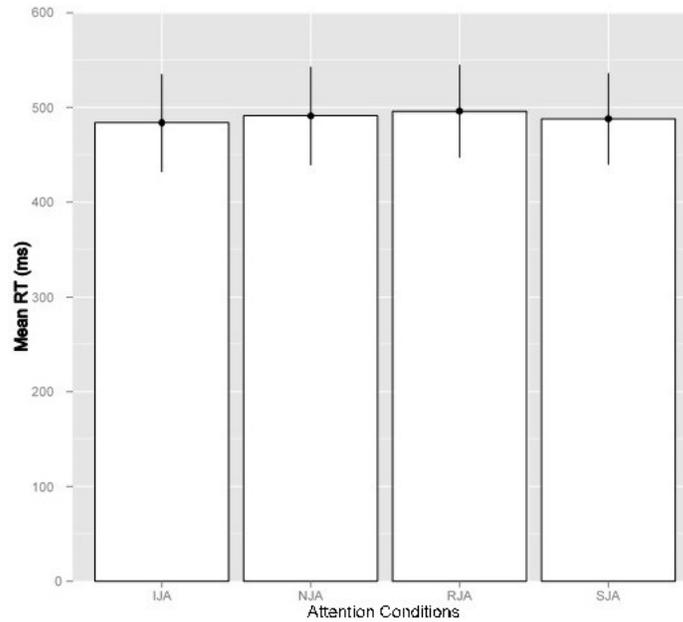


*Figure 5.* Mean response times per block. The error bars in this figure correspond to 95% confidence intervals.

Mean accuracy for each JA condition collapsed across blocks was not significantly different: 62.78% for IJA, 61.67% for NJA, 60.28% for RJA, and 62.78% for SJA,  $F(3, 24) = .159$ ,  $p = .92$ . Figure 5 shows the mean response performance accuracy across the four conditions collapsed across the blocks, with the error bars indicating 95% confidence intervals. Mean RT for each JA condition collapsed across blocks was not significantly different between conditions: 484 ms for IJA, 487 ms for NJA, 499 ms for RJA, and 492 ms for SJA,  $F(3, 24) = .264$ ,  $p = .85$ . Figure 6 shows the mean response times across the four conditions collapsed across the blocks, with the error bars indicating 95% confidence intervals.



*Figure 6.* Mean number of correct responses per joint attention condition. The error bars in this figure correspond to 95% confidence intervals.



*Figure 7.* Mean response times per joint attention condition. The error bars in this figure correspond to 95% confidence intervals.

Figure 7 shows the interaction between block and JA condition accuracy. Average performance per condition showed only minor improvements as Block 1 results stood at 56% IJA, 56% NJA, 46% RJA, and 61% SJA in comparison with Block 4 values of 67% IJA, 64% NJA, 58% RJA, 63% SJA (Block 2 results averaged at 57% IJA, 57% NJA, 70% RJA and 68% SJA and Block 3 results averaged at 68% IJA, 67% NJA, 65% RJA, and 57% SJA). The differences in conditions between Block 1 and Block 4 indicate that participants improved by the following ranges over the course of the experiment: 11% IJA, 8% NJA, 12% RJA, and 2% SJA, showing a maximum of 10% performance disparity between the conditions in the final test block. Nonetheless, no significant interaction between JA condition and block was observed,  $F(9, 72) = 1.44, p = .19$ .

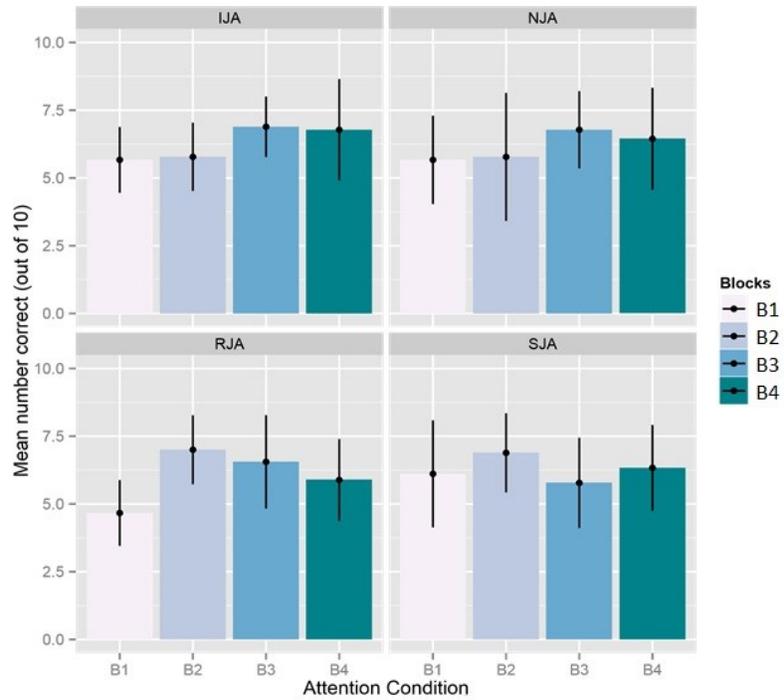
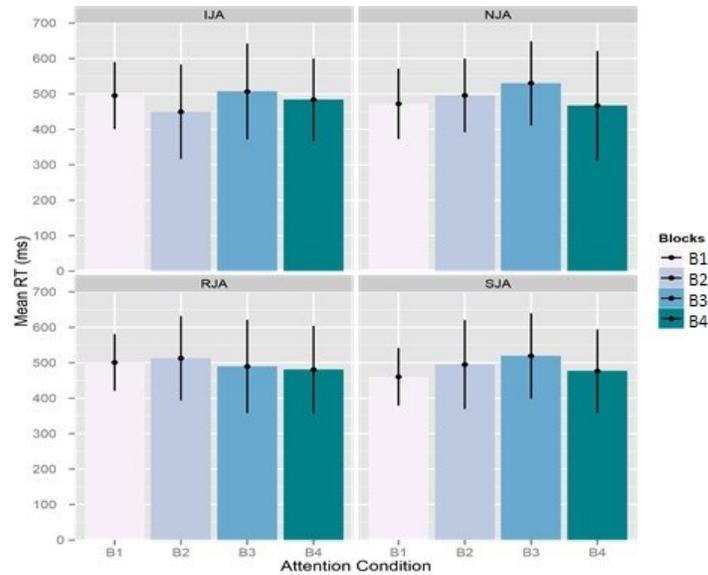


Figure 8. Mean correct responses as a function of JA condition and block. The error bars in this figure correspond to 95% confidence intervals.

Figure 8 shows the interaction between mean block and JA condition RTs. No clear pattern is apparent for response times as a function of block and attention conditions. This is borne out in the lack of a significant interaction between block and JA condition,  $F(9, 72) = .823, p = .59$ .



*Figure 9.* Mean response times as a function of JA condition and block. The error bars in this figure correspond to 95% confidence intervals.

A Language Background Questionnaire was administered to provide more detail about the verbal assertion made by participants that they spoke Canadian English as their dominant first language. This questionnaire also served to measure participants' previous exposure to any other language other than Canadian English and in what contexts. Results indicated that one subject was exposed to a bilingual household that spoke both English and Russian primarily until the age of 6, upon which time the home switched to English dominantly. This same participant was not born in Canada, but lived in Canada for the past 11 years and reported Canadian English as her dominant language. All of the other subjects were raised in monolingual Canadian English households. Seven of the participants reported being exposed to French in school after the age of 4. Two of these seven participants reported having been in French Immersion and learning in French for approximately 5-6 hours daily until the age of 16. One participant reported 4 hours of

exposure daily and four reported less than 2 hours. The remaining subject reported no exposure to other languages.

No relationship was found between language exposure and learning ability as measured in the pseudoword learning task.

