LANGUAGE LEARNING IN CONTEXT: AN INVESTIGATION OF THE PROCESSING AND LEARNING OF NEW LINGUISTIC INFORMATION

BY

ANNA TSIOLA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Linguistics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Kiel Christianson, Chair
Associate Professor Melissa Bowles, Co-Chair
Professor Tania Ionin
Professor Kim McDonough, Concordia University

ABSTRACT

Naturalistic language learning is contextually grounded. When people learn their first (L1) and often their second (L2) language, they do so in various contexts. In this dissertation I examine the effect of various contexts on language development.

Part 1 describes the effects of textual, linguistic context in reading. I employed an eye-tracking and a think-aloud experiment to examine how native and non-native speakers of English process new words presented in full sentences. The results from the mixed-methods approach indicate similar processes of semantic integration for both speaker groups, with the L2 group putting greater intentionality and effort into the task and engaging in deeper processing.

Subsequently, I operationalized context as additional information present in the learning environment, linguistic or visual. In two sets of related studies, I used self-paced reading (Part 3) and eye-tracking (Part 4) to track the learning process of L2 morphosyntax, as well as a series of offline receptive and productive tasks to evaluate learning outcomes. The results suggest a facilitative role for contextual information, both linguistic (L1 translations) and visual (images depicting sentence content). When no additional support was offered, learning was significantly diminished. The multi-method approach allowed me to operationalize ‘learning’ both as a process and as a product and to measure the various nuances of the construct. Results show how reading/reaction times gradually reduce as a result of learning; subsequent receptive and productive tasks reveal high accuracy and confirm that the L2 morphosyntax had been learned.

Taken together, the results of this dissertation projects underscore the importance of context for language learning and show that when we manipulate contextual information, we alter both the learning process and its outcomes.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors, Dr. Kiel Christianson and Dr. Melissa Bowles, who have guided me through this dissertation. I have appreciated their insights and willingness to help and support my efforts. I have been lucky to work under the supervision of professors who are not just leading experts in their fields, but also interesting and fun individuals, who have shown great support and understanding during the difficult moments of the PhD journey.

I would also like to thank Dr. Tania Ionin, who helped me shape my projects and also provided ample support and opportunities for development as the Linguistics Department Director of Graduate Studies. She has been a role model for organization and productivity.

Dr. Kim McDonough has offered me valuable feedback during my preliminary proposal, which helped me shape my arguments and develop my experimental design further. I value her feedback and our discussions immensely.

With all the academic help in the world, I would not have been able to complete this without the emotional support of a large number of people. First, my family, who supported me from abroad through daily calls. Secondly, my UIUC friends, especially Dr. Maria Goldshtein and soon-to-be-Dr. Andrew Armstrong, both of whom have consumed countless coffees and beers with me as we navigated the graduate student life together.

One final big thanks is due to my partner Michael Castle, who, especially at the end, tolerated frantic writing and numerous cycles of “I can’t do it” – “I am almost done”. During the dark times of last-minute edits, his encouraging words and actions helped me keep things in perspective and finish the manuscript.

TABLE OF CONTENTS

INTRODUCTION

Broadly, my research seeks to examine and understand how people process and learn language in context. With a greater focus on second language learning, I study how learners process new linguistic information and use context clues to make meaning. I focus on uninstructed second language acquisition, operationalized as learning that occurs naturally through interaction with the language and the world in a given context. Depending on the situation and the specific task, this could result in intentional or incidental learning for different people. Regardless, the central focus is on learning that occurs outside of the language classroom and not (necessarily) the result of a conscious attempt to learn a new language.

In uninstructed language learning, second language (L2) learners are exposed to input and must figure out the mapping between lexical or grammatical form and meaning/function. To do so, they must attend to cues in the environment in which language is presented, which entails that first, they need to identify which cues are informative. The goal of this dissertation is to examine the online processing of new linguistic information by taking into consideration the context in which linguistic input is presented.

Context is a general term that can include anything ranging from the text surrounding a specific word on a page, to actions and events in the world where the L2 is used. The 'real world' can be an actual physical location in a foreign country, or an online game where players use an L2 gaming platform and interact with other speakers of the language. People, even native or advanced L2 speakers, are continuously faced with new linguistic information in different situations. For example, in higher education, students encounter numerous new words while reading academic texts; they are expected to understand, learn and remember them for subsequent tests. This is true for students who attend an institution in their first language (L1) and for those who study in their L2. In both cases, learning new lexical items happens outside of a language classroom and thus falls under my operationalization of uninstructed language learning.

A different example of uninstructed L2 learning concerns learning that happens as a result of online interactions and gaming, which is in fact more frequent than physical travel to and study in a different country. Researchers have examined such learning from interactionist perspectives (see Peterson (2010) for a meta-analysis) and have largely overlooked how game elements such as visual scenes may facilitate learning. In online and gaming contexts, players engage through trial and error with the game mechanics: for instance, they encounter new lexical items along with their visual depictions and they must make the new lexical form-meaning mapping in order to progress in the game. As a result, they learn a new L2 word. Similar processes can lead to grammar learning as well: in order to decode a simple transitive sentence presented as part of the game, players need to correctly assign participant roles. Despite its significance, the surrounding visual context where L2 is presented has largely escaped researchers' attention. Only recently has research started to look at contextual cues such as interlocutor gestures or gaze, and in very few cases, visual scenes.

Context is thus ubiquitous in language learning, and it can vary greatly depending on the situation. My broad research question concerns how learners use context to learn new linguistic L2 information. This is approached in two ways, by examining both the process and the outcome of learning. In this dissertation, I focus on two different situations to examine word and grammatical structure learning. First, I examine learning new lexical items from sentence reading. Secondly, I examine morphosyntactic learning through integrating multimodal information. Part 1 investigates L1 and L2 word learning from reading; the remaining parts of the dissertation examine grammatical structure learning aided by contextual clues.

In Part 1, I present two experiments which focus on the processing of novel lexical items presented in sentence contexts. These studies simulate higher education situations where readers encounter new words and must learn their meaning from context. The general question is whether this process differs quantitatively and qualitatively for L1 and L2 readers. To answer this question, I conducted an eye-tracking and a think-aloud experiment with both L1 and L2 speaker groups. The eye-tracking experiment served to locate and quantify readers' attention to the new lexical items as well as the informative contextual cues. The think-aloud experiment aimed to quantify readers' processing depth and amount of effort put into understanding the new words. To anticipate the findings, the general conclusion is that L2 speakers process these items in a qualitatively similar way as L1 speakers and attempt to infer the word's meaning from the first available contextual cue, by making a thematic meaning connection between the word and the informative context. One difference between L1 and L2 groups is the level of engagement with the experiment, with L2 speakers engaging in deeper, more active processing in search of meaning. The goal of these experiments was not to compare the efficiency of processing at the end state of attainment (which has mostly been the case with L1-L2 comparison studies), but rather, the process of mapping a new form to a new conceptual representation. This process is shown to be largely the same in the L1 and the L2, once we use novel words that eliminate effects of familiarity, frequency or recency that typically place L2 speakers at a disadvantage.

The remaining parts of the dissertation focus on L2 grammatical structure learning and investigate the online processing of a novel L2 grammatical construction as well as the extent to which it is learned by true beginner participants. I attempted to answer these questions with two sets of related studies, one using self-paced reading and the other eye-tracking methodology. More specifically, I examined thematic role assignment in Modern Greek, a case marked and free word-order language. Morphological case marking is a topic that typically presents great difficulty for learners whose first language (e.g., English) lacks these morphosyntactic elements. The main focus of these studies is again on the surrounding context: I manipulated the modality of the environment in which the sentences were presented (linguistic-only or linguistic + visual) in order to examine how the learning process and outcome were affected by the modality of the information that was available in the learning environment.

Each set of studies (self-paced reading and eye-tracking) included a series of tasks in order to examine both the real-time process and the subsequent outcomes of learning. Tracking reaction times and especially eye-movements during stimuli presentation allowed me to examine how learners allocated their attention to different parts of the input. Attention and noticing are prerequisites for learning (Bowles & Leow, 2005; Ellis et al, 2009; Schmidt, 1995, 1993, 1990) and inability to attend to the informative parts can obstruct L2 morphosyntactic learning. To the best of my knowledge, no study to date has looked at the moment-by-moment processing and attention allocation to parts of the input, especially when the linguistic input is presented in environments of varying modalities of contextual support.

Besides tracking the learning process, I also used a number of tasks to evaluate the learning outcome. These tasks ranged in explicitness levels and were ordered from more implicit to more explicit. 'Learning' is a construct that can be operationalized in many different ways. Indeed, SLA studies find more or less successful 'learning,' depending on the measurement tasks that each study utilizes. In general, receptive and recognition tasks are easier than production tasks, as they pose fewer memory demands. To evaluate learning in the present studies, I used reaction times and eye-tracking during a sentence-picture matching task, which allowed me to compare reaction times/eye movements during and after learning; I also recorded accuracy scores. Two additional tasks involved production of the target structure, as well as explicitly verbalizing the learned morphosyntactic rule. When taken together, the results from all tasks provide converging evidence for improved L2 morphosyntactic learning when the surrounding context offered cues to the meaning of the Greek sentences.

To investigate the effects of context on learning, I have used mixed methods designs, combining online (eye-tracking, self-paced reading, think aloud protocols) with offline measures (vocabulary posttests, sentence-picture matching, sentence production, and qualitative open responses). Triangulating data from different measures allowed for more confidence in the obtained results. It also allowed for the research questions to be approached in different ways and to be answered with richer data. Specifically, in both parts of the dissertation, I have examined ‘learning’ both as a process and as a product, and I have attempted to establish connections between the two.

PART 1: L1 and L2 Novel Word Processing

Incidental vocabulary learning is an important way to build one's lexicon. Both L1 and L2 readers can learn new words from context during reading, especially in academic contexts (e.g., Landauer & Dumais, 1997). Although some research has characterized incidental learning as slow and inefficient (e.g., Laufer, 2003, 2005; Macaro, 2003; Read, 2004), it is particularly common for university students to encounter unknown words which they need to understand and integrate. This is especially true for L2 readers, who also have more explicit practice and have possibly developed strategies to discern a word's meaning.

L1 readers compute meaning incrementally during reading and make connections between words and their context as each word is encountered (e.g., Ehrlich & Rayner, 1981; Ellert & Holler, 2011; Wilson & Garnsey, 2009). L2 word learning studies initially focused on outcomes and on what conditions make incidental word learning more effective; for example, they examined the effects of word repetition on test scores (e.g., Rott, 1999). Only recently has the attention shifted to the cognitive processing underlying online L2 word learning.

In this study, we triangulate data from eye-tracking and think-aloud protocols to examine how L1 and L2 readers process novel words in sentence contexts. We use different eye-tracking measures to compare L1 and L2 reading patterns separately for early measures indexing lexical access and late measures indexing semantic integration (Libben & Titone, 2009). Specifically, we monitored eye movements and collected think-aloud data of L1 and L2 readers as they read sentences containing unknown words to determine whether the two groups' reading behaviors differ as they try to extract word meaning from context, and whether there are differences in depth of processing between the two groups, indicative of different amounts of effort and intentionality.

Novel word learning from reading

A number of eye-tracking studies have examined the online processing of novel words. Eye-tracking can be used to both locate and quantify the attention paid to stimuli during reading (Godfroid et al., 2013). It also allows us to tease apart early and late cognitive processing by analyzing different measures (e.g., first fixation duration/gaze duration are indicative of lexical identification/access, whereas later measures like total time and go-past time index semantic and sentence-level integration (Rayner et al., 2012). This approach allows us to examine whether L1 and L2 processing differs in terms of word recognition, establishing sentence-level meaning, or both.

Following typical familiarity effects in word processing, studies have found a well-established novel word effect, with novel or low-familiarity words taking longer to process than familiar controls (e.g., Chaffin et al., 2001; Pellicer-Sánchez, 2006; William & Morris, 2004). Inflated reading times on unknown words changed through exposure: fixation times decreased in subsequent exposures to novel words (or, similarly, as word familiarity increased). For example, Pellicer-Sánchez (2016) and Elgort et al. (2018) found that after eight exposures, L2 readers showed reading times of the novel/unfamiliar items similar to familiar words. Godfroid et al. (2018) further showed that this decrease is not linear, but rather s-shaped for both speaker groups.

Another important question concerns the relationship between eye movement data (reading times) and vocabulary gains. Some studies have found that more attention paid to novel stimuli results in higher recognition or recall rates. Studies with L1 (Chaffin et al., 2001; Williams & Morris, 2004) and with L2 (Godfroid et al., 2013) populations found that increases in later fixation time measures (second-pass reading time in Williams & Morris (2004) and total time in Godfroid et al. (2013) and Chaffin et al. (2001)) helped with word learning: the more attention paid to a new word, the better it was encoded in memory. Brusnighan and Folk (2012) in an L1 self-paced reading study similarly found that when sentences were read more slowly, subsequent word recognition was more likely. In Godfroid et al. (2013), novel words that had been fixated longer during reading were more likely to be recognized in a later form recognition test. This finding was replicated and extended in Mohamed (2018), who found longer reading times to result in better meaning recognition and recall in addition to better form recognition. Nevertheless, Mohamed (2018) found that repeated exposure supported form recognition but was not as significant for meaning recall and recognition, a finding echoed by Elgort et al. (2018). These later findings complicate the picture and point to individual variation in how readers process the novel items; Godfroid et al. (2018) found repetition as well as reading time effects on word retention, which they interpret as indicating that both quantity and quality of processing is important. In sum, despite early evidence for a positive relationship between reading times and subsequent recall, some studies have not found evidence for this relationship (e.g., Elgort et al., 2018); one explanation is that it is not merely amount of time spent with a stimulus, but rather what a reader choses to do with it that matters for meaning recall. This is especially relevant in unintentional learning conditions, where some readers may choose to treat the task as a learning opportunity, and others may not (Barcroft, 2015).

One study that compared L1 and L2 novel word processing was Pellicer-Sánchez (2016), who examined incidental vocabulary learning for L1 and L2 speakers using eye-tracking and three vocabulary posttests, each targeting different levels of lexical knowledge. Participants read a story containing six novel items, each presented eight times in the story context. There was a correlation between longer total reading times and better subsequent recall for both L1 and L2, and both groups performed similarly on the vocabulary posttests. There was a repetition effect; its magnitude did not differ for L1 and L2 participants, but its rate did: L1 participants needed fewer exposures than L2 for their novel word reading times to reach familiar item levels. The author argued that “integration of lexical and semantic information might happen earlier [after fewer exposures] for L1 readers than for L2 learners” (p. 118). One possible explanation is that L2 readers may have not noticed/attended to the informative context, which would help them establish the word's meaning, in the same way as L1 participants did. Another possibility is that the L2 group more actively tried to establish a semantic representation and learn the item, which would translate into longer reading times.

Finally, vocabulary retention from incidental learning has been found to be differentially successful, arguably as a function of study design. Studies that employed few novel words and many repetitions found higher retention rates (e.g., Pellicer-Sánchez, 2016); in contrast, one or a few exposures to a large number of novel words resulted in lower accuracy (e.g., Godfroid et al, 2013; Godfroid et al., 2018). This is especially true for tasks targeting meaning recall (Godfroid & Schmidtke, 2013); scores are generally higher for simple form recognition but low for form-meaning mapping. For example, in Mohamed (2018), accuracy was the highest for form recognition (42%), followed by 30% for meaning recognition, and 13% for meaning recall. These results are echoed by Godfroid et al. (2018), who found 30% accuracy for form and meaning recognition and 13% for meaning recall.

Contextual cues for word learning

In order to make a form-meaning connection, learners must pay attention to the word and its surrounding context. The studies described above offer valuable information about how fixation times on a lexical item can indicate changing levels of familiarity/knowledge. However, conclusions about readers using contexts to extract word meaning are indirect, as most studies have not analyzed reading times on contextual areas.

One exception from the L1 literature is Chaffin et al. (2001), who investigated the processing of novel items by L1 English speakers. The results showed that when the target item was presented within an informative context, L1 speakers learned the item on their first pass through the text, and they did not need to regress and reread it when they encountered a definition of the item in a subsequent sentence. However, if the context was not informative as to the novel word's meaning, readers regressed and reread it when they encountered definitional information in a subsequent sentence. In other words, readers established the meaning of the novel item when they first encountered definitive information, and this was inferred from longer fixations and more regressions out of the informative area. Learning was inferred from the fact that the novel items presented in informative contexts were read faster, equally fast as the familiar control items, as compared to novel items presented in non-informative contexts.

Similarly, Williams and Morris (2004) proposed that inflated second-pass reading times on novel words were indicative of an effort to connect contextual information to the novel item. The authors concluded that L1 readers were able to allocate their attention to the regions that were informative to the novel word's meaning. The question remains whether L2 speakers are also able to allocate their attentional and processing resources to the informative areas of text, and whether they do so with the same level of efficiency.

Adding to the L1 literature, Brusnighan and Folk (2012) tested novel compound word learning in informative and neutral contexts and found that readers showed processing advantages for novel items encountered in informative contexts, suggesting that they were able to use the context as a meaning cue and to establish a meaning for the novel word during reading. They analyzed whole sentence reading times between familiar-novel informative context conditions, but not at the contextual areas due to text length differences between conditions. This promising finding underlines the need to extend this design and analyze contextual cues as a region of interest. Finally, Lowell and Binder (2020) used self-paced reading and found that novel words presented in consistent contexts were read faster and retained better, adding further evidence that L1 readers use context to learn new words incrementally during reading. The question still remains how comparable L1 and L2 processing of context cues is, in terms of both efficiently locating the relevant cues, and how deeply they are processed.

In the L2 domain, studies have analyzed eye-tracking measures on target words, either looking at reading time decrease from repeated exposures, or comparing reading times between familiar and novel words. Even though studies have not analyzed reading time on contexts, there is some evidence suggesting that L2 readers process informative cues to establish the meaning of a new word. For example, Godfroid et al. (2013) used appositive cues (synonyms) placed either before or after the pseudo-word and found that the cues following the pseudo-words were fixated on longer, which the authors interpreted as reflecting processes of semantic integration at the sentence level (meaning-making). Nevertheless, Elgort et al. (2018) found that word integration into texts was not optimal even after multiple exposures to unfamiliar L2 words. This conclusion came from analyzing later eye-tracking measures on the target words, and it suggests possible difficulty with sentence-level semantic integration.

The above findings point to the need to include 'context' as an experimental manipulation in study designs and to examine how efficiently readers make use of the specific contextual clues offered in text. We should note that there are important methodological differences between studies examining L1 novel word processing and corresponding L2 word learning. Specifically, L1 studies have typically used traditional psycholinguistic designs with novel words embedded in sentence frames; in contrast, L2 word learning studies have opted for ecological validity and used paragraph or whole text reading, with some embedded novel words. Different tasks can induce different strategies in reading, which makes direct comparisons across studies difficult. Nevertheless, both types of studies suggest that L1 and L2 readers are able to use contextual cues to infer word meaning. Yet the question remains whether they show similar semantic integration processes, as no study has, to our knowledge, compared contextual cue processing.

Speed vs. depth of processing

Readers have longer reading times when they experience processing difficulty, such as when reading more difficult texts or when they are less skilled (Rayner, 2009). L2 populations are generally slower in reading than L1 (e.g., Roberts et al., 2008). This is reasonable to expect for less skilled readers, that is individuals who are less proficient in the L2, as proficiency modulates reading and processing speed in general (e.g., Sagarra & Herschenson, 2010). Secondly, task type, and the depth of processing it induces, also modulates L2 behavior; tasks that are designed to draw attention to a specific structure result in deeper processing and more native-like behavior (e.g., Jackson & Bobb, 2009; Lim & Christianson, 2015). Task type also influences depth of processing in L1 speakers: tasks that do not require deep processing can result in incomplete processing that is "good enough" to achieve the impression of comprehension (e.g., Christianson et al., 2001; Ferreira, et al., 2002; Ferreira & Patson, 2007).

In the case of novel words, both L1 and L2 readers slow down from their respective baselines. Pellicer-Sánchez (2016) indicated that L1 readers learned novel words faster than L2; that is, they needed fewer exposures for their reading times to resemble reading times on familiar words. There are two explanations for this. L1 groups may be able to more efficiently locate and use the contextual cues that offer information to the word's meaning and thus establish some sort of mental representation for the word earlier than L2 speakers. Alternatively, slower processing could also be indicative of deeper processing (a similar argument is made in Pellicer-Sánchez, 2020).

Depth of processing, operationalized as amount of cognitive effort and level of analysis and elaboration, has recently received attention in SLA theorizing and modeling the learning process (Leow, 2015). Due to their language learning experiences, it is possible that L2 speakers approach an incidental learning task as an intentional learning task, and consequently engage in deeper processing (see Barcroft (2015) for a discussion of researcher-induced incidental learning conditions and learner-internal processing). In Godfroid et al. (2018), there was in fact a difference between L1 and L2 speakers' perception of the task, which affected test performance. Specifically, a subset of L2 speakers suspected the unannounced test and treated the task as intentional learning; only one L1 speaker (bilingual) had the same suspicion. This confirms our hypothesis that L1 and L2 groups may approach experimental tasks differently, and namely that L2 groups may be more likely to treat them as learning tasks. This consideration becomes especially important if we consider that putting effort into deriving word meanings can lead to increased meaning learning (Elgort et al., 2018, pp. 347).

Inflated reading times on novel words can be the result of both processing difficulty and depth: unfamiliar words are inherently more difficult to process, but a reader's fixation times will also be influenced by how much effort they actively put into understanding these words and establishing their meaning. This can be thought of as a continuum from skipping unknown words to consciously searching text for their meaning. Godfroid and Schimdtke (2013) analyzed stimulated recall protocols and their relationship to vocabulary scores and found evidence that "language learners' subjective engagement with novel material (e.g., by invoking sound patterns, meaning associations, or personal experiences) may be especially facilitative of vocabulary learning" (pp. 200). Similarly, Mohamed (2018) also found individual attention to novel words to be a stronger predictor of retention than number of encounters with that word, which could also help explain variance in word learning (Godfroid et al., 2013). Beyond frequency of exposure, it is what participants chose to do with a word that better predicted learning and retention. Further evidence comes from the think-aloud methodology: Leow et al.(2008) concluded that what influenced comprehension and intake may be *the way* learners process a form. Since processing is participant-internal, it need not necessarily reflect task instructions, making research on depth of processing even more relevant and highlighting the need for mixed-methods approaches including think-aloud protocols.

Craik and Lockhart (1972) and Craik and Tulvin's (1975) seminal work on processing depth stresses the need to disentangle *more* from *deeper* processing. Quantitatively more processing (i.e. longer fixation durations) does not necessarily suggest deeper processing. Only a mixed-methods approach can disentangle these nuances. Another important point is task interpretation: word learning is better predicted by the qualitative nature of operations learners perform on the stimuli, rather than time spent processing them (Craik & Tulvin, 1975). Therefore, even though studies may create conditions for incidental learning, it is beyond researchers' control what operations learners choose to perform.

Data triangulation with eye-tracking and think-alouds

In an effort to disentangle amount and depth of processing, we consider the following from Craik and Tulvin (1975): processing time cannot always be taken as an absolute indicator of depth. Highly familiar stimuli can be rapidly analyzed to a complex meaningful level, but with less familiar stimuli, deeper processing is assumed to require more time.

Eye-tracking is better suited to measure cognitive effort and descriptively examine how much time participants spent in a given area; however, it cannot fully answer why. Typically, we assume the reason is processing difficulty, but as discussed above, task engagement and processing depth could also result in higher reading times. Think aloud protocols can examine processing depth and compare differences between speaker groups. A number of studies have used concurrent verbal reports to examine depth of processing in instructed SLA and have looked at how different instructional conditions affect processing depth (e.g., Adrada Rafael, 2017; Leow et al., 2008; Morgan-Short et al., 2012).

A mixed-methods between participants approach is well suited to test group differences in reading times, as well as whether these are the result purely of processing difficulty, or are also influenced by processing depth (cf. Pellicer-Sanchez, 2020; Godfroid, 2019). Godfroid and Schmidtke (2013) used a mixed-methods approach with eye-tracking and stimulated recall to examine the contributions of attention and awareness respectively, to word learning in a within-participants design. They argue that a "mixed methods approach will afford a richer perspective on L2 learners' cognitive processes than any single method could"; we agree and extend this line of reasoning to address L1 and L2 processing similarities and differences. Even though in practice attention and depth of processing co-occur, we operationalize them independently: we use eye-tracking to measure attention and concurrent think aloud protocols for processing depth. With eye-tracking data alone, it is not possible to determine to what extend slower reading is caused by difficulty and/or processing depth (a parallel argument is made in Godfroid & Schmidtke, 2013, regarding their operationalization of attention and awareness). Bowles (2019) similarly argues that eye-tracking relies on assumptions about what different measures index, whereas verbal reports allow a direct observation of thought processes; the two techniques can capture cognitive processes at different levels of awareness.

The present study triangulates multiple data sources to examine how L1 and L2 readers approach incidental word learning in sentence reading. We used sentence frames with embedded novel words and crucially, we manipulated context as an area of interest and analyzed reading times on this area- i.e., the place where subjects are expected to establish the meaning of the target word. We expected to replicate the novel word effect. We also predicted overall slower reading times for the L2 participants. We suspected that L2 readers might attend more to novel words, due to the fact that they likely encounter novel words relatively often as L2 speakers studying abroad. Yet we remained agnostic as to whether or how quantitative measures of reading (e.g., fixation durations) might be related to qualitative differences between the two groups that might emerge from analyses of the think-aloud data. As such, we address the following open questions:

1. Do L1 and L2 groups use context in the same way to establish the novel items' meaning?

2. Do longer reading times predict better word learning for L1 and L2?

3. How do reading times on the novel word change from first to second encounter?

4. Do L1 and L2 groups process novel items to the same depth?

5. Are L1 and L2 groups equally successful in word learning?

The study induced incidental learning, ‚operationalized as learning that occurs as a by-product of engaging in another activity, e.g., reading for meaning, and it is not the learners' intention to learn these words while they engage in the task (Hulstijn, 2003).

Experiment 1

Experiment 1 used eye-tracking to investigate how L1 and L2 participants allocate their attention to make meaning of sentences containing unfamiliar lexical items.

Method

Participants

There were two groups of participants: the first group consisted of 42 L1 American English speakers (26 F), ages 18-35, *M*=20.3. The second group had 41 L1 Chinese- L2 English speakers (22 F), ages 18-37, *M*=21.2. The L2 participants had intermediate to advanced proficiency in English, established with a multiple-choice cloze English proficiency task (adapted from O'Neill, Cornelius & Washburn (1981), Appendix A); in addition, they were studying in a US university. Each correct answer received a score of 1 and each incorrect a 0. The mean score was *M*=35.3 out of a maximum of 40 points with an *SD*=2.08, and range=31-38.

All participants had normal or corrected to normal vision, were university students recruited on campus and received either course credit or payment for their participation.

Materials

Each stimulus consisted of two sentences and belonged to one of four conditions created by crossing type of item and type of context (see Table 1). The 40 items were novel or familiar words (e.g., *barhep* and *guitar*) depicting a concrete noun/object. Novel and familiar words were matched pairwise for number of letters, syllables, and phonemes. Most of the novel words came from the ARC database (Rastle et al., 2002); the longer ones were constructed either by combining two shorter ARC pseudo-words, or by the researcher, followed by a Google search on each that returned no results. Sentence context was either informative or uninformative as to the meaning of the novel word. Informativeness refers to Sentence 1, in which the word first appeared. All stimuli had a second sentence in which the target item occurred again followed by the word's definitional hypernym (e.g., "instrument" for *barhep/guitar*). For the uninformative condition, the hypernym in the second sentence was the first cue to the novel word's meaning. Forty token sets were distributed over four lists in a Latin square so that each participant saw only one item from each set (all stimuli appear in Appendix B). The experiment also included 80 fillers, resembling experimental items but containing only real words, half of which were familiar (e.g., *cat*) and half less familiar (e.g., *civet*). The fillers were the same across lists. Comprehension questions followed ¾ of the fillers, focused on simple sentence comprehension rather than knowledge of items, and did not draw any attention to the target items. The presentation of the items within each list was pseudo-randomized.

Table 1.1. Stimuli token set example

	Informative Context	**Uninformative Context**
Novel Item	He picked up the **barhep** from the floor to **play some music** and immediately frowned. He had realized that the **barhep** is a difficult **instrument** to learn by yourself.	He picked up the **barhep** from the floor to **give to his friend** and immediately frowned. He had realized that the **barhep** is a difficult **instrument** to learn by yourself.
Familiar Item	He picked up the **guitar** from the floor to **play some music** and immediately frowned. He had realized that the **guitar** is a difficult **instrument** to learn by yourself.	He picked up the **guitar** from the floor to **give to his friend** and immediately frowned. He had realized that the **guitar** is a difficult **instrument** to learn by yourself.

The stimuli were normed on Amazon's MTurk with twenty-four L1 English participants who did not participate in the main experiment and were paid US $1.50. In the norming study, the first sentence of each stimulus in the Novel word-Informative context condition was presented, and MTurk workers were asked to provide the hypernym of the novel word. For instance, they saw the following text and they were asked to fill in the blank:

He picked up the barhep from the floor to play some music and immediately frowned.

The barhep is a type of _______.

Each response was given a score of 1 if it was acceptable and 0 if it was not. Exact matches (the word provided was the hypernym the researchers used, e.g. "instrument" in the example above) as well as synonyms or words belonging to the same semantic category (hypernyms, hyponyms, etc.) were coded as acceptable. Each item could have a maximum score of 24 if all participants guessed its meaning correctly. Results indicated that the contexts provided enough information to derive the meanings of the novel words, as the mean score across items was *M*=22.1, *SD*=2.4. The familiar control items were also normed for familiarity using a 5-point familiarity scale, with 1 indicating that an item was highly familiar and 5 highly unfamiliar. The mean familiarity score across the familiar control items was *M*=1.3, *SD*= 0.4. Hence, participants were indeed familiar with these items.

Apparatus

For the eye-tracking portion of the study, eye movements were recorded via an SR Research Eyelink 1000 eye tracker (spatial resolution of 0.01 degrees) sampling at 1000 Hz. Sentences were presented in 16-pt Courier New monospace font. Subjects were seated 60 cm away from a monitor with a display resolution of 1800 x 900, so that approximately 3 characters subtended 1 degree of visual angle. Head movements were minimized with a chin and forehead rest. Although viewing was binocular, eye movements were recorded from one eye (typically the right eye). The experiment was controlled with SR Research Experiment Builder™ software, and participants responded using a standard game controller.

Procedure

Participants first completed the eye-tracking portion of the experiment. They were instructed to read the sentences normally for meaning; they were told that the sentences include some very infrequent English words they may not have seen before but that they should just read normally for general meaning[1], in order to answer comprehension questions accurately. They

[1] To avoid the possibility that L1, but not L2, speakers suspect these words are novel, the instructions explained that readers might not know some words because they are very infrequent, thus minimizing the development of different awareness for the two groups.

completed five practice trials followed by the experimental trials. Each screen contained one stimulus consisting of two sentences on two separate lines; participants were instructed to push a button on the controller after they had finished reading the sentences on each screen to move to a comprehension question or a new item.

After the eye-tracking portion, participants completed a distractor Operation-Span task, followed by the vocabulary post-test, and a language history questionnaire. In the unannounced vocabulary test, participants were asked to match the 20 novel words from the reading task to their definitional hypernyms. Getting even one item correct by chance was highly unlikely.

Analysis and Discussion

To ensure participants were paying attention to the task, we checked accuracy scores on the filler comprehension questions. Participants had over 85% accuracy (M=93.35, SD=4.3, range: 75-100); only two L2 participants scored 75% and 83.3%, but these scores were deemed acceptable and were not excluded from analysis.

To capture word processing time, we report first-fixation duration (FFD; the duration of the first fixation on a word independent of the number of fixations on the word), gaze duration (GD; the sum of all fixations on a word prior to moving in either direction to another word), and total fixation time (TT; the sum of all fixations, including fixations resulting from regressions, on a word). To capture discourse processing (context), where the unit of analysis is larger than a single word, we report the total reading time (TRT; the sum of all fixations in the region) and go-past time (GPT; also known as regression path duration; the sum of all fixations on a region from first entering the region until exiting in the forward direction, including fixations on prior regions resulting from regressions originating in the current region).

Continuous reading measures were analyzed with linear mixed-effects models, and binary measures of regressions in/out of the interest area with generalized linear mixed models (with a logit link), using the lme4 package (Bates et al., 2015) in the R environment (R Development Core Team, 2013). Models were built for each measure in each region of interest. The dependent variable was the measure in question (FFD, GD, etc.) and predictors were type of word (novel or familiar), type of context (informative or uninformative), group, as well as the 3-way interaction. We used treatment contrasts (baseline: familiar, informative, L1). Subjects and items were included as random intercepts. Only statistical results that were significant are reported here in detail.

Eye-tracking Measures

The descriptive statistics of the eye-tracking measures reported in this section are found in Table 1.2 and Table 1.3 for the L1 and L2 groups, respectively. The reported coefficients are based on a log-transformed dependent variable in order to satisfy the assumption of normally distributed residuals (which was violated when the model was run on raw data) (James et al., 2015). In all models, there was a main effect of group, with L2 having consistently longer reading times than L1; this effect will not be repeated in the results below due to space limitations.

Table 1.2. Descriptive statistics for the eye-tracking measures of the L1 group

	Total Time	**First Fix Duration**	**Gaze Dur**	**Regression Path Dur**
WORD Fam-Inf	365 (278)	242 (77)	306 (141)	379 (264)
WORD Fam-Un	380 (257)	233 (83)	293 (139)	377 (234)
WORD Novel-Inf	667 (500)	283 (144)	430 (280)	593 (521)

Table 1.2 (cont.)

WORD Novel-Un	654 (436)	283 (140)	418 (256)	576 (395)
CONTEXT Fam-Inf	1020 (592)			977 (643)
CONTEXT Fam-Un	954 (580)			847 (444)
CONTEXT Novel-Inf	1065 (733)			1014 (709)
CONTEXT Novel-Un	962 (569)			901 (535)
WORD2 Fam-Inf	261 (194)	216 (69)	252 (115)	324 (249)
WORD2 Fam-Un	269 (220)	223 (71)	259 (147)	332 (211)
WORD2 Novel-Inf	332 (222)	240 (104)	295 (162)	394 (333)
WORD2 Novel-Un	353 (275)	241 (91)	299 (187)	397 (366)
HYPERNYM Fam-Inf	249 (222)	231 (73)	284 (122)	331 (222)
HYPERNYM Fam-Un	263 (198)	236 (73)	281 (123)	332 (270)
HYPERNYM Novel-Inf	264 (210)	235 (82)	290 (149)	323 (194)
HYPERNYM Novel-Un	300 (215)	238 (77)	298 (136)	361 (252)

The statistics reported are Mean (SD) and they are provided per interest area, per condition for the L1-English group. All values are rounded to the millisecond. The interest areas in question are WORD, CONTEXT, WORD2 and HYPERNYM. The conditions are FamINF (Familiar word, Informative context), FamUn (Familiar word, Uninformative context), NovelInf (Novel word, Informative context) and NovelUn (Novel word, Uninformative context). These are based on a sample size N=42.

Table 1.3. Descriptive statistics for the eye-tracking measures of the L2 group

	Total Time	First Fix Duration	Gaze Dur	Regression Path Dur
WORD Fam-Inf	724 (525)	296 (127)	481 (275)	572 (354)
WORD Fam-Un	764 (542)	290 (133)	479 (283)	576 (349)
WORD Novel-Inf	1027 (691)	317 (176)	568 (337)	725 (491)
WORD Novel-Un	1054 (691)	317 (171)	589 (384)	735 (465)
CONTEXT Fam-Inf	1717 (1115)		1159 (645)	1387 (706)
CONTEXT Fam-Un	1506 (910)		1080 (499)	1258 (674)
CONTEXT Novel-Inf	1864 (1252)		1163 (623)	1437 (781)
CONTEXT Novel-Un	1526 (929)		1052 (502)	1211 (616)
WORD2 Fam-Inf	425 (267)	259 (86)	360 (167)	418 (257)
WORD2 Fam-Un	461 (329)	268 (92)	350 (176)	425 (275)
WORD2 Novel-Inf	595 (447)	279 (111)	434 (224)	542 (537)
WORD2 Novel-Un	559 (389)	283 (103)	414 (206)	525 (466)
HYPERNYM Fam-Inf	399 (303)	279 (105)	387 (205)	441 (277)
HYPERNYM Fam-Un	442 (416)	275 (100)	402 (222)	450 (326)
HYPERNYM Novel-Inf	440 (397)	279 (91)	392 (221)	454 (343)
HYPERNYM Novel-Un	467 (341)	288 (107)	424 (218)	472 (294)

The statistics reported are Mean (SD) and they are provided per interest area, per condition for the L2-English group. All values are rounded to the millisecond. The interest areas in question are WORD, CONTEXT, WORD2 and HYPERNYM. The conditions are FamINF (Familiar word,

Table 1.3 cont.

Informative context), FamUn (Familiar word, Uninformative context), NovelInf (Novel word, Informative context) and NovelUn (Novel word, Uninformative context). These are based on a sample N=41.

Target words.

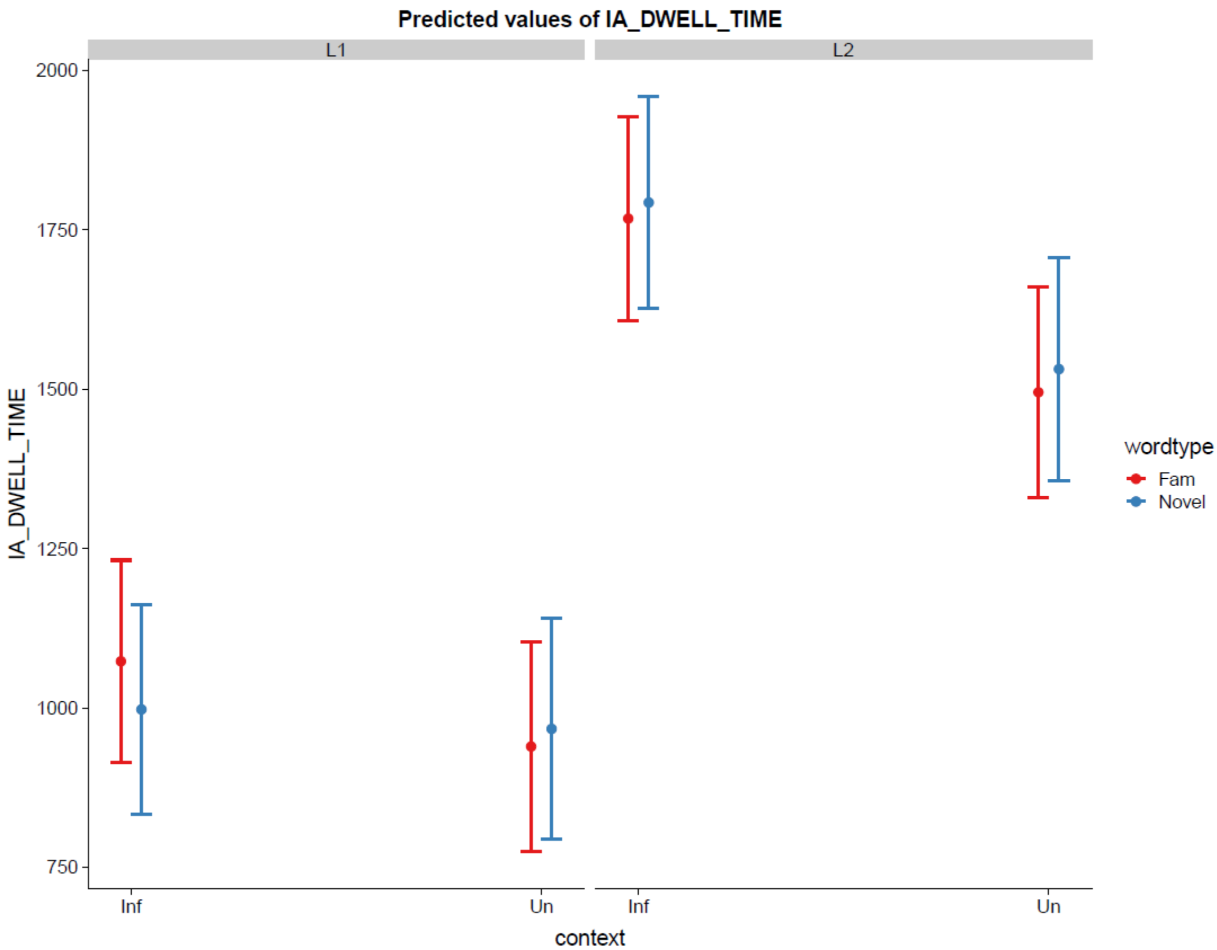
Seeking to replicate the novel word effect, we report on eye-tracking measures taken on the target words (novel or familiar). We indeed found that novel words had longer reading times than familiar ones, for both speaker groups, for all eye-tracking measures (FFD, GD, TT, GPT). This effect was observed in both the first and the second encounter with the target items. The delay in reading times for novel words reflects the search cost associated with a non-existent lexical entry (FFD, GD) and the difficulty integrating the item semantically (TT, GPT). Both groups had similar difficulties with semantic integration processes. The second occurrence of the word did not show the expected interaction between word type and context type, which Chaffin et al. (2001) found on the hypernym. This difference could be attributed to differences in the study stimuli: the present study included a repetition of the novel word, and it seems that a single exposure to a novel word + context was not enough to establish a lexical entry that could be accessed as fast as a more familiar item. Pellicer-Sanchez (2016) suggested that the rate of change in FFD was faster for the L1 group in that study, arguing that L1 participants learned the novel word after 1-2 encounters whereas the L2 group required 3-4. Although the present study included only two repetitions of the novel word, up to these two repetitions the groups patterned the same: neither L1 nor L2 established a lexical entry robust enough to be accessed equally fast as the familiar controls. Therefore, the two groups had similar difficulty with lexical processing when matched on experience (i.e., with novel words).

To answer whether L1 and L2 participants use contextual cues in the same way to infer the meaning of novel words (RQ1), we report on eye-tracking measures on the interest areas Context and Hypernym, as well as probability of regressions into Context, and out of the target Words and the Hypernym.

Context

There was a main effect of context informativeness on the TRT spent on the interest area Context, with informative contexts having longer total reading times than uninformative ones (β=0.04, t=3.01, $p < .001$); this was qualified by 2-way interactions illustrated in Figure 1. Pairwise comparisons revealed that for the L1 group, informative contexts had longer TRT than uninformative ones when the target word was familiar (β=0.15, $z.$ =4, $p < .001$), but this effect was not observed when the word was novel. The L2 group also had longer TRT on informative contexts (β=0.16, z = 4.25, $p < .001$), both with familiar and novel words (β=0.12, z = 3.29, $p <$ .05).

Figure 1.1. Total time spent on interest area "Context".



The fact that both groups had inflated TRT on informative contexts compared to uninformative ones for familiar words suggests that both groups were engaged in meaning-making and connecting the informative context to the preceding word. This process of semantic integration was similar for novel and familiar words for the L2 group. One possibility is that the

L2 readers incurred a cost in connecting the word and contextual information, regardless of whether or not the lexical item was novel. When the context was uninformative, then there is no such thematic connection to be made, resulting in reduced processing times. The L1 group's behavior is especially puzzling: they spent longer reading informative contexts only for familiar words, not novel. Following the reasoning for the L2 pattern above, it appears that the L1 group was less concerned with making a thematic connection when they encountered a novel item.

This group difference could indicate different strategies. The L2 group might be actively trying to establish meaning between the word and its informative context at the Context area, whereas the L1 group did so only for items they already knew. It does not seem reasonable that there is no processing difficulty for novel words for the L1, as there was a documented cost for the familiar words. It is possible that the L1 group is engaging in temporary semantic underspecification of the novel item, a well-established process in language processing (Frisson 2009; Sturt et al., 2004) when it results in "good-enough" task performance (Ferreira et al., 2002). If L1 readers are not actively attempting semantic integration at the context area for novel words, then they would not display the cost associated with processing difficulties. In other words, they might not be engaging in deep processing nor attempting full semantic integration from contextual cues; it is possible that L1 readers employed a reading strategy of not focusing much attention on the context but simply processed the stimuli quickly and shallowly in order to finish the task.

Similar patterns emerged for GPT. Informative contexts had longer GPT than uninformative ones (β= -0.18, t= -4.89, $p < .001$), suggesting an active integration process which is potentially costly in terms of time. Moreover, GPT on Context was longer for novel words (β=-0.1, t=-3.06, $p < .01$), and there was a significant interaction of context type and word type.

Post-hoc tests with a Tukey correction showed that when the preceding word had been familiar, informative contexts had longer GPT than uninformative ones (L1: β=0.18, t=4.89, $p < .001$; L2: β=0.14, t=3.81, $p < .01$), but this effect was not observed when the preceding word had been novel, which is especially puzzling. Semantic integration should be at least equally, if not more, costly when there is a new item involved. It is worth noting that the L2 group showed a numeric trend in that direction (following the pattern with TRT on context), but it did not reach significance. GPT involves re-reading triggered at the Context area; it is possible that when the word had been novel, both types of context triggered re-reading. However, when the word had been familiar, only informative contexts triggered re-reading.

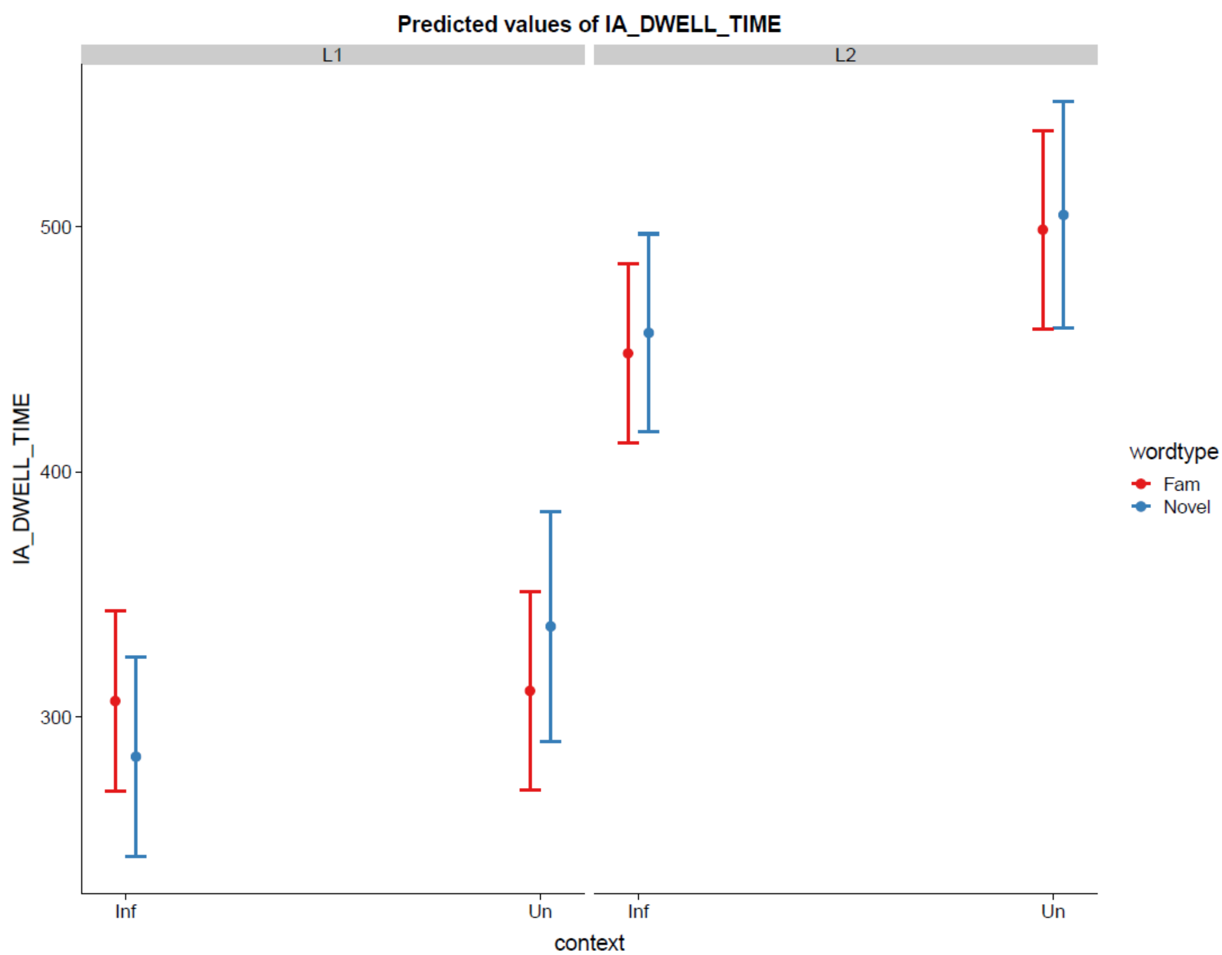
Finally, the L1 group had longer GPT at Context in informative contexts for familiar words compared to novel words (β=0.1, t=3.06, p < .05), showing some effort with semantic integration of familiar words; this effort is not observed when the word is novel, adding further support to the argument that they engage in semantic underspecification for novel words. The L2 group behaved more consistently and was actively engaged in meaning-making and semantic integration, as evidenced by their inflated GPT on informative contexts regardless of word type. This could suggest a difference in how the two groups approached the task, with L2 speakers being more actively involved in meaning-making, whereas L1 participants appeared to be more comfortable with underspecified or shallower processing of novel words.

In sum, semantic integration seems to follow generally similar patterns for both L1 and L2 speakers, with new lexical items taking more time to integrate. They also pattern similarly in making meaningful connections between familiar words and contextual information that is thematically linked to the preceding word. However, the groups demonstrate different strategies when it comes to novel words: the L2 participants still attempt to make meaningful connections, whereas the L1 group appear less engaged in this process and divert their attention from the context interest area.

Definitional hypernym

At the hypernym (e.g., “instrument”), there were no differences between conditions in FFD or GD for either group. This means that lexical identification and access of the hypernym was not influenced by experimental condition, suggesting that context or familiar target words did not prime faster lexical identification of the hypernym; readers retrieved the hypernym equally fast regardless of the type of preceding word or context. There was only a main effect of context informativeness on TT on the hypernym, with hypernyms having inflated reading times when the preceding context had been uninformative (β=-3.57E-02, t=-2.26, $p < .05$) (Figure 2). This finding indicates that readers (L1 and L2) found it easier to integrate the hypernym when the preceding context had been informative. This suggests online incremental processing and sentence-level integration for both groups. Finally, there were no differences between conditions for GPT on the hypernym for either group, suggesting that the hypernym was not a trigger for re-reading and revision.

Figure 1.2. Total time on interest area "Hypernym".



Regressions in and out

Model outputs are reported in Appendix A in log-odds ratios. Because binomial mixed effects models have difficulty converging, we used an optimizer from the optimx package (Nash & Varadhan, 2011).

Regressions into novel words were more likely than into familiar words both in their first and second occurrence; there were no effects of context, contra Chaffin et al. (2001). In addition, the L2 group was more likely to regress into Word1, Word2, and Context, than the L1 group, suggesting increased overall processing difficulty. In contrast, the L1 readers were more likely to regress out of Word2 than the L2 readers, which suggests more attentive processing of sentence1 by L2 readers, before moving on to sentence2. Regressions out of an interest area into previous parts of the text show increased cognitive load (Rayner & Pollatsek, 2006; Vasishth & Drenhaus, 2011) and Word2 is an area where L1 speakers show higher cognitive load than L2. This means that the L2 readers processed the previous input more attentively and were less "thrown off" when they re-encountered a novel word in sentence2. Finally, regressions out of Context and out of Hypernym were equally likely in all four conditions for both groups, contra Chaffin et al. (2001).

Overall, regressions seemed to index a general processing difficulty and not necessarily integration difficulty at the sentence level or a conscious attempt to make meaning. If the latter were the case, we would have expected word type and context type interactions. This is further supported by the finding that readers regressed *into* the context interest area equally frequently across conditions: regressions into an interest area did not seem to reflect a conscious search for meaning, or at least not a successful search.

In sum, the answer to RQ1 is that eye-tracking results indicate that L1 and L2 speakers patterned mostly similarly in lexical identification and lexical access processes. They diverged in semantic integration and processing difficulties at the sentence level. The L2 group seemed to actively attempt to derive sentence-level meaning from contextual cues, as shown by more time spent processing the informative contexts. With novel words, the L1 group displayed less processing by initially "ignoring" the novel word at first occurrence. This finding could indicate that differences in processing may reflect different strategies, that the groups differed in the way they approached the task, and in how actively they engaged in meaning-making.

Rate of learning

To address RQ3 (reading time change from first to second encounter with the novel word), we computed the rate of reduction in reading times, to examine whether the groups differ in learning speed, i.e., in their GD reduction from the first to the second occurrence of the novel word. First, we found that only about two-thirds of the trials in our dataset exhibited a reduction pattern, with an equal L1-L2 representation. Since a reduction and an increase in reading times portray two distinct strategies, we split the data accordingly. A *t*-test on the reduction data showed that there was no group difference in the rate of GD reduction, $t(1029)= -0.66$, $p >.5$, contra Pellicer-Sánchez (2016). We also computed the rate of increase in GD from Word1 to Word2 for the remaining 1/3 of trials. Interestingly, we found a group difference; the L2 group had larger increases than L1 ($t(507.28)=-2.47$, $p < .05$), which could suggest more difficulty in processing, or more attention paid to the task by the L2 group. The divergent patterns (reduction vs. increase in GD) suggest different reading strategies, which would be masked had we not split the dataset accordingly.

Vocabulary post-test scores

Regarding RQ5 (level of success in word learning), L1 and L2 performance on the unannounced vocabulary posttest was not statistically different, $t(78.26)=1.22$, $p=.22$; both groups retained a similar (very small) number of novel items ($M= 1.95$, $SD= 1.71$, range=0-7 for L1 and $M=1.53$, $SD= 1.38$, range=0-5 for L2). It is a nontrivial feat that participants were able to recall even a small number of items, considering the demanding nature of the task (20-option form-meaning matching). Previous studies have also reported smaller gains for meaning recall tasks (e.g., Godfroid & Schmidtke (2013), who included an 18-option matching test).

To address RQ2 (whether reading times predict word learning), we collapsed the data from both experiments into one dataset and ran two binomial mixed effects logistic regressions with vocabulary test accuracy as the binary dependent measure and total time on trial, or total time on the novel word, as the predictor, respectively. Group was not included in the models given the similar retention scores. There was no effect of total time spent on the trial or on the novel word. The prediction that longer reading times would result in higher accuracy was not born out, contra some previous studies (e.g., Godfroid et al., 2013; Williams & Morris 2004), but in accordance with others (Elgort et al., 2018). Task difficulty may have induced floor effects in participants' accuracy, whose scores ranged only 0-7 out of a maximum of 20. Finally, form-meaning mapping is more demanding than, for instance, simple form recognition. Previous studies that did find a correlation between longer reading times and better recall have utilized different designs; for instance, Pellicer-Sánchez (2016) contained only six novel items in the experiment, and each was repeated eight times, giving participants ample opportunity to rehearse the items in memory, opportunities to retrieve them during the experimental task, and presumably form a stronger representation. Our results indicate that even in a demanding form-meaning mapping task, some readers are able to remember a small number of items, and this seems not to be mediated by whether the text is in the reader’s L1 or L2, at least for the participants in this study.

Experiment 2

Experiment 2 was designed to further test whether L1 and L2 groups differ in the way they approach and engage in the given task, and similarly, in how comfortable they are with a shallower type of processing and semantic underspecification. We used think-aloud protocols which have been used to measure awareness in general SLA research (e.g., Bowles, 2010; Leow 1997; 2000), depth of processing (e.g., Adrada Rafael, 2017; Leow et al., 2008) and specifically in novel word-learning studies (Godfroid et al., 2013).

Materials and Methods

Participants

Participants were 19 L1-English speakers (13 F) and 19 L1-Chinese L2-English speakers (14 F) who received course credit for participation. None had participated in Experiment 1, but they were selected from the same pool and had similar demographic characteristics and linguistic backgrounds.

Procedure

Participants were instructed to read sentences aloud while vocalizing their thoughts, which were recorded using the Windows voice recorder. The stimuli were the same used in Experiment 1, presented in the same order. When participants did not verbalize, a researcher reminded them to keep saying their thoughts aloud. L2 speakers were instructed to use English, Chinese, or both. Subsequently, participants completed a distractor O-Span Task, followed by two vocabulary tasks of different demands. During the word identification task, they saw a list of nonce words containing the 20 novel items they saw during the experiment and 40 additional filler nonce items. They were instructed to circle the words they remembered seeing during the think-aloud reading task. This was followed by a word-hypernym matching task (the same as in Experiment 1), which targeted memory retention of word-meaning mappings, and a language history questionnaire.

Audio-recorded think-alouds were transcribed and subsequently coded by one of the researchers according to the detailed coding scheme in Appendix E. Twenty percent of the responses were also coded by a second research assistant to ensure consistency in coding. The inter-rater overlap (in percentages) in code assignment was M=96.12, SD=14.08.

The main goal of Experiment 2 was to examine whether the two speaker groups approached the task in different ways and engaged in deeper or shallower processing. If L2 participants had a harder time processing the stimuli, i.e., if the task was subjectively more difficult for them, then we expected to see *more* processing. This would mean more instances of every type of processing in the think-alouds in all conditions and with the fillers. If L1 and L2 differed in how *deeply* they processed new words, we expected the groups to pattern differently: the probability of deep/elaborate processing over all instances should be greater for the L2 than the L1.

Analysis and Discussion

The coded data were analyzed in R (R Development Core Team, 2013); initial detailed counts of codes were collapsed into 3 categories which refer to a shallow, medium, and deep level of processing (see Table 2 for examples). The categories loosely followed Leow et al.'s (2008) "noticing"/ "reporting"/ "interpreting" scheme and Leow's (2015) three-level operationalization of depth of processing for lexical items. as This categorization also follows Craik's (2002) suggestion that deeper/more elaborate processing involves semantic elaboration.

The count data were analyzed with a zero-inflated Poisson mixed-effects regression (with a log link) using the package glmmTMB. The model was selected because it minimized AIC, but all models can be found in the supplementary materials. The dependent variable was counts of utterances, and the predictors were group (L1 or L2), level of processing (Level1=shallow, Level2=medium, Level3=deep), and their interaction. Participants were included as a random intercept. L1 English speakers have a familiarity advantage with the language, the script, and the phonology, compared to L2 speakers. Even if novel words are equally novel for both groups, the context in which they are presented is not. For this reason, we ran two separate analyses of the think-aloud data: one on the familiar and filler stimuli, and a separate one on the novel items. The rationale was to establish a baseline pattern of processing for the two groups, and then compare the pattern for novel item processing against it, in order to address RQ4 (L1-L2 processing depth).

Table 1.4. Operationalization of three levels of processing depth

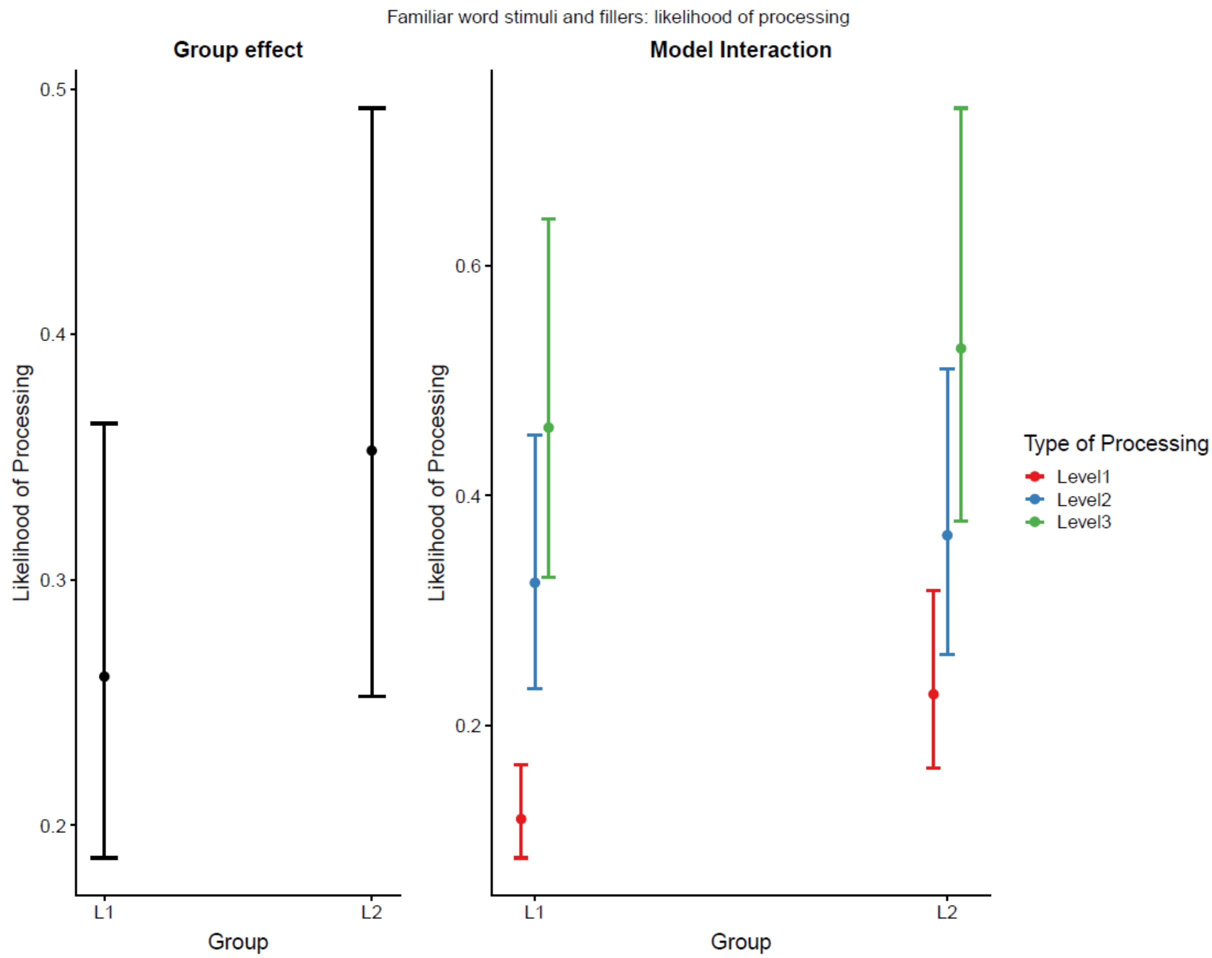
Level 1 (shallow)	Level 2 (medium)	Level 3 (deep)
Noticing	Reporting	Interpreting
Short pause, repetition, struggling to pronounce	long pause, wondering/questioning/stating ignorance about target word	linking target word and context or hypernym, realizing meaning, making meaning inferences, expressing emotion

Table 1.4 (cont.)

-"Sarah picked a …vewn" *-"Jonathan liked nupes-nupeswic"* *-"Lauren selected the guilk….guilk"*	*-"I don't really know what muzoom is"* *-"What is quixte?"*	*-"Yeah so it's like two types of pasta"* *-"[reading: She thought Esivel is a beautiful language] -Oh it is a language!"* *-"Aw cute"* *-[laughing]*

First, we report the results from the model on the familiar stimuli and fillers. All estimates reported are in log-odds. There was a main effect of group, with L2 participants having more verbalizations than L1 (β=0.65, *SE*=0.24, $p < .01$). There was also a significant interaction of group and level of processing. Pairwise comparisons with Bonferroni corrections showed that the two groups patterned similarly: both groups were more likely to engage in level 3 processing than level 2, which was turn was more likely than level 1, as shown in Figure 3.

Figure 1.3. Probability for the three levels of processing for familiar word stimuli and fillers.

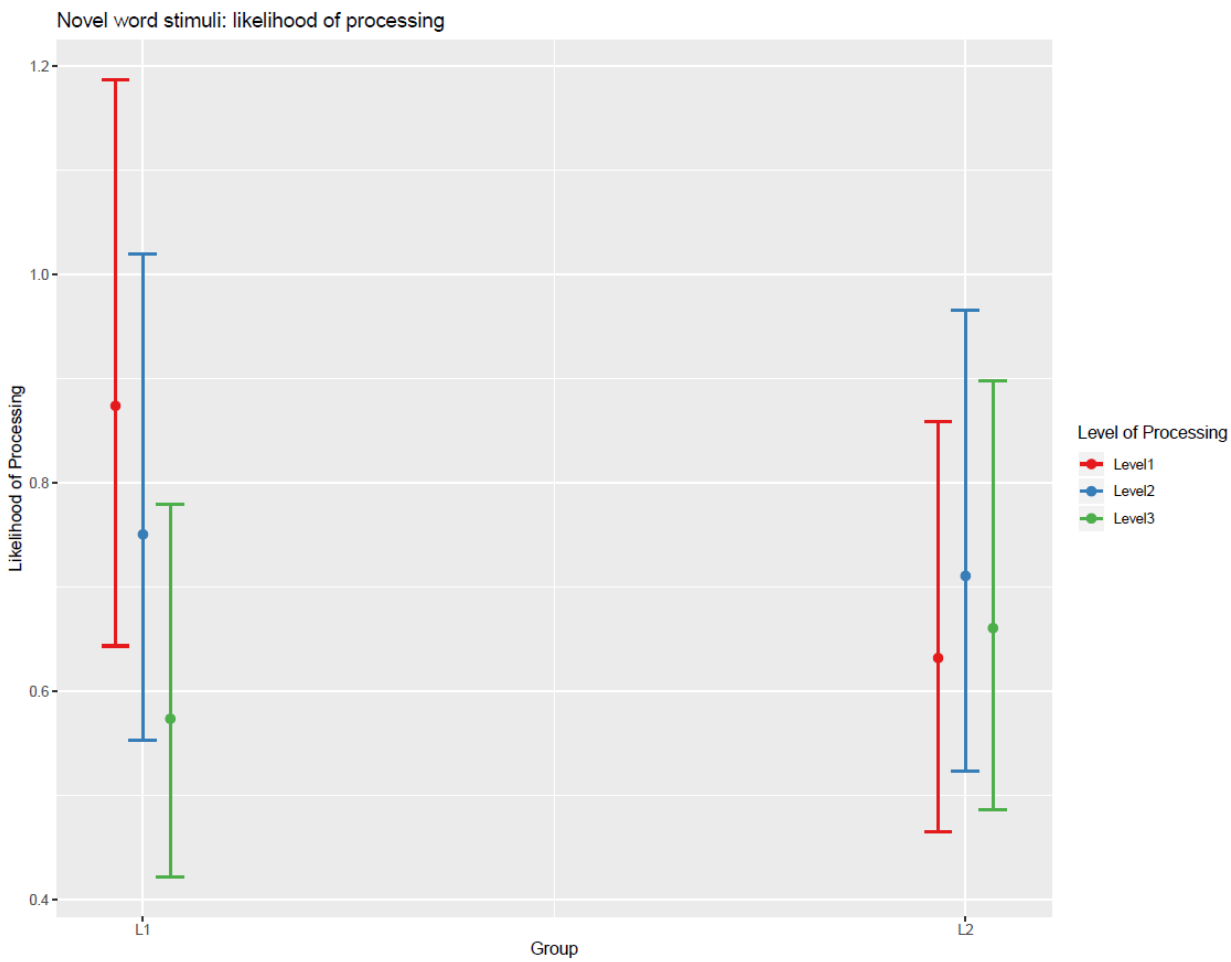


Next, we report the results on novel word stimuli. There was no main effect of group, but there was an interaction between group and level of processing. Pairwise comparisons with a Bonferroni correction revealed a different pattern for L1 and L2 speakers. L1 speakers were more likely to engage in Level 1 processing than Level 2 or Level 3 (β=0.15, *SE*=0.02, $p < .001$ and β=0.42, *SE*=0.02, $p < .001$, respectively). They were also more likely to engage in Level 2 than Level 3 processing (β=0.27, *SE*=0.02, $p < .001$). L2 speakers, in contrast, were more likely to engage in Level 2 than in Level 1 or Level 3 (β=0.12, *SE*=0.02, $p < .001$ and β=0.07, *SE*=0.02, $p < .001$, respectively). Figure 5 illustrates the group difference.

If we compare the patterns from familiar and novel stimuli, we see that upon encountering unfamiliar words, the groups (probabilistically) diverged in how they approached the task. In their baseline, they patterned similarly, with deeper and medium level processing more likely for both groups. In addition, the L2 Group had more verbalizations, which is to be expected, as the task might have been subjectively more difficult for them. That is, the L2 group put more effort into the task.

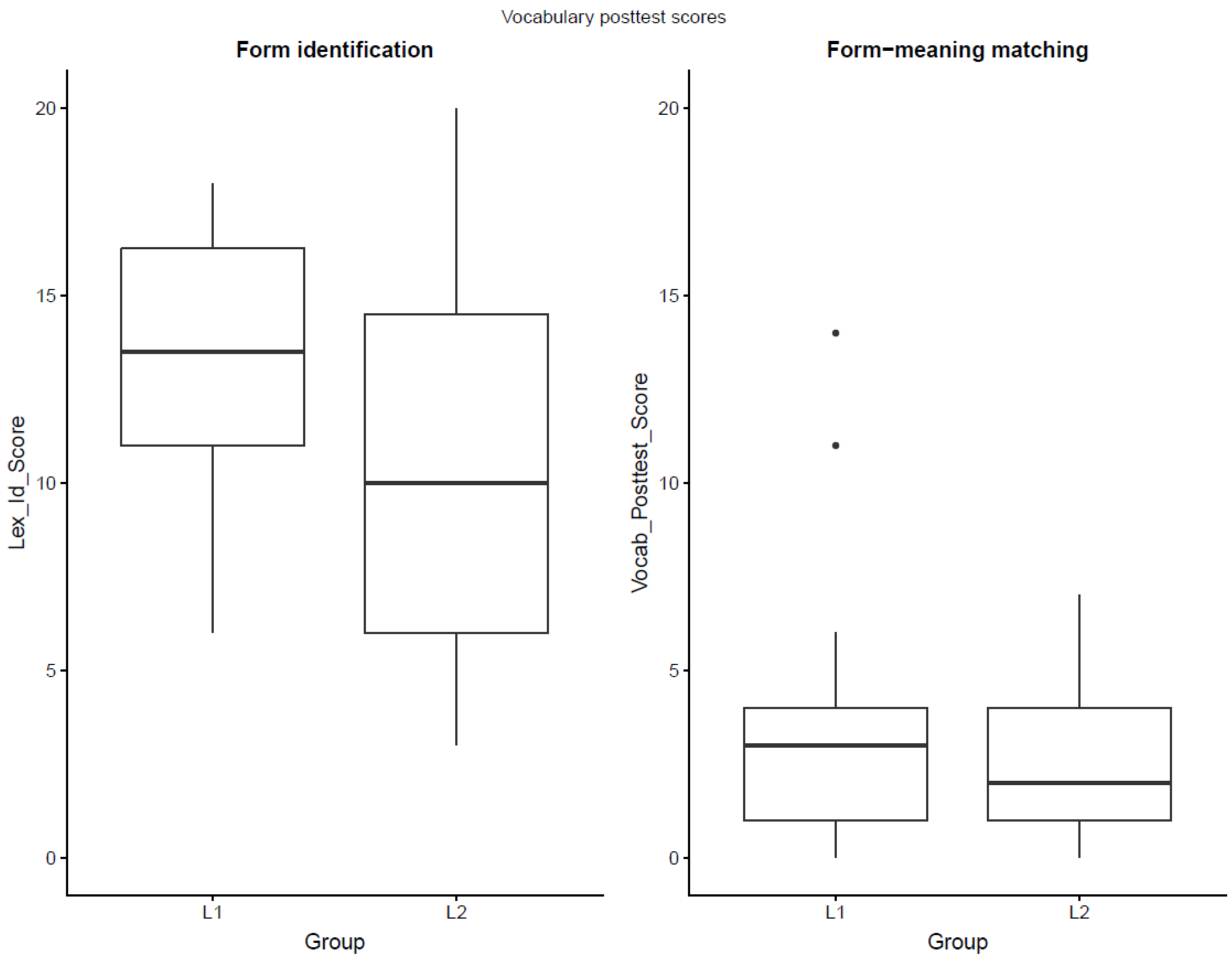
A different pattern can be observed in the novel word condition (Figure 4): the L1 participants were less likely to engage in medium and even less likely in deep processing. In contrast, the L2 group engaged in all three levels of processing fairly equally (and slightly more in medium level processing). This provides converging evidence that a nontrivial number of L1 speakers engaged in “ignoring” strategies when reading unfamiliar stimuli, whereas L2 speakers maintained a high level of effort in making meaning. It needs to be noted that there was substantial individual variation; it is beyond the scope of this study to account for individual differences factors that may account for it, but it is a promising area for future research.

Figure 1.4. Probability for the three levels of processing for novel word stimuli.



Regarding RQ5 (word learning success), L1 and L2 groups retained a similar number of novel items in the lexical identification and vocabulary matching tasks, similarly to experiment 1. An independent samples *t*-test showed that L1 identified a similar number of novel items (M=13.25, SD=3.85) as L2 (M=10.6, SD=5.2), $t(34.97)=1.83$, $p > .05$. Similarly, an independent samples *t*-test showed that L1 successfully matched a similar (much smaller) number of novel items to their hypernyms (M=3.25, SD=3.67) as L2 (M=2.45, SD=2.09), $t(30.15)=.85$, p>05. Figure 5 illustrates the group scores. Interestingly, L1 participants remembered a slightly larger number of items in both vocabulary posttests, although this difference was not statistically significant. One possibility is that the L2 group had lower proficiency compared to L1 and had to process more effortfully; in contrast, the L1 group required less effort in order to process and learn. This would be in line with Calderón (2013), who showed that as proficiency increases, less-deep processing can lead to learning. Finally, in line with previous studies, performance was better for form recognition than form-meaning mapping. Regarding the form-meaning mapping task, accuracy was higher in Experiment 2 than Experiment 1 for both groups, suggesting that the think-aloud task induced some deeper processing compared to silent reading in Experiment 1 and providing further evidence that the nature of processing operations performed on the stimuli influences retention (Elgort et al., 2018).

Figure 1.5. Vocabulary test scores



Summary and overall discussion

The experiments in this study compared the reading patterns of native and non-native speakers of English in how they process novel lexical items and engage in meaning-making. We expanded previous research on incidental vocabulary acquisition in two ways. First, we analyzed eye movement data both on individual words, and on contexts offering information to their meaning. In addition, we triangulated data from both eye tracking and think-alouds in order to examine how L1 and L2 differ in terms of processing speed and depth.

L1 – L2 processing profiles

Group similarities

The study replicated the well-attested novel word effect (see Godfroid (2020) for a review) for both L1 and L2 groups. By also analyzing reading times on contextual cues, we established that both groups make thematic meaning connections between words and thematically related information, something that had been previously indirectly deduced (Brusninghan & Folk, 2012; Elgort et al., 2018; Godfroid et al., 2013).

Group differences

The L1 group seemed to engage in underspecification and not attempt full semantic integration at the sentence level for novel words, whereas the L2 group more consistently and actively searched for meaning. This finding is supported by 1) reading times on the contextual cues, 2) regression patterns, and 3) by processing depth patterns. Specifically, the L1 group engaged in less deep processing in novel word trials, and they did not show TRT differences between informative-uninformative contexts when the preceding word had been novel (an effect that was observed for familiar words).

In addition, we find different processing strategies for some trials, which also suggests group differences. Even though previous studies found faster reading through word repetition (Elgort et al., 2018; Pellicer-Sánchez, 2006), this was confirmed in our experiment only for a proportion (2/3) of the trials. For the remainder trials, we observed slower reading (longer fixation durations), similarly to Godfroid et al (2018) who found an increase in fixation durations around the 7th to 10th exposure. This could be indicative of less/less deep processing of the first sentence of the stimulus. The rate of reduction of fixation durations did not differ for the two groups, but the rate of increase of fixation durations did: the increase was larger for L2 than for L1, adding further support to the argument that the L2 group is paying more attention and processing with intentionality.

In sum, L1 and L2 readers had similar processing profiles at the levels of lexical identification and access and semantic integration, but they differed in levels of engagement with the task. The two groups are likely to be habituated in different strategies, which they put to use in the experimental task. Our participants were studying in the US in L2 English, and experienced continued necessity to read and understand academic texts while encountering new words (Nagy & Townsend, 2012); therefore, it is likely that they were more used to the task of figuring out new words than the L1 participants were, and they were more committed to extracting meaning.

There are two strands of theory and research that are in line with our findings. The first one is 'good-enough' processing accounts from the L1 literature. (e.g., Christianson et al., 2001; Ferreira, et al., 2002; Ferreira & Patson, 2007). This line of theorizing has documented misinterpretations typically with syntactic manipulations in garden path sentences (Christianson, 2016 for an overview). In addition, some studies (Sanford et al., 2005; Sanford et al., 2006; Sturt et al., 2004) have demonstrated 'good-enough' lexical semantic processing by L1 readers. These studies used a change-detection task during reading (e.g., "the man with the hat [*cap*/*dog*]"; *hat* changed to either *cap* or *dog*) which examines the level of detail of representations. Results showed that readers were less likely to detect a change if the changed word was semantically related to the original, pointing to a shallower level of computed representation. In our study, L1 readers also seem to compute 'good-enough' lexical interpretations (presumably of the form "instance of category X") and to put less effort and intentionality into word learning.

The second relevant theory from the L2 domain is the Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006a, 2006b, 2006c). It posits shallower processing of morphosyntactic information by L2 speakers, who use primarily semantic information (to a greater extent than native speakers) as they attempt meaning making. Even though we did not test morphosyntactic processing, our findings do support a corollary of the SSH as L2 readers showed extra depth of semantic processing. Specifically, we found exactly what the SSH hypothesizes, "that L2 processing may prioritize semantic […] information, with L2 speakers potentially being more sensitive to these types of information compared to L1 speakers" (Clahsen & Felser, 2018, pp. 695). We cannot argue whether or not this resulted in diminished morphosyntactic processing in this experiment, but it does show intentionality in establishing word meaning and less willingness to semantically underspecify new, unknown lexical items, in contrast to L1 speakers as discussed above (e.g., Sturt et al., 2004).

Speed and depth in processing.

In the L1 reading literature, shorter fixations are associated with ease of processing (Rayner, 2009), or shallower processing (Christianson, 2016). Similarly, longer fixations can indicate processing difficulty or increased depth, both in L1 and L2, and this reading behavior can be modulated by task type (Jackson & Bobb, 2009; Lim & Christianson, 2015; Rayner, 2009). Therefore, the common finding that L2 reading is slower than L1 can be attributed to greater difficulty, greater effort, or both, and this is what our results support.

Our study addressed recent calls (e.g., Godfroid, 2020; Pellicer Sanchez, 2020) to mixed methods approaches to the study of incidental vocabulary learning and compared L1-L2 processing patterns in terms of speed and depth. First, L2 readers were consistently slower than

L1 in all eye-tracking measures, in all regions, in accordance with previous literature (Roberts et al., 2008). This difference can be attributed to their different backgrounds and levels of experience with the language as a whole. This would include, besides lexical and morphosyntactic knowledge, frequency effects, quality of phonological representations, and familiarity with the script (L2 participants were native speakers and readers of Chinese, which uses a different, non-alphabetic script) (e.g., Koda, 2005 for L1-L2 distance effects in L2 reading).

Besides quantitative reading time differences, our mixed-methods approach confirmed that reading times (especially later ones like TRT) can be at least partly dependent on task interpretation. L1-L2 divergent patterns in this study can be explained by different levels of processing depth. One caveat here is that, due to differences in linguistic experience, the two groups may *need* different levels of processing to achieve the same result, considering that the groups had similar vocabulary scores for form recognition and form-meaning matching. Nevertheless, we argue that the L2 group approached the experiment as a learning task (Barcroft, 2015) for the following reasons: 1) novel words were unknown to both groups, 2) the groups patterned similarly for familiar item trials and diverged for novel ones, and 3) the think-aloud data showed increased and decreased processing depth for novel words by L2 and L1 readers, respectively.

Our findings have implications for eye-tracking research in reading, where the general assumption is that longer reading times (e.g., total times, go-past times) are indicative of effortful cognitive processing due to difficult stimuli (Rayner, 2009, 2012). Our results suggest that in addition to cognitive difficulty, longer reading times can be the result of different levels of engagement with the task and depth of processing employed, which has been touched on in the novel word literature in terms of vocabulary attainment: Elgort et al. (2018) and Mohamed (2018), for instance, argued that individual attention and quality of processing is a strong predictor for novel word learning and retention (also Craik & Tulving, 1975). Our study showed how speed and depth of processing can be experimentally addressed to uncover L1 and L2 similarities and differences.

Conclusions

The results of this study are important for two main reasons. First, methodological triangulation of eye-tracking and think-aloud data addressed L1 and L2 processing in a new light; we considered and analyzed both processing difficulty and depth as factors contributing to L1 and L2 differences in novel word processing. Second, the study adds new findings to the literature of novel word learning; we manipulated the informativeness of context and directly showed that the two speaker groups were able to make meaning connections between a word and its surrounding contextual cues; interestingly, the L2 group appeared more committed to the process than the L1 group, who engaged in semantic underspecification of novel items. Further research should employ mixed-methods designs when comparing L1 and L2 processing; reading speed can be influenced not only by processing difficulty, but also by processing depth.

Even though we share the sentiments of researchers who emphasize ecological validity (Godfroid et al., 2018), in this study we opted for greater control which allowed us to manipulate context and analyze measures on those regions. Future studies can extend the design to more naturalistic language learning contexts.

Part 2: L2 Morphosyntactic Learning

It is widely accepted that most L2 learners lag in processing speed and proficiency attainment (e.g. Winke, Godfroid, & Gass, 2013), and many L2 speakers do not come close to native levels especially when it comes to morphological development (Bardovi-Harlig, 1992; Clahsen & Felser, 2006; VanPatten, 1996, 2006). There is also a consensus that the fact that L1 and L2 are learned differently at least partly explains the differential outcomes; linguists have focused on the differences of amount and quality of input, as well as amount of exposure or opportunities to practice, among other factors (Doughty & Long, 2003). Large scale modeling research (Hernandez, Li, & MacWhinney, 2005; Li, 2009) and smaller-scale experimental work (Ellis & Sagarra, 2010) have shown that such effects can be explained by the notions of cue competition, blocking, and L1 entrenchment (i.e., the mere fact that the L1 is already in place) with respect to aspects that differ from the L2 and thus prevent L2 learners from noticing and learning certain L2 cues.

These approaches have manipulated only linguistic cues and have so far not considered the role of the learning environment. Nevertheless, in real-world language comprehension, there are numerous linguistic and non-linguistic cues that compete for the attention of the listener (Kreysa, Knoeferle, & Nunneman, 2014). In addition, theoretical accounts argue for a facilitative role of the non-linguistic environment (Barsalou, 2008, 2012; Barsalou, Santos, Simmons & Wilson, 2008); multimodal input is theoretically argued -yet empirically understudied- to facilitate language processing and learning (Clark and Paivio, 1991; Paivio 1971, 1986; Sadoski & Paivio, 2004, 2013) by strengthening associations between the linguistic and visual modalities, resulting in better memory. These models view human cognition as separately processing information from distinct representational systems (e.g., linguistic and visual), which are however highly interconnected (Paivio 1971; 1990). No studies so far have systematically manipulated the non-linguistic learning environment to examine its effect on L2 learning. In this part of the dissertation, I extend the framework of cue competition to include non-linguistic cues that offer contextual support and to examine their effect on the learning of linguistic cues in the L2.