## **Discussion**

Overall, the results of the current study did not support those of previous studies that showed joint attention could contribute to efficient word-learning in adulthood. I expected that performance across blocks would be consistent with the learning results obtained by Hirotani et al. (under revision), and that response accuracy would increase significantly by the 4<sup>th</sup> test Block in comparison with the 1<sup>st</sup>, while displaying a linear progression of learning between all blocks. However, although the mean number of correct responses increased by 8% between Block 1 to Block 2, and by 2% between Block 2 and Block 3, the mean response accuracy between Block 3 and Block 4 decreased by 2%. Despite these modest differences, it is important to note that none were different from what would be expected by chance.

The decline in the last block could be due to fatigue or task habituation, as familiarization effects have been shown to occur between three to ten exposures to stimuli groups (Lubow, Alek & Arzy, 1975). Fatigue or boredom may have interfered with participants' learning, as the eye-tracking task is repetitive and the experiment involved an extended period of sitting in a stable position. However, as the duration of this portion of the experiment totalled at one and a half hours, it is more likely that loss of interest or distraction may have interfered with the participants' desire and ability to learn. The learning blocks were set to have a randomized order upon each presentation. This was done in part to counter predictability and loss of interest in the sequence

of item presentation. Yet, the continuous appearance of the same 40 picture-word pairs without novel filler stimuli to break the patterns may have resulted in habituation and desensitization to stimuli, as Lubow et al. (1975) noted. Not only has presenting the same stimuli in an unchanging environment been demonstrated to hinder learner interest and performance (Lubow, Rifkin & Alek, 1976), but interactive mindfulness and gaze interest have also been predicted to be heightened in novel settings as opposed to familiar ones (Langer, 1978). Should fatigue have been a culprit in interfering with performance, we can ask why these factors were not apparent in Hirotani et al.'s. (under revision) study? I speculate that their use of a live paradigm, in comparison with the prerecorded videos used in the current experiment, may have contributed to increased participant alertness. This is further discussed in the Future Directions section below.

In accordance with Hirotani et al.'s (under revision) and Kim and Mundy's (2012) findings, I expected that Simultaneous Joint Attention and Initiating Joint Attention would produce the highest learning results. However, contrary to expectations, none of the joint attention conditions produced differential effects on performance. Curiously, in contrast to Kim and Mundy's findings, RJA accuracy was the highest and RJA response timing was the fastest of all four conditions. Eye-tracking gaze time data reveals that RJA had the second longest opening shared eye gaze after SJA, however it had the longest average gaze time at the target objects. Although the results provided some marginal support that learning occurred in this condition and that the videos had some minor effects on the learner's attention, none of the learning scores yielded significant results. Overall, it is not clear what is responsible for these findings.

The presentation of a single, English pseudoword removed from a sentence and paired with only two pictures casts doubt over the possibility of lexical ambiguity or failed referent mapping in any JA trial. However, the difference in learning could be due to the difference in coloured

borders used for the RJA conditions. One could surmise that the yellow borders that surrounded the RJA target and non-target pictures countered target-object saliency in comparison to the stark contrast between the red and blue bordered trials. Due to the contrast in colour conditions, it is possible that the RJA condition may have been the only JA condition to establish a true JA event, as the yellow borders surrounding both the target and the non-target objects would have forced the participant to rely more heavily on the video participant, thus establishing greater awareness of the other, as opposed to simply turning to the red border when cued. If joint attention fosters a better learning environment for the learner, one would expect the Responding Joint Attention condition to yield the most successful results, as it simulates the joint attention experience from the learner's perspective where one's attention is being directed to the object of focus. However, it should especially be differentiated in a positive manner from the Non-Joint Attention condition. Unfortunately, the word-picture pairs learned in conjunction with social interaction conditions did not demonstrate faster identification over exposure nor a significant increase in recognition at all. In short, no JA effects were found. As was the case for the results considered by learning block, any systematic account proposed here is speculative given the insignificant differences between the JA conditions

### **Limitations**

Various differences between Hirontani et al.'s (under revision) experiment and the current study may account for the inconclusive results obtained in the present study. One major shortcoming of the recent study was the failure to recruit a large number of participants. Initially the project was expected to include a minimum of 20 subjects. However, due to time constraints and extenuating circumstances, recruitment was foreshortened. Participant recruitment ended after only one week, which resulted in useable data from a total of only nine subjects. As this number

of subjects is insufficient to make accurate statistical inferences, it may underlie the incongruity in results between the present study and previous JA experiments.

Another difference between Hirotsu et al.'s (under revision) experiment and the current study was the time allotted to the eye-tracking learning and testing phases. Due to time constraints placed on the behavioural phase by the non-behavioural phase (which is not discussed in this thesis), this portion of the experiment was restricted to 90 minutes to complete informed consent, practice trials and the training and preliminary test sessions. While Hirotsu et al. were able to perform five learning sessions and six testing sessions, the current study was only allotted enough time to complete four test sessions (including the pretest) and three learning sessions. Hirotsu et al.'s data indicates that the clear distinction between JA conditions did not emerge until the fourth test session, stipulating that the majority of participants required four learning trials in order to crystallize the information being learned. Perhaps had the current study included two or more extra learning sessions, the results may have better reflected those of previous studies.

Another limitation is that the shared gaze between participants and the video partner may have been insufficient to promote JA. As indicated in the Results section, analysis of the average dwell time on the video partner's eye interest areas across participants revealed that the duration of shared gaze was less than half of the designated minimum requirement. This shortened shared gaze at the beginning of trials is suspected to have prevented the subjects from subconsciously registering the subtleties between the JA conditions. In designating less time to sharing eye gaze, the participants may have failed to recognize a sense of "common ground" or "shared experience" which has been discussed as the predicate condition of JA. This may have been caused by the early appearance of the objects in addition to the later appearance of the borders. In this experiment, the non-object pictures appeared simultaneously along with the video participant. The

eyetracking data revealed that participants' attention often wandered away from the video participants' eyes in favour of viewing the two non-object pictures to each side of the video. To prevent this from happening in future experiments, the non-object pictures should appear at the time the borders were set to appear, alongside the borders rather than beforehand. This may compel the participant to share gaze with the video participant longer, which may contribute to establishing a better sense of joint attention and a shared common interest.

According to Patterson's (1982) Sequential Functional Model of Nonverbal Exchange, which analyzes the various levels of nonverbal interaction, behaviour is communicative when conscious intention can be identified. Patterson (1982) explains that alignment of behavior and intentions leads to successful communication, and so one's perception of an interlocutor subconsciously affects one's ability and desire to properly or fully align oneself in dialogue or action. As noted above, although gaze is commonly used to signal communicative intent, it is also arguably the most subtle communicative cue. It is thus possible that in the video context, the artificiality of the pre-recorded sessions or the nonideal social and spatial environment of the laboratory interfered with the natural interpretation of gaze shifts. Rather than facilitating JA, a synthetic social situation may have instead conflicted with unconscious pre-established models of JA instances, potentially causing a cognitive "misalignment" or disassociation between linguistic knowledge and situation models (Pickering & Garrod, 2004; 2006). It is possible that in their awareness of the recorded conditions, the participants felt decoupled from a sense of a true "shared" interpersonal experience.

Another factor contributing to the artificiality of the video conditions was the sizing of the videos. Although the clips were used specifically to control the particular JA parameters and were rated on believability, in only displaying the participant's eyes, the videos themselves added to the

unrealistic interactive atmosphere. As the majority of residents in North America leave their faces exposed to others, aside from speaking through the internet or the telephone, most North Americans are accustomed to being able to observe an interlocutor's facial features during quotidian interactions. Restricting exposure of the video participant's face to the eyes may have added a sense of artificiality to the interaction. It may have also made the participants uncomfortable, or unsettled, seeming as how they were unable to observe and interpret their partner's eye moments in the context of entire facial cues rather than isolated cues. The discomfort may have decreased subjects' willingness to align their actions, thoughts, and frame of mind to that of the video participant's. Future designs should reevaluate the size of the video, and take into consideration that displaying the entire face of the video participant may better imitate a more natural and expected social environment.