SLA research has generated an immense body of work on learning outcomes, predominantly focusing on the level of L2 learners' attainment, and comparing L2 to L1 knowledge and performance (Roberts & Siyanova-Chanturia, 2013). The main goal of studies on L2 processing is to find quantitative or qualitative similarities and differences compared to L1 (e.g., Portin & Laine, 2001; Roberts, Gullberg, & Indefrey, 2008). Very little research has looked at L2 processing from a developmental perspective, i.e. at the online encounter and processing of new linguistic information and the online mechanisms of acquisition (but see Pelicer-Sánchez, 2016, for a word-learning study), and even less so in multimodal environments.

The next sections of this dissertation are an attempt to address the understudied area of how L2 morphosyntactic knowledge develops from processing linguistic information in diversely rich contexts. More specifically, I ask how L2 learners process new morphosyntactic information when it is presented i) in isolation, ii) with L1 support, and iii) with visual scene support. Qualitative changes to the learning environment alter what cues become available to adult L2 learners, thus changing what gets noticed and learned (Arnon & Ramscar, 2012; Ellis & Sagarra, 2010). Therefore, it is imperative that studies begin incorporating a diverse set of cues, linguistic and non-linguistic, and explore their effect on the L2 learning process and outcome.

Parts 3 and 4 employ self-paced reading and eye-tracking methodology, respectively, and examine how different types of learning environments affect L2 morphosyntactic processing and learning. Expanding on the well-tested Cue Competition (MacWhinney, 1987, 2002) and Input Processing (VanPatten 1996, 2015) models, I propose and test a more holistic view of "cues" for L2 learning, both linguistic and visual. Under this view, adding non-linguistic cues will alter morphosyntactic cue availability to the learners, and thus influence how efficiently learners notice, process and integrate a novel L2 morphosyntactic structure. In other words, these cues might help circumvent L1 entrenchment and blocking. The first goal is to track L2 learners' attention to informative linguistic cues during online sentence comprehension in the presence of various types of contextual support, thus investigating the role of the learning environment (linguistic only or linguistic + pictorial) in morphosyntactic learning. Considering that the notion of cue competition has been largely supported in previous research (see Hernandez, Li, & MacWhinney, 2005), and that environment manipulation alters cue availability (Arnon & Ramscar, 2012; Ellis & Sagarra, 2010), it is important to extend the existing models to take into account extra-linguistic cues and their role in the learners' attention allocation processes. The second goal is to examine how such attention allocation during the learning process subsequently affects learning outcomes. It is attested that when a cue such as morphological case marking is absent in the L1, it is initially very difficult to learn to attend to that cue in the L2, as the learner often remains oblivious to the cue's existence (Fulga & McDonough, 2015; McDonough & Trofimovich, 2013). My studies examine what contextual conditions enhance or block attention to the relevant morphological markers and to what extent this attention can predict morphosyntactic learning. In sum, in the following sets of studies presented in Parts 3 and 4, I investigate the interplay of linguistic and non-linguistic cues and how different types of contextual support influence attention allocation as L2 learners of Greek process morphologically case-marked sentences and assign thematic roles, a necessary process for successful sentence comprehension in free word-order languages.

Second language processing as learning: linguistic cue competition

Successful L2 learning requires attention to form and meaning (Long, 1991; Robinson, Mackey, Gass, & Schmidt, 2002), but the view of adult L2 learners as limited-capacity processors (Just & Carpenter, 1992) suggests that they cannot attend to all parts of the input at the initial stages of learning, but rather must select parts for processing (Gass, Svetics, & Lemelin, 2003; MacWhinney, 1987; Sagarra, 2007; Schmidt, 1993). As Ellis (2006) aptly puts it, we do not use all the stimuli that are available in the environment, but instead we view them through the lens of prior experience.

Formal learning theory has established effects of competition and blocking in learning (Arnon & Ramscar 2012; McLaren & Mackintosh, 2000; Rescorla & Wagner, 1972). The principles of competition and blocking (the latter revisited as entrenchment and redundancy) have informed psycholinguistic SLA models, most notably the Cue Competition (Bates &MacWhinney, 1989; MacWhinney, 1989, 2001, 2002, 2005, 2008; MacWhinney, Bates, & Kliegl, 1984) and Input Processing models (VanPatten, 1996, 2015), which emphasize the role of competition between multiple linguistic cues in sentence comprehension. When the cues compete, learners compute each cue’s validity, which consists of its availability (the probability the cue will be present in a sentence) and its reliability (the probability of correctly indicating a function, such as participant role assignment).

These models view language learning as mapping form onto functions; they also view grammatical learning as an interaction between the learner, the input, and the context

(MacWhinney, 2002). They successfully account for various L2 learning patterns (Hernandez, Li, & MacWhinney, 2005), and specifically for error patterns emerging when learning morphosyntax (Ramscar & Yarlett, 2007, Ramscar, Dye, & Yarlett, 2009). These models claim that adult L2 processing is initially parasitic on L1, and they view L2 learning as a gradual readjustment of cue weights, which differ cross-linguistically. As L2 learning progresses, learners become sensitive to additional cues to which they were previously oblivious. Even though transfer of morphology from L1 to L2 is limited, there is transfer of the underlying functions, such as participant role mapping. Language learning, then, is seen as the product of language processing (VanPatten, 2015). Numerous studies providing support for these models follow similar designs that attempt to find the order of cue strengths in a given language, or an L1-L2 combination. Results show that in sentence comprehension, learning cue reliability takes place gradually. At first, L2 cue weights are similar to the L1 weights, and gradually they shift to those of a native speaker of the L2 (Bates & MacWhinney, 1981; de Bot & van Montfort, 1988; Gass, 1987; Harrington, 1987; Kilborn, 1989; Kilborn & Cooreman, 1987; Liu, Bates, & Li, 1992; McDonald, 1987a, b; McDonald & Heilenman, 1991; McDonald & MacWhinney, 1989). Learners first focus on one cue at a time, and gradually consider additional cues in order to reduce errors of understanding incurred by the insufficient previously learned cues (Matessa & Anderson, 2000).

Cue competition also involves blocking, a robust and well-researched phenomenon (Kamin, 1962; Rescorla & Wagner, 1972; Wills, 2005) which is a product of learned attention (Kruschke & Blair, 2000; Mackintosh, 1975). In language processing, when cues are redundant, they can be masked, i.e., blocked from learned attention (Ellis, 2006; Schmidt, 2001; Terrell, 1991; VanPatten, 1996) and therefore not get learned. The effect of blocking in SLA has been empirically attested: Ellis and Sagarra (2010) demonstrated that an early learned cue can block the learning of later cues (especially if they are redundant). In their study, learners of Latin failed to learn tense morphology when they had already learned time adverbials, and vice versa, failed to learn time adverbials when they had already learned tense morphology. They also showed that this can have a long-term effect: Chinese native speakers failed to learn tense inflectional morphology, which their L1 lacked, when it was available in the input along with adverbials. Ellis (2006) similarly discusses overshadowing, a combined result of blocking and salience: when two cues presented together both predict an outcome, the strength of each cue will depend on its salience. The more salient cue will become associated with the outcome and further strengthened, resulting in the less salient cue becoming overshadowed, escaping the focus of selective attention.

In L2 cue competition, cue salience can determine which part of the input will be attended to and selected for processing. Grammatical cues such as morphological markers have low salience (Ellis and Sagarra 2010; Ellis, 2008) which might explain adult L2 learner's increased difficulty in processing and producing L2 verbal morphology (Ellis, 2006; Goldschneider & DeKeyser, 2001; Jiang, 2004; Zobl & Liceras, 1994). In fact, Goldschneider and DeKeyser (2001) operationalized salience and found that it predicted L2 morpheme acquisition order. Cue salience can thus explain why L2 learners do not rely on grammatical cues but prefer lexical ones (Lee, Cadierno, Glass, & VanPatten, 1997; Sagarra, 2007; Ellis & Sagarra, 2010).

Both the theoretical accounts and the empirical studies so far have only manipulated cues that are linguistic in nature; no study to date has examined how altering the type of contextual environment and adding visual cues can affect L2 cue learning. There are reasons, however, to expand the main notion of cue redundancy and competition in initial L2 processing to contexts that include non-linguistic information and to examine how linguistic and visual cues interact, compete and influence what gets learned. Cues are not objectively available to all learners in all contexts, but they are dependent on the structure of the learning environment (Arnon & Ramscar, 2012) and the processing task/goal (Lim & Christianson, 2013). A natural question involves how the presence of non-linguistic contextual cues might alter the process of readjusting cue weights in the L2.

Testing cue competition: word order and thematic role assignment

Cue competition in L2 learning has been tested in participant role assignment, i.e., the mechanism that assigns NPs in a sentence the theta-roles required by the verb's syntactic argument structure. Thematic role assignment is achieved in different ways across different languages. Some languages, like English, use word order: the first NP in an active sentence is the agent and the second NP is the patient (which typically correspond to the Subject and the Object in active sentences). Other languages, however, do not have fixed word order and thus case marking is a more reliable cue for thematic role assignment. In languages like German, Korean, and Greek, both SVO and OVS patterns are attested, through a movement process of the object across the subject known as grammatical scrambling (Hopp, 2006, 2007). Even though SVO is typically the canonical, more frequent structure, and OVS patterns do incur processing costs (Hemforth & Konieczny, 2013), scrambling is attested in these languages. Case marking disambiguates the underlying syntactic structure and is a more reliable cue than word order.

In Greek active sentences, agents are marked with Nominative case on the Noun and through agreement, on their Determiner; patients are marked with Accusative case similarly on the Noun and its Determiner. Nouns and determiners bear additional features such as number and gender, but these are beyond the scope of this dissertation; only masculine singular nouns will be used as stimuli, and number/gender will not be marked. Sentences (1a, b) below involve the same proposition but 1(b) is expected to be more difficult for L2 learners, because in order to be correctly interpreted, the case marking needs to be noticed and integrated during parsing in order to compute the correct underlying syntactic representation.

1(a): O$_{\text{[Nom]}}$ skil-os$_{\text{[Nom]}}$ dagose ton$_{\text{[Acc]}}$ lik-o$_{\text{[Acc]}}$

The dog$_{\text{[agent]}}$ bit the wolf$_{\text{[patient]}}$

1(b): Ton$_{\text{[Acc]}}$ lik-o$_{\text{[Acc]}}$ dagose o$_{\text{[Nom]}}$ skil-os$_{\text{[Nom]}}$

The wolf$_{\text{[patient]}}$ bit the dog$_{\text{[agent]}}$

English has SVO word order, and this is a highly reliable cue; it has been found that when speakers of English learn other languages that mark semantic roles morphologically with case rather than with word order, they experience great difficulty. MacWhinney (2001) discusses how for a German native speaker of advanced L2 English proficiency, a “syntactic accent” was evident when he ignored word order in favor of agreement and animacy cues when the two competed. Similar findings were reported in Kempe and MacWhinney (1998) with native English speakers who learnt Russian or German. Studies focusing on the learning of morphosyntactic structures in transitive sentences have found that participants mistakenly interpret OVS as SVO structures (Fulga & McDonough, 2016; McDonough & Fulga, 2015; McDonough & Trofimovich, 2013, 2015, 2016; Papadopoulou, Varlokosta, Spyropoulos, Kaili, Prokou, & Revithiadou, 2011) due to L1 transfer and a tendency to interpret the first NP in the sentence as the Subject (MacWhinney 2012; VanPatten 1996; 2004). Learners assume that the first noun phrase available is the subject of the proposition, transferring the L1 pattern onto the

L2. Gradually, they learn that morphological case marking is a more reliable cue than word order in that language. Therefore, learning involves readjustment of cue weights in the L2. Papadoloupoou et al. (2011) tested Greek native speakers learning Turkish; even though both the L1 and the L2 have case marking and free word order (though Turkish exhibits a more complex system than Greek), learners were not “fully aware of the interaction between case morphology and word order”, which suggests they may fail to notice certain cues or their interaction (pp. 196). Very few studies to date have investigated attention allocation to relevant cues or the role of extralinguistic, visual cues in the process of cue weight readjustment.

McDonough, Trofimovich, Dao and Dion (2017) examined novel L2 Esperanto structure learning in a paradigm that involved aural presentation of sentences along with a set of two pictures depicting the same participants but with reversed role assignment. They found self-initiated eye-gaze to the correct picture to be a predictor of later test performance in OVS sentences. Interlocutor-initiated visual cues to the key form-meaning relationship were not helpful for understanding and learning the structure. In this study, looking at OVS pictures was *assumed* to result from noticing the case marking. Advancing this line of research further, we can, and should, test whether, when, and under what conditions participants notice case marking as an informative cue and then shift their eye-gaze towards the correct picture accordingly. In other words, we can extend this line of research into the time course of each cue’s integration by measuring attention to specific cues/parts of the input rather than only on the images as a whole. McDonough et al. (2017) found that longer self-initiated eye-gaze to the correct picture for OVS items during the comprehension learning activity predicted higher test scores. This means that the participants who looked more at the images that illustrated the meaning of OVS learning items were more accurate in the subsequent test. A natural next step would be to examine the circumstances that caused them to look longer, which might have been noticing the morphological markings and mapping these to the corresponding visual scene. It was only self-initiated gaze, and not interlocutor-initiated, that predicted learning, which suggests that looking because of understanding predicts learning, validating the use of eye-tracking in processing-as-learning approaches to SLA. As the authors point out, self-initiated gaze can "shed light on when learners make use of learning opportunities" (p. 864). The studies in Parts 3 and 4 track attention to fine grained areas (the case markers) and locate when, and under what contextual support, learners identify and use these cues for thematic role assignment.

The importance of visual context in language comprehension

Language comprehension is contextually situated, and research has shown that the surrounding environment influences processing. Arnon and Ramscar (2012) manipulated input presentation in an artificial language (whole sentence first, or single nouns first) and found that changing the learning environment qualitatively changed cue availability, processing and learning. In other words, cues are not objectively available to a learner (Ellis, 2006) but can be enhanced or masked by the type of learning environment they occur in. Expanding on this, I propose that changes to the modality of information (linguistic or visual) will similarly affect attention allocation, processing and learning.

The effect of non-linguistic context has been investigated in studies of visually situated language processing, most typically L1 processing. The results indicate that pictorial cues are considered, processed and integrated in comprehension incrementally. Tanenhaus, Spivey-Knowlton, Eberhart, and Sedivy's (1995) seminal paper showed that participants integrate visual and linguistic information during online spoken sentence comprehension. Numerous follow up studies have shown that semantic interpretation and syntactic processing involve early integration of contextual (visual) information as listeners resolve referential ambiguity with intersective and scalar adjectives (Sedivy, Tanenhaus, Chambers, & Carlson, 1999) and temporary syntactic ambiguity (Chamber, Tanenhaus, & Magnuson, 2004; Spivey, Tanenhaus, Eberhard, & Sedivy 2002). In particular, when a depicted scene and a corresponding sentence involve reversed (mismatched) thematic roles, there is an attested integration difficulty, evidenced by longer reading times on the part of the sentence that mismatched the picture (Knoeferle & Crocker, 2006), suggesting incremental integration of sentence constituents and picture parts. The authors propose an interactive and interpretative relationship of visual context and sentences in L1 comprehension, a relationship that seems reasonable to extend to L2 theory and research. Specifically in L2 processing and learning, visual cues to morphosyntax might be more successful in avoiding blocking effects, compared to L1 translations, because of the nonlinguistic nature of the pictorial information.

Regarding thematic roles, Knoeferle, Crocker, Schelpers, and Pickering (2005) showed that native German speakers used visual cues to assign thematic roles despite temporary linguistic ambiguity, before linguistic disambiguation through case markers. Therefore, they showed an influence of depicted events on how native speakers incrementally assigned thematic roles and resolved participant role ambiguity. The present work extends this relationship to L2 processing and learning under a cue competition framework and seeks to examine the interaction between linguistic and visual cues in thematic role assignment. As Ronderos, Münster, Guerra, Kreysa, Rodríguez, Kröger, Kluth, Burigo, Abashidze, Nunnemann, and Knoeferle, (2018) correctly point out, eye-tracking in integrated visual and language content can provide invaluable insights into how a second language is processed and learned.

In her unpublished dissertation, Palmer (2015) investigated the role of visual context for garden path ambiguity resolution with native and non-native speakers of English. She found facilitatory effects of images for L2 processing; when the visual and the linguistic codes matched in interpretation, reading times were shorter (faster processing) and accuracy in the subsequent questions was increased. However, the L2 readers appeared to over-rely on images: when a sentence was ambiguous, L2 readers relied on the image whether it was helpful or not. Her work confirms that both L1 and L2 readers integrate linguistic and pictorial information (e.g., when the two mismatched, this resulted in longer reading times suggesting integration difficulties). Moreover, when an image was presented first, it created expectations about upcoming linguistic materials. Interestingly, when the image was consistent with a garden path interpretation, it induced good-enough processing effects: participants ignored a disambiguating comma as the garden path interpretation (good enough) matched their expectations that the picture had created.

Learning occurs when there is a discrepancy between 1) what is already known and creates expectations, and 2) an environment that violates these expectations. However, prior learning can block learning of new information if it is made redundant (Rescorla & Wagner, 1972), something that has been attested in L2 learning of case morphology in Esperanto (McDonough & Fulga, 2015). A native English speaker learning Greek is expected to begin with the assumption that the language follows SVO word order, transferring the L1 pattern. When presented with an OVS sentence such as (1b) above, s/he may misinterpret it by failing to compute the correct underlying syntactic representation specified by case marking. In other words, s/he may reverse thematic role assignment. The presence of a picture, however, can draw attention to the thematic role mismatch and to the violated expectations, creating conditions conducive to learning. In turn, this can force learners to search for a more reliable cue, making case marking morphology more salient. However, it is also possible that case markers are not noticed, and learners only assume a free word order in the language, resulting in syntactic underspecification and "good enough" processing and understanding (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Lim & Christianson 2015, 2013a, b).

Summary

A large body of research confirms the theoretical view of L2 structure learning as the process of re-adjusting the weight of relevant cues in the L2, and gradually noticing and incorporating additional cues, when new input is not sufficiently explained by expectations generated by previously learned cues. In addition, changing the nature of the learning environment can mask or enhance the cues that are available to the learner, impeding or facilitating learning respectively. Considering the central role visual contexts have in L1 online sentence interpretation, it is crucial to investigate their role in L2 sentence processing and learning.

Part 3: L2 Morphosyntactic Learning: an investigation with SPR

This part focuses on the processing of a new linguistic structure in different contexts, namely textual (linguistic) and pictorial (non-linguistic). The goal is to examine how learners allocate attention to different cues during L2 comprehension, how the type of contextual environment (linguistic, visual) alters cue availability and weight, and how this affects the early stages of L2 learning of morphological case marking as a reliable cue for thematic role assignment.

This study uses self-paced reading (SPR) to investigate the effect of different types of contextual support on the attentional mechanisms of English-speaking adults at the initial stages of learning an L2 (Greek) with inflectional case morphology, as well as how the context influences which cues are attended to, enhancing or blocking morphosyntactic learning. Our hypothesis is tested with true beginner L1 English speaking adults who have no prior knowledge of any languages with case morphology or free word order. Previous studies have established that when such populations encounter case morphology, they experience difficulty due to L1 transfer and entrenchment (Hopp, 2010; Jiang, 2004, 2007; Li, 2009; MacWhinney, 2012; McDonough & Fulga, 2015; McDonough & Trofimovich, 2013), and potentially by avoiding building elaborate syntactic representations altogether and relying on heuristics (Ferreira, 2003). More specifically, MacWhinney (2012) suggests that when learners need to rely on morphological cues to assign participant roles, they experience difficulty. McDonough and Trofimovich (2013) and McDonough and Fulga (2015) reported that only a small number of their participants were able to detect the novel construction of Esperanto case marking. Even when participants' L1 had case marking, only half of them were able to detect the pattern in the aural input, suggesting variance in learners' ability to succeed at pattern detection[2]; the latter was judged indirectly through explicit testing of structure knowledge. It is worth tracking online attention to the relevant cues considering that L2 processing may rely on heuristics, especially when the task is cognitively challenging, and learners may not "see," i.e., process and understand, the morphological cues. In addition, no studies to date have factored in the role of the environment in which processing occurs, which leaves unanswered the question of how contextual support may facilitate or impede attention to case marking cues in the input. More specifically, the experiments in Part 3 investigate how contextual support (L1 translation equivalents or visual contexts) influences attention allocation to informative linguistic cues (word order, morphological case marking on nouns and determiners) and non-linguistic cues (goodness of agent) while reading and comprehending simple transitive sentences and assigning thematic roles in L2 Greek.

Experiments A, B and C

The present study includes three related experiments, each varying the type of contextual support available for L2 sentence comprehension (Experiment C: baseline experiment with no contextual support, Experiment B: additional English translation, Experiment A: visual contextual support [i.e., line drawings]).

Participants

Participants were English Native speakers recruited online who received payment or course credit for their participation. They were recruited through ads posted on the lab's website, Facebook, and Reddit. They were randomly assigned to one of the three experiments (Experiment A: N = 29, Experiment B: N = 27, Experiment C: N = 23) and completed all tasks

[2] Though, salience of a cue in aural and written input should not be presumed to be the same.

in a single online session. Data from 4 participants were excluded from analysis due to lack of attention during the task or not conforming to the study's requirements regarding their linguistic background.

Procedure and tasks

The three experiments included the same procedure and tasks, all completed in a single online session: 1) L2 vocabulary pretraining phase, 2) L2 structure learning phase 3) integrated SPR and sentence-picture matching task, 4) a structure production task, and 5) a brief exit survey (see Figure 3.7 below for a schematic illustration). The experiments were programmed on Ibex Farm. The learning phase is, to the best of my knowledge, the first attempt in the field to track the online processing and learning of morphological case marking through attention allocation to case morphology in different environments. The combination of online and offline outcome measures allowed us to take fine-grained measures of L2 knowledge, ranging from very implicit (measured by online reaction times) to explicit (measured by behavioral measures of comprehension and production) (Ellis, 2005).

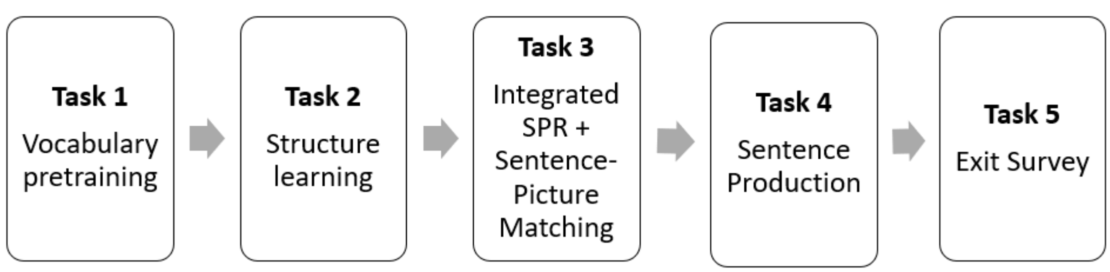


Figure 3.1. Schematic illustration of the five tasks participants completed in each experiment. Tasks 1, 3, 4, and 5 were identical in the three experiments; Task 2 manipulated type of environmental support where the L2 exposure took place.

Task 1 (Vocabulary pretraining): During the vocabulary pre-training phase, participants in all three experiments learned 15 Greek lexical items presented along with their English translations in a list. The list included two adverbials (one temporal and one locative), four verbs inflected in the present tense and the third person, and nine nouns presented in the nominative singular form. Participants were instructed to spend about 15 minutes and no less than 12 reciting and trying to learn them. Learning was confirmed with a subsequent vocabulary test that asked participants to match the Greek words and the English translations. These lexical items were combined to create the stimuli for the structure learning and test phases. The same adverbs and locative PP were used in all stimuli to keep the vocabulary learning load low. Task 1 was identical in experiments A, B and C.

Task 2 (Structure learning): During the structure learning phase, participants completed a cumulative self-paced reading task, where the previous words remained in view and new ones were added with each button press. The reason for choosing cumulative self-paced reading was that this was the first time participants were introduced to morphosyntactic elements in L2 Greek, and a non-cumulative reading task would be too difficult: they would have to both retain information in working memory and process new morphosyntactic markers simultaneously. Using SPR, participants read simple transitive sentences which contained two NPs in Greek, presented in phonetic “Greeklish” using the Latin alphabet (Androutsopoulos, 2012).

The stimuli for Task 2 were 32 simple items (presented twice in a fixed order) with a transitive verb and two NPs (agent and patient, nominative- and accusative-marked respectively on the noun and its determiner[3]). Each item began and ended with a time adverbial (*yesterday*) and a locative prepositional phrase (*in the park*) in order to avoid placing interest areas at the beginning or end of a sentence. The adverbials were the same across all items to reduce the vocabulary learning load. I manipulated (i) type of structure (SVO- OVS) and (ii) participant role plausibility (either both NPs equally plausible as agents or only one NP plausible as agent—there was be no implausible agent condition). This resulted in a 2 x 2 factorial design and stimuli were distributed over two lists, such that each participant saw only one stimulus from each token set (Table 3.1 presents an example of all conditions). No fillers were used to reduce the vocabulary learning load, a common option in SLA psycholinguistic research (McDonough & Fulga, 2015). No explicit, researcher generated feedback was offered.

During Task 2 in Experiment A, participants first saw an image, and following a space bar press, they completed cumulative SPR of the Greek sentence underneath the image. The images were drawn by an artist after the sentence stimuli had been created and they were simple line drawings. Examples of images for each condition are presented in Table 3.2 below. The images were normed with 20 Greek Native speakers who had completed the experiment and reported that all visual scenes were clear and corresponded to the sentences correctly. In Experiment B, participants first saw a whole English sentence and read it at their own speed; following a space bar press, they completed SPR of the Greek sentence presented underneath the English one. In Experiment C, participants saw the phrase "Next Sentence" and following a space bar press, they completed SPR of the Greek sentence presented in isolation.

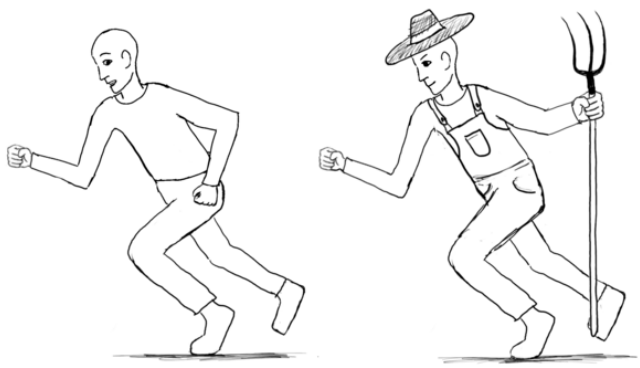
[3] During the vocabulary pretraining task, the nouns were presented without determiners, in the nominative case. During the structure learning task (and subsequent tasks), participants saw sentence stimuli that included nominative- and accusative-marked nouns and determiners (see Table 3.1 for examples).

Experiment C provided the baseline comprehension, when no additional contextual support is provided. In Experiment B, the English translation provided additional contextual support, as it allowed the creation of a situational model (Zwaan & Radvansky, 1998), against which the Greek sentence can be comprehended. In Experiment A, the visual scene provided concrete contextual support, illustrating the proposition of the sentence. Previous research suggests that images can be helpful in the creation of mental model representations (Glenberg & Langston, 1992).

Table 3.1. Stimuli conditions example.

Condition 1 [SVO, 1 good agent]	*Htes o[Nom] skilos[Nom] dagose ton[Acc] antra[Acc] sto parko.* Yesterday the[Nom] dog[Nom] bit the[Acc] man[Acc] in the park.
Condition 2 [OVS, 1 good agent]	*Htes ton[Acc] antra[Acc] dagose o[Nom] skilos[Nom] sto parko.* Yesterday the[Acc] man[Acc] bit the[Nom] dog[Nom] in the park.
Condition 3 [SVO, 2 good agents]	*Htes o[Nom] skilos[Nom] ide ton[Acc] kokora[Acc] sto parko.* Yesterday the[Nom] dog[Nom] saw the[Acc] rooster[Acc] in the park.
Condition 4 [OVS, 2 good agents]	*Htes ton[Acc] kokora[Acc] ide] o[Nom] skilos[Nom] sto parko.* Yesterday the[Acc] rooster[Acc] saw the[Nom] dog[Nom] in the park.

Table 3.2. Example images accompanying sentence stimuli for Experiment A for each Agent condition (1 good or 2 good agents).

[1 good agent]	
[2 good agents]	

When the word order presented follows SVO, participants should not experience any problems as L2 and L1 align. When, however, the stimulus includes an OVS structure, participants' initial interpretation of the sentence was expected to be incorrect, as an assumed SVO syntactic structure will result in reversed thematic role assignment, following, e.g., McDonough & Fulga (2015). When both NPs are equally plausible agents, this reversal results in a plausible, unproblematic (though inaccurate) interpretation. When only one NP is a plausible agent and it is presented as NP2, this could trigger the learner to re-evaluate sentence interpretation and theta-role assignment. Therefore, in Condition 4 (OVS, 2 good agents), contextual support may prove especially facilitative: in cue competition, learning occurs when there is a discrepancy between 1) what is already known and creates expectations, and 2) an environment that violates these expectations (Rescorla & Wegner, 1972). In the presence of a picture and in expectation of an SVO sentence, the OVS sentence can be interpreted as mismatching the thematic roles in the picture (flipping them), thus violating participants' (incorrect) expectations and creating the conditions necessary for noticing the case markers. Thus, a picture may make this violation more salient, as opposed to reading only sentences in the L2 or even translation equivalents, as it provides an unambiguous mental model, a real-world situation that the proposition refers to.

In contrast, it is also possible that visual scenes encourage good enough processing. In L1 sentence processing, "good enough" processing phenomena are attested, e.g., plausibility or top down effects on online sentence comprehension, whereby under certain circumstances and tasks the representations computed during on-line processing can be underspecified (e.g., Christianson, 2016; Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Ferreira, Ferraro, & Bailey, 2002; Stoops, Luke, & Christianson, 2014); L2 speakers have also been shown to employ good-enough processing strategies in certain tasks (Lim & Christianson, 2015; 2013a, b). Therefore, a condition presenting learners with two NPs and a corresponding visual scene could result in a good-enough approach: because the NPs match the depictions, the actual syntactic thematic role assignment might be blocked from being attended to and learned. If participants do gradually learn and are able to use the morphological marking to assign participant roles, attention to case markers is predicted to change throughout the learning task; we predict that initially, participants will not pay attention to these areas, but after they notice them, and understand their function and high reliability, they should allocate more attention to them (as measured by testing for interactions between presentation order and fixed effects, e.g., Christianson, Mestre, & Luke, 2012). If, in contrast, participants engage in good enough processing, attention to case markers is expected to be minimal and not change throughout the learning task, as participants will not have identified case markers as informative cues for role assignment.

Task 3: SPR with sentence-picture matching

Task 3 was a combination of cumulative self-paced reading and sentence-picture matching; for brevity, I will refer to it as sentence-picture matching. The stimuli for this task were the sentences taken from the counterbalanced list from the structure learning task, so that they involved the same words but not identical items that participants saw during the learning phase. For example, if a participant saw List 1 in the learning phase, s/he saw items from List 2 in the sentence-picture matching task. The stimuli for Task 3 additionally included an implausible agent condition (e.g., *The man bit the dog*), not present in the structure learning phase. The 48 items in this task crossed type of structure (SVO-OVS) and type of agent (both NPs as good agents, only 1 NP as a good agent, implausible agent NP) resulting in a 2 x 3 design.

In all three studies, participants first saw a screen with two visual scenes depicting two NPs as either agent or patient performing the same action (e.g., *a pelican looking at a rooster* vs. *a rooster looking at a pelican*). The positioning of the correct image on screen (right or left), as well as the positioning of the agent relative to the patient (to the right or to the left) was counterbalanced. Following a space bar press, participants completed cumulative SPR of the Greek sentence added under the images. Table 3.3 provides examples of the stimuli for each condition. Participants were asked to read the sentence and select the picture that illustrated the meaning of the sentence as fast as possible. This way we were able to measure both selection accuracy, and reading speed, which can indicate confidence and success in learning the morphosyntactic elements in the previous task. If participants have learned that case markers are informative, we expect similar performance between conditions in Task 3.

Table 3.3. Example stimuli for the sentence-picture matching task.

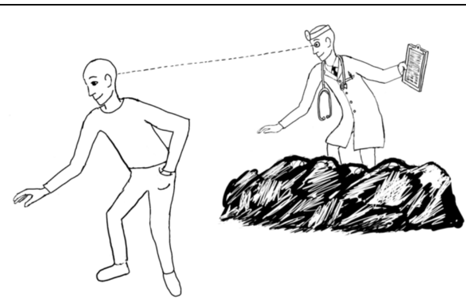
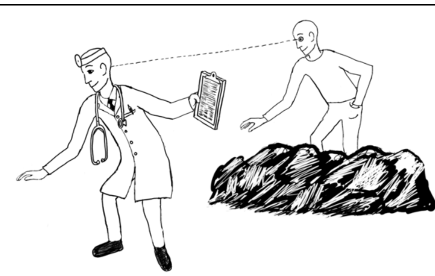
1 good agent [SVO]		
	O skilos dagose ton agroti.	
[OVS]	Ton agroti dagose o skilos.	
	The dog bit the man	
2 good agents [SVO]		
	O giatros ide ton adra.	
[OVS]	Ton adra ide o giatros.	
	The doctor saw the man	

Table 3.3 (cont)

Implausibl e agent [SVO]		
	O kokoras haidepse ton agroti.	
[OVS]	Ton agroti haidepse o kokoras	
	The rooster pet the farmer.	

Task 4: Structure production task Subsequently, during the structure production task, participants were shown six images (2 images with both NPs as good agents, 2 images with only 1 NP as a good agent, and 2 images with an implausible agent); they were instructed to write a sentence in Greek to describe the picture.

Task 5: Exit survey The final task was an exit survey, which involved open-ended questions targeting explicit noticing and learning of the Greek morphosyntactic markers. Participants were asked to indicate by typing in a box what they learned and noticed during the experiment. In addition, they saw three items from the picture selection task, and they were asked again to choose the correct image and to explain/justify their choice.

Predictions

Extending the cue competition model to extra-linguistic contextual cues suggests that visual scenes may facilitate noticing and learning morphosyntactic L2 cues such as case marking, resulting in improved processing and comprehension. Since determiners are more salient than noun endings, and in this experiment are reliable (nouns are always presented with a determiner), we may expect identification of case marking on determiners but not noun endings. In contrast, it is also possible that images encourage a good-enough processing approach and distract attention away from the case morphology, making it redundant for comprehension.

During the learning phase, if visual scenes are facilitative for learning morphological cues, we expect participants to demonstrate better ability to use the morphological cues in Experiment A compared to B and C. Participants should also be more effective in using those cues to recover from violated expectations in OVS sentences. If it is not an effect of input modality (visual) but simply of additional unambiguous information, we expect no differences in reading behaviors between Experiments A and B.

During the sentence-picture matching task, if visual scenes were facilitative for noticing case markers, we predict more accurate performance in Experiment A. Across studies, if participants have identified case morphology as a reliable cue, they may be able to use it predictively in OVS sentences to identify the right picture before reading the whole sentence.

Finally, more attention to morphological markers is expected to predict higher accuracy in the behavioral scores, and the latter should be higher in Experiment A, if visual scenes facilitate the allocation of attention to case morphology.

Analysis and Results

Task 2: Learning Phase Self-Paced Reading Results

The sentence frame for all stimuli was: Adverbial/ Determiner 1/ Noun 1/ Verb/ Determiner 2/ Noun 2/ Last Word (also an adverbial). To analyze reaction time data in each experiment and examine how conditions influenced reaction times, I built one linear mixed effects model per area using the lme4 (Bates et al., 2014) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in the R environment (Team R, 2013). Specifically, I built one model for Determiner 1, Noun 1, Determiner 2, Noun 2, Verb, Last word, and whole Sentence reading times. I included Last Word to examine sentence wrap-up effects in the different conditions. In each model, the dependent variable was the log transformed reaction times and the predictors were Structure (SVO-OVS), Agent (one good - two good), the log-transformed trial order, and their 3-way interaction. Subjects and Items were included as random intercepts. In a few cases where the model did not converge, the Item random intercept was removed. Only reaction times longer than 200 msec were entered for analysis. The results are presented below separately for each experiment. In all models, there was always a main effect of trial order; participants became faster in their reactions as they progressed through the experiment. Because this result is consistent, I will not discuss it in any more detail.

Experiment A (Images)

Determiner 1. In this experiment, there was a main effect of Structure on Determiner 1 reaction times: Determiner 1 had longer reaction times in OVS trials than in SVO ($\beta = 0.73$, $t = 2.9$, $p < .01$). This effect was qualified by an interaction between Structure and trial order, such that reaction times on Determiner 1 decreased for OVS trials, but they remained fairly similar (there was only a small decrease) for SVO trials (β = -0.16, t = 02.09, p < .05) (Figure 3.1 illustrates this effect).

Noun 1. There were no significant differences between conditions in the reaction times for Noun 1.

Determiner 2. There were no significant differences between conditions in the reaction times for Determiner 2.

Noun 2. There were no significant differences between conditions in the reaction times for Noun 2.

Last Word. There were no significant differences between conditions in the reaction times for Last Word.

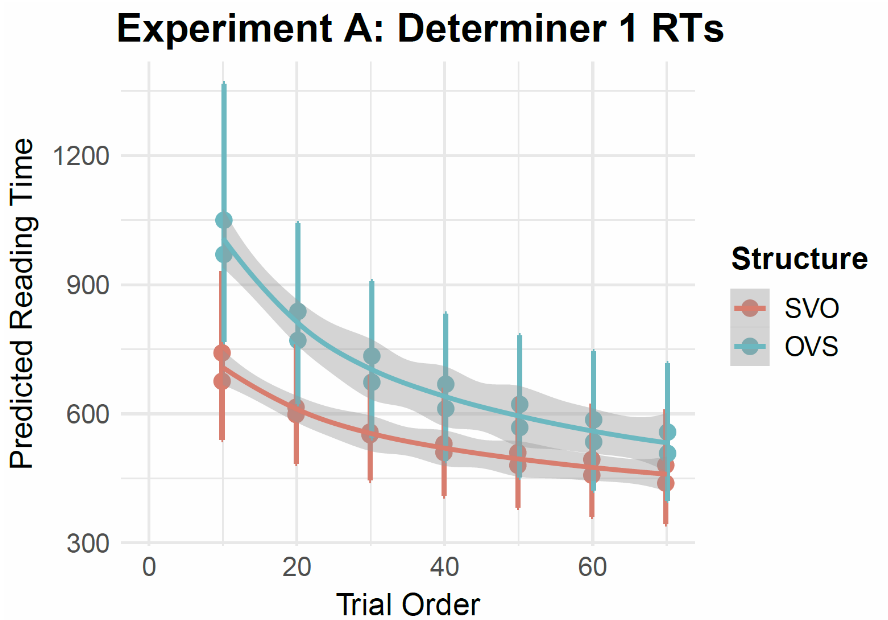


Figure 3.2. Reaction times on Determiner 1: interaction between Structure and trial order. The plot includes a smooth line for predictions from a linear model; the gray area represents 95% confidence bands.

Verb. There was a main effect of Structure on the Verb reaction times: the verb had longer reaction times in OVS compared to SVO trials ($\beta = 0.92$, $t = 3.91$, $p < .001$). This effect was qualified by an interaction between Structure and trial order, such that reaction times for the verb decreased as the experiment progressed in OVS trials but remained fairly similar (there was a smaller decrease) in SVO trials ($\beta = -0.27$, $t = -3.77$, $p < .001$). Figure 2 illustrates this effect.

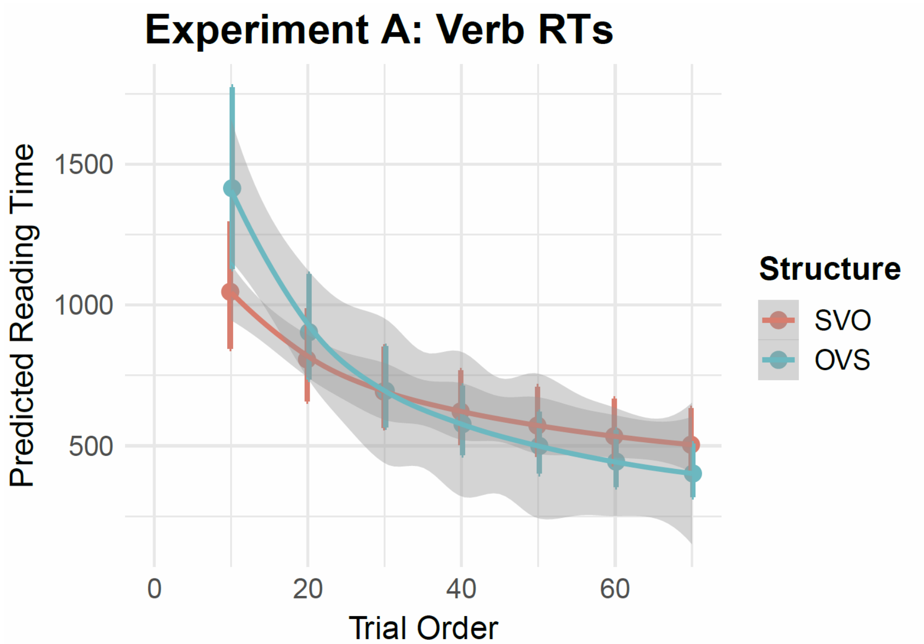


Figure 3.3. Reaction times on the verb: interaction between Structure and trial order. The plot includes a smooth line for predictions from a linear model; the gray area represents 95% confidence bands.

Sentence. The same main effect and interaction were observed (as a numerical trend) for whole sentence reading times. OVS trials took longer to read than SVO ($\beta = 0.43$, $t = 1.9$, p = .059). This was again qualified by an interaction between Structure and trial order, such that sentence reading times were reduced more as the experiment progressed for OVS trials, compared to SVO ($\beta = -0.13$, $t = -1.97$, p = .051). These effects are only numerical trends but they are worth noting as they are in line with the results on Determiner 1 and Verb.

Both determiners. I also ran a model with Structure, Determiner (1 or 2) and trial order, as well as their interaction as the predictors, and reading time on the Determiners as the dependent variable. There was a main effect of Structure, with OVS trials having longer reaction times than SVO ($\beta = 0.6$, $t = 3.58$, p < .001). In addition, the second determiner had longer reaction times than the first (β = 0.34, t = 3.05, p < .01). These were qualified by 2- and 3-way interactions, such that in SVO trials, reaction times between the two determiners did not differ much; in contrast, in OVS trials, Determiner 1 had much longer reaction times than Determiner 2 (Figures 3.3 and 3.4).

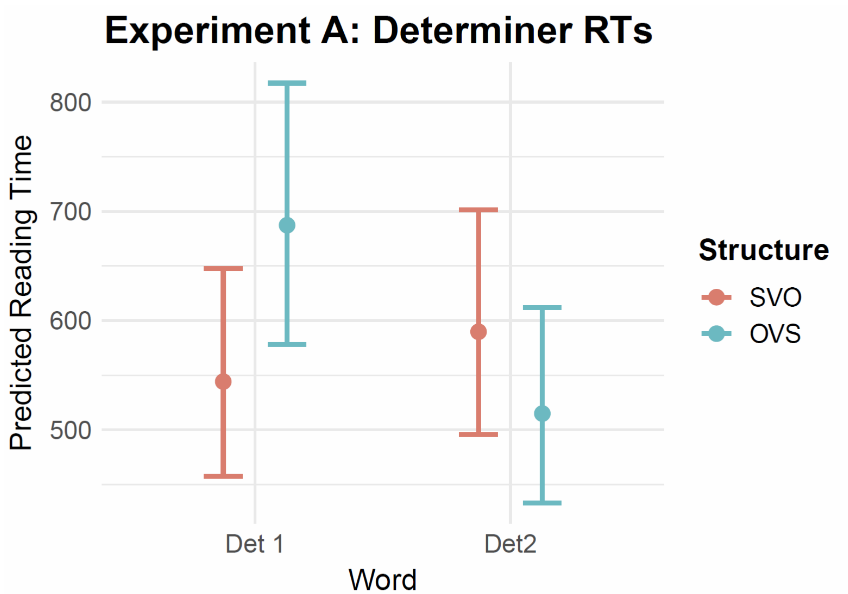


Figure 3.4. Reaction times on Determiner 1 & 2.

Figure 3.5. Reaction times on Determiner 1 & 2 by trial order. The plot includes a smooth line for predictions from a linear model; the gray area represents 95% confidence bands.

Both Nouns. I also ran a model with Structure, Determiner (1 or 2) and trial order, as well as their interaction as the predictors and reading time on the Nouns as the dependent variable. There were no differences in the reading times of Noun 1 and 2, and no differences between conditions.

Experiment B

Determiner 1. There were no significant differences between conditions in the reaction times for Determiner 1.

Noun 1. There was a main effect of Structure, with longer reaction times on Noun 1 in OVS trials compared to SVO (β = 0.69, t = 3.59, p < .001). This effect was qualified by an interaction between Structure and trial order, such that reaction times on Noun 1 decreased significantly as the experiment progressed for OVS trials but showed only a small change for SVO (β = -0.19, t = -3,18, p < .01)

Determiner 2. There were no significant differences between conditions in the reaction times for Determiner 2.

Noun 2. There were no significant differences between conditions in the reaction times for Noun 2.

Last Word. There were no significant differences between conditions in the reaction times for Last Word.

Verb. There was a main effect of Structure on the Verb reaction times, with longer reaction times for OVS compared to SVO trials ($\beta = 0.45$, $t = 2.27$, $p < .05$).

Sentence. There were no significant differences between conditions in the reading times for the whole sentence.

Both Determiners. There was only an interaction between Structure and Determiner (1 or 2) ($\beta = -0.39$, $t = -2.13$, $p < .05$), such that in SVO sentences, Determiner 2 had longer reaction times but in OVS, Determiner 1 had longer reaction times. I note that this involves the same determiner 'ton' having longer reaction times in both cases.

Both Nouns. There was a main effect of Structure, with OVS trials having longer reaction times than SVO ($\beta = 0.47$, $t = 3.1$, $p < .001$). In addition, the second Noun had shorter reaction times than the first ($\beta = -0.33$, $t = -2.73$, $p < .01$). These were qualified by 2- and 3-way interactions, such that reaction times decreased more for Noun 1 compared to Noun 2, and especially so in OVS trials (Figure 3.5).

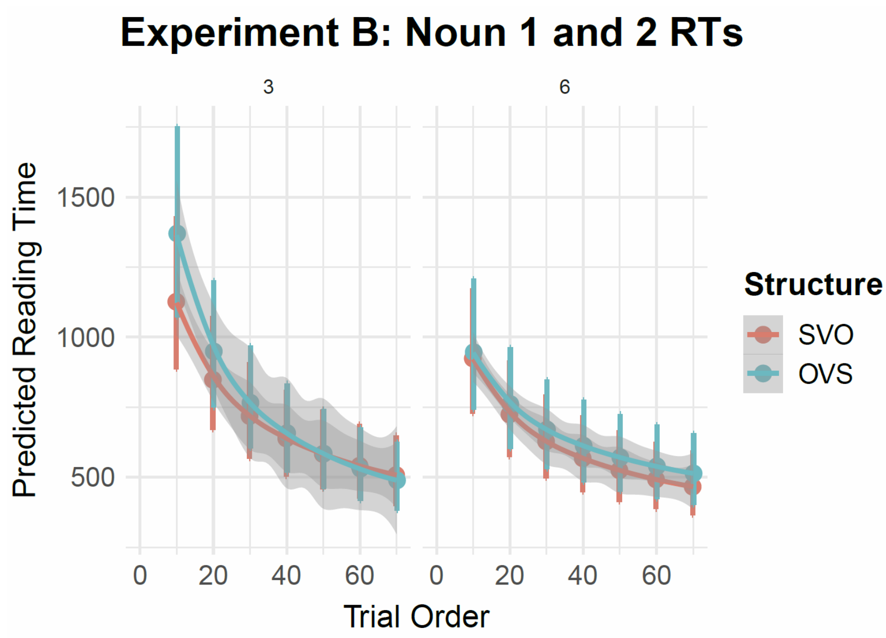


Figure 3.6. Reaction times on Noun 1 & 2.

Experiment C

Determiner 1. There were no significant differences between conditions in the reaction times for Determiner 1.

Noun 1. There were no significant differences between conditions in the reaction times for Noun 1.

Determiner 2. There were no significant differences between conditions in the reaction times for Determiner 2.

Noun 2. There was a main effect of Agent on Noun 2 reaction times, such that in trials where the two nouns were both good agents, reaction time on Noun 2 was longer, compared to trials where only one noun was a good agent (β = 037, t = 1.98, p < .05).

Last Word. There were no significant differences between conditions in the reaction times for the Last Word.

Verb. There were no significant differences between conditions in the reaction times for the Verb.

Sentence. There were no significant differences between conditions in the reading times for the whole sentence.

Both determiners. There were no differences in the reading times of Determiner 1 and 2, and no differences between conditions.

Both Nouns. There were no differences in the reading times of Noun 1 and 2, and no differences between conditions.

Task 2: Learning Phase Discussion

The results from the cumulative self-paced reading task provide preliminary evidence that initial processing of L2 Greek case morphology is influenced by the contextual support available in the learning environment. In the baseline Experiment C, which included decontextualized Greek sentences, the generally null results are particularly informative: naïve learners seemed unable to locate informative parts in the sentences and thus showed similar reaction times between conditions. The only difference in reaction times for Noun 2 involved the Agent manipulation. When both nouns were possible good agents for the given verb, reaction times on the second noun increased. This can reasonably be taken to indicate confusion, or inability to confidently assign thematic roles. When only one of the two Nouns was a good Agent, then reaction times decreased, possibly because learners employed heuristics. The fact that there was no interaction with trial order further supports that longer reaction times in this case indicate confusion; if they had decreased through the task, this would have indicated a process of learning, or committing to some interpretation. Instead, we observe uncertainty to assign participant roles when both nouns present were equally plausible agents.

In Experiment B, which included translations, learners were using the English information while reading the Greek sentence. We see this as structure effects on Noun 1 and the Verb, and changes throughout the experiment of reaction times on the Nouns. OVS trials had longer reaction times on Noun 1 and the Verb, which suggests that participants experienced a slow-down in reading as their expectations of word order (influenced by their L1) were violated. Interestingly, we did not see real-time evidence of noticing the determiners, as their reaction times did not differ between conditions, neither did they change over the course of the task. Instead, we observed differences between conditions and changes in reaction times through the task on the Nouns. Translations offered concrete, unambiguous information about the meaning of the Greek sentences. The fact that we do not have evidence for attention paid to the determiners could indicate good-enough approaches where learners identified the correct vocabulary items and ignored less salient, and at this point redundant, case morphology. A different explanation could relate to the study design: since this was a cumulative self-paced reading task, at least a number of participants may have been pressing the buttons without fully processing each word, and then rereading previous parts once the whole sentence was revealed.

Experiment A, which included visual information, is the only place where we have some evidence on real time noticing of the determiners (Determiner 1). Specifically, the interaction between structure and trial order of Determiner 1 reaction times suggests some type of 'learning.' Learners appeared to notice the different determiners and to change the amount of attention paid to them throughout the task, possibly as they understood their function and learned a new form-meaning morphological mapping. In the ‘difficult’ OVS trials, they initially paid more attention (i.e., took longer to react) to the first determiner, but this effect decreased as they progressed through the task. The same effect was found on the Verb and marginally on the whole sentence reading times. The fact that OVS reaction times were higher but through repeated exposure got reduced suggests an attentional shift, presumably the result of noticing and learning relevant linguistic cues in the input, whereas attention in the SVO trials remained fairly constant (besides smaller habituation effects) throughout the task.

Taken together, the different results from the 3 SPR experiments suggest that the context where the L2 is presented can influence which elements in the input get noticed. In Experiment A (images), attentional shifts on the first Determiner, the Verb, and whole sentence reading times suggest a process of noticing linguistic cues to participant role assignment. This learning process is theorized to be the result of error signals generated when an initial misinterpretation of OVS trials clashes with the available images and forces reanalysis of the Greek sentences. Then, learners need to search for alternative cues to word order as more reliable indicators of participant roles. In Experiment B, we had similar findings, but reaction times changed on the Nouns and Verb, not on the Determiners. However, I view this result with caution considering that there may have been rereading involved not captured by button press reaction time data. This becomes more plausible if we consider the high accuracy scores learners achieved in Experiment B, presented in the next section.

Task 3: SPR and sentence-picture matching: Analysis and Results

During this task, participants saw two images and read a Sentence in Greek presented underneath them with cumulative SPR. The sentences were similar to those in the learning phase, but they were taken from the opposite list. Participants were instructed to select the correct image which corresponded to the sentence. Reaction time results and accuracy are presented below. I report reaction times on Determiners 1 and 2, Nouns 1 and 2, the Verb, the Last Word, as well as whole sentence reading times. The model syntax was the same as in the Learning phase SPR models: log-transformed reaction times were the dependent variable, and predictors were Agent (one good, two good, implausible[4]), Structure (SVO-OVS), log-transformed trial order, and their 3-way interaction. The same predictors were used in a binomial mixed effects logistic regression with a log link and Accuracy as the dependent variable. The results are reported below in log-odds. Trial order was always significant, such that participants became faster as they progressed through the task. I will not report the main trial order effect separately for each measure.

Experiment A (Images during the learning phase)

Determiner 1. There were no significant differences between conditions in the reaction times for the Determiner 1.

Noun 1. There was a main effect of Agent on Noun 1 reaction times, with implausible agents having longer reaction times ($\beta = 0.48$, $t = 2.21$, $p < .05$) than one or two good Agents. This effect was qualified by an interaction between Agent and trial order ($\beta = -0.15$, $t = -2.05$, p <

[4] As a reminder, there were 2 Agent conditions during the learning phase (one good, two good) but 3 Agent conditions during the sentence-picture matching (one good, two good, implausible).

.05), such that reaction times to implausible agents reduced as participants progressed through the experiment. This decrease was sharper than in the other Agent conditions.

Determiner 2. There were no significant differences between conditions in the reaction times for the Determiner 2.

Noun 2. There were no significant differences between conditions in the reaction times for the Noun 2.

Last Word. There were no significant differences between conditions in the reaction times for the Last Word.

Verb. There were no significant differences between conditions in the reaction times for the Verb.

Sentence. There were no significant differences between conditions in the reaction times for the Sentence.

Accuracy. There was a main effect of Structure on Accuracy, with OVS trials being less likely to be accurately answered than SVO ($\beta = -2.55$, $z = -2.56$, $p < .05$). There was also an effect of Agent, with implausible agents being less likely to be answered accurately than the one or two good agent conditions ($\beta = -1.98$, $z = -2.4$, $p < .05$). There was also a 3-way interaction illustrated in Figure 3.7. Interestingly, accuracy improved through the trials for the implausible agent condition but deteriorated sharply for the two good agents condition.

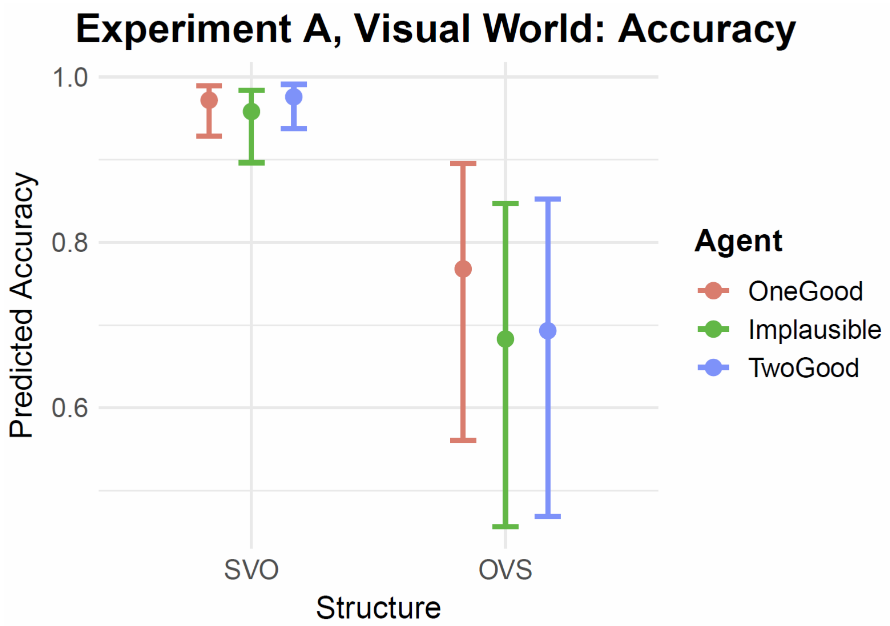


Figure 3.7. Experiment A: Accuracy.

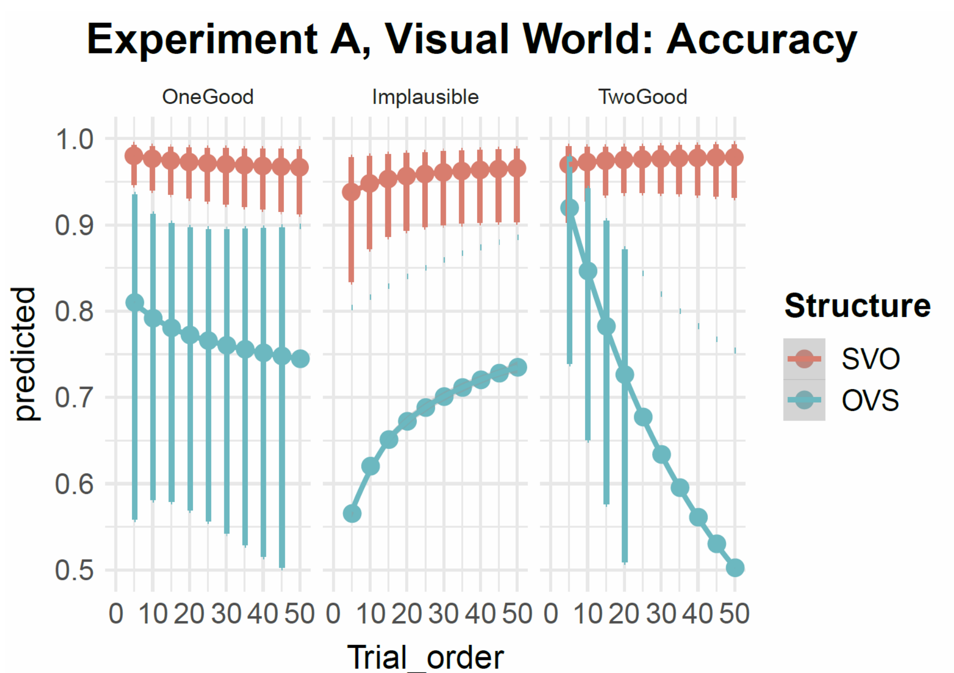


Figure 3.8. Accuracy in sentence-picture matching, by structure, Agent condition, and trial order.

Experiment B

Determiner 1. There were no significant differences between conditions in the reaction times for the Determiner 1.

Noun 1. There was a main effect of Structure on Noun 1 reaction times, with OVS trials having shorter reaction times than SVO ($\beta = -0.56$, $t = -2.79$, $p < .01$). This was qualified by an interaction between Structure and trial order ($\beta = 0.17$, $t = 2.61$, $p < .01$), such that Noun 1 reaction times decreased as a function of trial order in SVO sentences but were fairly stable (and lower initially) in OVS, a surprising finding.

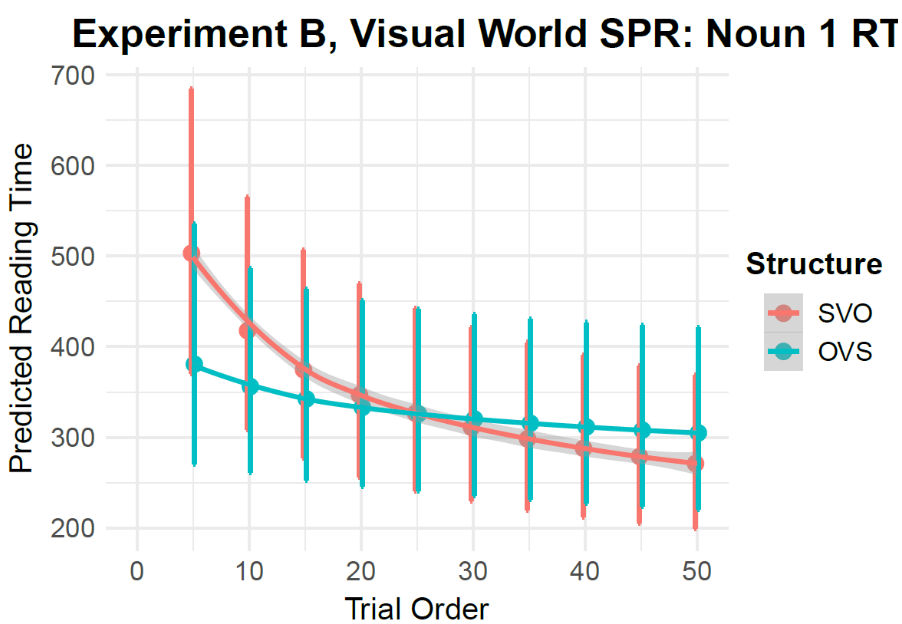


Figure 3.9. Noun 1 RTs: interaction between Structure and trial order

Determiner 2. There was an interaction of Structure and Agent on Determiner 2 reaction times ($\beta = -0.66$, $t = -2.08$, $p < .05$), such that in the OVS condition, reaction times on Determiner 2 did not differ between agent conditions; however, in SVO trials, the two good agent condition had inflated reaction times.

Noun 2. There were no significant differences between conditions in the reaction times for the Noun 2.

Last Word. There were no significant differences between conditions in the reaction times for the Last Word.

Verb. There were no significant differences between conditions in the reaction times for the Verb.

Sentence. There was an effect of Agent on whole sentence reading times, with two good agent trials having longer reading times than the other conditions ($\beta = 0.46$, $t = 2.09$, $p < .05$).

Accuracy. There was an effect of Structure on Accuracy, with OVS trials being less likely to be answered correctly than SVO ($\beta = -2.57$, $z = -2.17$, $p < .05$). This was qualified by an interaction between Structure and Trial order ($\beta = 0.97$, $z = 2.44$, $p < .05$) such that accuracy in OVS trials was lower in earlier trials and improved, whereas accuracy in SVO started and remained high.

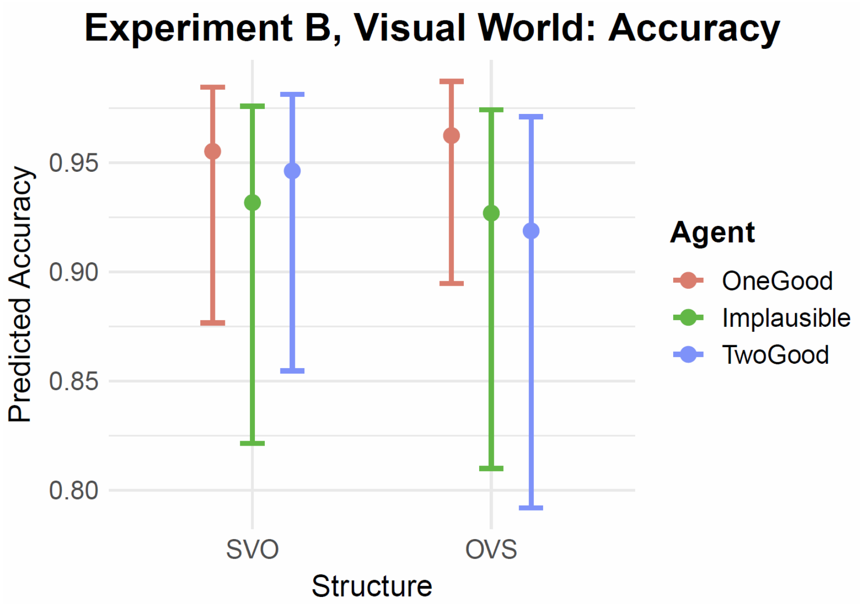


Figure 3.10. Experiment B accuracy

Experiment C

Determiner 1. There were no significant differences between conditions in the reaction times for the Determiner 1.

Noun 1. There were no significant differences between conditions in the reaction times for the Noun 1.

Determiner 2. There was an effect of Agent, with two good agents having longer reaction times on Determiner 2 than the other two agent conditions ($\beta = 0.53$, $t = 2.11$, $p < .05$).

Noun 2. There was an effect of Agent, with two good agents having longer reaction times on Noun 2 than the other two agent conditions ($\beta = 0.83$, $t = 2.61$, $p < .05$).

Last Word. There were no significant differences between conditions in the reaction times for the Last Word.

Verb. There were no significant differences between conditions in the reaction times for the Verb.

Sentence. There were no significant differences between conditions in the reading times for the whole sentence.

Accuracy. There was an effect of Structure on Accuracy, with OVS trials being less likely to be answered correctly than SVO (β = -2.69, z = -3.21, p < .01).

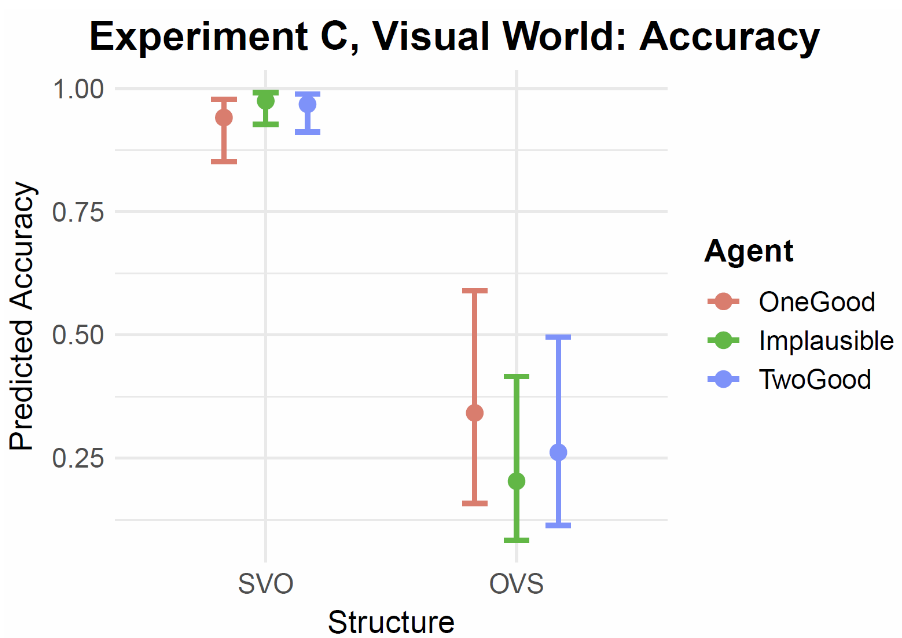


Figure 3.11. Experiment C Accuracy

Task 3: SPR and sentence-picture matching: Discussion

Starting with the baseline Experiment C which offered no contextual support during the learning phase, there is evidence that learners, as a group, generally failed to notice and learn the morphological case markers. The only effects are found on the second NP in the sentence: reaction times are inflated on the second determiner and noun when the sentence nouns are both equally good agents. This suggests possible confusion for the learners, who are not able to use the heuristic “goodness of agent” in this case and, presumably, do not rely on the morphology because they have not noticed it or understood its function. This is supported by the very low accuracy in the OVS condition, which is below chance levels and suggests that they interpret OVS sentences as SVO. Participants are still very accurate in the SVO condition, suggesting that in all conditions they rely on word order to assign participant roles.

In Experiment B, reaction time results are not straightforward to interpret. We observe effects of the two good agents condition at the sentence level, and on the Determiner 2 in SVO trials. This indicates an increased processing cost when both nouns are equally good as agents. Possibly, this is the condition that ‘stands out’ the least: one good agent adheres to heuristics, whereas an implausible agent violates them. Two good agents, however, are equally viable options and could incur a delay in the decision-making process. An alternative explanation would involve confusion, but this is unlikely considering the high levels of accuracy achieved. It is worth noting, however, that there is substantial variation in the accuracy scores in both SVO and OVS conditions, which suggests that some participants achieved high and some low accuracy. Finally, Noun 1 reaction times decreased throughout the task only for SVO trials; in OVS trials, reaction times started lower than SVO and remained at the same level throughout. Taken together with the previous findings, there is some evidence in support of a good-enough approach in the OVS trials, which, however, did not negatively affect accuracy. Finally, accuracy started lower for OVS trials and improved through the task, suggesting learning during the sentence-picture matching task, at least for some learners. It is likely that some participants used this task (two images and a sentence) to notice the determiners and started formulated hypotheses, even though they had failed to do so during the learning phase.

In Experiment A, we observe no differences between conditions on the reaction times for determiners or nouns, with the exception of inflated reaction times on the first noun when it is an implausible agent. Considering that this effect is reduced as trials progress, it can be interpreted as surprisal due to the violation of the “plausible agent” heuristic. In this experiment, attention is allocated on determiners during the learning phase but not during the test phase, which suggests that learners have committed to an interpretation of determiner function (correct or not). Accuracy scores in picture selection are at ceiling for SVO trials and vary for OVS, indicating that some individuals correctly learned and applied case morphology to assign participant roles, but some did not. The interaction illustrated in Figure 3.7 is especially puzzling regarding how accuracy changed as a function of trial order: it remained stable for one good agent trials, improved for implausible agent trials, and deteriorated for two good agent trials. For implausible agents, this pattern suggests learning, or recovery from initial surprise/heuristics violation. For two good agents, however, it suggests a possible reconsideration of an initial hypothesis. It appears that at least some participants initially correctly identified the morphological form-meaning mapping but erroneously reconsidered and revised it during the picture selection task. Considering that this was an online (web-based) experiment, this could also suggest fluctuations in attention or distractions.

Structure production task Analysis and Results

Participant productions were coded with 1 or 0 for correct production of the Agent and Patient determiners, as well as the correct Agent and Patient Noun endings. Cases where nouns were missing altogether were coded as N/A and cases where it was not clear which noun was produced as the agent/patient were coded as ‘unclear.’ Other parts of the sentence, including stem noun spelling, were ignored. I then calculated the percentages of correct production of those 4 morphemes.

Agent determiners were produced correctly 71.1% of the time in Experiment A (Images), 86.4% in B (Translations), and 68.8% in C (No support). Patient determiners had similar production rates: 65.5% in Experiment A, 92% in Experiment B, and 70% in C. Agent endings were produced correctly 80% of the time in A, 85.2% in B, and 82.6% in C. Production rates of the patient determiner were lower: 35.5% in A, 41.4 % in B, and 29.7% in C. These results are summarized visually in Figure 3.11 below.

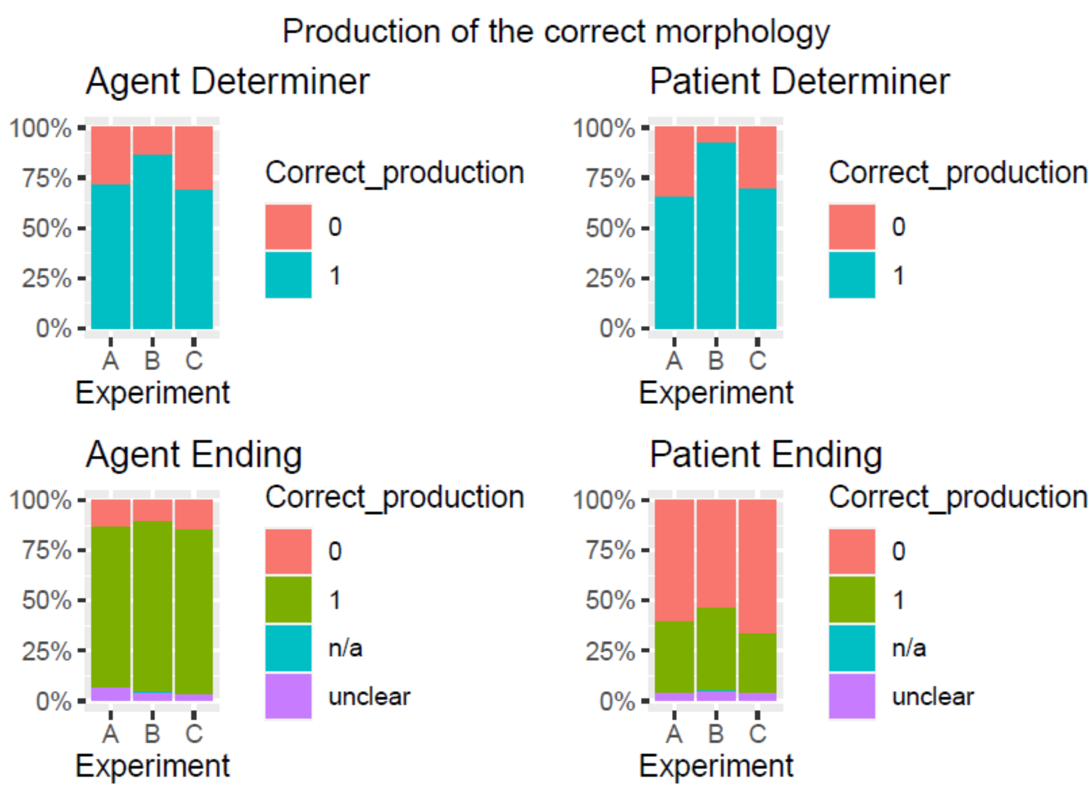


Figure 3.12. Production rates of agent determiner (Top left panel), patient determiner (Top right panel), agent ending (Bottom left panel), and patient ending (Bottom right panel).

Structure production task Discussion

Generally, participants in all experiments noticed the determiners and produced them during the sentence production task. They also correctly produced the agent ending; this was expected, as this was the form that they had seen during the vocabulary pretraining phase. Therefore, accurate production of the final '-s' nominative case marker on the noun probably indicates not structure learning, but accurate memory retrieval of a form. This argument is strengthened by the fact that patient endings, which required an operation to be performed (i.e., removal of the final -s from the already learned nominative form), were produced correctly at a much lower rate.

Task 5: Exit survey Analysis and Results

This task explicitly asked participants what they thought they learned during the Experiment. Participants were first asked to provide detailed information about anything they had noticed. Next, they completed 3 sentence-picture matching questions where they were asked to justify and explain how they made their choices (i.e., how they decided which was the correct picture for the given sentence). Two sentences were taken from each of the following conditions: SVO-one good agent, OVS- two good agents, and OVS-implausible agent. The purpose was to evaluate what each participant noticed and learned through the experiment.

The open-question qualitative responses from the exit survey were coded with 1 or 0 for the following two variables: noticing the determiners and noticing the noun endings. They were also coded for understanding the function and articulating a form-function mapping for the morphosyntax indicating participant roles. In this case, there were 4 categories: 1 for understanding the function, 0 for not understanding (e.g., using only word order to assign participant roles), "Almost" for formulating a hypothesis but an incorrect one, and finally "Implicit," for participants who showed evidence of learning but were unaware of this knowledge or were unable to justify their correct choices. Below I present the percentages of participants who noticed the determiners and noun endings, as well as those who understood the function, by experiment.

56.7% of participants noticed the determiners in Experiment A, 66.7% in B, and 60.9% in C. Noticing of noun endings was lower: 26.7% in Experiment A, 33.3% in Experiment B, and 13% in C. Regarding understanding the function, 46.7% of participants in A showed evidence of understanding, 55.6% in Experiment B, and 34.8 in C. There findings are visualized in Figure 3.11 below.

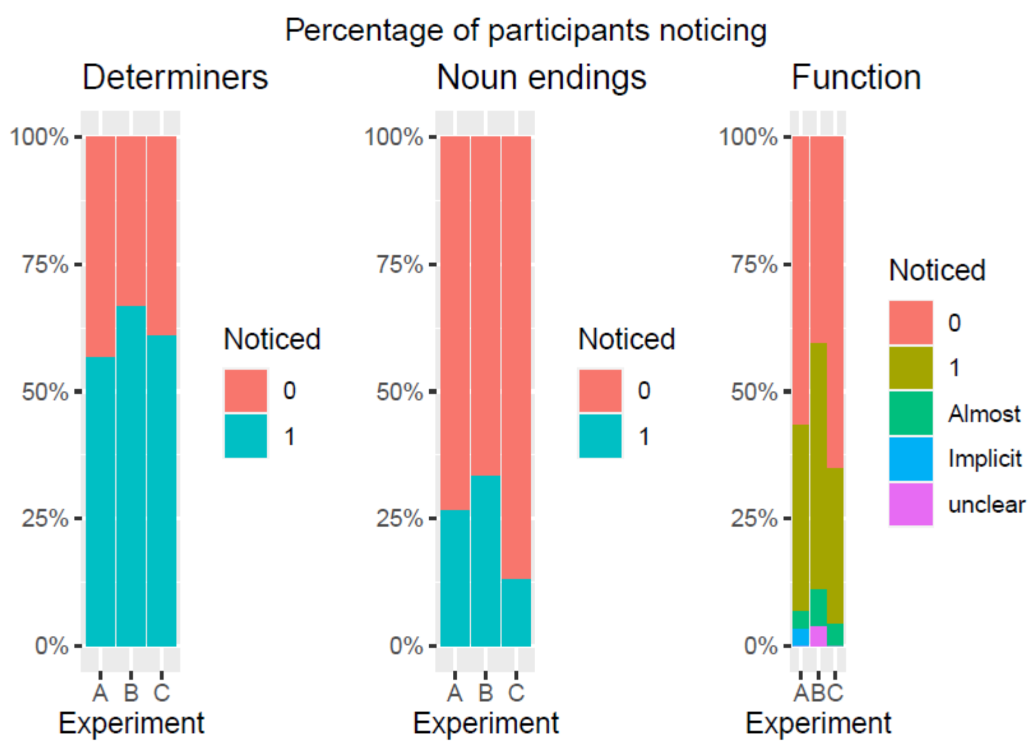


Figure 3.13. Noticing of determiners (Left panel), noun suffixes (Middle panel), and understanding of the form-function mapping (Right panel).

Task 5 Exit survey Discussion

A note of caution needs to be made about these findings. As a result of the experiment being online, the quality of many responses was not ideal: a large number of participants gave laconic responses which were conservatively coded as 0 but may underestimate actual learning. Indeed, if we consider the sentence-picture matching accuracy scores, the percentages of participants who reported noticing case morphology seem too low for Experiments A and B, and too high for C. Regarding experiments A and B, this pattern might be explained by the fact that many participants offered short, even single word accounts of what they learned (e.g., "Greek words").

The results from Experiment C are especially puzzling: participant scores were at floor levels for OVS sentences during the sentence picture matching task, but about a third of participants showed evidence of having understood the function during the exit survey. If that was the case, why did they fail to apply their knowledge during the sentence picture matching task? It is possible that this knowledge was acquired later during the experiment, as the tasks became progressively more explicit. For instance, those successful participants may have started to notice case morphology during the sentence picture matching task. The presence of two images with reversed thematic roles might have focused attention to relevant parts of the stimuli. In addition, the final sentence production and even survey task itself may have triggered a last-minute understanding for those participants.

SPR studies: Summary

Taken together, the results from all tasks (reaction times during structure learning, reaction times and accuracy during sentence-picture matching, sentence production, and exit survey questions), suggest modest to high levels of learning. Results from these tasks were more consistent in

Experiments A and B. Specifically, in those experiments, we observe trial order effects, indicative of participants gradually noticing and using relevant parts of the stimuli to interpret the Greek sentences. This is followed by high accuracy in the picture selection task, as well as modest to high morpheme production accuracy, and awareness of the L2 morphosyntactic rule.

The results from the same tasks were less consistent in Experiment C. First, absent trial order effects and structure effects during the learning phase suggest that participants in this experiment did not differentiate between OVS and SVO trials and processed all with similar levels of effort/difficulty. In addition, the absence of trial order effects, in contrast to Experiments A and B, suggests that attention to morphological markers did not change over the course of the learning task, i.e., there was no learning. This is confirmed by subsequent very low scores in the OVS trials in the sentence-picture matching task. However, the results from the structure production task and the qualitative exit survey suggest that some participants noticed and were able to produce the morphemes. This is especially puzzling, considering that production is a more difficult task than recognition (Ellis, 2005), which is what the sentence-picture matching task required. It is possible that participants in Experiment C did not, in fact, learn the morphosyntactic pattern during the learning phase, but that some participants started learning during the later tasks. This is also possibly for Experiment B, where accuracy scores in the sentence picture matching task started low for OVS trials but improved through the task.

In general, web data collection to study L2 learning was not ideal but it was a practical necessity due to lab closure during the pandemic. Online data collection introduced limitations: it reduced experimenter control and introduced confounds in measuring attention. The results from the SPR studies are thus viewed with caution; however, they offer some preliminary support for improved L2 morphology learning and participant role assignment when the input is presented with contextual support, images or translations. This conclusion is supported by different amounts of attention paid to parts of the stimuli during the learning process under the different conditions, as well as shifts in attention during the ‘difficult’ OVS trials with task progression. In addition, these effects were not observed during the sentence picture matching task, and accuracy scores were high for experiments A and B, but showed reduced learning of OVS participant roles in Experiment C.

The tentative findings from the SPR studies, even with the limitations imposed by online data collection, offer some preliminary support for our hypotheses that contextual support can facilitate L2 morphosyntactic learning. The same questions are explored with eye-tracking methodology and in-person data collection in the next section (Part 4).

Part 4: L2 morphosyntactic learning: an investigation with eye-tracking

The studies in Part 4 replicate the design of the studies in Part 3 (Experiments A and B only, with Images and Translations respectively) using eye-tracking instead of self-paced reading in order to obtain richer data. The in-person data collection also ensures greater control during the learning process.

Eye-tracking is a fitting methodology to study attention to various cues, as it allows us to both locate and quantify the amount of attention paid to parts of the stimulus (Godfroid, Boers, & Housen, 2013) during both sentence comprehension and during sentence-picture integration (Ronderos et al., 2018). Reading times reflect relative ease or difficulty in comprehension (Rayner et al., 2012), and the method relies on the assumed eye-mind link, which suggests that overt attention (where the eyes are fixated) is indicative of covert attention (what gets processed) (Just & Carpenter, 1980; Rayner, 2009; Reichle, Pollatsek, & Rayner, 2006).

Regarding the integration of language and visual context, previous L1 research has suggested a close link between eye movements related to language comprehension and reference assignment (Bock, Irwin, Davidson, & Levelt, 2003; Cooper 1974; Griffin, 2004; Griffin & Bock, 2000; Kuchinsky, Bock, & Irwin, 2011; Meyer & Lethaus, 2004; Meyer, Roelofs & Levelt, 2003). Not many studies have used eye-tracking to examine the unfolding of learning novel L2 structures as a result of processing, but the results of McDonough et al. (2017) indicate that eye movements, as an index of attention, can be successfully used to predict L2 structure comprehension and learning.