In combination with creating more naturalistic communicative settings, future research could begin to explore how JA with adults can be elicited through other forms of pragmatic gestures in live settings similar to that of Hirotsu et al. (under revision), Saito et al. (2010), and Redcay et al. (2013). The present results should then be compared the potential advantages or disadvantages on learning effects in these various contexts.

### **Future Directions**

Previous child language research has documented the heightened learning effects of interpersonal interactions in the development of phoneme categories compared to learning in digital interactive scenarios such as televisions, radios, and computerized animations (Kuhl, Tsao, & Liu, 2003). As for adult language learning, the advantages of videos for comprehension and grammar enhancement have been documented (Herron, York, Corrie, & Cole, 2006) as well as the use of captioned or subtitled television in lexical recognition and retrieval (Yuksel & Tanriverdi,

2009). However, the benefits are mostly driven by studies using learners with prior exposure to the language in question, comparisons to audio material only (Bal-Gezegin, 2014), or conclusions that a combination of live interactive learning and computer-driven learning result in the best learning environment for language acquisition (Rezaee & Ahmadzadeh, 2012). Redcay et al. (2010) showed that joint attention is most effective in a live social setting (i.e., in an in-person, face-to-face interaction) as opposed to an offline interaction (e.g., recorded video), however it remains unclear exactly why the live interaction conditions are differentiated from the recorded conditions. One could speculate that awareness of the presence of the other may cause the participant to be more alert or attentive, although the potential reasons remain numerous: a feeling of being watched, a sense of competition or an underlying drive to perform better, a desire to interact, or engrained cultural expectations about appropriate social interaction, the neurological sense of ‘reward’ as noted by Redcay et al. (2010), etc.

Despite these findings highlighting the benefits of live interaction to information encoding, currently a great emphasis is being placed on the role of enhancing language learning through digital resources. In light of this, future joint attention research would benefit from trials that simulate a natural joint attention occurrence, like the eye-tracking experiment of Hirotsani et al. (under revision) where two or more adult subjects participate in a digital on-line interaction, and then next in a non-digitized on-line interactive space. Whether in mobile interfaces or classrooms equipped with computers and relevant software, a comparison of the success rates of these digital and interpersonal interactive environments could help shed light on how video on demand, or other digital classrooms may help or hinder language learning.

Participants in the present study were told they were interacting with a video of a teaching assistant, which was intended to make them consider that the video corresponded to a person

relatively close in age to themselves. In the planned second recruitment phase of this experiment, participants were going to be told that they would be interacting with a professor, making a distinction in both age and authority. The aim was to compare the results of the effectiveness of joint attention with a person of perceived similar age or a “peer” with a person of authority in order to explore whether joint attention is more effective within certain relationship dynamics where there is less or more of a power difference. Reactions to gaze, willingness to partake in shared gaze, and effects of gaze have been proposed to be contingent on subjective evaluation of the other (see Kleinke, 1986 for an elaboration). In light of this, it would be useful to expand the video choices by creating gender differences, so that there are both male and female teacher assistant and professor pairs in order to analyze whether a gender bias towards the same sex or the opposite sex during joint attention.

A crucial limitation of the present study was the lack of a questionnaire like that administered by Kendall and Schussler (2012) in which participants rated perceived differences between professor and TA teaching styles, attitudes, knowledge base, authority, approachability and communicative competence. This suggests an opportunity for future research to explore the role of the explicit or perceived knowledge of a JA partner and how this may internally motivate or encourage language learners. Bakeman and Adamson (1984) suggested that infant JA is more successful with known or very familiar partners (i.e., with one’s mother as opposed to a family friend or a stranger). They speculate that because the patterns of behavior are familiar, less effort and time is required to interpret the other’s behavior and properly coordinate one’s attention, thus perhaps allocating more neural resources to encoding the experience and consolidating new information (Bakeman & Adamson, 1984). As far as I am aware, currently no literature is dedicated to observing the effects of relationship dynamics or perceived prior knowledge on

adult JA interactions, leaving a wide area open for exploration in JA studies. After completing the analysis of instructors, a next step would be to recruit friend groups, couples, siblings or family members to measure if there are affective factors that influence the efficacy of JA contexts. Potential confounds include huge variability in subjective measures of relationship status and dynamics between proposed JA pairings (professors, significant others, good friends, siblings) and methodological challenges in situating or pinpointing affective variables on learning.

It could be pertinent to design an experiment with a similar focus on joint attention in addition to an emphasis on personality nuances. It has yet to be documented whether self-reported extrovert type personalities outperform others in an interactive joint attention learning situations, or whether distinguished personality types may effect one's willingness and ability to establish a sense of conscious interaction and shared goals. This could be accomplished through the use of a stringent and reliable personality questionnaire, such as the Myers-Briggs Type Indicator, or other tests based on the Five-Factor Model.

The experiment design in the present study required learners to correctly map picture-word pairs, but it did not require verbal or written production of the words. Although the tasks may have targeted implicit memory, recognition, and comprehension, it did not examine whether joint attention contributes toward the type of full acquisition that would entail explicit recall and production. Future work could involve a longitudinal study where the learners could be asked to produce the words themselves immediately following the task, in addition to a week or two later to observe to what degree the content was learned or acquired. However, before undertaking as elaborate of an experiment as the above mention, it may be best to start with a more simplistic longitudinal design similar to the current experiment. Instead of exposing learners to all of the

learning and testing sessions in a single day, it may be more efficient, more naturalistic, and less taxing on the participants if they were asked to frequent the lab several times throughout a week or a month. During each visit they would be exposed to a single learning session as well as a single testing session, through which it would be possible to measure the efficiency of numerous shorter exposures in comparison to a single, prolonged session.

All four joint attention conditions have yet to be thoroughly tested with adults using electroencephalography technology. Given that infants show an early negativity for the congruent picture-word pairings (200-600 ms) and a late (800-1200ms) N400-like semantic effect in response to incongruent presentations of stimuli after JA teaching phases (Hirotsu et al., 2009), similar responses may exist in adults. The temporal results available from EEG data will also help determine the degree of recognition of the word-sound pairs even when the participants' processing time showed hesitation or when confusion of the congruent and incongruent matches manifest in response performance. This information will help refine the more subtle cognitive effects that joint attention may have on subconscious learning.

## **Conclusion**

In contrast to expectations, the results of the current study were not consistent with those of previous joint attention experiments. Although few conclusions can be drawn from the data obtained in the present study, several limitations and methodological shortcomings are suspected to have affected the final outcome of the project. In light of these limitations, many avenues for future research in the area of joint attention remain. At most, this experiment exemplifies a test of short-term advantages and relative rate of acquisition rather than ultimate attainment potential in adults. Evidently this experiment does not represent a global approach to language learning. However, in considering the well documented support for joint attention, as discussed above, JA

could presumably be applicable to many facets of language teaching and learning, beyond the current academic focus on lexical development. The best approach to flush out all of the intricacies surrounding the inherent value of true joint attention contexts would be to devise a systematic design to a series of experiments, each of which test minor aspects of JA one by one with all three types of partners: avatars, pre-recorded videos, and live. Of particular interest to the field of applied linguistics is how social perceptions may affect interactions, and in turn the unconscious willingness or desire of the language student to learn. In conclusion, it remains to be discovered whether joint attention can guide or enhance adult language learning and thus serve as a step in combating the apocryphal belief that adults are fated to experience overwhelming difficulties as they try to acquire a new language.