Finally, it is important to begin accounting for the fact that L2s are learned in rich environments where linguistic and non-linguistic input co-occur. Only recently has the need to consider non-linguistic cues in L2 development been acknowledged (Ronderos et al., 2018). Understanding the end state of L2 attainment can be better achieved by examining not just the product, but crucially the process of learning in a multimodal, more ecologically valid environment. The combination of eye-tracking and offline measures employed in this study attempts to study 'learning' both as a process and a product, and to establish links between the two.

In two eye-tracking experiments, I compare the online processing of Greek case morphology by naïve learners with visual (Experiment A) or linguistic (L1 translations) contextual support (Experiment B). The two experiments have the same series of tasks presented in Figure 4.1: vocabulary learning, structure learning phase, adapted visual world, sentence production, exit survey, all completed in this order in a single session. All tasks were identical between experiments except for the structure learning phase, where, in addition to Greek sentences, Experiment A offered visual scene support and Experiment B direct English translations. The SPR experiments described in the previous section suggested that when Greek sentences were presented in isolation, without any contextual support, during the structure learning phase, learning was largely not achieved. For this reason, I did not include the "no contextual support" experimental condition for the eye-tracking studies, which are substantially resource-demanding.

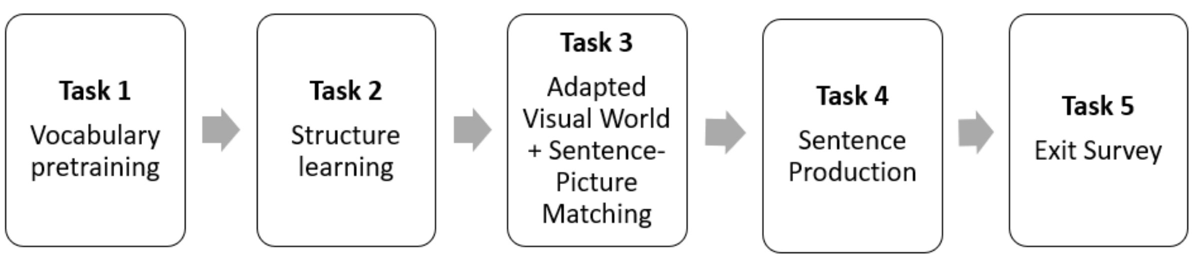


Figure 4.1. Sequence of experimental tasks. All tasks were completed in a single session.

Participants

Participants were 81 L1 English speakers (40 in experiment A and 41 in Experiment B) with no knowledge of languages with case morphology or free word order. Eight additional participants were not included in the analysis due to tracker calibration problems. Participants were recruited on the University of Illinois community and received payment or course credit for their participation. They also received an additional monetary bonus—they were told that the bonus was contingent on performance in the experiment in order to motivate them to pay attention. However, all participants received it regardless of test performance, and they were debriefed at the end of the experiment. They all had normal or corrected to normal vision.

Materials and Procedure

The same materials were used as in the self-paced reading experiments. First, participants learned 15 words in Greek. The words were presented on PowerPoint slides, one at a time, along with the English translation and an image depicting the word or action. To ensure the words had been learned, participants took a vocabulary test (matching the Greek words to the English translations). Participants were allowed to continue with the experiment if they had at the most two errors, to ensure they could indeed recognize the lexical items during reading.

Structure learning phase. During this task, participants read 32 simple transitive Greek sentences presented twice (total 64 trials) while their eye movements were being recorded with an SR Research Eyelink 1000 Plus eye tracker (spatial resolution of 0.01 degrees) sampling at 1000 Hz. Sentences were presented in 36-pt Courier New monospace font. The larger than usual font was used to minimize landing errors between the Greek sentences and the images/translations. In addition, because the morphological case markers are short (for example, the nominative marked determiner is a single letter "o" and the noun suffixes are 2- or 3-letter long, the larger font might discourage processing of these in the parafovea. Pilot testing confirmed that the font size did not disrupt normal reading or cause surprise; in fact, it was reported that it made reading easier and put less strain on the eyes. Subjects were seated 70 cm away from a monitor with a display resolution of 1800 x 1200. Head movements were minimized with a chin and forehead rest. Viewing was binocular but eye movements were recorded only from the right eye. The experiment was controlled with SR Research Experiment Builder software, and participants responded using a button box.

During the structure-learning phase, participants were instructed to read Greek sentences presented one at a time on a single line either under an image (Experiment A[5]) or under an equivalent English translation. (Experiment B). The stimuli crossed Structure (SVO-OVS) and Agent (one good- two good) in a 2 x 2 Latin square design, resulting in 32 items distributed in 2 lists and presented twice in a fixed order. Participants first looked at the image or read the English sentence at their own speed, and then pressed a button to read the Greek sentence, which was added on the screen underneath the image/translation. Data analysis involved only viewing times from the button press onward. When they were done reading, participants pressed a button to move to a new item. It was emphasized to participants that the images always matched the Greek sentences, or that the English sentences were faithful translations. No researcher-generated feedback was given during the task, and participants did not perform any comprehension checks.

[5] The images were the same as in the SPR Experiment.

Adapted visual world. After the structure-learning phase, participants completed a sentence-picture matching task while their eye movements were again being recorded. In this task, they first saw two images on the left and right top of the screen. The images depicted the same event but had reversed participant roles (e.g., *a doctor chasing a farmer* and *a farmer chasing a doctor*) and they were the same as in the SPR experiments. Participants looked at the pictures at their own speed and pressed a button to reveal the Greek sentence underneath the pictures. They were instructed to read the sentence and using the button box, to select the correct picture. This is an adaptation of the visual world paradigm, which allowed the recording of both eye movement and accuracy data. The stimuli in this task were similar to those in the learning phase with the addition of an implausible agent condition. The stimuli were taken from the counterbalanced list of the learning phase. The stimuli crossed Structure (SVO- OVS), Agent (one good, two good, implausible) in a 2 x 3 design and a total of 48 items distributed over 2 lists. Presentation order was fixed. I note that the implausible agent condition was not present during the structure learning phase, but all other conditions were.

Sentence production. After the visual world task, participants completed a sentence production task (the same as in the self-paced reading experiments) where they were shown 6 images one at a time and were asked to write a sentence in Greek to describe it. Two images depicted actions with one good agent, two with two good agents, and two with an implausible agent.

Exit survey. Finally, participants completed an open-ended survey task which explicitly targeted what they had learned during the experiment. This task included two general, open-ended questions asking participants to describe what they had noticed during the previous tasks. They were also shown pairs of pictures taken from the visual world task along with a Greek sentence underneath (taken from the following conditions: SVO + one good agent, OVS + two good agents, OVS + implausible agent; three in total). Participants were asked to explain in detail their thinking process in choosing the correct picture.

Analysis

Structure Learning Phase

The data from the structure learning phase were analyzed separately for the two experiments. The main goal was to identify differences in attention between conditions as well as patterns in attentional shifts through the task, which necessitate modeling 3-way interactions. Adding Experiment as a predictor would result in 4-way interactions, which are notoriously difficult to interpret and would require a much larger number of participants for statistical power. I analyzed eye movement data on the following interest areas: Determiner 1, Noun 1, Verb, Determiner 2, Noun 2, Contextual support (this was either the Image in Experiment A, or the English sentence in B).

Considering that participants were reading sentences in an entirely new language, I analyzed only late eye movement data: total reading time (TT: the sum of all fixations, including fixations resulting from regressions, on a word) and go-past time (GPT: the sum of all fixations on a word from first entering the word until exiting in the forward direction, including fixations on prior regions resulting from regressions originating in the current word). These measures reflect semantic processing and information integration at the sentence level as they include rereading of the word, in the case of TT, and rereading of previous areas, for GPT (Libben & Titone, 2009; Rayner, 2012). I also report the probability of regressions out of, and into the interest areas. Continuous reading measures were analyzed with linear mixed effects models using the lme4 and lmerTest packages in the R environment. One model was built for each measure in each interest area. The dependent variable was the log-transformed measure (TT or GPT) and the predictors were Structure, good agents (one good or two good), the log-transformed trial order and their 3-way interaction. I used treatment coding for the categorical predictors with SVO and one good agent as the baseline conditions. Subjects and Items were included as random intercepts; adding random slopes either did not allow for model convergence or it did not result in a significantly improved model. All model outputs can be found in the appendices.

Experiment A

Analysis/Results

In all the models, there was always a main effect of trial order, such that total times decreased as participants progressed through the task. This effect was consistent within and across experiments and will not be presented separately for each model unless it enters into an interaction. The descriptive statistics for the Experiment A structure learning task are reported in table 4.1 below.

Table 4.1. Experiment A: Descriptive statistics (mean, sd) of reading times during the learning phase in milliseconds.

Total time			**Go past time**	
		Determiner 1		
	1 good agent	2 good agents	1 good agent	2 good agents
Structure				
SVO	587.31 (671.55)	593.45 (659.40)	510.95 (469.28)	488.07 (527.03)
OVS	1064.22 (998.28)	1029.92 (1049.00)	645.15 (775.61)	590.92 (598.51)
		Noun 1		

Table 4.1 (cont.)

SVO	1706.86 (1488.50)	1817.80 (1805.99)	1344.16 (1255.78)	1319.72 (1326.22)
OVS	1667.57 (1608.17)	1674.45 (1532.76)	1190.80 (1159.50)	1206.69 (1300.97)
		Verb		
SVO	1409.66 (1348.35)	1296.12 (1390.91)	1212.00 (1357.96)	1189.39 (1732.87)
OVS	1765.25 (1857.30)	1453.73 (1473.91)	1348.75 (1474.95)	1092.08 (1279.48)
		Determiner 2		
SVO	847.75 (937.96)	872.98 (897.96)	684.23 (954.51)	641.37 (938.53)
OVS	591.74 (656.89)	627.45 (748.98)	670.49 (1259.72)	651.79 (963.61)
		Noun 2		
SVO	1227.49 (1114.47)	1286.19 (1309.45)	1155.85 (1875.86)	1285.32 (2038.95)
OVS	1622.91 (1521.50)	1609.96 (1616.02)	1445.30 (1932.69)	1596.72 (2191.51)
		Picture		
SVO	982.34 (1477.41)	1148.52 (1535.65)		
OVS	1072.41 (1362.73)	1318.09 (1661.18)		
		Whole trial		
SVO	8548.25 (6336.67)	8749.51 (7197.44)		
OVS	9350.10 (7581.05)	9159.76 (7560.32)		

Total Time

Determiner 1. There was an effect of Structure, with OVS trials having longer total time on Determiner 1 than SVO (β = 1.2, t = 7.17, p < .001). This was qualified by an interaction between Structure and trial order (β = -0.15, t = -2.98, p < .01) such that TT started higher for OVS but decreased throughout the experiment; for SVO trials, TT on the first determiner remained the same throughout the task (Figure 4.1).

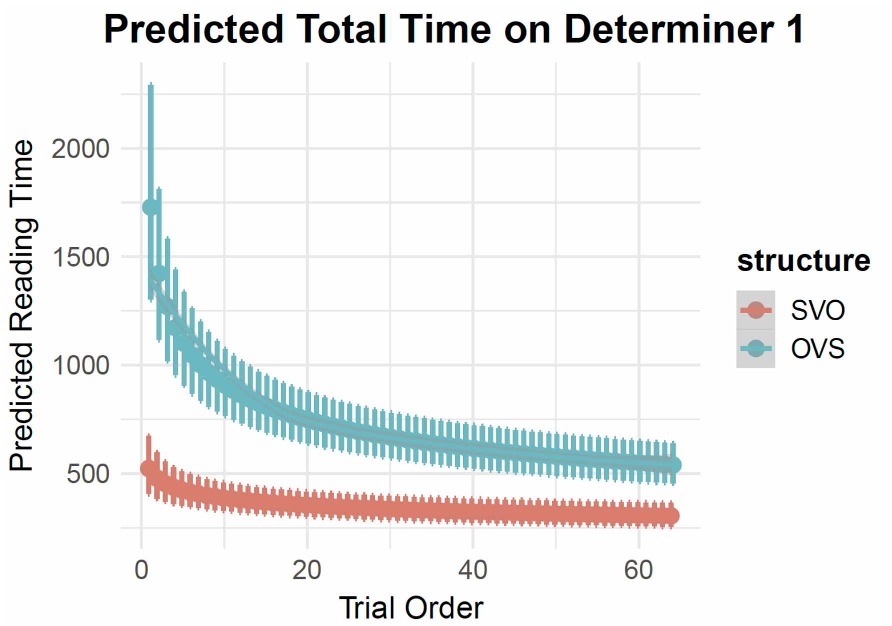


Figure 4.2. Total times on Determiner 1.

Noun 1. There was an effect of Structure with OVS trials having longer TT on Noun 1 than SVO (β = 0.48, t = 2.94, p < .01), qualified by an interaction between Structure and trial order (β = -0.16, t = -3.42, p < .001) such that the OVS TT decreased through the experiment whereas SVO TT was constant. There was also an effect of Agent: the two good agents condition had longer TT than the one good agent condition (β = 0.31, t = 2.04, p < .05), again qualified by an interaction between Structure and trial order (β = -0.09, t = -2.07, p < .05).

Verb. There was an effect of Structure, with OVS trials having longer total time on the Verb than SVO (β = 0.72, t = 4.34, p < .001). This was qualified by an interaction between Structure and trial order (β = -0.16, t = -3.43, p < .001) such that TT started higher for OVS but decreased throughout the experiment; for SVO trials, TT on the verb remained the same throughout the task.

Determiner 2. There were no differences between conditions on the TT spent on the second determiner.

Noun 2. There were no differences between conditions on the TT spent on the second noun.

Trial. There was an effect of Structure with OVS trials having longer TT than SVO ($\beta = 0.41$, $t = 4.0$, $p < .001$), qualified by an interaction between Structure and trial order ($\beta = -0.2$, $t = -4.18$, $p < .001$) such that the OVS trial TT decreased through the experiment whereas the SVO was fairly constant. There was also an effect of Agent: the two good agents condition had longer TT than the one good agent condition ($\beta = 0.28$, $t = 2.76$, $p < .01$), again qualified by an interaction between Agent and trial order ($\beta = -0.09$, $t = -2.98$, $p < .001$) such that the two good agents condition had a sharper decrease in TT than the one good agent condition. These results are illustrated in Figure 4.2.

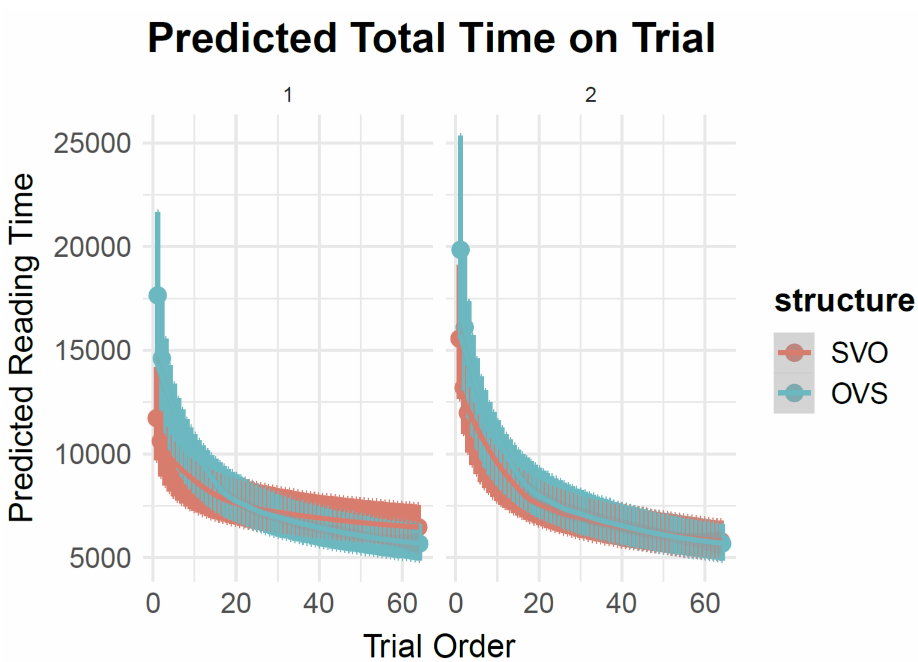


Figure 4.3. Total time on the trial. Left panel: 1-good Agent. Right panel: 2-good Agents.

Image. The TT on the image includes only fixations when both the image and the Greek sentence were on the screen; it does not include viewing times prior to the appearance of the Greek sentence. There was an effect of Agent: the two good agents condition had longer TT on the image than the one good agent condition ($\beta = 0.52$, $t = 2.73$, $p < .01$), qualified by an interaction between Agent and trial order ($\beta = -0.11$, $t = -1.97$, $p < .05$), such that TT on the image in the two good agents condition decreased through the experiment.

Go-past time

Go-past time (GPT) is a measure that includes rereading of previous areas, i.e., it is the sum of fixations on the target word or region prior to moving off of it to the right and including all fixations on previously read text resulting from leftward saccades out of the target region. Analyses of GPT thus allows us to compare how much rereading was triggered by different conditions.

Determiner 1. There was only an effect of Structure, with OVS trials having longer GPT than SVO ($\beta = 0.47$, $t = 3.1$, $p < .01$).

Noun 1. There were no differences between conditions on the GPT on the first noun.

Verb. There was an interaction between Structure and Agent ($\beta = -0.7$, $t = -2.24$, $p < .05$), with OVS trial verbs having shorter GPT in the two good agents condition. This was qualified by a 3-way interaction with trial order ($\beta = 0.18$, $t = 2.01$, $p < .05$).

Determiner 2. There were no differences between conditions on the GPT on the second determiner.

Noun 2. There was an effect of Structure, with OVS trials having longer GPT on Noun 2 than SVO (β = 0.57, t = 2.87, p < .01). There was also an effect of Agent (β = 0.43, t = 2.31, p < .05), with two good agents having longer GPT than one good agent.

Regressions

Regressions into the image were more frequent in the two good agents condition (β = 0.37, t = 2.55, p < .05).

Regressions into as well as out of Determiner 1 were more frequent in the OVS condition, but this decreased with trial order, i.e., as participants progressed through the task. The same pattern emerged for Noun 1, Verb, Determiner 2, and Noun 2.

Regressions into Noun 1 and the Verb were also more frequent in the two good agents condition, an effect that again decreased as the experiment progressed.

Skipping

I also analyzed the probability of skipping the two nouns' endings, which contain the morphological case markers. The noun ending was defined as the last character of the noun's stem, plus the additional 2 (nominative) or 1 (accusative) characters. For the first Noun, there was only an effect of trial order: skipping became less likely as the experiment progressed (β = -0.24, t = -2.55, p < .05). For Noun 2, there were no differences between conditions and no trial order effects. As shown in Figure 3, the probability of skipping the Noun's case morphology is very high (from about 55% - 75%).

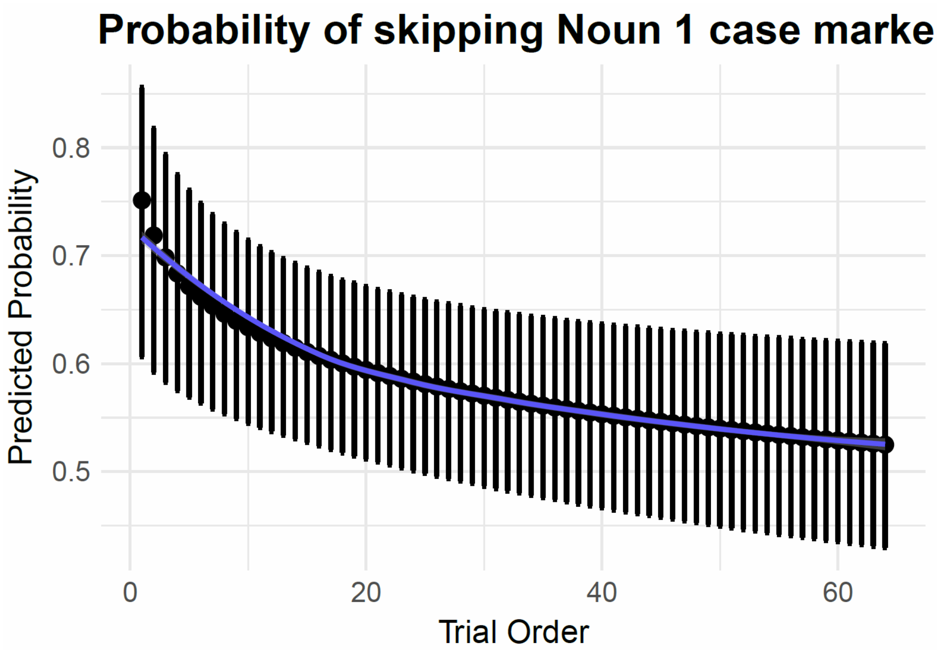


Figure 4.4. Probability of skipping the Noun case morphology.

Discussion of Experiment A

The learning phase results illustrate the learning process in terms of quantifying attention paid to relevant parts of the stimuli, and, more importantly, in terms of showing how the attention changed as the task progressed. TTs on the first NP and the verb were initially inflated in OVS trials but reading times decreased as participants were exposed to additional stimuli. This follows the prediction that such trials would be difficult for L1 English speakers, as they violate the L1 syntactic pattern. The interaction with trial order is particularly interesting as it suggests that through continued exposure, learners' attention to the relevant language features became similar for both structure types. This indicates some type of learning, in the sense of committing to a hypothesis about the function of these linguistic features. The accuracy of their hypotheses/learning will be evaluated in the following sections.

The Agent effects on the verb and the whole trial TTs indicate a similar process of learning as described above. When both nouns in the sentence were good (plausible) agents, reading times on the verb and the whole trial in general, increased. This effect, however, went away with continued exposure. Inflated reading times in the two good agents condition points to the inability to rely on heuristics for participant role assignment. When only one of the two nouns was a good agent, learners were able to use heuristics and world knowledge and were not slowed down (which would be reasonable to expect in the OVS + one good agent condition which violated their L1-induced expectations). Instead, the slow-down occurred when reliance on this heuristic was no longer possible. In fact, this is the condition where viewing times on the Image were also inflated (at the beginning of the experiment), presumably as learners searched for more reliable cues than goodness-of-agent to assign participants roles. The difference in the viewing times on the image between Agent conditions diminished throughout the task, which supports the argument that learners found and relied on different cues after continued exposure to stimuli.

GPT is another measure that yielded interesting results. It is a measure that includes rereading, and this analysis illustrates which areas triggered rereading and in which conditions. Determiner 1 triggered more rereading in the OVS condition. Considering that the only regions prior to the determiner are one adverbial and the image, it is safe to assume that 'rereading' in this case is equivalent to revisiting the image. GPT on the verb was shorter in the OVS + two good agents condition, an effect that went away with continued exposure. Interestingly, even though this was the most difficult condition, it triggered less rereading originating from the verb. One possible explanation is that learners relied on different parts of the sentence to make meaning. In fact, the second noun appears to be that area, as it triggered more rereading in OVS and in two good agent trials.

Analysis of regression frequency adds further support to the previous arguments. Regressions into and out of the regions of interest were generally more frequent in the OVS condition; the interaction with trial order suggests that as they progressed through the task, learners were less surprised. Reduced frequency of regressions indicates that cognitive effort diminished with continued exposure, assuming that regressions are an indicator of cognitive effort (Vasishth & Drenhaus, 2009). Regressions into the image were more frequent in the two good agents condition, and for this effect, there was no interaction with trial order. It seems that when linguistic processing became more complex because it was not supported by heuristics, learners relied on the visual cues. This is not necessarily a learning process, but rather a processing or inferential strategy, as it remained consistent throughout the task, whereas other effects disappeared after some initial exposure. Specifically, the two good agents condition had more frequent regressions into the first noun and the verb (possibly originating from the second noun), but this effect was reduced after some trials. In sum, in the two good agents condition, there was a difference between regressions into the image and into the Noun1 and the Verb. Regressions into the image remained consistently more frequent in this condition throughout the task, but regressions into the Noun1 and the Verb were reduced. This suggests that reliance on the image was a processing strategy, whereas the shift of attention on the early sentence parts indicates initial confusion, search for meaning, and learning.

Finally, the very high probability of skipping Noun case morphology is in line with SLA accounts of salience, redundancy and shadowing. Reading times on the determiners were substantial, even though the nominative/Agent case marker is a single letter. Determiners are more salient (they are separate words) and once they have been noticed, noun case markers are redundant; therefore, it is unsurprising that the latter were not fixated. Another possible explanation might be that they got processed in the parafovea; nevertheless, results from the qualitative survey suggests that most learners remained unaware of their existence (see following section for analysis of the qualitative data and noticing). Therefore, these skipping patterns add support to the argument that one reason L2 case morphology is difficult for learners is that it is non-salient and redundant, which leads to learners not noticing it. The current experiment included written stimuli in a large font, yet noun endings were still generally not fixated. In aural language presentation, the salience of case morphology would be even further diminished.

Experiment B

Analysis/Results

The descriptive statistics for the reading measures for the learning phase of Experiment B are presented in table 4.2 below.

Table 4.2. Experiment B: Descriptive statistics (mean, sd) of reading times during the learning phase in milliseconds.

Total time			**Go past time**	
		Determiner 1		
	1 good agent	2 good agents	1 good agent	2 good agents
Structure				
SVO	433.75 (381.37)	426.40 (342.76)	477.03 (557.50)	454.08 (688.91)
OVS	931.82 (1246.90)	852.20 (786.40)	611.58 (660.25)	576.57 (678.20)
		Noun 1		
SVO	1381.86 (1426.51)	1389.54 (1207.34)	1141.07 (1101.31)	1117.10 (1148.91)
OVS	1312.22 (1152.89)	1246.30 (856.73)	1007.87 (1109.70)	1076.50 (1174.40)

Table 4.2 (cont.)

		Verb		
SVO	1031.76 (896.92)	909.45 (746.52)	1141.07 (1101.31)	1117.10 (1148.91)
OVS	1232.35 (1137.64)	1006.57 (854.73)	1007.87 (1109.70)	1076.50 (1174.40)
		Determiner 2		
SVO	611.34 (538.10)	612.52 (509.10)	600.55 (1113.99)	683.94 (1117.63)
OVS	410.09 (349.37)	378.39 (309.35)	645.63 (992.05)	676.59 (1061.93)
		Noun 2		
SVO	915.67 (852.33)	947.06 (796.98)	977.35 (1400.87)	1130.08 (1603.12)
OVS	1161.75 (996.27)	1141.34 (912.32)	1460.98 (2272.27)	1355.39 (1692.89)
		English Sentence		
SVO	1962.35 (2260.70)	2186.03 (2662.44)		
OVS	2292.95 (2760.73)	2294.60 (2739.00)		
		Whole trial		
SVO	7697.18 (5597.45)	7851.28 (5680.70)		
OVS	8685.14 (6800.05)	8137.39 (6165.28)		

Total Time

Determiner 1. There was an effect of Structure, with OVS trials having longer total time on Determiner 1 than SVO (β = 1.18, t = 7.9, $p < .001$). This was qualified by an interaction between Structure and trial order (β = -0.14, t = -3.12, $p < .01$) such that TT started higher for OVS but decreased throughout the experiment; for SVO trials, TT on the first determiner remained the same throughout the task.

Noun 1. There was an effect of Structure, with OVS trials having longer total time on Noun 1 than SVO (β = 0.55, t = 3.85, p < .001). This was qualified by an interaction between Structure and trial order (β = -0.2, t = -4.47, p < .001) such that in OVS trials, TT on the first noun decreased as the experiment progressed, whereas in SVO they remained the same. There was also a 2-way interaction between Structure and Agent, qualified by a 3-way interaction with trial order. These interactions are illustrated in Figure 4.4 and show that the difference between SVO-OVS was larger with one good agent, and the decrease was the sharpest in the OVS one-good agent condition.

Figure 4.5. Total Time on Noun 1. Left panel: 1-good Agent. Right panel: 2-good Agents.

Verb. There was an effect of Structure, with OVS trials having longer total time on the verb than SVO (β = 0.76, t = 5.16, p < .001). This was qualified by an interaction between Structure and trial order (β = -0.19, t = -4.37, p < .001) such that in OVS trials, TT on the verb decreased as the experiment progressed, whereas in SVO it remained the same. There was also a 2-way interaction between Structure and Agent, qualified by a 3-way interaction with trial order. These interactions are illustrated in Figure 4.5 and show that there was a difference between SVO-OVS verb TT only in the one-good agent condition, and TT decreased for OVS trials but remained constant for SVO as the experiment progressed.

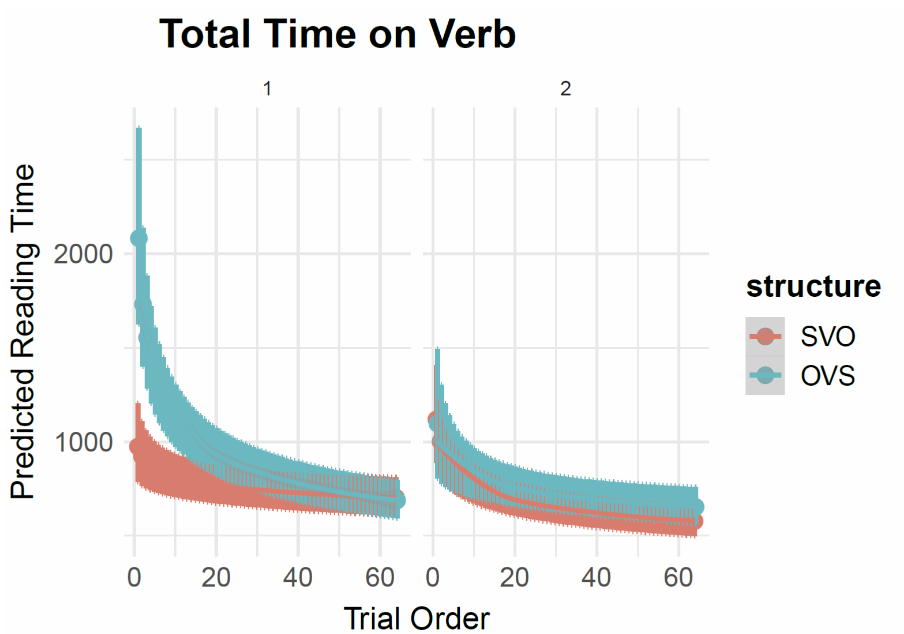


Figure 4.6. Total Time on the Verb. Left panel: 1-good Agent. Right panel: 2-good Agents.

Determiner 2. There was an effect of Structure, with OVS trials having shorter TT on the second determiner than SVO (β = -0.48, t = -2.95, p < .01).

Noun 2. There was an effect of Structure, with OVS trials having longer TT on the second noun than SVO (β = 0.43, t = 2.93, p < .01).

Trial. There was an effect of Structure, with OVS trials having longer TT than SVO (β = 0.48, t = 4.45, p < .001). This was qualified by an interaction between Structure and trial order (β = -0.14, t = -4.39, p < .001) such that TT in OVS trials decreased as the experiment progressed, whereas TT in SVO trials remained the same.

English translation. There was an effect of Structure, with OVS trials having longer TT on the English translation than SVO ($\beta = 0.94$, $t = 4.78$, $p < .001$). This was qualified by an interaction between Structure and trial order ($\beta = -0.24$, $t = -4.01$, $p < .001$) such that TT on the English translation in OVS trials decreased as the experiment progressed, whereas TT in SVO trials remained the same. There was also an effect of agent, with two good agents having longer TT than one good agent ($\beta = 0.53$, $t = 2.86$, $p < .01$), also qualified by a 2-way interaction with trial order ($\beta = -0.13$, $t = -2.34$, $p < .05$), suggesting that the inflated TT with two good agents decreased through the task. These are illustrated in Figure 4.6.

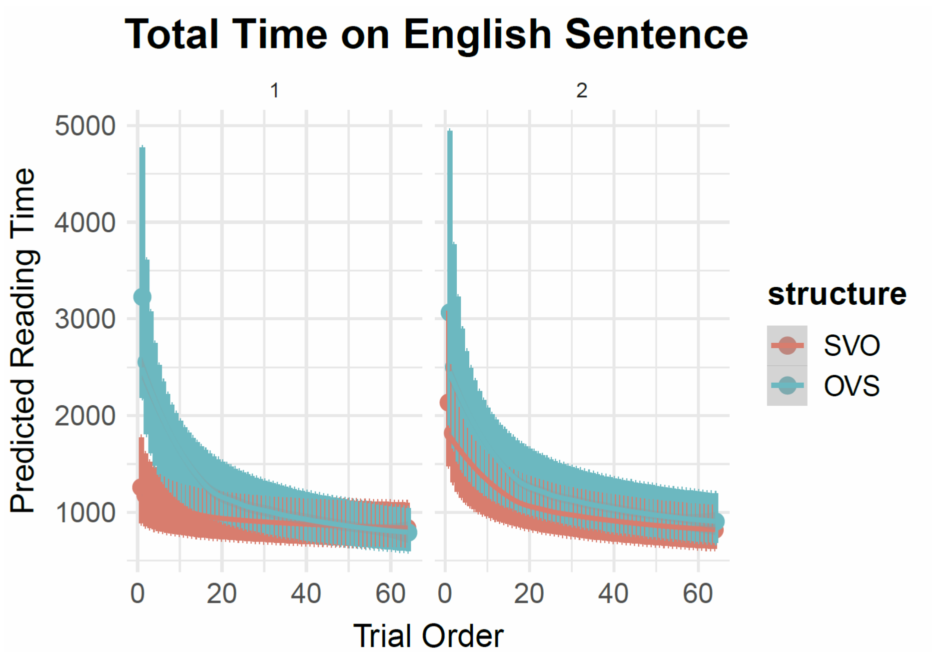


Figure 4.7. Total time on English translations. Left panel: 1-good Agent. Right panel: 2-good Agents.

Go Past Time

Determiner 1. There was an effect of Structure, with OVS trials having longer GPT on the first determiner ($\beta = 0.47$, $t = 2.94$, $p < .01$).

Noun 1. There were no differences between conditions on the GPT on the first noun.

Verb. There was an interaction between Agent and Structure (β = -0.73, t = -2.43, p < .05), qualified by an interaction with presentation order. Figure 4.7 shows that there was a difference between SVO and OVS verb GPT only in the two good agent condition; OVS started lower than SVO but changed as the experiment progressed.

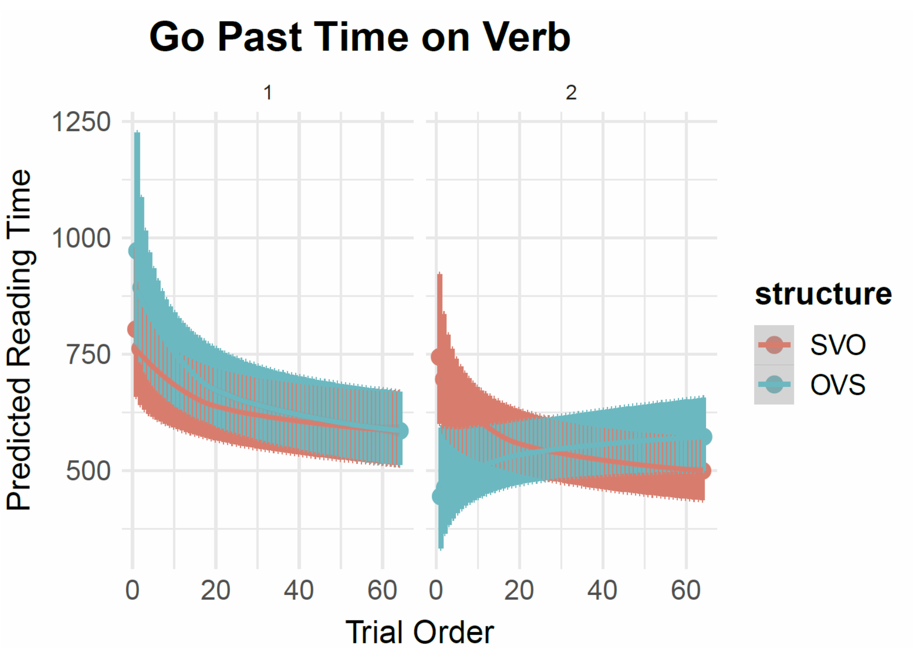


Figure 4.8. GPT on the Verb. Left panel: 1 good Agent. Right panel: 2 good Agents.

Determiner 2. There were no differences between conditions on the GPT on the second determiner.

Noun 2. There was an effect of Structure, with OVS trials having longer GPT on the second noun (β = 0.39, t = 1.99, p < .05).

Regressions

Regressions into the English sentence were more likely in the OVS and the 2 good agents conditions, but these were reduced as the experiment progressed.

Regressions into and out of Determiner 1, Noun 1, Verb, were more likely for OVS compared to SVO trials, but this effect decreased as the task progressed. Regressions into Noun 1 and into and out of the Verb were also more likely for two good agents compared to one, and this effect also decreased through the experiment.

Regressions into Determiner 2 and Noun 2 were equally likely in all conditions. Regressions out of Determiner 2 were less likely for OVS trials and more likely for two good agents. Regressions out of Noun 2 were more likely for OVS trials, and this effect was reduced as trials progressed.

Skipping

Similar to Experiment A, I calculated the probability of skipping the two nouns' endings, which contain the morphological case markers. There were no differences between conditions and no trial order effects for either noun ending. As shown in Figure 8, the probability of skipping the Noun's case morphology is high, in line with Experiment A results.

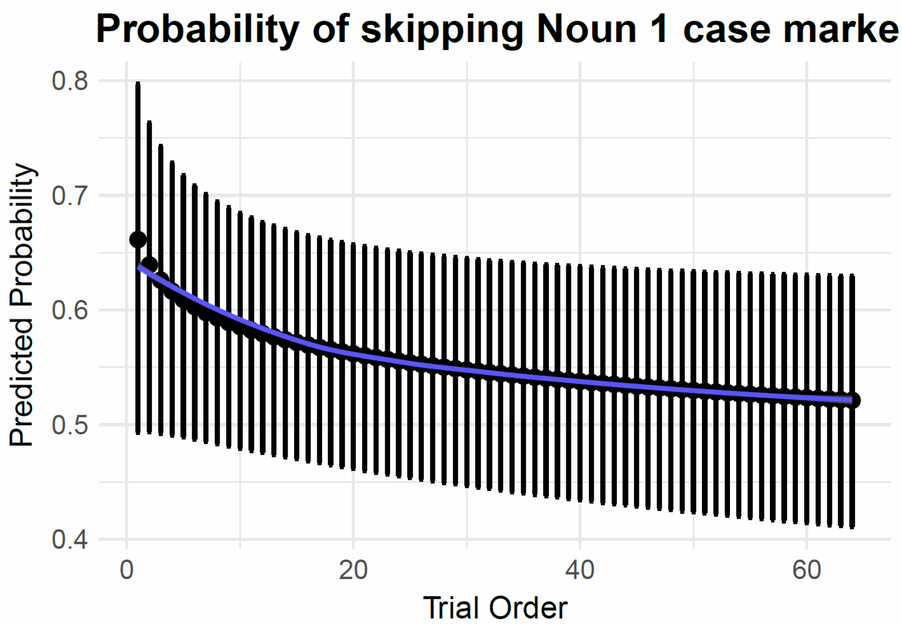


Figure 4.9. Skipping of Noun 1 ending.

Discussion of Experiment B

The results of the learning phase of Experiment B show similar processes to those in Experiment A. Analysis of TT shows that OVS trials had inflated reading times on the first NP, the Verb, and the whole trial. Similar to Experiment A, these inflated reading times were reduced after continued exposure to stimuli, indicative of some type of learning as a result of noticing the relevant language elements. Inflated reading times in OVS trials were also observed on the second NP, but here there was no interaction with trial order: reading times on the second NP started and remained inflated throughout the task.

There was also an interaction between Structure and Agent on the first noun and the verb: these regions had inflated reading times in OVS trials with two good agents, suggesting again that this might reflect a processing cost associated with inability to rely on the heuristic “goodness of agent,” and searching for more reliable information. This processing cost was diminished through continued exposure and reading times became similar halfway through the task. This shift in the amount of attention suggests possible learning, similar to Experiment A.

Regarding TT spent on the English sentence (as a result of regressions out of the Greek sentence), it was higher in OVS and two good agent trials, but both effects were reduced as the experiment progressed. In the ‘difficult’ conditions, when the Greek sentence violates expectations in word order, or when heuristics are useless for sentence interpretation, learners revisited the English sentence and compared the Greek one against it. The fact that the English sentence was revisited less and less through the task suggests an online and successful learning process.

One interesting difference between Experiments A and B is this trial order effect. Results from other tasks which will be presented in detail below show that learners in both experiments succeeded in learning. However, image viewing times remained stable whereas English sentence reading times were reduced (both viewing/reading times are the results of regressions out of the Greek sentence). This finding suggests different processes for integrating either L1-L2 or Visual-L2 information, but both appear to have been equally successful.

Analysis of GPT illustrates the rereading patterns triggered at different areas in different conditions. GPT is longer on Determiner 1 and Noun 2 in OVS trials, indicating additional rereading triggered by those regions. Rereading triggered by Determiner 1 can be assumed to include the English translation, as the only other possible region would be only an adverbial. This result is consistent with Experiment A. When Noun 2 triggered rereading of previous areas, and this did not change with task progression, it can indicate confirmatory rereading to confidently assign participant roles. Finally, GPT on the Verb showed an interaction between

Structure and Agent which also changed through the task. When only one noun was a good agent, both SVO and OVS Verb GPTs decreased at a similar rate. When both nouns were good agents, however, OVS trials initially had *lower* GPTs on the verb than SVO; OVS GPT increased and SVO decreased till they converged. The two good agents condition proved again a more difficult condition, especially in the OVS structure. This interaction suggests an attentional shift through the task, possibly with initially reduced attention on the verb for OVS + two good agents trials. One possible reason might be that they allocated their attention elsewhere in this condition, e.g. on the NPs, or the English sentence (evidenced by the increased frequency of regressions into the English sentence in this condition). Over the course of the learning session, however, attention to the verb converged to similar levels across conditions.

Regressions indicate which conditions incurred a higher cognitive load. Regressions into the English translation were more frequent in the 'difficult' conditions (OVS, two good agents), in line with Experiment A findings. This result also supports the hypothesis that when the L2 stimuli violate expectations such as word order, or when heuristics such as goodness of agent are unavailable, learners use information in the context (in this case L1 information) to make meaning of the L2 data, in this case to confidently assign participant roles. In addition, the trial order effects provide indirect evidence that the contextual information guided learner attention to the reliable L2 cues, as they relied less and less on the L1 sentence.

In general, regressions in and out of most regions were more frequent in the OVS trials but decreased with continued exposure, suggesting reduced cognitive load, presumably as a result of noticing and learning the morphosyntactic patterns of the L2 sentences. Regressions into Noun 1 and out of the verb were also more frequent in the two good agents condition, but this effect again decreased throughout the task. We observe initial uncertainty with two good agents, but participants appeared to become more confident in assigning participant roles as they got exposed to more stimuli.

Finally, the probability of skipping noun case morphology was quite high (roughly 60% throughout the task. This finding echoes Experiment A and again follows from SLA accounts of salience and redundancy. Given availability and subsequent noticing of the determiners, neither images nor English translations guided attention on the noun endings. This is reasonable, considering that noun endings are not necessary for making meaning and correctly assigning participant roles, as determiners are obligatory in Greek and were always present and a reliable cue in the experiment.

In sum, results from the learning phase of Experiments A and B suggest an active learning process, which is fairly similar for both experiments. Images and English translations raised error signals to initial misinterpretations of the stimuli in the more ‘difficult’ conditions (OVS structure, two good agents) and presumably guided learners to form hypotheses about the morphosyntactic patterns of Greek. Robust trial order effects provide support to an argument for ‘learning,’ defined as forming a hypothesis about Greek morphosyntax and committing to this interpretation that fits the available data (whether it is correct or not, which will be discussed below).

In these first two tasks, there is evidence for shifts in the amount of attention paid to different parts of the input as a function of exposure. The fact that attention changed *only* in some conditions, i.e., those conditions that violated expectations or precluded reliance on heuristics, suggest that it is in fact evidence for learning, rather than simply habituation to the task. Until now, I have presented findings that describe the process of learning and integrating L2 either with visual or with L1 textual information in real time. The results provide indirect evidence for learning in both experiments. In the next section, I present results from a subsequent task (adapted Visual World) designed to evaluate the outcome of this learning process and its relative success.

Adapted Visual World

As a reminder, in this task participants read a Greek sentence presented underneath two images which involved the same two participants performing the same action but with reversed roles (e.g., *a pelican chasing a chicken* vs. *a chicken chasing a pelican*). Participants also selected the correct image out of the two options. Their eye movements were being recorded throughout the task. The 48 stimuli crossed Structure (SVO or OVS) and good Agents (one good agent, two good agents, implausible agent) in a 2 x 3 design.

I present analyses from both eye movement data and accuracy scores. The data from both Experiments A and B were analyzed together in order to examine whether reading times and attention to different parts of the stimuli differed as a result of the two different learning processes. The descriptive statistics of reading times (TT) are presented in table 4.3 below. I analyzed TT on the following areas: Determiner 1, Noun 1, Verb, Determiner 2, Noun 2, Correct Image. Predictors were Structure, Agent, log-transformed trial order, their 3-way interaction, and Experiment (A or B).

Table 4.3. Descriptive statistics of TT during the visual words task for experiments A and B.

Total time		
Determiner 1		
	Experiment A	Experiment B
Structure		
SVO	318.89 (267.83)	325.02 (281.07)
OVS	598.51 (660.15)	633.69 (657.57)
Noun 1		
SVO	1343.36 (1109.75)	1321.82 (1065.84)
OVS	1113.66 (1087.31)	1129.29 (994.40)
Verb		
SVO	951.08 (852.16)	980.82 (876.99)
OVS	888.37 (890.56)	904.74 (848.01)
Determiner 2		
SVO	559.21 (477.96)	571.13 (580.12)
OVS	390.79 (441.02)	333.68 (384.51)
Noun 2		
SVO	740.75 (697.55)	815.39 (715.60)
OVS	1130.85 (1071.38)	942.70 (803.71)
Whole trial		
SVO	5663.71 (3922.09)	6069.43 (4297.44)
OVS	5751.60 (4174.99)	5826.55 (4093.49)
Correct Picture		
SVO	1096.96 (845.67)	1083.60 (774.88)
OVS	1095.74 (816.14)	999.74 (751.95)

Determiner 1. There was only an effect of Structure, with OVS trials having longer TT on the first determiner than SVO (β = 0.62, t = 2.86, $p < .01$).

Noun 1. There was an effect of Agent, with implausible agents having longer TT than one good agents (β = 0.35, t = 2.05, $p < .05$). There was also a trial order effect; TT on Noun 1 was reduced through the task.

Verb. There was a trial order effect, with later trials having shorter TT on the verb. There was also a numerical trend of an interaction between agents and presentation order (β = -0.14, t = -2.03, $p < .05$).

Determiner 2. There was only a trial order effect, with later trials having shorter TT.

Noun 2. There was only a trial order effect, with later trials having shorter TT.

Correct Image. There was only a trial order effect, with later trials having shorter TT.

Accuracy. A binomial mixed effects regression with a logit link was run with accuracy as the dependent measure; predictors were Structure, Agent, Experiment, and their 3-way interaction. There was an effect of Structure, with OVS trials being less likely to be accurate (β = -0.57, t = -2.16, $p < .05$). There was also an effect of Agent: two good and implausible agent trials were less likely to be accurate than one good agent trials (two good: β = -0.54, t = -2.05, $p < .05$; implausible: β = -1.07, t = -4.17, $p < .001$). Overall, however, accuracy was high, as evident in Figure 4.9. It is worth emphasizing that there was no statistically significant difference in accuracy between the two experiments.

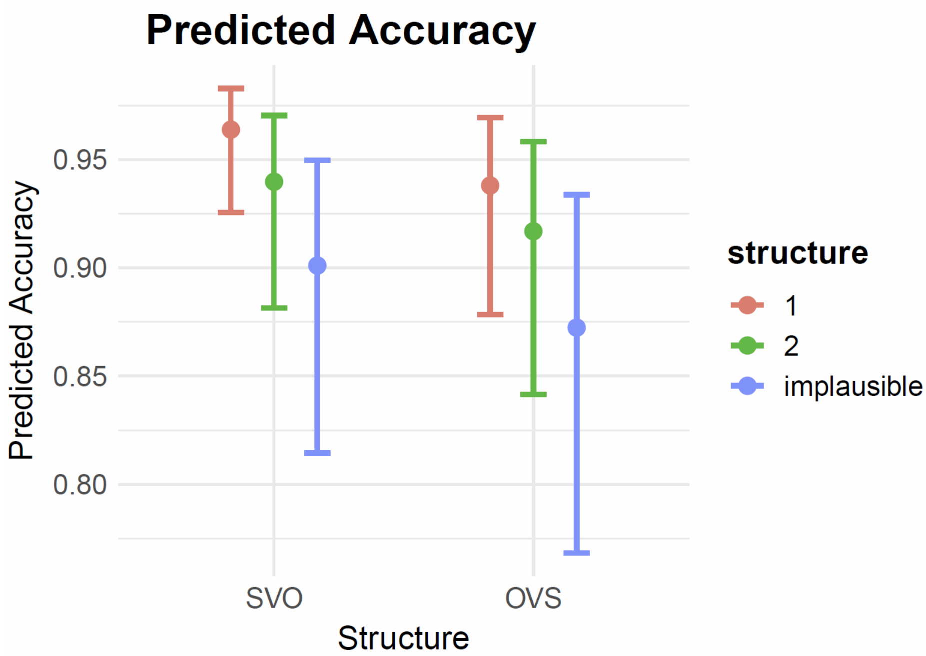


Figure 4.10. Predicted Accuracy.

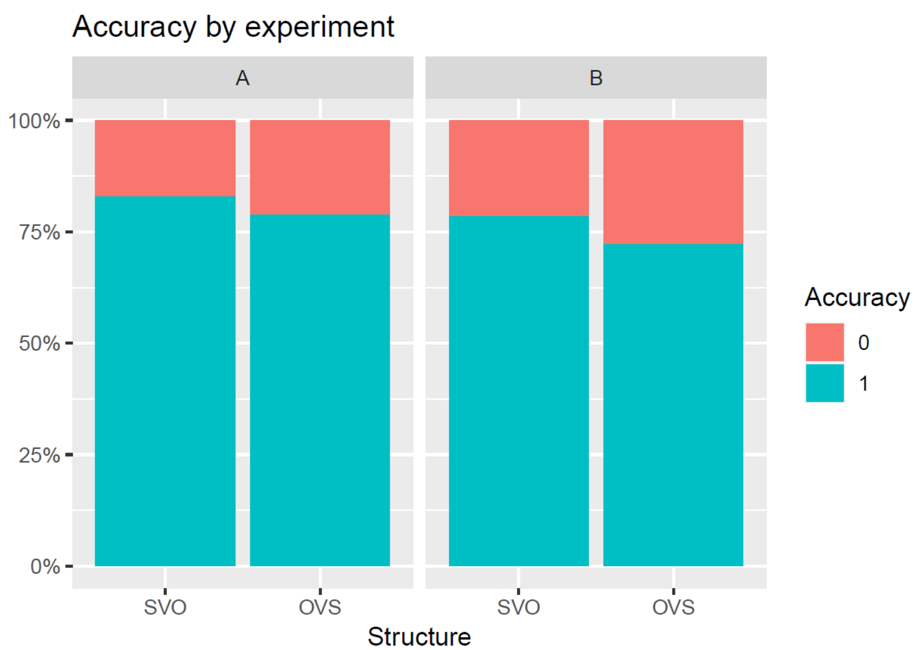


Figure 4.11. Accuracy scores in percentages.

An exploratory scanpath analysis using the package scanpath (von der Malsburg & Vasishth, 2011; 2013) was also run, in an effort to validate the adapted visual world methodology by showing a relationship between looking target and accuracy. Scanpaths are sequences of fixations in a trial and can be defined to fit a pattern of interest. A full scanpath would contain the information from the full trial, but a sub-scanpath of interest could include, for instance, the first fixation on region 4, followed by all fixations until a fixation is made on region 2 (hypothetical regions to illustrate how a pattern can be defined in scanpath analysis). From the full trials, I selected only the last pair of fixations that conformed to the following pattern: the first fixation was on the Greek sentence and the second fixation was on one of the two images, and the trial ended. In other words, I selected the last regression into one of the images which was followed by a button press to indicate picture selection. This pattern search yielded 435 scanpaths, which is only a subset of the complete dataset. Many trials included an additional fixation elsewhere (possibly even on the same image) before the button press and were thus excluded. Regressions into the correct images correlated with accuracy (see Table 4.4 below) and there were no differences between experiments or conditions.

Table 4.4. Percentage of trials with the final regression into the Correct and Incorrect images, for Accurate and Inaccurate trials.

	Accurate		**Inaccurate**	
Image	SVO	OVS	SVO	OVS
Correct	96.1 %	94.9 %	8.8 %	5.1 %
Incorrect	3.9 %	5.1 %	91.2 %	94.9 %
	N = 206	N = 156	N = 34	N = 39

I ran a binomial mixed effects regression (logit link) with Accuracy as the dependent variable; predictors were the image where the last fixation landed (Correct or Incorrect), structure, and their interaction. Subjects were entered as random intercepts. There was an effect of Image: when the regression landed on the incorrect image, the probability for an accurate response was much lower than when the regression landed on the correct image (β = -0.7.82, t = -5.0, p < .001). There was no effect of structure. The predicted probability for accuracy is presented in Figure 4.11 below, and the descriptive plot in figure 4.12.

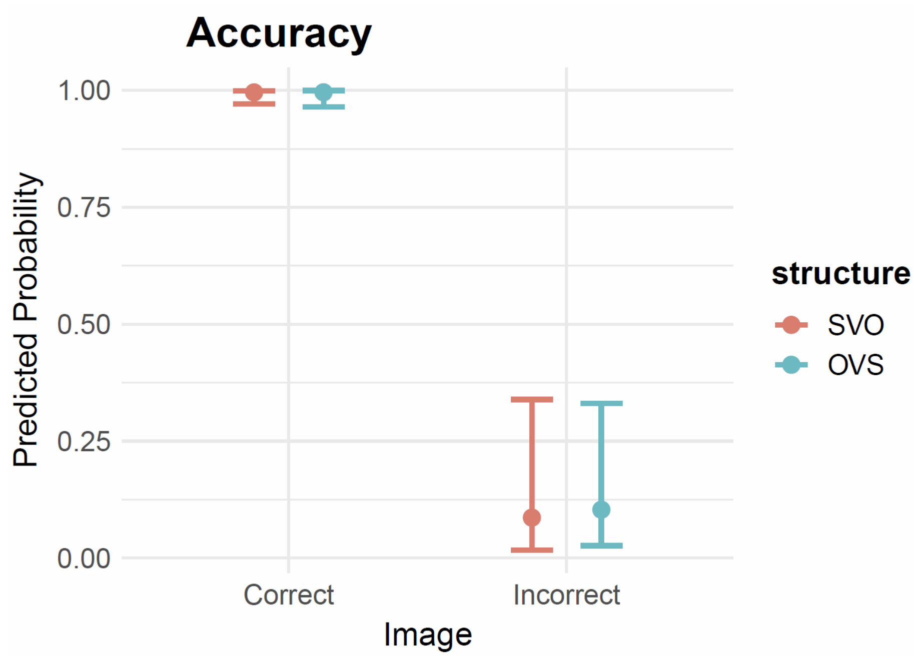


Figure 4.12. Probability of Accurate responses by the last regression being into the correct or incorrect image.

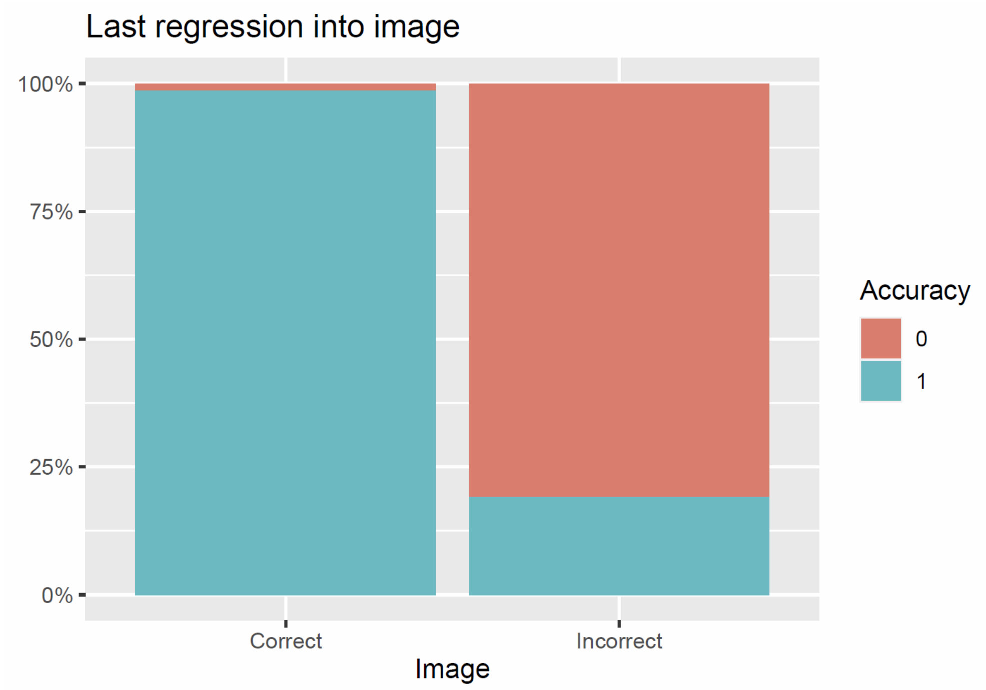


Figure 4.13. Percentage of accurately answered trials, by the final regression being in the correct or incorrect image.

Visual World Discussion

In this task, there were main effects of trial order but no interactions with other conditions, which suggests that reduced TTs were in this case indicative of only habituation to the task. OVS trials had longer TT only on the first determiner and on no other regions. This could indicate a slightly longer reading time due to a 3-letter word as compared to the 1-letter SVO determiner. Another possibility is that learners were alerted to the OVS structure at the first determiner and reading times were not affected beyond that. At the first Noun, reading times were inflated for implausible agents. This is a reasonable finding, as implausible agents violate heuristics and common world knowledge (Christianson, Luke, & Ferreira, 2010; Ferreira, 2003; Lim & Christianson, 2013a,b; Zhou & Christianson, 2016; Zhou, Yao, Christianson, Zwaan, &

Coltheart, 2018). In addition, implausible agents was a new condition, which participants had not been previously exposed to during the learning phase. In other words, these inflated reading times (which were reduced through exposure) can be taken to indicate surprisal.

Regarding viewing times on the correct image, these were similar across conditions, which suggests increased levels of confidence through the task. Unlike the learning phase, learners approached all trials similarly in the adapted visual world task and showed similar levels of cognitive effort for both SVO and OVS trials, and for one good agent, two good agents, and, crucially, implausible agent trials. This pattern consolidated the arguments made in the previous section regarding indirect evidence for learning; the fact that during the testing phase, all conditions had similar reading times on the sentence and similar viewing times on the correct image suggests that participants had, in fact, learned the morphosyntactic patterns of Greek transitive sentences.

It is important to note that there were no differences in reading times between experiments. Learners from both learning phases exhibited the same reading speed on all parts of the Greek sentence as well as on the correct image. In addition, they were both equally accurate in picture selection. There were some small effects of Structure and Agent on accuracy, the result of a very small number of individuals who did not learn the correct morphosyntactic pattern. The majority of participants, however, performed at ceiling. This last piece of evidence further supports the arguments made above regarding indirect evidence for learning. High levels of accuracy provide direct evidence in support of real time morphosyntactic learning as a result of integrating L2 with L1 or visual information.

An important point is the role of feedback during the structure learning phase. Typically, feedback in SLA is discussed in terms of being implicit or explicit, but it is always instructor-generated (e.g., corrections, recasts). During the learning phase of these experiments, there was no researcher-generated feedback, but there *was* system-generated feedback available for noticing. Specifically, if a Greek sentence had been misinterpreted (e.g., it had been interpreted as SVO instead of OVS), the images or the English translations would contradict the initial (mis)interpretation and force reanalysis. This is exactly what integrative saccades (indexed by regressions into the image/translation) between the Greek sentences and the images/translations indicated, which is further strengthened by the fact that they became less frequent through continued exposure and presumably learning. Once an error signal to the misinterpretation had been raised, learners allocated attention to seeking other, more reliable cues, i.e., the determiners. Indeed, reading times on the determiners also showed initially increased attention to them, which again decreased as their form-function mapping was understood. Because attention to the determiners sufficed to assign participant roles, noun endings remained, to a large extent, unnoticed.

This idea of system-generated feedback is well-established in game design theory (Schell, 2004) but has eluded attention in SLA. In this series of experiments, I have shown that this type of feedback can guide learner attention to relevant linguistic forms and, in conjunction with the learning context, help understand their function. The learning process was successful even though it only included exposure and no additional action from the part of the learners. Imagine a scenario where some action is required to progress to the next sentence (e.g., selecting the correct picture). In this case, inability to progress would provide even stronger feedback, in the form of an error signal, and additional motivation to allocate attention to the correct cues. This would be especially useful for less salient cues such as the noun morphological case markers. Therefore, manipulating the learning environment to constrain the hypothesis space is a very promising premise for the design of L2 learning games and materials, as well as for further research.

Finally, the exploratory scanpath analysis validated the use of the adapted visual world task. There was a clear relationship between looking target and image selection: when participants looked at the correct image, they also selected it. In contrast, when they looked at the incorrect image (much less frequently), their response was inaccurate, meaning that they in fact selected the incorrect image.

Sentence Production Task

Analysis

In this task, participants wrote 6 sentences to describe 6 images. The sentences were presented one at a time and they wrote their answers by typing in a box underneath the images. Participant productions were coded with 1 or 0 for correct production of the Agent and Patient determiners, as well as the correct Agent and Patient Noun endings. Cases where nouns were missing altogether were coded as N/A, and some cases where it was not clear which noun was produced as the agent and which as the patient were coded as "unclear." Other parts of the sentence, including spelling of the nouns, were ignored. I then calculated the percentages of correct production of those 4 morphemes in each experiment.

Determiners. In Experiment A, participants produced the agent determiner correctly 80.5% of the time and in Experiment B 69.4%. They were slightly more accurate with the patient determiner, producing it 84.1% of the time in Experiment A and 71.4% in Experiment B.

Noun suffixes. The agent morphological case marking (nominative case nouns end in -s to denote agent) was produced correctly 69.5% of the time in Experiment A and 57.5% in Experiment B. The patient suffix (no final -s) was less accurate and it was produced correctly 44.3% of the time in Experiment A and 32.5% in Experiment B. Overall, Experiment B had more trials than A where productions were not available altogether (e.g., they only wrote a verb) or they were unclear regarding case morphology. The results presented above are summarized visually in Figure 4.13.

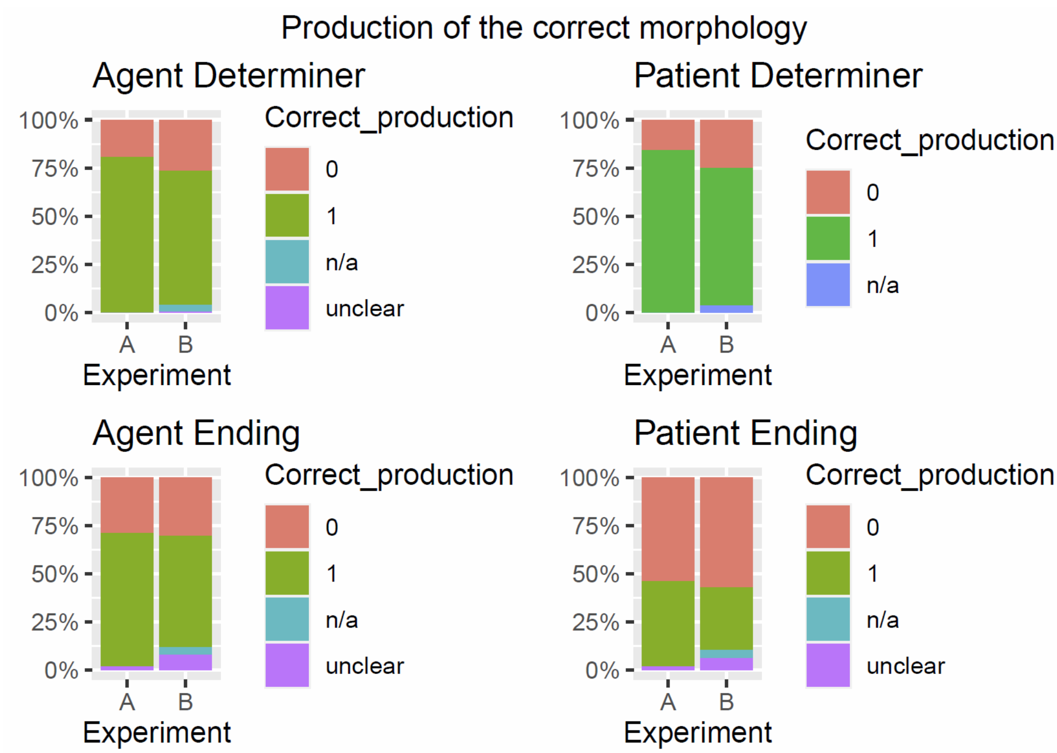


Figure 4.14. Percentage of trials with correct morphology. Top left panel: production of the agent determiner "o". Top right panel: production of the patient determiner "ton". Bottom left panel: production of the agent noun ending "-s". Bottom right panel: production of the patient noun ending (no final -s).

Discussion

Production tasks are generally considered more difficult than recognition tasks. Production accuracy ranged depending on the target linguistic element. Determiners were produced accurately most of the time, and slightly more so in Experiment A. Regarding noun suffixes, the agent suffix was produced accurately fairly frequently, which was expected as it was the noun form presented to participants during the vocabulary learning phase. Therefore, accurate production of the agent final -s does not necessarily indicate understanding that the suffix marks the noun as agent. Instead, it could simply suggest successful recall of the individual noun, as learned during vocabulary learning. Production of the patient ending is more informative about learning outcomes, as successful production involves a rule operation of removing the final "-s" from the noun. 44% of trials in Experiment A and a third of trials in Experiment B showed evidence for participants applying this operation.

The results from the sentence production task provide concrete evidence that most participants noticed the determiners, a smaller number additionally noticed the noun endings and understood their function as indexes to participant roles. Considering that they were able to produce accurate responses (written task) suggests that their learning extends from receptive to productive knowledge. These results are conservative as a number of cases (especially in Experiment B) were coded as "n/a" or "unclear"; in these cases, participants may have been unable to produce anything due to memory limitations. If they could not remember the lexical item, then they did not produce a suffix, even if they were actually aware of the rule. The next section provides some additional information about participants' noticing and awareness of the morphological case markers on the nouns.

Exit Survey

Analysis

This task explicitly asked participants what they thought they learned during the experiment. They were encouraged to provide detailed information about anything they noticed. Next, they completed 3 sentence-picture matching questions which included an open-form response where they were asked to justify their choices. The sentences were SVO-one good agent, OVS- two good agents, and OVS-implausible agent. The purpose of this task was to evaluate what each participant noticed and learned through the experiment, and it targeted different levels of awareness.

The open-question qualitative responses from the exit survey were coded with 1 or 0 for the following two variables: noticing the determiners and noticing the noun endings. They were also coded for understanding the function and articulating a form-function mapping for the morphosyntax as indicating participant roles. In this case, responses were assigned a code from one of 4 categories: 1 for understanding the function, 0 for not understanding (e.g., using only word order to assign participant roles), "almost" for formulating a hypothesis but an incorrect one, and finally "Implicit," for participants who showed evidence of learning but were unaware of this knowledge or could/did not justify their correct choices. Below I present the percentages of participants who noticed the determiners and noun endings, as well as those who understood the function, by experiment.

92.7% of participants in Experiment A and 82.5% in Experiment B noticed the determiners and mentioned them during the exit survey. In contrast, only 36.6% and 12.5% noticed and mentioned the noun endings in Experiments A and B, respectively. Regarding understanding the function, 65.9% of participants in Experiment A showed an understanding that case morphology and not word order is a reliable indicator for participant roles in Greek. An additional 1.5% were able to formulate a reasonable hypothesis and correctly used the morphological cues to assign participant roles; however, their interpretation was not entirely accurate. Instead, they interpreted the Greek OVS sentences as passive voice, which yields correct participant role assignment, but is not the accurate underlying syntactic representation. In Experiment B, 47.5% of participants correctly understood the function of the morphological case markers, and an additional 15% formulated a reasonable, but not fully accurate hypothesis. An additional two participants (one in each experiment) were coded as indicating implicit knowledge. They were not aware of having acquired any rule, but their sentence-picture matching choices were accurate. This information is summarized in Figure 4.14 below. Overall, about 2/3 of participants in experiment B and 3/4 in Experiment A were able to verbalize explicitly about the Greek morphosyntactic pattern they identified and used to assign participant roles.

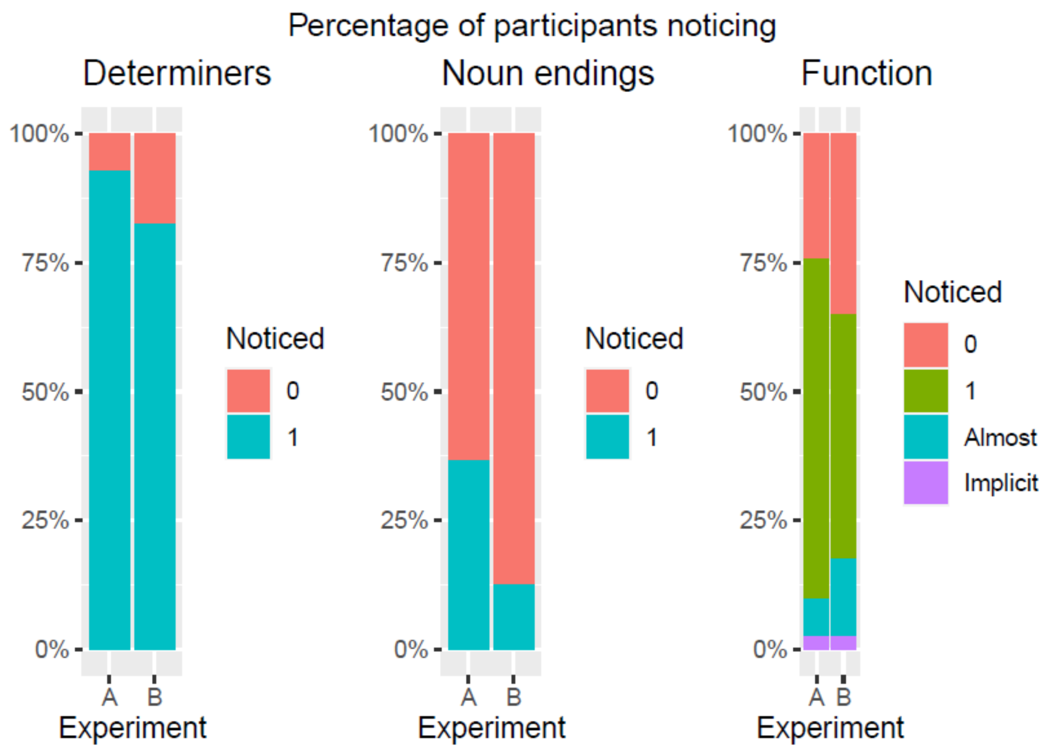


Figure 4.15. Percentages of participants who noticed the determiners (Left panel), the noun endings (Middle panel) and understood the function (Right panel).

Discussion

The results from this task strengthen the conclusions drawn from the two eye-tracking tasks. In those analyses, reading times suggested targeted attention paid to determiners, and less so to noun endings, which were very likely to be skipped. Participant verbalizations confirm this: the majority of the naïve learners in both experiments noticed the determiners and mentioned them during the exit survey. At this point, noticing is not equivalent to understanding; in fact, some participants made incorrect hypotheses about the determiners' meaning and function. Nevertheless, the determiners were salient enough to attract attention and encourage hypothesis formation about their role in participant role assignment. In contrast, a much smaller number of participants (curiously, even smaller in Experiment B) mentioned noun endings, which supplements well the finding that noun endings were very likely to be skipped during the structure learning phase.

Taken together, the results from both tasks suggest a correlation between reduced fixations during reading, indicative of less attention, and diminished learning. This supports the predictions of the noticing hypothesis for L2 morphosyntactic learning and suggests that input not directly attended to is not learned, at least in the early stages of the learning process when parts of the input compete with one another for the learner's attention. Even though it is possible that noun endings got processed in the parafovea, this apparently did not suffice for noticing and learning; instead, direct, focused attention appears to be necessary. Such attention is indeed paid to the determiners, even to the single letter nominative case marker "o," resulting in the determiners turning into intake. Noticing of noun endings was even lower in Experiment B, even though skipping rates had been roughly similar during the structure learning phase. It appears that concurrent L1-L2 processing may have temporarily improved comprehension during reading, but reliance on L1 seems to have detrimental effects on general learning that can be subsequently applied to different tasks. In other words, relying on L1 translations to decode the Greek sentences seems to have impeded, for some learners, identification of the L2 morphosyntactic pattern.

The majority of participants were also able to formulate a hypothesis around the function of case morphology (typically the determiners) as denoting participant roles. One interesting finding is that some participants interpreted the OVS sentences as passive voice, which allows for correct identification of participant roles, but erroneously transfers an L1 syntactic pattern that allows to preserve the expected word order. In Experiment A, this was a reasonable hypothesis for the limited dataset participants were exposed to. Based on these sentences, there was no information that violated this hypothesis. For Experiment B, however, participants read direct English translations of the Greek sentences during the learning phase. They were instructed that the sentences were direct, faithful translations; yet, instead of entertaining the possibility that word order in Greek is flexible, they persisted in an L1-influenced syntactic interpretation, disregarding the instructions. This is a very interesting finding that reveals the complexity of L2 syntax at the interface of syntax and morphology, even when provided with unambiguous information. This finding points to the need to constrain the hypothesis space during the learning process in a way that incorrect hypotheses would violate the contextual cues.

The results from the exit survey and the sentence production task show a relationship between awareness (operationalized here as ability to verbalize) and production. Most participants were aware of the determiners, and able to accurately produce them. In contrast, awareness was lower for noun endings, and so were accuracy scores for those. I use ‘awareness’ as evidence from verbalizations during the written exit survey; of course, lack of verbalizations does not necessarily indicate lack of awareness. Still, the similar trends and percentages between verbalized awareness and production of linguistic forms is suggestive of a deeper connection between the two.

Part 5: General Discussion

Summary of findings

The results from the self-paced reading experiments and especially from the eye-tracking experiments provide strong evidence for learning, even though it has been well-attested that L2 morphology is notoriously difficult to attain (e.g., Clahsen & Felser, 2006; VanPatten, 2006). During the structure learning phase, we observed inflated reading times in the conditions that were predicted to be 'confusing' for learners: sentences with OVS word order and sentences with two nouns as good agents. In the first case, OVS sentences violated the L1-generated expectation that the agent noun would appear first, followed by the patient noun (Ferreira, 2003; Townsend & Bever, 2001). This is in line with previous studies that have found that OVS sentences were mistakenly interpreted as SVO, even when learners' L1s have free word order and case morphology (Fulga & McDonough, 2016; McDonough & Fulga, 2015; McDonough & Trofimovich, 2013, 2015, 2016; Papadopoulou et al., 2011). In the present experiments, inflated reading times reflected this confusion, or surprisal, as learners regressed to the image or the English translation and attempted to make meaning. In the case of two good agents, longer reading times reflected a search cost that was the result of learners' inability to use world knowledge and the 'goodness-of-agent' heuristic. The most crucial finding, however, which departs from the findings in previous literature, is that these initially longer reading times declined as a function of task progression. Because this decline occurred *only* for the difficult conditions, it is an indicator of learning, rather than merely task habituation.

The interaction between trial order and structure is a sign that participants noticed the relevant morphological markers (namely, the determiners) and the flexible word order, and learned this pattern. This argument is strengthened by subsequent performance in the adapted visual world task. In this (testing) task, the only effects involved the implausible agent condition and thus can be reasonably thought of as indexing surprisal at the violation of world knowledge and heuristics (i.e., men tend not to bite dogs in the real world). Use of heuristics in L1 sentence processing follows from probabilistic constraint-based models (MacDonald et al., 1994) and particularly from well-attested good-enough processing accounts (e.g., Christianson et al, 2001, 2006; Lim & Christianson 2015, 2013a, b; Zhou et al, 2018). Multiple studies have shown that speakers are sensitive to thematic role fit or plausibility (Bicknell, Elman, Hare, McRae & Kutas, 2010; Kamide et al, 2003; McRae et al, 1998).

Besides heuristic violation, the implausible agent condition was not present during the learning phase of the experiment, which added to participants' surprise as they reacted to a new type of stimuli. The surprisal effect was reduced after some trials, as participants became accustomed to the peculiar sentences. However, reading times on the parts of the input that are of interest were similar across conditions, and trial order did not influence differently how each part was read (i.e., no interaction between trial order and conditions), besides a general habituation effect for all parts of the sentence. Therefore, the learning phase was enough for participants to commit to a hypothesis and a way to interpret the Greek sentences, which they applied consistently during the testing (adapted Visual World) phase. They were equally confident in all conditions and largely accurate in their picture selections, which provides even stronger evidence for learning during the first (Structure Learning) task.

When provided with no contextual support (SPR experiment C), nothing violated learners' expectations, so they continued to transfer the L1 pattern and to erroneously rely on word order to assign participant roles. This result is in line with previous SLA work that has found that when

L2 morphological cues are absent in the L1, they remain unnoticed and consequently unlearned in the L2 (Fulga & McDonough, 2015; McDonough & Trofimovich, 2013). Based on the present findings, the reason for this failure seems to be that no part of the system during the learning phase provides feedback indicating that the Greek sentence has been misinterpreted, so learners are not encouraged to revise their incorrect hypotheses. In the other two conditions (Images, Translations), both in the SPR and the eye-tracking experiments, contextual cues prompted learners to revise their interpretations and to identify more reliable cues than word order, in line with theoretical accounts viewing L2 learning as gradual readjustment of cue weights (e.g., Hernandez, Li & MacWhinney, 2005; MacWhinney, 2002; 2005; 2008, inter alia).

Both images and translations seemed equally useful in helping participants decode the Greek sentences and perform accurately in the various subsequent testing tasks. Participants in Experiment A (images) appeared to be more aware of the learning process when asked to verbalize. It is possible that Participants in Experiment B over-relied on L1-L2 translations during the learning process and that they were less aware of patterns in the L2 morphosyntax when asked explicitly about it. Therefore, integration of information was successful both between L1-L2 and between Image-L2. In the first case, it is possible that at least for some participants, the same modality (linguistic) interfered with processing, evidenced by slightly lower performance at the later, testing stages. Integration of L2 and visual information was rapid, evidenced through integrative saccades between text and image, and resulted in successful learning, following our predictions. These results suggest that the multimodal input in Experiment A strengthened associations between the visual and linguistic domains and facilitated processing, learning, and memory. This process has been theorized (Barsalou 2008;

2012; Barsalou et al., 2008; Clark & Paivio, 1991; Paivio 1971, 1986, 1990; Sadoski & Paivio, 2004, 2013) but not extensively studied in SLA.

Cue competition extended: multimodal cue integration

The present studies successfully extended the Cue Competition (MacWhinney, 1987; 2002) and the Input Processing Models (VanPatten, 2006) to include non-linguistic and non-L2 cues, in an attempt to simulate real world multimodal cue availability and competition (Kreysa, Knoeferle, & Nunneman, 2014). L2 and visual/L1 cues were successfully integrated to form a situation model reflected dually by the Greek sentence and the additional context. Even though the information in those two sources was the same, their integration was not redundant. It was instead an integral part of the learning process, in support of general theories of grounded cognition and the superiority of multimodal information processing for memory and learning (Sadoski & Paivio, 2004, 2013).

The L2 linguistic cues that entered the competition were word order, determiners, and noun endings, and the additional, non-L2 cues were the images or translations. The function of the latter was to reduce the strength of word order and render it unreliable, causing learners to readjust cue weights (MacWhinney, 2002; Ramscar & Yarlett, 2007; VanPatten, 2015). In search of a new, more reliable cue, learners' attention shifted to the determiners. In SPR Experiment C, which did not include any type of contextual support, the reliability of word order was never challenged for most participants, and therefore case morphology as a cue remained largely unnoticed and unused, i.e., it never entered the competition. Following Ellis (2006), the available cues were not processed because they were filtered through prior (L1) experience. Persistent transfer of the L1 structure resulted in blocking subsequent learning (Rescorla & Wagner, 1972).

In other words, these results show that the additional contextual support surrounding the L2 input subjectively altered for participants which cues entered the competition, which did not, and which were reevaluated, similarly to Arnon and Ramscar (2012). I would like to note that word order was still considered and was still part of the competition. Especially considering that some participants interpreted OVS sentences as passive voice, it becomes clear that word order persisted as a cue, but its reliability was reduced (in Cue Competition model terms, its weights were re-adjusted to a lower level (Bates & MacWhinney, 1981; Gass, 1987; Liu, Bates, & Li, 1992)). Instead, after the learning phase, the determiners appeared to be the strongest, winning cue for the task of assigning participant roles.

Cue competition has been typically studied with linguistic-only stimuli where the main focus was on how the features of each cue affected its relative strength. Arnon and Ramscar (2012) and Ellis and Sagarra (2010) showed that order of exposure and learning of a linguistic cue can influence subsequent learning of additional cues, suggesting that external manipulations can affect cue salience and learning. In the present studies, I followed their paradigm and extended it to multimodal cues. I showed that context manipulation, i.e., manipulating what type of information is included in the context of L2 exposure, can also influence which linguistic cues become more salient and attract the learners' attention. When language was presented in a contextualized way, context (in this case visual scenes or translations) had a direct and strong influence on how cue salience and strength were altered for the learners.

Therefore, cue strength can be thought of as both objective and relative: noun endings are objectively less salient than determiners, because when presented in print, the first are suffixes whereas the latter are separate words. It has been well attested that L2 learners rely less on grammatical cues and more on lexical ones (Lee, Cadierno, Glass, & VanPatten, 1997; Sagarra,

2007) and that cue salience can predict difficulty and morpheme prioritization in L2 acquisition (Goldschneider & DeKeyser, 2001). To use Ellis's (2006) terminology, determiners overshadowed noun suffixes because both cues predicted the same outcome but the first was more salient. At the same time, determiner salience and strength as a cue signaling participant roles can increase or decrease depending on the conditions in which language is presented, or the specific task characteristics and goals (Lim & Christianson, 2013). Cue strength and reliability, then, need to be reconceptualized to accommodate both objective and subjective differences. The first are cue-internal and linguistic in nature (e.g., length, position, prosodic stress for spoken language); the second are cue-external and relate to how much the context in which language is presented constrains the hypothesis space for learners. Based on the present findings, one reason case morphology is notoriously difficult for L2 learners may be because certain cues never enter the competition process, as was the case for noun endings for many participants in the present studies.