## References

- Baddeley, A.D. (1966). Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362-365.
- Baddeley, A.D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575-589.
- Bakeman, R., & Adamson, L. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interactions. *Child Development*, *55*, 1278–1289.
- Baldwin, D. & Markman, E.M. (1989). Establishing word-object relations: A first step. *Child Development*, *60*, 381-398.
- Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875-890.
- Baldwin, D. (1993). Infant's ability to consult the speaker for clues to word reference. *Child Language*, *20*, 395-418.
- Baldwin D.A. (1995). Understanding the link of joint attention and language. In: C. Moore, P. Dunham (Eds.), *Joint attention: Its origin and role in development* (pp. 131–158). Hillsdale, NJ: Lawrence Erlbaum.
- Baldwin, D., Markman, E.M., Bill, B., Desjardins, R., Irwin, J., & Tidball, G. (1996). Infant's reliance on a social criterion for establishing word-object relations. *Child Development*, *67*, 3135-3153.
- Bal-Gezegin, B. (2014). An investigation of using video vs. audio for teaching vocabulary. *Procedia - Social and Behavioral Sciences*, *143*, 450 – 457.

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bayliss, A.P., Paul, M.A., Cannon, P.R., & Tipper, S.P. (2006). Gaze cuing and affective judgments of objects: I like what you look at. *Psychonomic Bulletin & Review*, 13(6), 1061-1066.
- Bloom, K. (1974). Eye contact as a setting event for infant learning. *Journal of Experimental Child Psychology*, 17, 250-263.
- Bloom, K. (1975). Social elicitation of infant vocal behaviour. *Journal of Experimental Psychology*, 20, 51-58.
- Breed, G., & Colaiuta, V. (1974). Looking, blinking, and sitting: Nonverbal dynamics in the classroom. *Journal of Communication*, 24, 74-81.
- Brooks, R., & Meltzoff, A.N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535-543.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Bruinsma, Y., Koegel, R.L., & Koegel, L.K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews*, 10, 169-175.
- Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition*, 3, 255-287.
- Bruner, J. (1981). The social context of language. *Language & Communication*, 1(2/3), 155-178.
- Campoy, G. (2008). The effects of word length in short term memory: Is rehearsal necessary? *The Quarterly Journal of Experimental Psychology*, 61(5), 724-734.

- Campoy, G. (2011). Retroactive interference in short-term memory and the word-length effect. *Acta Psychologica, 138*(1), 135-142.
- Caproni, V., Levine, D., O'Neal, E., McDonald, R., & Garwood, G. (1977). Seating position, instructor's eye contact availability, and student participation in a small seminar. *Journal of Social Psychology, 103*, 315-316.
- Carpenter, M., Nagell, K., & Tomasello, T. (1998). Social cognition, joint attention and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4, Serial No. 225).
- Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of the mind, and neuroscience* (pp. 159-182). Cambridge, MA : MIT Press.
- Creel, S., Aslin, R., & Tanenhaus, M. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language, 54*, 1-19.
- Dörnyei, Zoltán. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Egbert, M. (1996). Context-sensitivity in conversation analysis: Eye gaze and the German repair initiator 'Bitte.' *Language in Society, 25*(4), 587-612.
- Ervin-Tripp, S.M. (1974). Is second language learning like the first. *TESOL Quarterly, 8*(2), 111-127.
- Friederici, A.D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Science, 9*(10), 481-488.
- Fry, R., & Smith, G. F. (1975). The effects of feedback and eye contact on performance of a digit-coding task. *Journal of Social Psychology, 96*, 145-146.

- Gosling, S., Rentfrow, P., & Swann, W. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504-528.
- Herron, C., York, H., Corrie, C., & Cole, S.P. (2006). A comparison study of the effects of a story-based video instructional package versus a text-based instructional package in the intermediate-level foreign language classroom. *Computer Assisted Language Instruction Consortium Journal, 23*(2), 281-307.
- Hirotsani, M., Stets, M., Striano, T., & Friederici, A. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *Developmental Neuroscience, 20*(6), 600-605.
- Hirotsani, M., Shimada, K., Okazaki, S., Tanabe, H., & Sadato, N. (under revision). *The role of social context in adults word learning.*
- Kendall, K.D., & Schussler, E.E. (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE- Life Sciences Education, 11*(2), 187-199.
- Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience, 6*, 172-183.
- Kinzler, K.D., Corriveau, K.H., & Harris, D.L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106-111.
- Kleinke, C.L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*(1), 78-100.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

- Krashen, S., Long, M., & Scarcella, R.C. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly*, 13(4), 573-582.
- Kroll, J., & Potter, M. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23, 29-66.
- Kuhl, P., Tsao, F-M., & Li, H-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15), 9096-9101.
- Kuhl, P. (2007). Is speech learning 'gated' by the social brain? *Developmental Science*, 10(1), 110-120.
- Langer, E. J. (1978). Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2, pp. 35-58). Hillsdale, NJ: Erlbaum.
- Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*. Boston, MA: Wadsworth Cengage Learning.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
- Long, M. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition* 12(3).251–85.
- Long, M. (2005). Problems with supposed counter-evidence to the Critical Period Hypothesis. *International Review of Applied Linguistics in Language Teaching*, 43(4), 287-317.
- Lubow, R.E., Alek, M., & Arzy, J. (1975). Behavioral decrement following stimulus pre-exposure: Effects of number of preexposures, presence of a second stimulus, and

- interstimulus interval in children and adults. *Journal of Experimental Psychology: Animal Behavior Processes*, 1, 178–188.
- Lubow, R.E., Rifkin, B., & Alek, M. (1976). The context effect: The relationship between stimulus pre-exposure and environmental pre-exposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 2, 38–47.
- Morgan, J.L., Edwards, S., & Wheeldon, L.R. (2014). The relationship between language production and verbal short-term memory: The role of stress grouping. *The Quarterly Journal of Experimental Psychology*, 67(2), 220-246.
- Mundy, P., Block, C., Delgado, C., Pomares, Y., Vaughan Van Hecke, A., & Venezia Parlade, M. (2007). Individual differences and the development of joint attention in infancy. *Child Development*, 78(3), 938-954.
- Mundy, P., Sullivan, L., & Mastergeorge, A. (2009). A parallel and distributed-processing model of joint attention, social cognition, and autism. *Autism Research*, 2, 2-21.
- Mundy, P., & Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Networks*, 23, 985-997.
- Otteson, J. P., & Otteson, C. R. (1980). Effect of teacher's gaze on children's story recall. *Perceptual and Motor Skills*, 50, 35-42.
- Oyama, S. (1976). A sensitive period for the acquisition of nonnative phonological system. *Journal of Psycholinguistic Research*, 5, 266-282.
- Oyama, S. (1978). The sensitive period and comprehension of speech. *Working Papers on Bilingualism*, 16, 1-17.
- Patterson, M.L. (1982). A sequential functional model of nonverbal exchange. *Psychological Review*, 89(3), 231-249.

- Pickering, M., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioural and Brain Sciences*, 27(2), 169-190.
- Pickering, M., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, 4, 203-228.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. *Quarterly Journal of Experimental Psychology*, 55A, 1339-1362.
- Redcay, E., Dodell-Feder, D., Pearrow, M.J., Mavros, P.L., Kleiner, M., Gabrieli, J. & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage*, 50, 1639-1647.
- Redcay, E., Dodell-Feder, D., Mavros, P., Kleiner, M., Pearrow, M.J., Triantafyllou, C.,...Saxe, R. (2013). Atypical brain activation patterns during a face-to-face attention game in adults with autism spectrum disorder. *Human Brain Mapping*, 24(10), 2511-2523.
- Rezaee, A., & Ahmadzadeh, S. (2012). Integrating computer mediated with face-to-face communication and EFL learners' vocabulary improvement. *Journal of Language Teaching and Research*, 3(3), 346-352.
- Riecker, A., Brendel, B., Ziegler, W., Erb, M., & Ackermann, H. (2008). The influence of syllable onset complexity and syllable frequency on speech motor control. *Brain and Language*, 2, 102-113.
- Rockwell, P. (1996). The effects of vocal variation on listener recall. *Journal of Psycholinguistic Research*, 25(3), 431-441.
- Roodenrys, S., & Hulme, C. (2002). Word-frequency and phonological-neighbourhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1019-1034.