Cue salience, redundancy and blocking

What causes some cues to not enter the competition? The present findings support the notions of cue salience, blocking, redundancy and overshadowing. Once the more salient determiners were noticed and successfully used to identify participant roles, noun endings became redundant. Numerous studies in various fields, including SLA, have observed such blocking effects in learning: a previously learned cue 'blocks' the learning of a new, semantically redundant one (e.g., Ellis, 2006; Ellis & Sagarra, 2010; Rescorla & Wagner, 1972; Terrell, 1991; VanPatten 1996). For example, in Ellis and Sagarra (2010), when learning tense morphology, learners were unable to notice and learn the past tense suffix if they had first been exposed to the more salient adverbials. Interestingly, some participants in the current studies did notice both cues.

Examination of individual differences was beyond the scope of this study, but it is a promising area for future research. The current results suggest overshadowing effects for most, but not all participants. It would be extremely promising for future research to identify individual cognitive differences that allowed those few learners to overcome blocking effects and to consider both morphological cues of equal validity and reliability in participant role assignment. Still, it may be the case that one cue is prioritized during fast online processing even when both have been identified.

Another important point stemming from the current results concerns the role of noticing/awareness for cues entering the competition process. The role of attention, noticing and awareness for SLA has been extensively theorized and empirically studied (e.g., Ellis et al, 2009; Leow & Bowles, 2005, for a review; Schmidt, 1995, 1993, 1990). When considered in tandem, the results from the exit survey (explicit verbalizations of awareness) and the sentence production task reveal a relationship between higher levels of awareness and production accuracy. High levels of awareness about the determiners went hand in hand with high accuracy in their production; in contrast, lower levels of awareness for the noun suffixes were accompanied by similarly lower production accuracy of those suffixes. When participants noticed the noun endings and verbalized about them during the exit survey, they were also more likely to correctly produce them in the sentence production task. Higher levels of awareness related to understanding were facilitative during the initial stages of the learning process, adding to previous studies that found an important role for this higher level of awareness in L2 development (Bowles, 2003; Leow, 2001; Rosa & Leow; 2004, Rosa & O'Neill, 1999). Even though lack of verbalization does not necessarily entail lack of noticing of a form, the present experiments have provided evidence for a strong relationship between the two for the initial learning of L2 case morphology.

System feedback as an attention guiding mechanism

Central in the way cue competition changes as a function of the surrounding learning environment is the idea of system-generated feedback. This notion can also explain the high rates of learning and accuracy in the present studies, which departed significantly from previous, more modest gains (e.g., McDonough et al., 2017; Papadopoulou et al., 2011). Cue salience can change depending on the implicit feedback learners receive from the learning system itself. In the SPR Experiment C without any contextual support, learners interpreted the sentences presumably by transferring the L1 pattern, and then did not receive any feedback (dis)confirming their initial hypotheses about how Greek grammar works. As a result, they largely failed to develop a form-meaning mapping that allowed them to use morphological cues to assign participant roles in the sentence. Instead, they continued to interpret the sentences by relying on word order, and generally remained oblivious to case morphology, or at least to its function (similar findings in, e.g., Fulga & McDonough, 2016; Kempe & MacWhinney, 1998; Papadopoulou et al., 2011; VanPatten, 1996; 2004). When, however, the learning environment presented information that contradicted their initial hypotheses, learners were forced to revise them. During this process, the images or translations served as implicit feedback mechanisms, alerting learners to their misinterpretations and guiding their attention to alternative, more reliable cues. The amount and quality of feedback are directly related to how much context constrains the possible interpretations. For instance, a number of participants interpreted OVS sentences as passive voice. This interpretation was plausible given the limited stimuli and corresponding images. The images guided the learners to use the case-marked determiners to assign participant roles but left the syntactic structure unconstrained to various plausible interpretations. Interestingly, the same result was found in the translations experiment, even though in principle the syntax was constrained by the English sentence. It is likely that participants allowed for the same semantic meaning but different underlying syntactic structures between the Greek and English sentences, even though the instructions emphasized that they were direct and faithful translations, and participants had even been pre-taught the active voice verbs.

Participants were generally able to identify this implicit, system-provided feedback and to apply it during the learning process. This is especially promising, considering that the learning task did not include any comprehension checks such as questions after the sentences. Even if they had not paid a lot of attention or used the available feedback, learners would still have been able to complete the sentence reading task. Conceptualizing the system (i.e., learning context + stimuli) as an implicit feedback generator would predict even more power in guiding learner attention in cases where some action was required by the learners. Inability to progress through a task (experiment, learning task, or game) would necessarily alert learners to a mistake such as a linguistic misinterpretation. In order to progress, they would have to reconsider their hypotheses, revise, and retest them.

Implications for L2 learning and gamification

Trial and error approaches are the result of system-generated feedback-based learning. This concept is central in game design theory and practice, as well as player experience (see Whitton (2009) and Schell (2004) for a discussion on the nature of digital educational and game feedback, respectively). In SLA, however, explicit or implicit feedback has been considered mostly as originating from an instructor or other interlocutor (Ellis, Loewen, & Erlam, 2006; Li, 2010; Lyster & Saito, 2010; Sanz, 2004). Even in Computer Assisted Language Learning (CALL) applications, computer-generated feedback simulates human feedback to learner utterances (Cerezo, Caras, & Leow, 2016; Cerezo, 2013, for a meta-analysis; Ware & Kessler, 2013). For example, feedback may involve marking a sentence as correct or incorrect, or playing a buzzer sound to an incorrect production. What I describe in these experiments is distinctly different and very promising for shaping gamification practices in L2 learning and game design for language learning. Specifically, L2 gamification designers should manipulate the learning environment intelligently so that only a few (and preferably one) interpretations of the L2 input are allowed in that context. In other words, using context can constrain the possible linguistic interpretations and increase the salience of the linguistic forms that signal the correct interpretation. This process will encourage noticing of new linguistic forms and, in the current experiments, increased awareness and understanding was shown to be related to higher production accuracy.

As learners engage with the L2 input in a specific context, they can use that context to decode linguistic meaning and notice new linguistic forms that signal specific functions. In Part 1, readers used textual context to find clues to the meaning of new words. In certain cases, e.g., with concrete nouns, an image can easily facilitate a word form-meaning mapping. In a similar way, with visual scenes and simple transitive sentences, case morphology-participant role mappings can also be facilitated. In the current experiments in Parts 3 and 4, I have shown that participants were largely successful in understanding and even producing case morphology even after limited exposure when it was presented within a visual and an informationally rich L1 context. Admittedly, the stimuli were simple and repetitive; yet, when presented in a decontextualized way, they proved difficult to understand and precluded learning of case morphology (see results from SPR Experiment C).

Methodological implications

There are two important methodological implications that stem from the current experimental design. Firstly, participants were able to integrate multimodal information in real time to construct mental representations and, during learning, to map the L2 input onto those representations. Visual world studies with L1 populations have demonstrated incremental online linguistic processing and concurrent integration of linguistic and visual information (Knoeferle et al, 2005; Sedivy et al., 1999; Spivey et al, 2002; Tanenhaus et al., 1995). To the best of my knowledge, no studies thus far have examined how linguistic and visual information are integrated as part of the L2 learning process, besides Palmer (2015), who found facilitatory effects of pictures for (non-novel, garden path) L2 sentence processing. The present findings confirm that such integration occurred in real time with naïve learners, evidenced by integrative saccades between the image and the Greek text. This learning subsequently resulted in accurate performance in a variety of tasks (picture selection, sentence production, verbalizations about learning gains). Therefore, this methodology is promising for investigating how learners process information from multiple sources and domains (e.g., linguistic, visual), how they integrate it to create meaning and mental situational models, as well as how information processing changes throughout the learning task. Previous L1 work suggested that eye movements can be used to test reference assignment during L1 language comprehension (Bock et al., 2003; Griffin & Bock, 2000; Meyer & Lethaus, 2004; Meyer et al., 2003). The current results extend this method to the L2 domain and validate the use of eye movements to examine participant role assignment even with naïve learners.

The second methodological implication concerns the adapted visual world task. A typical visual world experiment with aural presentation of stimuli would have been too difficult for the present study involving naïve learners (see, for instance, McDonough et al. (2017) for difficulties in processing novel L2 stimuli). Instead, I used written L2 stimuli presented under the images, and a forced choice picture selection task. As shown clearly by the exploratory scanpath analysis, participants regressed into the image they selected, showing a direct relationship between the looking target and picture selection. This finding extends the concept of the eye-mind link (Reichle et al., 2012): besides locating attention during reading, eye movements can also reflect higher level processes in terms of integrating linguistic and visual information as well as engaging in decision-making. This has been the assumption in L1 visual world studies with aural presentation of stimuli. The current results suggest that the method can reliably be extended to designs that utilize written stimuli and images.

The adapted visual world works similarly to a typical visual world task where looks are time-locked to presentation of different parts of the stimuli. The adapted version allows us to examine at which point during reading participants looked at the correct image. In other words, it allows us to track the regions of the sentence that triggered looks to the corresponding image. This alternative is promising as it extends the use of the visual world to study designs and populations that would otherwise find aural presentation of L2 input too difficult, such as naïve or beginner learners.

Conclusion

Besides classrooms, language learning takes place or continues ‘in the wild’: in a foreign country, in an L2 university, in a computer game, in an online L2 website. Many of these contexts have not attracted enough research attention; admittedly, there are important methodological difficulties in studying learning in context. Psycholinguistic methodologies have been developed to fit lab-based designs with mostly decontextualized language presentation and they are not readily applicable to contextualized language learning. Nevertheless, it is important to transcend traditional methodological boundaries and to create study designs that allow us to study context-bound language learning while maintaining experimental control. A second important point is the operationalization of ‘learning’; it is necessary to approach learning both as a process and as a product and to document learning gains from a variety of tasks in order to more accurately represent the various levels of linguistic knowledge, ranging from implicit to explicit and from receptive to productive.

In this dissertation, I have studied how language is processed and learned in context. I have examined different situations and operationalizations of context: first, I focused on learning new lexical items from full sentences in both the L1 and the L2. The main finding was that the processing of new words was qualitatively similar in the L1 and the L2 and that readers were largely able to use the first available sentence (textual) cue to infer the word’s meaning. In addition, L2 readers engaged in deeper processing of the experimental items, which could at least partly explain their overall slower reading speed.

Secondly, I examined the learning of L2 morphological case marking, and the effect of different environments in which the L2 is presented. In general, visual scenes and translations provided mental models and were facilitative for the interpretation of the L2 sentences. The contextual support helped to guide learners' attention to the parts of the sentence that provided cues to form-meaning mappings (participant role assignment). Learning was operationalized as both a process and an outcome and was accordingly measured with a series of online and offline tasks. When taken together, these tasks provide strong converging evidence for successful learning of case morphology as a result of the additional supporting information available in the learning environment.

The current studies add significant findings to the field of SLA. Specifically, they focus on the less-studied learning processes of uninstructed SLA situations. I attempted to simulate real world language learning situations outside of the L2 classroom while maintaining experimental rigor and control. The majority of SLA studies focus on one of the following paradigms: some studies examine the outcome of learning in terms of what L2 learners can achieve at a given stage in L2 development. Other studies are concerned with teaching and classroom practices and their effect and relative success for students' development. However, not much work has been done on uninstructed SLA that takes place outside of a classroom, at least not on the process of this learning. In the present dissertation, I have attempted to create experimental designs to simulate uninstructed L2 learning situations and to dually examine the learning process and outcomes.

Second language learning is a messy process. Its success can vary greatly for different learners, especially for adults. When researchers formulate theories to evaluate and account for L2 learning success, it is necessary to consider the context where such learning occurs. In a series of controlled experiments targeting the very first stages of learning, I have shown that the surrounding context when the L2 is presented can alter which language features are attended to and subsequently get learned. To conclude, since the available context can alter both the process and the outcomes of learning, it needs to become an integral part of SLA theorizing. Language learning does not happen in a vacuum, but it is grounded in context. In the present dissertation, context has been shown to alter the course and success of early L2 morphosyntactic development.

REFERENCES

Adrada-Rafael, S. (2017). Processing the Spanish imperfect subjunctive: Depth of processing under different instructional conditions. *Applied Psycholinguistics*, *38*, 477–508.

Androutsopoulos, J. (2012). Greeklish: Transliteration practice and discourse in the context of computer-mediated digraphia. *Orthography as social action: Scripts, spelling, identity and power*, 359-392.

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24*(4), 477-492.

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., and Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, *39*(2), 187–205.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292-305.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive psychology*, *17*(3), 364-390.

Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam: John Benjamin's Publishing Company.

Bardovi-Harlig, K. (1992). The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language. *Applied Psycholinguistics*, *13*(3), 253-278.

Barsalou, L.W. (2012). The human conceptual system. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge handbook of psycholinguistics* (pp. 239-258). New York: Cambridge University Press.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-645.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. *Symbols, embodiment, and meaning*, 245-283.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. *The crosslinguistic study of sentence processing*, *3*, 73-112.

Bates, E., & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences*, *379*(1), 190-214.

Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. (2003). Minding the clock. *Journal of Memory and Language*, *48*(4), 653-685.

Bowles, M. A. (2019). Verbal reports in instructed SLA research: opportunities, Challenges, and Limitations. In R. P. Leow (Ed). *The Routledge Handbook of Second Language Research in Classroom Learning: Processing and Processes*, 31-43.

Bowles, M. A. (2010). *The think-aloud controversy in second language research*. New York, NY: Routledge.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of memory and language*, *63*(4), 489-505.

Brusnighan, S. M. & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial processing. Reading Research Quarterly, 47, 172–190.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. *Cognitive processes in eye guidance*, 53-77.

Calderón, A. M. (2013). The effects of L2 learner proficiency on depth of processing, levels of awareness, and intake. *Noticing and second language acquisition: Studies in honor in Richard Schmidt*, 103-121.

Cerezo, L. (2013). Beyond Hybrid Learning: A Synthesis of Research on E-tutors Under the Lens of Second Language. *AAUSC 2012 Volume--Issues in Language Program Direction: Hybrid Language Teaching and Learning: Exploring Theoretical, Pedagogical and Curricular Issues*, 50.

Cerezo, L., Caras, A., & Leow, R. P. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish gustar structures: An analysis of learning outcomes and processes. *Studies in Second Language Acquisition*, *38*(2), 265-291.

Chaffin, R., Morris, R. K., & Seely, R. E. (2001). Learning new word meanings from context: A study of eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 225-235.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of experimental psychology: Learning, memory, and cognition*, *30*(3), 687.

Chapelle, C. & Roberts, C. (1986). Ambiguity tolerance and field independence as predictors of proficiency in English as a second language. *Language Learning, 36*, 27–45.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified, or shallow language processing. *The Quarterly Journal of Experimental Psychology*, *69*(5), 817-828.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, *42*(4), 368-407.

Christianson, K., Luke, S. G., & Ferreira, F. (2010). Effects of plausibility on structural priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 538.

Christianson, K., Mestre, J., & Luke, S. G. (2012). Practice makes (nearly) perfect: Solving "students & professors"-type algebra word problems. *Applied Cognitive Psychology, 26,* 810-822.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics*, *27*(1), 3-42.

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

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*.

Craik, F. I. (2002). Levels of processing: Past, present... and future?. *Memory*, *10*(5-6), 305-318.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of experimental Psychology: general*, *104*(3), 268.

De Bot, K., & Van Montfort, R. (1988). Cue-validity in het Nederlands als eerste en tweede taal. *Interdisciplinair Tijdschrift voor Taal en Tekstwetenschap*, *8*, 111-120.

Dewaele, J. M., & Wei, L. (2013). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition*, *16*(1), 231-240.

Doughty, C. J., & Long, M. H. (2003). The scope of inquiry and goals of SLA. *The handbook of second language acquisition*, 3-16.

Eckerth, J., & Tavakoli, P. (2012). The effects of word exposure frequency and elaboration of word processing on incidental L2 vocabulary acquisition through reading. *Language Teaching Research*, *16*(2), 227-252.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of verbal learning and verbal behavior*, *20*(6), 641-655.

Elgort I., Brysbaert M., Stevens M. & Van Assche E. (2018). Contextual word learning during reading in a second language: an eye movement study. *Studies in Second Language Acquisition* 40: 341-66.

Ellert, M., & Holler, A. (2011). Semantic and structural constraints on the resolution of ambiguous personal pronouns-a psycholinguistic study. In *Discourse Anaphora and Anaphor Resolution Colloquium* (pp. 157-170). Springer: Berlin.

Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *The modern language journal*, *92*(2), 232-249.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164-194.

Ellis, N. C., & Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and learned attention. *Studies in Second Language Acquisition*, *32*(4), 553-580.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition*, *27*(2), 141-172.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in second language acquisition*, *28*(2), 339-368.

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision research*, *42*(5), 621-636.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 164-203.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science*, *11*(1), 11-15.

Ferreira, F., & Patson, N. D. (2007). The ‘good enough’approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71-83.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass*, *3*(1), 111-127.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Fulga, A., & McDonough, K. (2016). The impact of L1 background and visual information on the effectiveness of low variability input. *Applied Psycholinguistics*, *37*, 265–283.

Gass, S. M. (1987). The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics*, *8*(4), 329-350.

Gass, S. M., Svetics, I., & Lemelin, S. (2003). Differential effects of attention. *Language Learning*, *53*, 495 – 543.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of memory and language*, *31*(2), 129-151.

Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners’ eye-tracking data. In: Leow, R. P. (ed) *The Routledge handbook of second language research in classroom learning*. New York: Routledge, pp. 44–57.

Godfroid A., Ahn J., Choi I., Ballard L., Cui Y. & Johnston S. (2018) Incidental vocabulary learning in a natural reading context: an eye-tracking study. *Bilingualism: Language and Cognition* 21(3): 563–84.

Godfroid, A., Boers, F., & Housen, A. (2013). Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35* (3).

Godfroid, A. & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In: Bergsleithner, JM, Frota, SN, Yoshioka, K. (eds) *Noticing and second language acquisition: Studies in honor of Richard Schmidt*. Honolulu, HI: National Foreign Language Resource Center, pp. 183–205.

Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning*, *51*, 1 – 50.

Grace, C. (1998). Personality type, tolerance of ambiguity, and vocabulary retention in CALL. *CALICO journal*, 19-45.

Griffin, Z. M. (2004). The eyes are right when the mouth is wrong. *Psychological Science*, *15*(12), 814-821.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274-279.

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

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

Hamrick, P., & Rebuschat, P. (2014). Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L. Amoroso (Eds.), *Measured language: Quantitative approaches to acquisition, assessment, processing and variation* (pp. 125 – 139). Washington, DC: Georgetown University Press.

Harrington, M. (1987). Processing transfer: Language-specific processing strategies as a source of interlanguage variation. *Applied Psycholinguistics*, *8*(4), 351-377.

Hemforth, B., & Konieczny, L. (Eds.). (2013). *German sentence processing* (Vol. 24). Springer Science & Business Media.

Herman, J. L., Stevens, M. J., Bird, A., Mendenhall, M., & Oddou, G. (2010). The tolerance for ambiguity scale: Towards a more refined measure for international management research. *International Journal of Intercultural Relations*, *34*(1), 58-65.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in cognitive sciences*, *9*(5), 220-225.

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901-931.

Hopp, H. (2007). *Ultimate attainment at the interfaces in second language acquisition: Grammar and processing*. Groningen, The Netherlands: Grodil Press.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369-397.

Huettig, F. (2015). Four central questions about prediction in language processing. *Brain research*, *1626*, 118-135.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research*, 18(*3*), 193-223.

Hulstijn, J. (2003). Incidental and intentional learning. In C. Doughty & M. H. Long (eds.), *The handbook of second language acquisition* (pp. 349–381). Oxford: Blackwell.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431-439.

Jackson, C. N., & Bobb, S. C. (2009). The processing and comprehension of wh-questions among second language speakers of German. *Applied Psycholinguistics*, *30*(4), 603-636.

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2015). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language learning*, *57*(1), 1-33.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603-634.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of working memory: Individual differences in comprehension. *Psychological Review*, *99*, 122 – 149.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 276 – 296). New York: Appleton-Century-Crofts.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case marking by adult learners of Russian and German. *Studies in second language acquisition*, *20*(4), 543-587.

Kilborn, K. (1989). Sentence processing in a second language: The timing of transfer. *Language and Speech*, *32*(1), 1-23.

Kilborn, K., & Cooreman, A. (1987). Sentence interpretation strategies in adult Dutch–English bilinguals. *Applied Psycholinguistics*, *8*(4), 415-431.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481-529.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95-127.

Kreysa, H., Knoeferle, P., & Nunnemann, E. (2014). Effects of speaker gaze versus depicted actions on visual attention during sentence comprehension. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.* 2513-2518. Quebec City, Canada.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review*, *7*, 636 – 645.

Kuchinsky, S. E., Bock, K., & Irwin, D. E. (2011). Reversing the hands of time: Changing the mapping from seeing to saying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 748.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Laufer, B. (2005). Instructed second language vocabulary learning: The fault in the "default hypothesis." In A. Housen, & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 286–303). Amsterdam, The Netherlands: Mouton de Gruyter.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? *Canadian Modern Language Review*, 59, 565–585.

Lee, J. F., Cadierno, T., Glass, W. R., & VanPatten, B. (1997). The effects of lexical and grammatical cues on processing past temporal reference in second language input. *Applied Language Learning*, *8*, 1 – 23.

Lehtonen, M., and Laine, M. (2003). How word frequency affects morphological processing in mono- and bilinguals. *Bilingualism: Language and Cognition, 6*, 213–225.

Leow, R. P. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. New York, NY: Routledge.

Leow, R. P. (2001). Attention, awareness, and foreign language behavior. *Language learning*, *51*, 113-155.

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, *22*, 557–584.

Leow, R. P. (1997). Attention, awareness, and foreign language behavior. *Language Learning*, *47*(3), 467–505.

Leow, R. P., & Bowles, M. A. (2005). Attention and awareness in SLA. *Mind and context in adult second language acquisition: Methods, theory, and practice*, 179-203.

Leow, R. P., Hsieh, H. C. & Moreno, N. (2008). Attention to form and meaning revisited. *Language Learning*, *58*(3), 665-695.

Libben, M. R. & Titone, D. A. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381.

Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*(6), 1283-1315.

Lim, J-H. & Christianson, K. (2013a). Second language sentence processing in reading for comprehension and translation. *Bilingualism: Language and Cognition, 16*, 518-537.

Lim, J-H. & Christianson, K. (2013b). Integrating meaning and structure in L1-L2 and L2-L1 translations. *Second Language Acquisition, 29*, 233-256.

Li, P. (2009). Lexical organization and competition in first and second languages: Computational and neural mechanisms. *Cognitive science*, *33*(4), 629-664.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language learning*, *60*(2), 309-365.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451-484.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign language research in cross-cultural perspective*, *2*(1), 39-52.

Lowell, R., Pender, K.W. & Binder, K.S. (2020). Impact of Informative Context's Meaning Consistency During Incidental Vocabulary Acquisition. *Reading Research Quarterly*, 55(4), 679– 697.

Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in second language acquisition*, *32*(2), 265-302.

Macaro, E. (2003). *Teaching and learning a second language: A review of recent research*. London, UK, and New York, NY: Continuum.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82*, 276 – 298.

MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass & A. Mackey (Eds.), *Routledge handbook of second language acquisition* (pp. 211–227). New York, NY: Routledge.

MacWhinney, B. (2008). Cognitive precursors to language. *The evolution of communicative flexibility*, 193-214.

MacWhinney, B. (2005). A unified model of language acquisition. *Handbook of bilingualism: Psycholinguistic approaches*, *4967*.

MacWhinney, B. (2002). Extending the competition model. *Bilingual sentence processing*, 31-58.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69 – 90). New York: Cambridge University Press.

MacWhinney, B. (1989). Competition and lexical categorization. *Linguistic categorization*, 195-242.

MacWhinney, B. (Ed.). (1987). *Mechanisms of language acquisition* (Vol. 20). Psychology Press.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of verbal learning and verbal behavior*, *23*(2), 127-150.

Matessa, M., & Anderson, J. R. (2000). Modelling focused learning in role assignment. *Language and Cognitive Processes*, *15*(3), 263-292.

McDonald, J. L. (1987a). Assigning linguistic roles: The influence of conflicting cues. *Journal of Memory and Language*, *26*(1), 100.

McDonald, J. (1987b). Sentence interpretation processes: The influence of conflicting cues. *Journal of Memory and Language*, *26*, 100-117.

McDonald, J., & MacWhinney, B. (1989). Maximum likelihood models for sentence processing. *The cross-linguistic study of sentence processing*, 397-421.

McDonald, J. L., & Heilenman, L. K. (1991). Determinants of cue strength in adult first and second language speakers of French. *Applied Psycholinguistics*, *12*(3), 313-348.

McDonough, K., & Fulga, A. (2015). The detection and primed production of novel constructions. *Language Learning*, *65*, 353–384.

McDonough, K., Trofimovich, P., Dao, P., & Dion, A. (2017). Eye gaze and production accuracy predict English L2 speakers' morphosyntactic learning. *Studies in Second Language Acquisition*, *39*(4), 851-868.

McDonough, K., & Trofimovich, P. (2016). The role of statistical learning and working memory in L2 speakers' pattern learning. *The Modern Language Journal*, *100*, 428–445.

McDonough, K., & Trofimovich, P. (2015). Structural priming and the acquisition of novel form-meaning mappings. In S. Eskildsen & T. Cardierno (Eds.), *Usage-based perspectives on second language learning* (pp. 105–123). Berlin: Mouten De Gruyter.

McDonough, K., & Trofimovich, P. (2013). Learning a novel pattern through balanced and skewed input. *Bilingualism: Language and Cognition*, *16*, 654–662.

McLaren, I. P. L., & Mackintosh, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, *28*(3), 211-246.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283-312.

Mohamed A. A. (2018) Exposure frequency in L2 reading: an eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition* 40: 269–93.

Moreno, E. M., and Kutas, M. (2005). Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish–English bilinguals. *Brain Research Cognitive Brain Research, 22(2)*, 205–220.

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

Morgan-Short, K., Heil, J., Botero-Moriarty, A. & Ebert, S. (2012). Allocation of attention to second language form and meaning: Issues of think alouds and depth of processing. *Studies in Second Language Acquisition*, *34*(4), 659–685.

Meyer, A. S., & Lethaus, F. (2004). The use of eye tracking in studies of sentence generation. *The interface of language, vision, and action: Eye movements and the visual world*, 191-211.

Meyer, A. S., Roelofs, A., & Levelt, W. J. (2003). Word length effects in object naming: The role of a response criterion. *Journal of Memory and Language*, *48*(1), 131-147.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. *Journal of Statistical Software*, *43*(9), 1-14.

Oxford, R. L., & Ehrman, M. (1992). Second language research on individual differences. *Annual review of applied linguistics*, *13*, 188-205.

Paivio, A. (1990). Dual coding theory: Retrospect and current status. *Canadian Journal Of Psychology*, *45*(3), 255-287. Canadian Psychological Assn.

Paivio, A. (1986). *Mental representations: A dual coding approach* (Vol. 9, p. x, 322 p.). Oxford University Press. Retrieved from http://www.loc.gov/catdir/enhancements/fy0635/85018743-d.html

Paivio, A. (1971). Imagery and verbal processes. Leonardo (Vol. 5). Holt, Rinehart and Winston. doi:10.2307/1572599

Palmer, C. M. (2015). *The interaction of images and text during comprehension of garden-path sentences: is integration better than good enough?* Doctoral dissertation, University of Illinois at Urbana-Champaign.

Papadopoulou, D., Varlokosta, S., Spyropoulos, V., Kaili, H., Prokou, S., & Revithiadou, A. (2011). Case morphology and word order in second language Turkish: Evidence from Greek learners. *Second Language Research*, *27*(2), 173-204.

Pellicer-Sánchez, A. (2020). Expanding English Vocabulary Knowledge through Reading: Insights from Eye-tracking Studies. *RELC Journal*, *51*(1), 134-146.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading. *Studies in Second Language Acquisition*, 38 (*1*), 97-130.

Portin, M., and Laine, M. (2001). Processing cost associated with inflectional morphology in bilingual speakers. *Bilingualism: Language and Cognition, 4,* 55–62.

Ramscar, M., Dye, M. & Yarlett, D. (2009). No representation without taxation: The costs and benefits of learning to conceptualize the environment. In *Proceedings of the Analogy 09 Conference*, Sofia, Bulgaria.

Ramscar, M., & Yarlett, D. (2007). Linguistic self-correction in the absence of feedback: A new approach to the logical problem of language acquisition. *Cognitive Science, 31*, 927–960.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC Nonword Database. Quarterly Journal of Experimental Psychology, 55A, 1339-1362.

Rayner, K. (2009). Eye Movements in Reading: Models and Data. *Journal of Eye Movement Research*, 2 (5):2, 1-10.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124* (3), 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191-201.

Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review.Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504-509.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton Jr, C. (2012). *Psychology of reading*. Psychology Press.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1290.

Read, J. (2004). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, 24, 146–161.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4-22.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological review*, *105*(1), 125.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*(03), 333-357.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*(02), 213-235.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition* (pp. 630–678). Oxford: Blackwell.

Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45*, 283–331.

Robinson, P., Mackey, A., Gass, S.M. & Schmidt, R. (2012). *Attention and awareness in second language acquisition.* In: The Routledge handbook of second language acquisition, 247-267. Routledge, London.

Ronderos, C. R., Münster, K., Guerra, E., Kreysa, H., Rodríguez, A., Kröger, J., Kluth, T., Burigo, M., Abashidze, D., Nunnemann, E., & Knoeferle, P. (2018). Eye Tracking During Visually Situated Language Comprehension: Flexibility and Limitations in Uncovering Visual Context Effects. *JoVE (Journal of Visualized Experiments)*, (141), e57694, doi:10.3791/57694.

Rosa, E. M., & Leow, R. P. (2004). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics*, *25*(2), 269-292.

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in second language acquisition*, *21*(4), 511-556.

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, 17(*2*).

Sadoski, M., & Paivio, A. (2013). *Imagery and text: A dual coding theory of reading and writing*. Routledge.

Sadoski, M., & Paivio, A. (2004). A dual coding theoretical model of reading. *Theoretical models and processes of reading*, *5*, 1329-1362.

Sagarra, N. (2007). From CALL to face-to-face interaction: The effect of computer-delivered recasts and working memory on L2 development. *Conversational interaction in second language acquisition: A collection of empirical studies*, 229-248.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*(8), 2022-2039.

Sanz, C. (2004). Computer delivered implicit vs. explicit feedback in processing instruction. *Processing instruction: Theory, research, and commentary*, *241*.

Schell, J. (2014). *The Art of Game Design: A book of lenses*. USA, PA: CRC Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3 – 32). New York: Cambridge University Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*, *9*, 1-63.

Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206 – 226.

Schmidt, R. W. (1990). The role of consciousness in second language learning1. *Applied linguistics*, *11*(2), 129-158.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109-147.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive psychology*, *45*(4), 447-481.

Stoops, A., Luke, S. G., & Christianson, K. (2014). Animacy information outweighs morphological cues in Russian. *Language, Cognition, & Neuroscience, 29*, 584-604.

Sturt, P., Sanford, A.J. & Stewart, A. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review,* 11**,** 882–888.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632-1634.

Team, R. (2013). R development core team. *R: A Language and Environment for Statistical Computing*, *55*, 275-286.

Terrell, T. (1991). The role of grammar instruction in a communicative approach. *Modern Language Journal*, *75*, 52 – 63.

VanPatten, B. (2015). Input processing in adult SLA. *Theories in second language acquisition: An introduction*, 113-134.

VanPatten, B. (2006). Input processing. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 115 – 136). Mahwah, NJ: Erlbaum.

VanPatten, B. (Ed.) (2004). Processing instruction: Theory, research, and commentary. Mahwah, NJ: Lawrence Erlbaum.

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Westport, CT: Ablex.

Vasishth, S., & Drenhaus, H. (2011). Locality in German. *Dialogue and Discourse*, *1*, 59 – 82.

von der Malsburg, T., Kliegl, R., & Vasishth, S. (2015). Determinants of scanpath regularity in reading. *Cognitive science*, *39*(7), 1675-1703.

von der Malsburg, T., & Vasishth, S. (2013). Scanpaths reveal syntactic underspecification and reanalysis strategies. *Language and Cognitive Processes*, *28*(10), 1545-1578.

von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis?. *Journal of Memory and Language*, *65*(2), 109-127.

Ware, P., & Kessler, G. (2013). CALL and digital feedback. *Contemporary computer-assisted language learning*, 323-341.

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

Whitton, N. (2009). *Learning with digital games: A practical guide to engaging students in higher education*. USA, NY: Routledge.

Williams, R.S., & Morris, R.K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology, 16,* 312-339.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language, 60*(3), 368-392.

Wills, A. J. (Ed.). (2005). *New directions in human associative learning*. Mahwah, NJ: Erlbaum.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue. Eye-movement recordings in second language research. *Studies in Second Language Acquisition, 35*(2), 205-212.

Zhou, P., & Christianson, K. (2016). I "hear" what you're "saying": Auditory perceptual simulation, reading speed, and reading comprehension. *Quarterly Journal of Experimental Psychology*, *69*(5), 972-995.

Zhou, P., Yao, Y., Christianson, K., Zwaan, R., & Coltheart, M. (2018). When structure competes with semantics: reading Chinese relative clauses. *Collabra: Psychology*, *4*(1).

Zobl, H., & Liceras, J. (1994). Functional categories and acquisition orders. *Language Learning*, *44*(1), 159-180.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

Appendix A: Model Outputs for Part 1 Analyses

Appendix A contains the model outputs for all eye-tracking measures, presented by interest area.

Interest area Word1

Table A.1. First fixation duration on interest area "Word 1"

FFD on Word1				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.44107	0.02932	185.584	< 2e-16 ***
Uninformative	-0.04157	0.02973	-1.398	0.16232
Novel	0.10501	0.0293	3.584	0.00035 ***
L2	0.16884	0.0401	4.21	3.85e-05 ***
Uninformative: Novel	0.03622	0.04101	0.883	0.37734
Uninformative: L2	0.01337	0.03963	0.337	0.73592
Novel: L2	-0.06125	0.03952	-1.55	0.12135
Uninformative:Novel:L2	-0.01457	0.05584	-0.261	0.79414

Table A.2. Gaze duration on interest area "Word 1"

Gaze duration on Word1				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.61653	0.03948	142.278	< 2e-16 ***
Uninformative	-0.03647	0.04014	-0.909	0.3638
Novel	0.27037	0.03914	6.908	7.32e-12 ***
L2	0.40451	0.05219	7.751	4.94e-13 ***
Uninformative: Novel	0.02475	0.05449	0.454	0.6497
Uninformative: L2	0.04077	0.05096	0.8	0.4238
Novel: L2	-0.10611	0.05083	-2.088	0.0369 *
Uninformative:Novel:L2	-0.03013	0.07181	-0.42	0.6748

Pairwise comparisons (tukey correction) revealed only a difference in the novel word effect magnitude (L1: -0.27 and -0.30 for informative and uninformative contexts, L2: -.16 for both contexts)

Table A.3. Total time on interest area "Word 1"

Total time on Word1				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.81E+00	6.13E-02	94.76	< 2e-16 ***
Uninformative	1.53E-03	4.22E-02	0.036	0.97114
Novel	5.27E-01	4.03E-02	13.06	< 2e-16 ***
L2	5.58E-01	8.21E-02	6.798	5.38e-10 ***
Uninformative: Novel	-8.71E-03	5.56E-02	-0.157	0.87557
Uninformative: L2	4.83E-02	4.97E-02	0.971	0.3315
Novel: L2	-1.51E-01	4.95E-02	-3.043	0.00237 **
Uninformative:Novel:L2	-3.78E-03	6.98E-02	-0.054	0.95685

Table A.4. Go-past time on interest area "Word 1"

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.76E+00	4.95E-02	116.367	< 2e-16 ***
Uninformative	1.31E-02	4.23E-02	0.31	0.756
Novel	3.83E-01	4.09E-02	9.35	< 2e-16 ***
L2	4.02E-01	6.60E-02	6.089	1.06e-08 ***
Uninformative: Novel	-1.18E-03	5.68E-02	-0.021	0.983
Uninformative: L2	1.80E-02	5.21E-02	0.344	0.731
Novel: L2	-1.32E-01	5.20E-02	-2.544	0.011 *
Uninformative:Novel:L2	-3.15E-02	7.35E-02	-0.429	0.668

Same; only differences in the magnitude of the novel word effect between L1 and L2.

Interest Area Word2

Table A.5. First fixation duration on interest area "Word 2"

FFD on Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.33E+00	2.43E-02	219.625	< 2e-16 ***
Uninformative	2.81E-02	2.54E-02	1.108	0.268317
Novel	9.37E-02	2.48E-02	3.782	0.000163 ***
L2	1.75E-01	3.31E-02	5.296	2.98e-07 ***
Uninformative: Novel	-2.76E-02	3.50E-02	-0.788	0.430863
Uninformative: L2	-3.20E-03	3.36E-02	-0.095	0.924212
Novel: L2	-3.55E-02	3.33E-02	-1.069	0.285236
Uninformative:Novel:L2	2.18E-02	4.72E-02	0.461	0.644902

Table A.6. Gaze duration on interest area "Word 2"

Gaze duration on Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.43076	0.03327	163.219	< 2e-16 ***
Uninformative	0.02598	0.03507	0.741	0.459
Novel	0.14872	0.03387	4.391	1.21e-05 ***
L2	0.34912	0.04361	8.006	8.40e-14 ***
Uninformative: Novel	-0.02632	0.0475	-0.554	0.58
Uninformative: L2	-0.05634	0.04405	-1.279	0.201
Novel: L2	0.01595	0.04355	0.366	0.714
Uninformative:Novel:L2	0.02961	0.06181	0.479	0.632

Table A.7. Total time on interest area "Word 2"

Total time on Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.54E+00	4.45E-02	124.547	< 2e-16 ***
Uninformative	5.61E-02	4.03E-02	1.391	0.165
Novel	2.13E-01	3.86E-02	5.525	3.80e-08 ***
L2	4.07E-01	5.82E-02	6.997	8.45e-11 ***
Uninformative: Novel	-3.11E-02	5.37E-02	-0.579	0.562
Uninformative: L2	-3.97E-03	4.86E-02	-0.082	0.935
Novel: L2	3.66E-02	4.82E-02	0.76	0.447
Uninformative:Novel:L2	-3.11E-02	6.82E-02	-0.457	0.648

Table A.8. Go past times on interest area "Word 2"

Go-past times on Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.61E+00	4.31E-02	130.214	< 2e-16 ***
Uninformative	5.30E-02	4.14E-02	1.282	0.2
Novel	1.78E-01	3.97E-02	4.481	7.95e-06 ***
L2	2.90E-01	5.62E-02	5.165	6.94e-07 ***
Uninformative: Novel	-6.22E-02	5.54E-02	-1.122	0.262
Uninformative: L2	-4.31E-02	5.05E-02	-0.854	0.393
Novel: L2	2.70E-03	4.99E-02	0.054	0.957
Uninformative:Novel:L2	4.04E-02	7.08E-02	0.571	0.568

Interest Area Context

Table A.9. Total time on interest area "Context"

Total time on Context				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.7396	0.06148	109.617	< 2e-16 ***
Uninformative	0.0409	0.01357	3.015	0.00259 **
Novel	0.0165	0.01293	1.276	0.20218
L2	0.4857	0.07775	6.247	1.85e-08 ***
Uninformative: Novel	0.03311	0.0118	2.806	0.00505 **
Uninformative: L2	0.02771	0.01397	1.984	0.04731 *
Novel: L2	-0.01191	0.01397	-0.853	0.39399
Uninformative:Novel:L2	-0.02243	0.01397	-1.606	0.10838

Table A.10. Go-past time on interest area "Context"

Go-past time on Context				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.76179	0.05894	114.714	< 2e-16 ***
Uninformative	-0.17991	0.03676	-4.894	1.04e-06 ***
Novel	-0.10473	0.03418	-3.064	0.0022 **
L2	0.38669	0.07113	5.436	3.69e-07 ***
Uninformative: Novel	0.18596	0.04686	3.968	7.41e-05 ***
Uninformative: L2	0.03851	0.03926	0.981	0.3267
Novel: L2	0.01263	0.03923	0.322	0.7475
Uninformative:Novel:L2	-0.10752	0.05548	-1.938	0.0527 .