- Roseberry, S., Hirsh-Pasek, K., Golinkoff, R.M. (2014), Skype me! Socially contingent interactions help toddlers learn language. *Child Development*, 85(3), 956-970.
- Rothman, J. (2008). Why all counter-evidence to the critical period hypothesis in second language acquisition is not equal or problematic. *Language and Linguistics Compass*, 2(6), 1063–1088.
- Saito, D.N., Tanabe, H.C., Izuma, K., Hayashi, M.J., Morito, Y., Komeda, H.,...Sadato, N. (2010). “Stay tuned”: Inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience*, 4, 1-12.
- Scaife, M., & Bruner, J.S. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265-266.
- Schegloff, E.A., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics*, 22, 3-31.
- Schumann, J. H. (1976). The acculturation model for second language acquisition. In R. C. Gingras, (Ed.), *Second Language Acquisition and Foreign Language Learning*. Washington; Center for Applied Linguistics.
- Service, E. (1998). Effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology*, 51A(2), 283-304.
- Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics*, 43, 269-286.
- Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context . *Journal of Memory and Language*, 48, 103-130.

- Snedeker, J., & Yuan, S. (2008). Effects of prosodic and lexical constraints on parsing in young children (and adults). *Journal of Memory and Language*, 58(2), 574-608.
- Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, 25(7-9), 1234-1264.
- Snow, C., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, 49(4), 1114-1128.
- Tomasello, M., & Ferrar, M.J. (1986). Joint attention and early language. *Child Development*, 57, 1454-1463.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
- VanPatten, B., & Williams, J. (2007). Early theories in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp.17- 35). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Verga, L., & Kotz, S. (2013). How relevant is social interaction in second language? *Frontiers in Human Neuroscience*, 7, 1-7.
- Vitevich, M., & Luce, P. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioural Research Methods, Instruments, & Computers*, 36(3), 481-487.
- Wartenburger, I., Heekeren, H.R., Abutalebi, J., Cappa, S.F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron*, 37(1), 159-170.

Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *Turkish Online Journal of Education Technology*, 8(2), 48-54.

## Appendix A SONA Announcement

Study title: "How adults learn new words in social contexts?"

Description: Canadian English speakers will learn novel words by interacting with a video-taped instructor or her teaching assistant. While doing so, their eye gaze will be recorded with an eye-tracker. After a few learning sessions, their EEG will be recorded. In addition, they will fill out a questionnaire regarding their language background.

Eligibility Requirements: We are looking for right-handed Canadian English speakers between the ages of 18 and 22. Their dominant language must be Canadian English. The participants must have good or corrected vision (with soft contact lenses or glasses) and must not be colorblind.

They must not have any diagnosed speech or hearing impairments, psychiatric or neurological disorders.

Duration: 3 hours (including instructions and debriefing)

Locale: Language and Brain Lab (Paterson Hall 168)

Percentage: 3% towards a psychology course (PSYC 1001, 1002, 2001, or 2002)

Researchers: Dr. Masako Hirotsu (Principal Researcher)

Phone: 613-520-2600 ext. 2805, Email: [masako.hirotsu@carleton.ca](mailto:masako.hirotsu@carleton.ca)

Krista Elliott, Researcher

Phone: 613-520-2600 ext. 3268, Email: [krista.elliott@carleton.ca](mailto:krista.elliott@carleton.ca)

Kelsey Meagher, Researcher

Phone: 613-520-2600 ext. 3268, Email: [kelsey.meagher@carleton.ca](mailto:kelsey.meagher@carleton.ca)

This study has been approved by the Carleton University Ethics Committee for Psychological Research (Study Number #15-XXX).

## Appendix B-1 Pseudowords

Sorted by group. Taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002)

Group 1	Group 2	Group 3	Group 4
Ded	bil	bEz	d@k
gYn	dIt	fRt	fYd
Huk	l^n	gil	lel
Lim	g@k	hIn	pod
p^z	tRz	sYl	tud
b@l	bok	bRl	d^t
Fod	ged	gYt	flk
Hik	kYn	kik	koz
kEt	p^d	l^t	pYt
Ten	sRn	put	s@n

Appendix B-2 List of words removed from original 80

Pseudowords with real-word English or French equivalents: 9

**Bik** (bic, pen company); **daɪl** (dial); **soʊt** (f “saute”); **foʊt** (f “faute”); **dɪk** (sport term); **tɛt** (f “tete”);  
**keɪd** (cade); **sɛn** (hockey team); **huz** (whose, who’s);

Pseudowords that sound like proper nouns: 5

**Sid**; **hæɪ** (Hal); **dæɪn** (Dan); **bɛɪt** (Bert); **lɛɪn** (Len);

Pseudowords that sounded too similar to a real-word: 14

**Fɛt** (sounded like vet); **kaɛk** (sounded like a colloquial term for penis); **tɔɪt** (sounds like turd); **lɔɪd**  
(sounds like lured); **pʌɪ** (sounds like pull); **hʊd** (who’d); **fʊz** (foozeball); **lɪɪ** (colloquial ‘lil’); **pʊz**;  
(colloquial term for bowel movement pluralized); **fæk** (dialectal variation of the obscenity ‘fuck’); **tɪd**  
(tidbit); **bʌɪ** (ball); **ɡɪt** (sounds like a dialectal variation of ‘get’); **sɔɪz** (pluralized ‘sir’ “sirs”);

Pseudowords that sound too similar to other pseudoword choices: 4

**tʌd** (p^d); **hɛt** (kɛt); **ɡɪd** (ɡɪl); **tɛɪd** (tɛɪn); (note: see **kit** below)

Pseudowords eliminated to avoid an abundance of similar onset-nuclei pairs: 1

**Guz** (too many onset g’s);

Pseudowords eliminated to avoid an abundance of similar nuclei-coda pairs: 5

**kaɪd**; **seɪd**; **looɡ**; **doʊd**; **heɪd**

Latin root affix that sounds very familiar: 1

**Ped**

Words removed in order to fix Between Group SD: 1

**Kit** (\*kɪk was originally removed on the basis of an abundance of similar sounding pseudoword choices as it sounded too similar to **hɪk** and **kit**, but was chosen to replace kit in order to account for an uneven SD in Group 3)

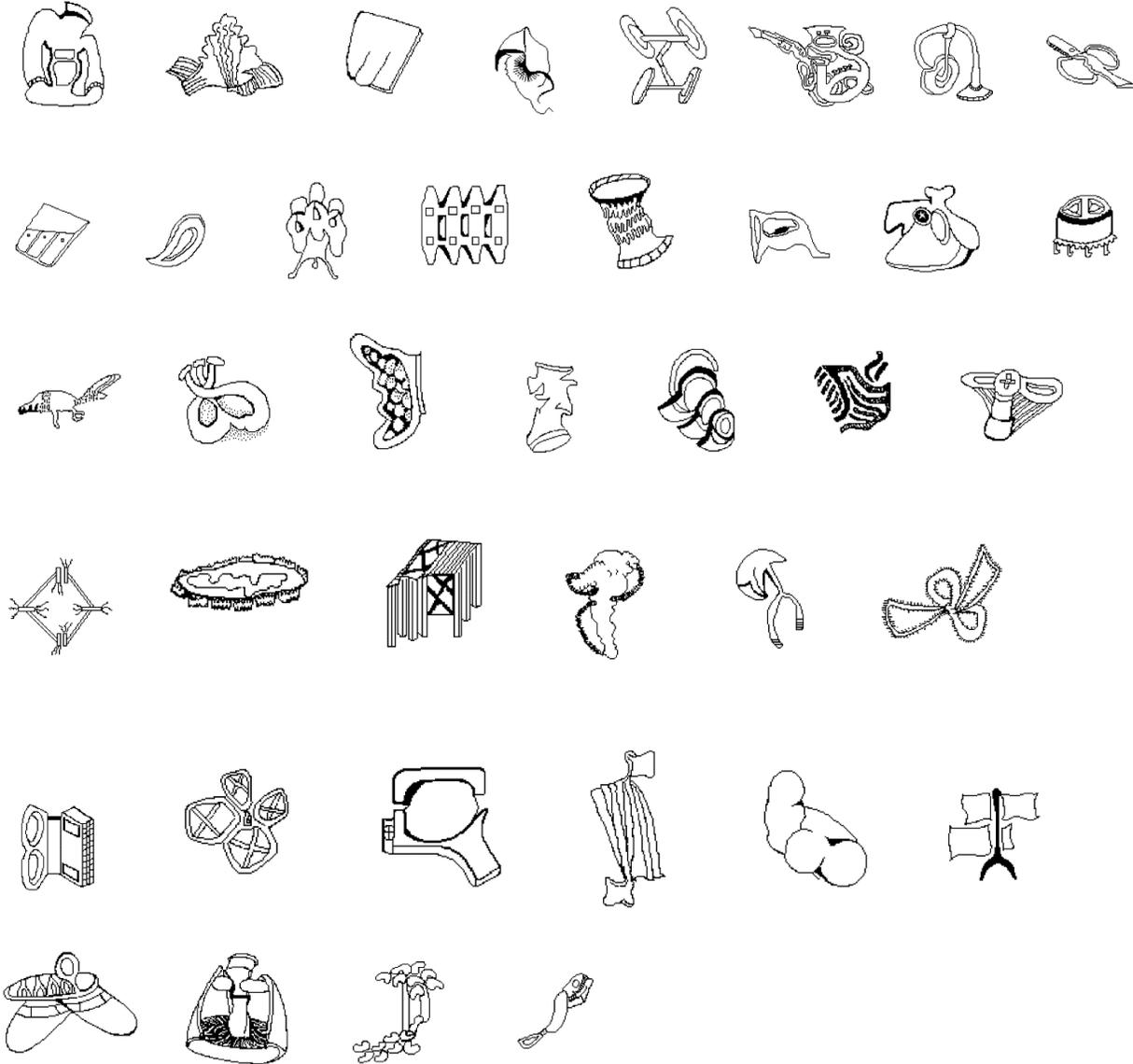
### Appendix B-3 Wordlist deletion justifications

The original list of 80 pseudowords was further refined to 40. In order to avoid a potential for increased recall due to familiarity, the first 9 words to be eliminated were words that were found to have real word equivalents belonging to the same or different word classes (i.e., the noun “huz” could be conflated with the possessive pronoun “whose” or the contracted pronoun or possessive pronoun “who’s”). Given that the experiment was conducted in a highly French-English bilingual environment, any pseudowords that sounded like French real-word equivalents were also eliminated (i.e., “soot” could be confused with the French verb, “saute”). The following eliminations were: 5 which sounded like proper nouns; 14 sounded too similar to real words or common dialectal and colloquial alternates of real words; 4 were too similar to other pseudoword choices and were easily mistaken; 5 were eliminated on the basis of an abundance of similar nuclei-coda pairs; 1 was eliminated to decrease an abundant combination of the onset-nuclei pair “gu-“, 1 was eliminated on the basis of it being a Latin affix that sounded very familiar even if it wasn’t a standalone real word and the last word was eliminated and exchanged in group 3 after all of the others in order to account for a high SD in pitch between the words and between the groups.

Appendix C-1 Unfamiliar target picture objects

From Kroll and Potter (1984)

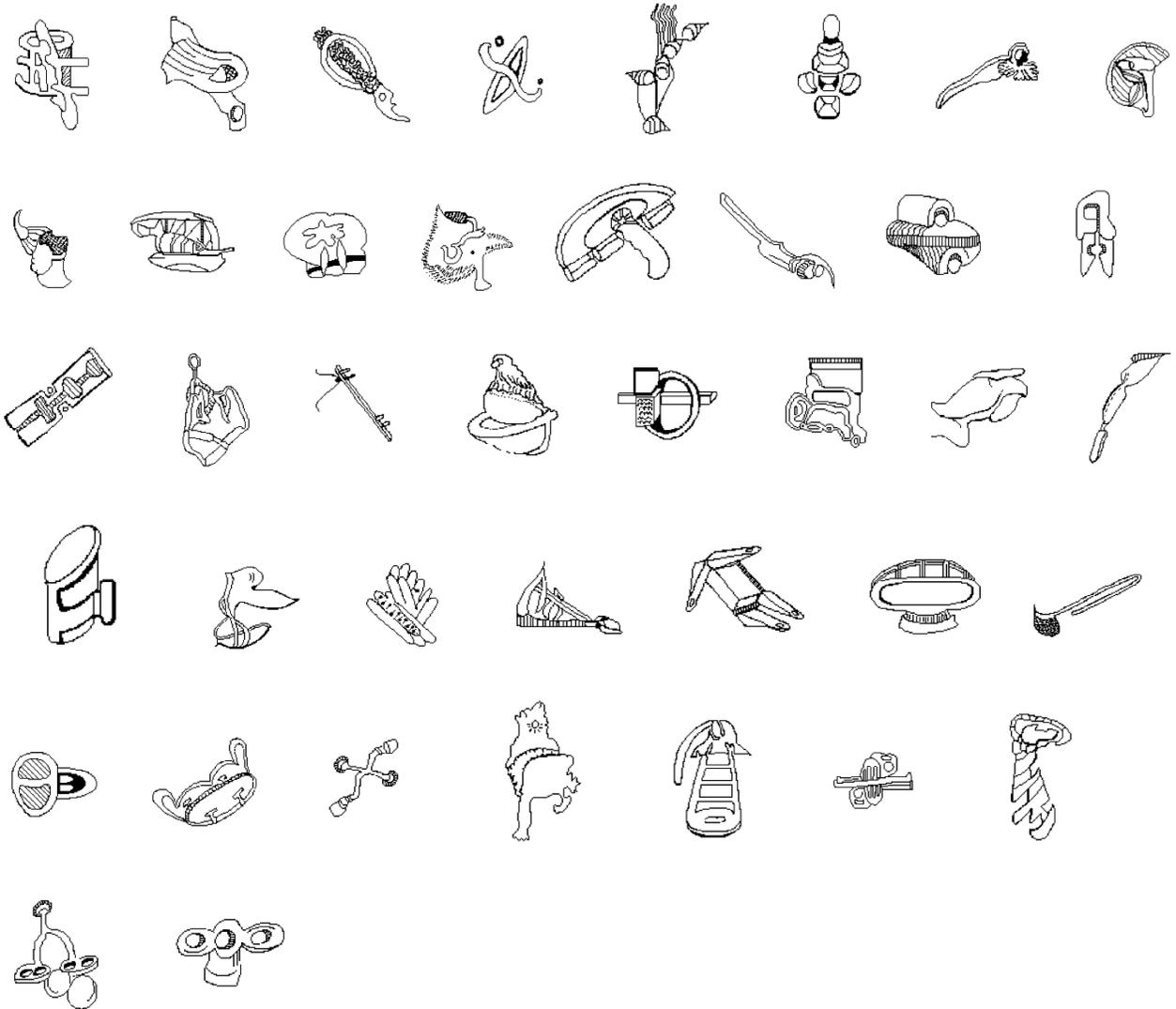
Pictures are listed in order from p01 to p40.



## Appendix C-2 Unfamiliar non-target picture objects

From Kroll and Potter (1984)

Pictures are listed in order from p41 to p80.



Appendix C-3 Unfamiliar target picture object ratings Hirotani et al. (under revision)

Group 1	Object Similarity Rating		Group 2	Object Similarity Rating
p01n.bmp	5.4		p11n.bmp	5.5
p02n.bmp	6.2		p12n.bmp	2.7
p03n.bmp	3.9		p13n.bmp	3.6
p04n.bmp	5.7		p14n.bmp	4.9
p05n.bmp	3.6		p15n.bmp	5.2
p06n.bmp	3.9		p16n.bmp	2.9
p07n.bmp	3.3		p17n.bmp	4.3
p08n.bmp	3.6		p18n.bmp	4.4
p09n.bmp	2.6		p19n.bmp	3.6
p10n.bmp	3.3		p20n.bmp	5.9
Group 3	Object Similarity Rating		Group 4	Object Similarity Rating
p21n.bmp	3.6		p31n.bmp	5.6
p22n.bmp	5.9		p32n.bmp	4.1
p23n.bmp	4		p33n.bmp	5.6
p24n.bmp	3.4		p34n.bmp	3.4
p25n.bmp	5.3		p35n.bmp	5.4
p26n.bmp	2.9		p36n.bmp	3.7
p27n.bmp	5.6		p37n.bmp	4.6

p28n.bmp	4.6		p38n.bmp	5.1
p29n.bmp	4.3		p39n.bmp	4.5
p30n.bmp	3.2		p40n.bmp	5.5

Appendix C-4 Unfamiliar target picture object ratings Hirotani et al. (under revision)

Group 1	Object Similarity Rating		Group 2	Object Similarity Rating
p41.bpm	5.3		p51.bmp	5.4
p42.bmp	5.6		p52.bmp	5.9
p43.bmp	2.8		p53.bmp	4
p44.bmp	5.9		p54.bmp	3.2
p45.bbp	5.2		p55.bmp	4.3
p46.bmp	4.3		p56.bmp	3.1
p47.bmp	4.9		p57.bmp	3
p48.bmp	5.7		p58.bmp	5.3
p49.bmp	5.4		p.59.bmp	3.1
p50.bmp	5.1		p60.bmp	3.4
Group 3	Object Similarity Rating		Group 4	Object Similarity Rating
p61.bmp	3.8		p71.bmp	2.8
p62.bmp	5.7		p72.bmp	3.7
p63.bmp	5.2		p73..bmp	2.9
p64.bmp	4.8		p74.bmp	3.1
p65.bmp	3.3		p75.bmp	5
p66.bmp	5.9		p76.bmp	3.4
p67.bmp	4.4		p77.bmp	4.8

p68.bmp	4.8		p78.bmp	6.1
p69.bmp	4.2		p79.bmp	3.7
p70.bmp	3.1		p80.bmp	3.2

Appendix D-1 Sample visual stimuli sequence of word learning sessions



Appendix D-2 Example audio-visual stimuli used to test learned words in test sessions



## Appendix E Informed Consent Form

The purpose of this informed consent form is to ensure that you understand both the purpose of the study and the nature of your participation. The informed consent must provide you with sufficient information so that you have an opportunity to determine whether you wish to participate in the study. Please read this form carefully and ask any questions you may have before consenting to participate in the study.

Study Title "How adults learn new words in social contexts?"

### Purpose

The purpose of this study is to investigate the human cognitive and information (language) processing system by examining how Canadian English speakers learn new words in different social contexts. This study makes use of an eye-tracker as well as EEG (brain wave) recording and behavioral measures (e.g., questionnaires).

The study is designed to answer some basic questions about the relation between the brain and behavior and should in no way be perceived as therapy, nor will it benefit the participant directly.

### Task Requirements

In this study you will be asked to learn novel words by interacting with a video-taped instructor or her teaching assistant to be presented on a computer screen. (Before the experiment begins, you will be informed who you will be interacting with, "the instructor" or her "teaching assistant".) In some trials you will be asked to look at one of the two unfamiliar objects (the target object) by following the eye gaze of the video-taped instructor or her assistant, and in some other trials you will be asked to shift your eye gaze towards the target object before the instructor or her assistant looks at the same object. There are also trials in which you and the instructor or her assistant look at the target object at the same time or look at different target objects. You will be prompted by the color of the object frames regarding where to look or how to interact with the video-taped instructor or her assistant. (You will be given the detailed instructions as well as a practice session to become familiarized with the procedures before the

experiment starts.) You will be asked to learn about 40 words. The learning sessions with the same words will be repeated several times. During the learning sessions, your eye gaze will be recorded with an eye-tracker. After each learning session, you will be tested for how many words you have mastered. At that time, you will be presented a sound and an object and will be asked to judge whether or not the pairing of the sound and object is consistent with what you have learned during the learning sessions. You will be asked to indicate your answer by using a response pad. The last test session will take place in an EEG room in which a computer monitor and EEG recording equipment are placed, and your brain activity (EEG) will be recorded through an electroencephalogram machine while you are being tested for the learned words. To record your EEG we will use recording electrodes. Recording electrodes are small sensors that act like microphones, picking up and recording electrical activity in your brain. The electrode holders are inserted into an elastic cap that is similar to a nylon swimming cap. In order to help the electrodes accurately measure your brain activity, we will insert a small amount of gel in each holder. Finally, you will be asked to fill out a language background questionnaire.

#### Eligibility Requirements

The study requires right-handed Canadian English speakers between the ages of 18 and 22. Their dominant language must be Canadian English. They must have good or corrected vision (with soft contact lenses or glasses) and must not be colorblind. They must not have any diagnosed speech or hearing impairments, psychiatric or neurological disorders.

#### Duration and Locale

The study will take place in the Language and Brain Laboratory (168 Paterson Hall) and will last approximately 3 hours, including preparation of electrodes, instructions and debriefing.

#### Remuneration

You will receive 3% towards a psychology course (PSYC 1001, 1002, 2001, 2002) for your participation in this study.

### Potential Risk or Discomfort

There are no known risks associated with the behavioral part of this study. For the EEG part, there are minimal risks posed by the procedures being used. The electro-gel may cause some discomfort in your hair, but once the cap is removed after the experiment is completed, the gel will come out quite easily and a well-equipped hair-washing station will be available for your use. Some discomfort may also occur when removing small stickers which hold electrodes to the cheek and forehead. This discomfort would be similar to removing a Band-Aid. Should any discomfort occur, it generally lasts no longer than a few seconds. Breaks will be available to you throughout the experiment. The experiment is divided into blocks, and you may take a short break between the blocks. If additional breaks are needed, you are free to do so.

### Anonymity and Confidentiality

Any personal information you provide will be kept confidential, and any data collected in the study will be coded with random numbers so that the data remain anonymous. Only the research personnel involved in this study will have access to the data.

### Data Security and Management

Any data collected in the study (e.g., filled out questionnaires, electronic files) will be either kept in a locked cabinet or stored in a password protected computer or hard drive in the Language and Brain Laboratory (168 Paterson Hall). All the data will be destroyed within five years after the completion of the study.

### Right to Withdraw

If at any point you wish to terminate your participation you may do so without any penalty and without being asked for a reason.

### Contact information

For questions about this study, please contact:

Dr. Masako Hirotsu, Principal Researcher

(email: masako.hirotsu@carleton.ca, phone: 613-520-2600 ext. 2805)

Krista Elliott, Researcher

(email: krista.elliott@carleton.ca, phone: 613-520-2600 ext. 3268)

Kelsey Meagher, Researcher

(email: kelsey.meagher@carleton.ca, phone: 613-520-2600 ext. 3268)

For any ethical concerns, please contact:

Dr. Shelley Brown, Chair, Carleton University Ethics Committee for Psychological Research (email:

shelley.brown@carleton.ca, phone: 613-520-2600 ext. 1505)

For any other concerns, please contact:

Dr. Jo-Anne LeFevre, Director, Institute of Cognitive Science (email: jo-anne.lefevre@carleton.ca,

phone: 613-520-2600 ext. 2693)

This study has been approved by the Carleton University Ethics Committee for Psychological Research

(Study Number #15-XXX).

Statement of consent

I acknowledge that my participation in this research project is voluntary. I have read the above information and the study has been explained to my satisfaction. I understand that I have the right to withdraw from the study at any time and that my identity will be kept confidential and will not be associated with any of the information or data I provide. My signature indicates that I consent to participate in the experiment.

---

Print your name

---

Researcher's name

---

Signature

---

Date

---

Researcher's signature

---

Date

Appendix F Language background questionnaire (based on Wartenburger et al., 2003)

Please answer the following questions to the best of your knowledge.

1. Date of birth: Day/Month/Year
2. Sex (circle one): Male/Female
3. Please indicate your program and current year of study (e.g. 4th year, Mechanical Engineering)
4. Country of origin: \_\_\_\_\_
5. Country of residence: \_\_\_\_\_

If the country of origin is different from the country of residence, how long (years) have you been in the country of residence?

---

---

6. What was the primary language spoken in your household throughout your childhood?

---

7. Please fill out the following table.

Age	During each of the time periods indicated below, please answer the following questions:		
	Which language(s) were you exposed to?	Approximately how many hours a day were you exposed to the language(s)?	What was the source of your exposure? (home, school, work, tv and radio, job etc.)
0 to 1 years			
1 to 2 years			
2 to 4 years			
4 to 6 years			
6 to 8 years			
8 to 10 years			
10 to 12 years			
12 to 14 years			
14 to 16 years			

16 to 18 years			
18 to 20 years			
20 to 22 years			

## Appendix G Debriefing Form

Study Title "How adults learn new words in social contexts?"

What are we trying to learn with this research?

Starting at a very young age, children map objects in their surroundings to sounds that label them. Mastering this arbitrary, yet predetermined, association between objects and sounds constitutes an important part of language learning. Although particularly prolific in childhood, word learning continues throughout our lives. New words are constantly being coined, and in the age of the global village, many of us are exposed to words in languages other than our own. The present research focuses on adults' learning of new words. Specifically, it investigates whether or not particular social interactions between two adults help them learn new words, and if so, how. Should the role or roles of social context in adult word learning be identified, it would have enormous implications.

What are our hypotheses and predictions?

Social context assists children's learning of new words. In particular, joint attention, which arises when an adult directs and shares a child's attention towards an outside object or event, has been shown to facilitate word learning. While the significance of social contexts in infants' word learning is clear, its role in adults' word learning has yet to be determined. We hypothesize that the positive effect of joint attention on word learning is not restricted to infants and children, but rather extends into adulthood. Specifically, we hypothesize that joint attention strengthens the linkage between objects and their labels during adult word learning. We further hypothesize that it does so by allocating additional resources to the word learning process. In the experiment you have just participated in, you were asked to interact with either the "instructor" or "teaching assistant" during the learning of new words. We are interested in investigating who you interacted with, the instructor or assistant, matters in learning new words.

Why is this important to scientists or the general public?

The practical importance of this research is far reaching. For one, a deeper understanding of vocabulary learning methods can only serve to enhance adult second language learning and teaching. This, in turn, has positive implications for the training of newcomers to Canada, along with professionals and foreign students at our universities and colleges. Equally important, this knowledge can help improve language training for government officials and business persons, contributing to the bilingual nature of Canada and enhancing our ability to communicate with the world at large.

If you would like to know the details about the experiment in which you have just participated, please let the experimenter know. She may be able to answer your questions, or will refer you to the principal researcher, Dr. Masako Hirotsani (613-520-2600 ext. 2805).

Where can I learn more?

You can read some of the papers below to learn more about word or object learning in social contexts: Hirotsani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: event-related potential evidence. *Neuroreport*, 20(6), 600–605.

Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience*, 6, 172.

The following website provides information about the subfield concerning joint attention:

<http://www.anthropoetics.ucla.edu/ap1301/1301vanno.htm>.

What if I have questions later?

For questions about this study, please contact:

Principal researcher: Dr. Masako Hirotsani (email: [masako.hirotsani@carleton.ca](mailto:masako.hirotsani@carleton.ca), phone: 613-520-2600 ext. 2805)

Researchers: Krista Elliott and Kelsey Meagher (613-520-2600 ext. 3268)

For any ethical concerns, please contact:

Dr. Shelley Brown, Chair, Carleton University Ethics Committee for Psychological Research

(email:

shelley.brown@carleton.ca, phone: 613-520-2600 ext. 1505)

For any other concerns please contact:

Dr. Jo-Anne LeFevre, Director, Institute of Cognitive Science (email: jo-

anne.lefevre@carleton.ca, phone: 613-520-2600 ext. 2693)

Thank you for your participation!

This study has been approved by the Carleton University Ethics Committee for Psychological

Research (Study Number #15-XXX).

## Glossary

**Active engagement:** an instance in which an observer directs attention to an object, event, or person, with awareness that another person is also directing attention to the same referent

**Acquisition** (c.f. learning): unforced, spontaneous, implicit linguistic development which occurs through organic interaction with -and direct exposure- to the language.

**Alignment:** a state in which interlocutors mirror each other's speech and movement patterns

**Block:** one full learning or testing list composed of forty trials

**Coda:** the end portion of a syllable

**Condition:** one of the four types of joint attention used to distinguish a trial

**Critical Period Hypothesis:** The proposition that language cannot be acquired after puberty

**Joint Attention:** condition in which two or more partners direct their focus to the same object, event, or person with the shared conscious understanding of participating in the directed focus together

**Initiating Joint Attention (IJA):** joint attention context in which one cues the other to focus on an object or event

**Learning** (c.f. acquisition): linguistic knowledge developed through explicit and conscious effort to identify a language's rules and patterns of usage

**Nucleus:** the center portion of a syllable

**Non—Joint Attention (NJA):** condition void of a joint attention; participants r interlocutors direct their focus to different events, objects, or people, or are unconscious of the other's directed focus

**Non-object:** pictures of non-existent objects, used as stimuli to associate the previously unknown pseudolanguage word-object nominal pairs

**Non-target object** (c.f. target-object): non-object pictures used as the picture participants were not meant to focus on during trials

**Onset:** the beginning of a syllable

**Passive engagement:** an instance in which one observer focuses on an object, event, or person, with little recognition of whether another is directing focus to the same referent

**Pseudoword:** a possible, but non-existent word resembling those of an existent language (for this current experiment, the existent language is English)

**Responding Joint Attention (RJA):** joint attention condition in which one partner cues the other's focus; the partner who does not initiate the cue is the partner who partakes in the responding joint attention

**Saccade:** rapid eye movement in which focus transition from one point to another

**Simultaneous Joint Attention (SJA):** joint attention condition in which both partners attend to the object of focus in synchronicity

**Target object (c.f. non-target object):** non-object pictures used as the pairs matched to the pseudowords

**Trial:** one sequence in which the picture-word pairs are presented