Interest Area Hypernym

Table A.11. First fixation duration on interest area "Hypernym"

FFD on Hypernym				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.40E+00	2.45E-02	219.976	< 2e-16 ***
Uninformative	1.95E-02	2.57E-02	0.758	0.449
Novel	7.52E-03	2.55E-02	0.295	0.768
L2	1.73E-01	3.31E-02	5.218	4.18e-07 ***
Uninformative: Novel	9.03E-04	3.54E-02	0.026	0.98
Uninformative: L2	-3.47E-02	3.39E-02	-1.023	0.306
Novel: L2	3.54E-03	3.38E-02	0.105	0.917
Uninformative:Novel:L2	3.75E-02	4.74E-02	0.791	0.429

Table A.12. Gaze duration on interest area "Hypernym"

Gaze duration on Hypernym				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.54E+00	3.94E-02	140.647	< 2e-16 ***
Uninformative	-4.95E-03	3.83E-02	-0.129	0.897
Novel	-2.79E-02	3.65E-02	-0.766	0.444
L2	2.99E-01	4.14E-02	7.239	8.47e-12 ***
Uninformative: Novel	7.95E-02	4.95E-02	1.605	0.109
Uninformative: L2	1.57E-02	4.20E-02	0.374	0.708
Novel: L2	-8.54E-03	4.19E-02	-0.204	0.838
Uninformative:Novel:L2	3.32E-02	5.88E-02	0.565	0.572

Table A.13. Total time on interest area "Hypernym"

Total time on Hypernym				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.62E+00	3.94E-02	142.748	< 2e-16 ***
Uninformative	-3.57E-02	1.58E-02	-2.262	0.0238 *
Novel	-1.88E-03	1.52E-02	-0.124	0.9012
L2	3.65E-01	4.03E-02	9.065	7.56e-14 ***
Uninformative: Novel	2.71E-02	1.39E-02	1.946	0.0518 .
Uninformative: L2	-1.25E-02	1.65E-02	-0.758	0.4488
Novel: L2	6.10E-04	1.65E-02	0.037	0.9705
Uninformative:Novel:L2	-4.73E-03	1.65E-02	-0.287	0.7742

Table A.14. Go-past time on interest area "Hypernym"

Go-past time on Hypernym				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.66E+00	3.70E-02	152.777	< 2e-16 ***
Uninformative	-5.73E-03	3.95E-02	-0.145	0.885
Novel	-2.41E-03	3.95E-02	-0.061	0.951
L2	2.74E-01	5.15E-02	5.324	2.38e-07 ***
Uninformative: Novel	8.23E-02	5.51E-02	1.494	0.135
Uninformative: L2	1.40E-02	5.45E-02	0.257	0.797
Novel: L2	6.17E-03	5.43E-02	0.113	0.91
Uninformative:Novel:L2	-9.95E-03	7.63E-02	-0.13	0.896

Probability of regressions

Table A.15. Probability of regressions in interest area "Word 1"

Probability of regressions in Word1				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.53679	0.211	-7.284	3.25e-13 ***
Uninformative	-0.20683	0.20604	-1.004	0.315461
Novel	0.71189	0.18829	3.781	0.000156 ***
L2	0.58211	0.27779	2.096	0.036126 *
Uninformative: Novel	0.1376	0.26461	0.52	0.603048
Uninformative: L2	0.31505	0.25222	1.249	0.211627
Novel: L2	-0.03255	0.23913	-0.136	0.891743
Uninformative:Novel:L2	-0.10919	0.33771	-0.323	0.746463

Table A.16. Probability of regressions in interest area "Context"

Probability of regressions in Context				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.63013	0.200265	-8.14	3.96e-16 ***
Uninformative	0.194243	0.197251	0.985	0.325
Novel	0.027497	0.196097	0.14	0.888
L2	0.747548	0.259743	2.878	0.004 **
Uninformative: Novel	-0.08208	0.269135	-0.305	0.76
Uninformative: L2	-0.30917	0.242147	-1.277	0.202
Novel: L2	0.001102	0.242786	0.005	0.996
Uninformative:Novel:L2	-0.0627	0.341338	-0.184	0.854

Table A.17. Probability of regressions in interest area "Word 2"

Probability of regressions in Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3.1698	0.276	-11.486	<2e-16 ***
Uninformative	0.5199	0.289	1.799	0.0720 .
Novel	0.5608	0.2843	1.973	0.0485 *
L2	1.0408	0.3457	3.011	0.0026 **
Uninformative: Novel	-0.2053	0.3739	-0.549	0.5829
Uninformative: L2	-0.4701	0.3583	-1.312	0.1895
Novel: L2	-0.1598	0.3487	-0.458	0.6468
Uninformative:Novel:L2	0.194	0.4693	0.413	0.6794

Table A.18. Probability of regressions out of interest area "Context"

Probability of regressions out of Context				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.37985	0.21814	-10.91	<2e-16 ***
Uninformative	-0.30877	0.26492	-1.166	0.244
Novel	0.27755	0.23707	1.171	0.242
L2	0.18151	0.27253	0.666	0.505
Uninformative: Novel	0.14506	0.34548	0.42	0.675
Uninformative: L2	0.06808	0.32841	0.207	0.836
Novel: L2	-0.37776	0.30279	-1.248	0.212
Uninformative:Novel:L2	-0.21869	0.4543	-0.481	0.63

Table A.19. Probability of regressions out of interest area "Word 2"

Probability of regressions out of Word2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.84557	0.17445	-10.58	<2e-16 ***
Uninformative	0.19516	0.20294	0.962	0.3362
Novel	0.13812	0.20219	0.683	0.4945
L2	-0.59839	0.26417	-2.265	0.0235 *
Uninformative: Novel	-0.26822	0.28319	-0.947	0.3436
Uninformative: L2	0.02581	0.31404	0.082	0.9345
Novel: L2	-0.047	0.31708	-0.148	0.8822
Uninformative:Novel:L2	0.28473	0.43621	0.653	0.5139

Table A.20. Probability of regressions out of interest area "Hypernym"

Probability of regressions out of Hypernym				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.7555	0.2575	-10.699	<2e-16 ***
Uninformative	-0.1954	0.3129	-0.624	0.532
Novel	0.1903	0.2906	0.655	0.513
L2	-0.0245	0.3461	-0.071	0.944
Uninformative: Novel	0.3437	0.4088	0.841	0.401
Uninformative: L2	-0.1055	0.4371	-0.241	0.809
Novel: L2	-0.1054	0.4021	-0.262	0.793
Uninformative:Novel:L2	-0.2492	0.5849	-0.426	0.67

Appendix B. Model Outputs for Part 3 Analyses

Appendix B contains the model outputs for the statistical analyses performed for Part 2 (self-paced reading experiments A, B, C). The results are organized by task and by experiment.

Learning phase

Table B.1. Learning phase, Experiment A. Reaction time on Determiner 1.

Reaction time on Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.92	0.18	38.00	< 0.001	***
OVS	0.73	0.25	2.92	<0.01	**
good_agents:2	0.31	0.23	1.34	0.183	
log(presentation_order)	-0.17	0.05	-3.68	< 0.001	***
OVS:good_agents:2	-0.25	0.40	-0.62	0.534	
OVS:log(presentation_order)	-0.16	0.08	-2.09	0.039	*
good_agents:2:log(presentation_order)	-0.09	0.07	-1.33	0.186	
OVS:good_agents:2:log(presentation_order)	0.10	0.12	0.85	0.396	

Table B.2. Learning phase, Experiment A. Reaction time on Noun 1.

Reaction time Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.79	0.21	37.58	< 0.001	***
OVS	0.34	0.28	1.22	0.226	
good_agents:2	0.13	0.26	0.49	0.624	
log(presentation_order)	-0.35	0.05	-6.69	< 0.001	***
OVS:good_agents:2	0.19	0.44	0.44	0.663	
OVS:log(presentation_order)	-0.11	0.08	-1.32	0.191	
good_agents:2:log(presentation_order)	-0.05	0.08	-0.69	0.491	
OVS:good_agents:2:log(presentation_order)	0.01	0.13	0.04	0.969	

Table B.3. Learning phase, Experiment A. Reaction time on Determiner 2.

Reaction time Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.26	0.14	52.79	< 0.001	***
OVS	-0.29	0.18	-1.60	0.112	
good_agents:2	0.26	0.17	1.53	0.128	
log(presentation_order)	-0.27	0.04	-7.65	< 0.001	***
OVS:good_agents:2	-0.18	0.30	-0.62	0.535	
OVS:log(presentation_order)	0.06	0.06	1.11	0.271	
good_agents:2:log(presentation_order)	-0.05	0.05	-0.92	0.362	
OVS:good_agents:2:log(presentation_order)	0.02	0.09	0.24	0.812	

Table B.4. Learning phase, Experiment A. Reaction time on Noun 2.

Reaction time Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.64	0.17	43.97	< 0.001	***
OVS	0.36	0.24	1.53	0.129	
good_agents:2	0.23	0.22	1.04	0.301	
log(presentation_order)	-0.34	0.05	-7.49	< 0.001	***
OVS:good_agents:2	0.02	0.38	0.05	0.963	
OVS:log(presentation_order)	-0.10	0.07	-1.39	0.167	
good_agents:2:log(presentation_order)	-0.05	0.07	-0.75	0.455	
OVS:good_agents:2:log(presentation_order)	-0.02	0.11	-0.13	0.893	

Table B.5. Learning phase, Experiment A. Reaction time on Last Word.

Reaction time last_word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.88	0.23	38.27	< 0.001	***
OVS	0.33	0.31	1.09	0.278	
good_agents:2	-0.22	0.29	-0.79	0.434	
log(presentation_order)	-0.60	0.06	-10.34	< 0.001	***
OVS:good_agents:2	-0.12	0.49	-0.25	0.806	
OVS:log(presentation_order)	-0.09	0.09	-1.00	0.321	
good_agents:2:log(presentation_order)	0.06	0.09	0.73	0.470	
OVS:good_agents:2:log(presentation_order)	0.04	0.15	0.29	0.770	

Table B.6. Learning phase, Experiment A. Reaction time on Verb.

Reaction time Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.82	0.17	45.08	< 0.001	***
OVS	0.92	0.24	3.91	< 0.001	***
good_agents:2	0.13	0.22	0.60	0.547	
log(presentation_order)	-0.38	0.05	-8.33	< 0.001	***
OVS:good_agents:2	-0.55	0.38	-1.44	0.152	
OVS:log(presentation_order)	-0.27	0.07	-3.77	< 0.001	***
good_agents:2:log(presentation_order)	-0.07	0.07	-1.06	0.291	
OVS:good_agents:2:log(presentation_order)	0.19	0.11	1.66	0.100	.

Table B.7. Learning phase, Experiment A. Reaction time on Sentence.

Sentence Reaction time					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	10.18	0.19	54.99	< 0.001	***
OVS	0.43	0.22	1.91	0.059	.
good_agents:2	-0.09	0.21	-0.44	0.662	
log(presentation_order)	-0.41	0.04	-9.50	< 0.001	***
OVS:good_agents:2	-0.07	0.36	-0.18	0.855	
OVS:log(presentation_order)	-0.13	0.07	-1.97	0.051	.
good_agents:2:log(presentation_order)	0.01	0.06	0.19	0.847	
OVS:good_agents:2:log(presentation_order)	0.06	0.11	0.55	0.581	

Table B.8. Learning phase, Experiment A. Reaction time on Both Determiners.

Reaction time on both determiners					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.04	0.13	54.62	< 0.001	***
OVS	0.60	0.17	3.58	< 0.001	***
good_agents:2	0.34	0.11	3.05	<0.01	**
log(presentation_order)	-0.21	0.03	-7.05	< 0.001	***
OVS:good_agents:2	-0.99	0.19	-5.23	< 0.001	***
OVS:log(presentation_order)	-0.10	0.05	-2.09	0.037	*
good_agents:2:log(presentation_order)	-0.08	0.03	-2.19	0.028	*
OVS:good_agents:2:log(presentation_order)	0.18	0.06	3.14	<0.01	**

Table B.9. Learning phase, Experiment A. Reaction time on Both Nouns.

Reaction time on both Nouns					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.88	0.15	51.72	< 0.001	***
OVS	0.38	0.19	1.96	0.051	.
good_agents:2	-0.19	0.12	-1.51	0.132	
log(presentation_order)	-0.38	0.04	-10.77	< 0.001	***
OVS:good_agents:2	-0.03	0.21	-0.13	0.894	
OVS:log(presentation_order)	-0.09	0.06	-1.59	0.114	
good_agents:2:log(presentation_order)	0.02	0.04	0.61	0.539	
OVS:good_agents:2:log(presentation_order)	-0.01	0.06	-0.17	0.867	

Table B.10. Learning phase, Experiment B. Reaction time on Determiner 1.

Reaction time on Determiner1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.01	0.15	46.63	< 0.001	***
OVS	0.13	0.17	0.76	0.448	
good_agents:2	0.12	0.16	0.73	0.464	
log(presentation_order)	-0.26	0.03	-7.77	< 0.001	***
OVS:good_agents:2	-0.07	0.28	-0.24	0.811	
OVS:log(presentation_order)	0.00	0.05	0.04	0.969	
good_agents:2:log(presentation_order)	0.02	0.05	0.33	0.738	
OVS:good_agents:2:log(presentation_order)	-0.01	0.08	-0.11	0.914	

Table B.11. Learning phase, Experiment B. Reaction time on Noun 1.

Reaction time on Noun1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.87	0.19	42.46	< 0.001	***
OVS	0.69	0.19	3.59	< 0.001	***
good_agents:2	0.27	0.18	1.53	0.126	
log(presentation_order)	-0.38	0.04	-10.21	< 0.001	***
OVS:good_agents:2	-0.54	0.31	-1.76	0.079	.
OVS:log(presentation_order)	-0.19	0.06	-3.18	< 0.01	**
good_agents:2:log(presentation_order)	-0.09	0.06	-1.67	0.095	.
OVS:good_agents:2:log(presentation_order)	0.17	0.09	1.81	0.071	.

Table B.12. Learning phase, Experiment B. Reaction time on Determiner 2.

Reaction time on Determiner2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.29	0.14	51.80	< 0.001	***
OVS	-0.34	0.18	-1.88	0.062	.
good_agents:2	-0.15	0.17	-0.87	0.384	
log(presentation_order)	-0.30	0.04	-8.53	< 0.001	***
OVS:good_agents:2	0.08	0.29	0.28	0.782	
OVS:log(presentation_order)	0.08	0.06	1.44	0.152	
good_agents:2:log(presentation_order)	0.04	0.05	0.79	0.434	
OVS:good_agents:2:log(presentation_order)	-0.02	0.09	-0.23	0.821	

Table B.13. Learning phase, Experiment B. Reaction time on Noun 2.

Reaction time on Noun2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.63	0.15	51.58	< 0.001	***
OVS	0.02	0.18	0.13	0.901	
good_agents:2	0.00	0.17	-0.01	0.994	
log(presentation_order)	-0.35	0.03	-10.22	< 0.001	***
OVS:good_agents:2	-0.22	0.29	-0.77	0.444	
OVS:log(presentation_order)	0.01	0.06	0.16	0.871	
good_agents:2:log(presentation_order)	0.01	0.05	0.19	0.851	
OVS:good_agents:2:log(presentation_order)	0.07	0.09	0.81	0.418	

Table B.14. Learning phase, Experiment B. Reaction time on Last Word.

Reaction time on last_word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.05	0.23	40.03	< 0.001	***
OVS	0.11	0.31	0.35	0.724	
good_agents:2	-0.27	0.29	-0.93	0.354	
log(presentation_order)	-0.60	0.06	-9.99	< 0.001	***
OVS:good_agents:2	-0.09	0.50	-0.18	0.854	
OVS:log(presentation_order)	-0.03	0.09	-0.35	0.725	
good_agents:2:log(presentation_order)	0.10	0.09	1.09	0.279	
OVS:good_agents:2:log(presentation_order)	0.02	0.15	0.13	0.898	

Table B.15. Learning phase, Experiment B. Reaction time on Verb.

Reaction time on Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.57	0.16	47.25	< 0.001	***
OVS	0.45	0.20	2.27	0.025	*
good_agents:2	0.08	0.19	0.45	0.653	
log(presentation_order)	-0.36	0.04	-9.29	< 0.001	***
OVS:good_agents:2	-0.13	0.32	-0.41	0.686	
OVS:log(presentation_order)	-0.10	0.06	-1.57	0.119	
good_agents:2:log(presentation_order)	-0.01	0.06	-0.18	0.856	
OVS:good_agents:2:log(presentation_order)	0.00	0.10	0.01	0.991	

Table B.16. Learning phase, Experiment B. Reaction time on Sentence.

Reaction time on Sentence					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	10.37	0.17	61.33	< 0.001	***
OVS	0.23	0.23	1.04	0.303	
good_agents:2	0.02	0.21	0.09	0.925	
log(presentation_order)	-0.45	0.04	-10.38	< 0.001	***
OVS:good_agents:2	0.06	0.36	0.15	0.879	
OVS:log(presentation_order)	-0.06	0.07	-0.88	0.381	
good_agents:2:log(presentation_order)	0.00	0.07	0.03	0.976	
OVS:good_agents:2:log(presentation_order)	-0.02	0.11	-0.21	0.837	

Table B.17. Learning phase, Experiment B. Reaction time on Both Determiners.

Reaction time on both determiners					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.05	0.13	55.58	< 0.001	***
OVS	0.09	0.15	0.65	0.519	
good_agents:2	0.17	0.11	1.57	0.117	
log(presentation_order)	-0.24	0.03	-9.08	< 0.001	***
OVS:good_agents:2	-0.39	0.18	-2.13	0.033	*
OVS:log(presentation_order)	0.00	0.04	-0.03	0.973	
good_agents:2:log(presentation_order)	-0.04	0.03	-1.12	0.265	
OVS:good_agents:2:log(presentation_order)	0.07	0.06	1.27	0.206	

Table B.18. Learning phase, Experiment B. Reaction time on Both Nouns.

Reaction time on both Nouns					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.97	0.15	54.15	< 0.001	***
OVS	0.47	0.15	3.10	< 0.01	**
good_agents:2	-0.33	0.12	-2.73	< 0.01	**
log(presentation_order)	-0.41	0.03	-14.53	< 0.001	***
OVS:good_agents:2	-0.53	0.21	-2.59	< 0.01	**
OVS:log(presentation_order)	-0.12	0.05	-2.58	0.010	*
good_agents:2:log(presentation_order)	0.06	0.04	1.52	0.129	
OVS:good_agents:2:log(presentation_order)	0.16	0.06	2.50	0.012	*

Table B.19. Learning phase, Experiment C. Reaction time on Determiner 1.

Reaction time on Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.86	0.15	45.52	< 0.001	***
OVS	0.11	0.17	0.65	0.519	
good_agents:2	-0.05	0.16	-0.31	0.758	
log(presentation_order)	-0.22	0.03	-6.80	< 0.001	***
OVS:good_agents:2	0.01	0.28	0.04	0.966	
OVS:log(presentation_order)	0.03	0.05	0.58	0.560	
good_agents:2:log(presentation_order)	0.01	0.05	0.30	0.767	
OVS:good_agents:2:log(presentation_order)	-0.02	0.08	-0.26	0.797	

Table B.20. Learning phase, Experiment C. Reaction time on Noun 1.

Reaction time on Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.87	0.22	35.92	< 0.001	***
OVS	0.01	0.23	0.05	0.957	
good_agents:2	-0.40	0.21	-1.87	0.065	.
log(presentation_order)	-0.39	0.04	-8.87	< 0.001	***
OVS:good_agents:2	-0.05	0.36	-0.14	0.886	
OVS:log(presentation_order)	0.00	0.07	0.06	0.953	
good_agents:2:log(presentation_order)	0.12	0.07	1.83	0.070	.
OVS:good_agents:2:log(presentation_order)	0.00	0.11	0.03	0.973	

Table B.21. Learning phase, Experiment C. Reaction time on Determiner 2.

Reaction time on Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.08	0.15	48.25	< 0.001	***
OVS	-0.33	0.17	-1.92	0.055	.
good_agents:2	-0.19	0.16	-1.15	0.249	
log(presentation_order)	-0.23	0.03	-6.72	< 0.001	***
OVS:good_agents:2	0.24	0.28	0.87	0.386	
OVS:log(presentation_order)	0.06	0.05	1.20	0.231	
good_agents:2:log(presentation_order)	0.07	0.05	1.35	0.179	
OVS:good_agents:2:log(presentation_order)	-0.08	0.08	-0.96	0.337	

Table B.22. Learning phase, Experiment C. Reaction time on Noun 2.

Reaction time on Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.38	0.20	36.28	< 0.001	***
OVS	0.17	0.20	0.88	0.377	
good_agents:2	0.37	0.18	1.98	0.048	*
log(presentation_order)	-0.25	0.04	-6.48	< 0.001	***
OVS:good_agents:2	-0.37	0.31	-1.18	0.239	
OVS:log(presentation_order)	-0.05	0.06	-0.85	0.397	
good_agents:2:log(presentation_order)	-0.08	0.06	-1.41	0.160	
OVS:good_agents:2:log(presentation_order)	0.10	0.09	1.10	0.273	

Table B.23. Learning phase, Experiment C. Reaction time on last word.

Reaction time last_word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.56	0.24	36.37	< 0.001	***
OVS	-0.10	0.28	-0.37	0.714	
good_agents:2	0.15	0.27	0.55	0.584	
log(presentation_order)	-0.49	0.05	-9.02	< 0.001	***
OVS:good_agents:2	-0.71	0.46	-1.56	0.122	
OVS:log(presentation_order)	0.05	0.09	0.55	0.586	
good_agents:2:log(presentation_order)	-0.06	0.08	-0.70	0.484	
OVS:good_agents:2:log(presentation_order)	0.19	0.14	1.35	0.179	

Table B.24. Learning phase, Experiment C. Reaction time on Verb.

Reaction time Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.83	0.21	36.70	< 0.001	***
OVS	0.14	0.22	0.62	0.535	
good_agents:2	-0.32	0.20	-1.59	0.113	
log(presentation_order)	-0.33	0.04	-7.62	< 0.001	***
OVS:good_agents:2	-0.09	0.35	-0.26	0.793	
OVS:log(presentation_order)	-0.07	0.07	-0.98	0.329	
good_agents:2:log(presentation_order)	0.06	0.06	0.91	0.362	
OVS:good_agents:2:log(presentation_order)	0.05	0.11	0.49	0.627	

Table B.25. Learning phase, Experiment C. Reaction time on Sentence.

Sentence Reaction time					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.97	0.19	53.42	< 0.001	***
OVS	0.07	0.19	0.38	0.704	
good_agents:2	-0.11	0.17	-0.65	0.515	
log(presentation_order)	-0.40	0.04	-11.03	< 0.001	***
OVS:good_agents:2	-0.28	0.30	-0.92	0.359	
OVS:log(presentation_order)	0.01	0.06	0.24	0.812	
good_agents:2:log(presentation_order)	0.05	0.05	0.89	0.378	
OVS:good_agents:2:log(presentation_order)	0.05	0.09	0.53	0.595	

Table B.26. Learning phase, Experiment C. Reaction time on Both Determiners.

Reaction time both determiners					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.85	0.13	52.79	< 0.001	***
OVS	0.13	0.14	0.95	0.344	
good_agents:2	0.13	0.11	1.12	0.264	
log(presentation_order)	-0.22	0.03	-8.69	< 0.001	***
OVS:good_agents:2	-0.35	0.19	-1.83	0.067	.
OVS:log(presentation_order)	0.02	0.04	0.37	0.712	
good_agents:2:log(presentation_order)	0.02	0.04	0.64	0.525	
OVS:good_agents:2:log(presentation_order)	0.01	0.06	0.17	0.869	

Table B.27. Learning phase, Experiment C. Reaction time on Both Nouns.

Reaction time both Nouns	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.69	0.19	40.45	< 0.001	***
OVS	0.03	0.17	0.16	0.873	
good_agents:2	-0.15	0.13	-1.12	0.265	
log(presentation_order)	-0.34	0.03	-10.68	< 0.001	***
OVS:good_agents:2	-0.02	0.22	-0.11	0.914	
OVS:log(presentation_order)	-0.01	0.05	-0.12	0.906	
good_agents:2:log(presentation_order)	0.06	0.04	1.35	0.177	
OVS:good_agents:2:log(presentation_order)	0.00	0.07	0.05	0.959	

Sentence-picture matching

Table B.28. Sentence-picture matching, Experiment A. Reaction times on Determiner 1.

Determiner 1	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.98	0.09	63.93	< 0.001	***
OVS	-0.11	0.14	-0.78	0.437	
Implausible	0.08	0.11	0.76	0.449	
TwoGood	0.21	0.13	1.53	0.127	
log(Trial_order)	-0.17	0.02	-9.42	< 0.001	***
OVS:Implausible	0.29	0.20	1.43	0.154	
OVS:TwoGood	0.24	0.24	1.01	0.314	
OVS:log(Trial_order)	0.03	0.05	0.71	0.475	
Implausible:log(Trial_order)	-0.04	0.04	-0.97	0.335	
TwoGood:log(Trial_order)	-0.06	0.05	-1.31	0.190	

Table B.28. (cont.)

OVS:Implausible:log(Trial_order)	-0.08	0.07	-1.29	0.199	
OVS:TwoGood:log(Trial_order)	-0.08	0.08	-0.99	0.322	

Table B.29. Sentence-picture matching, Experiment A. Reaction times on Noun 1.

Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.47	0.15	43.46	< 0.001	***
OVS	0.13	0.27	0.46	0.644	
Implausible	0.48	0.22	2.21	0.030	*
TwoGood	0.40	0.26	1.52	0.132	
log(Trial_order)	-0.31	0.04	-8.81	< 0.001	***
OVS:Implausible	-0.12	0.39	-0.31	0.754	
OVS:TwoGood	-0.21	0.47	-0.44	0.659	
OVS:log(Trial_order)	-0.01	0.09	-0.12	0.905	
Implausible:log(Trial_order)	-0.15	0.07	-2.05	0.044	*
TwoGood:log(Trial_order)	-0.06	0.09	-0.65	0.517	
OVS:Implausible:log(Trial_order)	0.01	0.13	0.11	0.913	
OVS:TwoGood:log(Trial_order)	-0.02	0.16	-0.11	0.915	

Table B.30. Sentence-picture matching, Experiment A. Reaction times on Determiner 2.

Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.62	0.14	47.36	< 0.001	***
OVS	-0.44	0.27	-1.64	0.105	
Implausible	-0.18	0.22	-0.84	0.401	
TwoGood	-0.04	0.26	-0.17	0.867	

Table B.30. (cont.)

log(Trial_order)	-0.26	0.03	-7.40	< 0.001	***
OVS:Implausible	0.35	0.39	0.92	0.363	
OVS:TwoGood	-0.06	0.47	-0.13	0.896	
OVS:log(Trial_order)	0.14	0.09	1.59	0.115	
Implausible:log(Trial_order)	0.06	0.07	0.77	0.441	
TwoGood:log(Trial_order)	0.00	0.09	0.06	0.955	
OVS:Implausible:log(Trial_order)	-0.13	0.13	-1.04	0.302	
OVS:TwoGood:log(Trial_order)	-0.02	0.15	-0.16	0.874	

Table B.31. Sentence-picture matching, Experiment A. Reaction times on Noun 2.

Noun 2

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.76	0.18	37.56	< 0.001	***
OVS	-0.22	0.37	-0.59	0.558	
Implausible	0.11	0.29	0.37	0.711	
TwoGood	-0.23	0.35	-0.64	0.522	
log(Trial_order)	-0.27	0.05	-5.59	< 0.001	***
OVS:Implausible	0.72	0.53	1.37	0.175	
OVS:TwoGood	0.52	0.64	0.81	0.418	
OVS:log(Trial_order)	0.10	0.12	0.85	0.397	
Implausible:log(Trial_order)	0.00	0.10	0.03	0.974	
TwoGood:log(Trial_order)	0.19	0.12	1.55	0.125	
OVS:Implausible:log(Trial_order)	-0.28	0.17	-1.62	0.109	
OVS:TwoGood:log(Trial_order)	-0.27	0.21	-1.28	0.203	

Table B.32. Sentence-picture matching, Experiment A. Reaction times on Last Word.

Last Word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.37	0.30	31.35	< 0.001	***
OVS	0.37	0.58	0.63	0.528	
Implausible	-0.25	0.47	-0.54	0.594	
TwoGood	0.24	0.57	0.43	0.671	
log(Trial_order)	-0.52	0.08	-6.91	< 0.001	***
OVS:Implausible	-0.02	0.84	-0.02	0.982	
OVS:TwoGood	-0.02	1.02	-0.02	0.988	
OVS:log(Trial_order)	-0.10	0.19	-0.52	0.602	
Implausible:log(Trial_order)	0.14	0.16	0.90	0.373	
TwoGood:log(Trial_order)	0.03	0.19	0.16	0.875	
OVS:Implausible:log(Trial_order)	-0.07	0.28	-0.26	0.794	
OVS:TwoGood:log(Trial_order)	-0.06	0.33	-0.17	0.867	

Table B.33. Sentence-picture matching, Experiment A. Reaction times on Verb.

Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.49	0.17	38.58	< 0.001	***
OVS	0.05	0.31	0.15	0.880	
Implausible	0.34	0.25	1.36	0.177	
TwoGood	0.31	0.30	1.01	0.314	
log(Trial_order)	-0.28	0.04	-6.88	< 0.001	***
OVS:Implausible	0.02	0.45	0.04	0.972	
OVS:TwoGood	-0.88	0.54	-1.64	0.105	
OVS:log(Trial_order)	0.03	0.10	0.27	0.789	

Table B.33. (cont.)

Implausible:log(Trial_order)	-0.10	0.08	-1.14	0.257	
TwoGood:log(Trial_order)	-0.08	0.10	-0.74	0.460	
OVS:Implausible:log(Trial_order)	0.00	0.15	0.00	0.999	
OVS:TwoGood:log(Trial_order)	0.23	0.18	1.32	0.192	

Table B.34. Sentence-picture matching, Experiment A. Reaction times on Sentence.

Sentence					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.90	0.17	56.77	< 0.001	***
OVS	0.07	0.33	0.22	0.824	
Implausible	-0.14	0.27	-0.53	0.600	
TwoGood	0.03	0.32	0.08	0.934	
log(Trial_order)	-0.38	0.04	-8.85	< 0.001	***
OVS:Implausible	0.33	0.48	0.70	0.488	
OVS:TwoGood	0.09	0.57	0.16	0.871	
OVS:log(Trial_order)	-0.02	0.11	-0.20	0.846	
Implausible:log(Trial_order)	0.07	0.09	0.80	0.429	
TwoGood:log(Trial_order)	0.06	0.11	0.51	0.613	
OVS:Implausible:log(Trial_order)	-0.13	0.16	-0.83	0.412	
OVS:TwoGood:log(Trial_order)	-0.05	0.19	-0.26	0.795	

Table B.35. Sentence-picture matching, Experiment A. Reaction times on Accuracy.

Accuracy					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	4.26	0.63	6.78	< 0.001	***
OVS	-2.55	0.99	-2.56	0.010	*

Table B.35 (cont.)

Implausible	-1.98	0.82	-2.40	0.016	*
TwoGood	-1.03	1.04	-0.99	0.323	
log(Trial_order)	-0.23	0.14	-1.62	0.105	
OVS:Implausible	0.00	1.42	0.00	0.999	
OVS:TwoGood	3.45	1.77	1.95	0.051	.
OVS:log(Trial_order)	0.06	0.33	0.19	0.847	
Implausible:log(Trial_order)	0.50	0.28	1.79	0.074	.
TwoGood:log(Trial_order)	0.38	0.36	1.06	0.291	
OVS:Implausible:log(Trial_order)	0.00	0.47	-0.01	0.992	
OVS:TwoGood:log(Trial_order)	-1.27	0.59	-2.15	0.032	*

Experiment B

Table B.36. Sentence-picture matching, Experiment B. Reaction times on Determiner 1.

Determiner 1

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.16	0.11	57.36	< 0.001	***
OVS	-0.15	0.15	-1.02	0.313	
Implausible	-0.22	0.12	-1.89	0.063	.
TwoGood	-0.06	0.14	-0.44	0.664	
log(Trial_order)	-0.18	0.02	-9.68	< 0.001	***
OVS:Implausible	0.35	0.21	1.67	0.099	.
OVS:TwoGood	0.04	0.25	0.18	0.862	
OVS:log(Trial_order)	0.04	0.05	0.86	0.391	
Implausible:log(Trial_order)	0.06	0.04	1.40	0.166	
TwoGood:log(Trial_order)	0.02	0.05	0.33	0.743	
OVS:Implausible:log(Trial_order)	-0.09	0.07	-1.37	0.174	

Table B.36. (cont.)

OVS:TwoGood:log(Trial_order)	-0.01	0.08	-0.15	0.885	

Table B.37. Sentence-picture matching, Experiment B. Reaction times on Noun 1.

Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.65	0.17	39.23	< 0.001	***
OVS	-0.56	0.20	-2.79	0.005	**
Implausible	-0.04	0.16	-0.25	0.803	
TwoGood	-0.10	0.19	-0.51	0.613	
log(Trial_order)	-0.27	0.03	-10.37	< 0.001	***
OVS:Implausible	0.35	0.29	1.21	0.228	
OVS:TwoGood	0.08	0.35	0.24	0.808	
OVS:log(Trial_order)	0.17	0.07	2.61	< 0. 01	**
Implausible:log(Trial_order)	0.00	0.05	0.09	0.930	
TwoGood:log(Trial_order)	0.05	0.07	0.69	0.488	
OVS:Implausible:log(Trial_order)	-0.11	0.09	-1.12	0.262	
OVS:TwoGood:log(Trial_order)	-0.02	0.11	-0.17	0.862	

Table B.38. Sentence-picture matching, Experiment B. Reaction times on Determiner 2.

Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.33	0.10	65.70	< 0.001	***
OVS	0.07	0.18	0.39	0.694	
Implausible	0.17	0.15	1.17	0.242	
TwoGood	0.07	0.18	0.37	0.712	
log(Trial_order)	-0.18	0.02	-7.60	< 0.001	***

Table B.38. (cont.)

OVS:Implausible	-0.21	0.26	-0.80	0.422	
OVS:TwoGood	-0.66	0.32	-2.08	0.038	*
OVS:log(Trial_order)	-0.01	0.06	-0.09	0.930	
Implausible:log(Trial_order)	-0.04	0.05	-0.86	0.393	
TwoGood:log(Trial_order)	0.01	0.06	0.11	0.914	
OVS:Implausible:log(Trial_order)	0.05	0.09	0.62	0.538	
OVS:TwoGood:log(Trial_order)	0.17	0.10	1.67	0.095	.

Table B.39. Sentence-picture matching, Experiment B. Reaction times on Noun 2.

Noun 2

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.66	0.18	37.99	< 0.001	***
OVS	0.11	0.36	0.31	0.758	
Implausible	0.12	0.29	0.41	0.684	
TwoGood	0.36	0.34	1.05	0.299	
log(Trial_order)	-0.12	0.05	-2.68	< 0. 01	**
OVS:Implausible	0.37	0.51	0.71	0.478	
OVS:TwoGood	0.15	0.62	0.25	0.807	
OVS:log(Trial_order)	-0.03	0.12	-0.28	0.780	
Implausible:log(Trial_order)	-0.03	0.10	-0.33	0.741	
TwoGood:log(Trial_order)	-0.08	0.12	-0.72	0.475	
OVS:Implausible:log(Trial_order)	-0.11	0.17	-0.68	0.499	
OVS:TwoGood:log(Trial_order)	-0.08	0.20	-0.39	0.697	

Table B.40. Sentence-picture matching, Experiment B. Reaction times on Last Word.

Last Word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.90	0.24	36.48	< 0.001	***
OVS	0.00	0.37	-0.01	0.992	
Implausible	-0.09	0.30	-0.29	0.772	
TwoGood	0.38	0.36	1.06	0.290	
log(Trial_order)	-0.28	0.05	-5.77	< 0.001	***
OVS:Implausible	0.01	0.53	0.02	0.981	
OVS:TwoGood	-0.55	0.64	-0.86	0.393	
OVS:log(Trial_order)	0.05	0.12	0.44	0.661	
Implausible:log(Trial_order)	0.10	0.10	0.97	0.334	
TwoGood:log(Trial_order)	-0.06	0.12	-0.51	0.610	
OVS:Implausible:log(Trial_order)	-0.08	0.18	-0.45	0.651	
OVS:TwoGood:log(Trial_order)	0.04	0.21	0.17	0.866	

Table B.41. Sentence-picture matching, Experiment B. Reaction times on Verb.

Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.93	0.15	46.25	< 0.001	***
OVS	-0.44	0.24	-1.89	0.063	.
Implausible	0.07	0.19	0.38	0.708	
TwoGood	-0.29	0.23	-1.30	0.198	
log(Trial_order)	-0.30	0.03	-9.86	< 0.001	***
OVS:Implausible	0.26	0.34	0.76	0.447	
OVS:TwoGood	0.36	0.41	0.88	0.384	
OVS:log(Trial_order)	0.08	0.08	1.06	0.294	

Table B.41. (cont.)

Implausible:log(Trial_order)	-0.06	0.06	-1.00	0.318	
TwoGood:log(Trial_order)	0.04	0.08	0.56	0.578	
OVS:Implausible:log(Trial_order)	-0.03	0.11	-0.30	0.769	
OVS:TwoGood:log(Trial_order)	-0.07	0.13	-0.55	0.583	

Table B.42. Sentence-picture matching, Experiment B. Reaction times on Sentence.

Sentence					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.67	0.14	69.01	< 0.001	***
OVS	-0.04	0.23	-0.16	0.876	
Implausible	0.08	0.18	0.46	0.650	
TwoGood	0.46	0.22	2.09	0.040	*
log(Trial_order)	-0.24	0.03	-7.95	< 0.001	***
OVS:Implausible	0.16	0.33	0.50	0.616	
OVS:TwoGood	-0.66	0.39	-1.66	0.100	
OVS:log(Trial_order)	0.04	0.08	0.50	0.620	
Implausible:log(Trial_order)	0.00	0.06	0.05	0.963	
TwoGood:log(Trial_order)	-0.11	0.07	-1.54	0.129	
OVS:Implausible:log(Trial_order)	-0.08	0.11	-0.70	0.485	
OVS:TwoGood:log(Trial_order)	0.13	0.13	0.99	0.326	

Table B.43. Sentence-picture matching, Experiment B. Reaction times on Accuracy.

Accuracy					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	3.68	0.70	5.28	< 0.001	***
OVS	-2.57	1.18	-2.17	0.030	*

Table B.43. (cont.)

Implausible	-1.49	0.86	-1.72	0.085	.
TwoGood	-0.56	1.17	-0.48	0.632	
log(Trial_order)	-0.24	0.15	-1.53	0.127	
OVS:Implausible	0.69	1.61	0.43	0.668	
OVS:TwoGood	-0.46	2.03	-0.23	0.820	
OVS:log(Trial_order)	0.97	0.40	2.44	0.015	*
Implausible:log(Trial_order)	0.39	0.30	1.31	0.189	
TwoGood:log(Trial_order)	0.15	0.40	0.37	0.710	
OVS:Implausible:log(Trial_order)	-0.39	0.54	-0.73	0.466	
OVS:TwoGood:log(Trial_order)	-0.11	0.68	-0.16	0.869	

Experiment C

Table B.44. Sentence-picture matching, Experiment C. Reaction times on Determiner 1.

Determiner 1

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.97	0.10	57.45	< 0.001	***
OVS	-0.08	0.15	-0.51	0.610	
Implausible	0.01	0.12	0.09	0.928	
TwoGood	0.15	0.14	1.07	0.287	
log(Trial_order)	-0.15	0.02	-7.84	< 0.001	***
OVS:Implausible	0.20	0.21	0.95	0.344	
OVS:TwoGood	0.07	0.26	0.29	0.773	
OVS:log(Trial_order)	0.02	0.05	0.40	0.689	
Implausible:log(Trial_order)	-0.02	0.04	-0.59	0.554	
TwoGood:log(Trial_order)	-0.04	0.05	-0.91	0.362	
OVS:Implausible:log(Trial_order)	-0.04	0.07	-0.52	0.603	

Table B.44 (cont.)

OVS:TwoGood:log(Trial_order)	-0.01	0.09	-0.09	0.932	

Table B.45. Sentence-picture matching, Experiment C. Reaction times on Noun 1.

Noun 1

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.30	0.17	37.05	< 0.001	***
OVS	-0.04	0.21	-0.18	0.855	
Implausible	0.09	0.17	0.55	0.581	
TwoGood	0.09	0.21	0.45	0.658	
log(Trial_order)	-0.21	0.03	-7.47	< 0.001	***
OVS:Implausible	0.05	0.31	0.15	0.879	
OVS:TwoGood	0.25	0.37	0.68	0.501	
OVS:log(Trial_order)	0.04	0.07	0.55	0.581	
Implausible:log(Trial_order)	-0.05	0.06	-0.79	0.432	
TwoGood:log(Trial_order)	0.02	0.07	0.28	0.782	
OVS:Implausible:log(Trial_order)	0.00	0.10	-0.03	0.977	
OVS:TwoGood:log(Trial_order)	-0.11	0.12	-0.94	0.348	

Table B.46. Sentence-picture matching, Experiment C. Reaction times on Determiner 2.

Determiner 2

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.22	0.14	45.82	< 0.001	***
OVS	0.04	0.26	0.15	0.883	
Implausible	0.21	0.21	1.01	0.314	
TwoGood	0.53	0.25	2.11	0.038	*
log(Trial_order)	-0.11	0.03	-3.36	< 0. 01	**

Table B.46. (cont.)

OVS:Implausible	0.04	0.37	0.10	0.918	
OVS:TwoGood	-0.52	0.45	-1.15	0.255	
OVS:log(Trial_order)	-0.03	0.09	-0.29	0.770	
Implausible:log(Trial_order)	-0.06	0.07	-0.83	0.410	
TwoGood:log(Trial_order)	-0.16	0.09	-1.89	0.062	.
OVS:Implausible:log(Trial_order)	-0.02	0.12	-0.13	0.901	
OVS:TwoGood:log(Trial_order)	0.15	0.15	0.99	0.324	

Table B.47. Sentence-picture matching, Experiment C. Reaction times on Noun 2.

Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.40	0.19	34.19	< 0.001	***
OVS	0.45	0.33	1.38	0.172	
Implausible	0.31	0.26	1.17	0.247	
TwoGood	0.83	0.32	2.61	0.011	*
log(Trial_order)	-0.11	0.04	-2.50	0.014	*
OVS:Implausible	0.05	0.47	0.11	0.913	
OVS:TwoGood	-0.15	0.57	-0.26	0.795	
OVS:log(Trial_order)	-0.12	0.11	-1.11	0.269	
Implausible:log(Trial_order)	-0.08	0.09	-0.87	0.388	
TwoGood:log(Trial_order)	-0.20	0.11	-1.84	0.069	.
OVS:Implausible:log(Trial_order)	-0.06	0.15	-0.36	0.719	
OVS:TwoGood:log(Trial_order)	-0.02	0.19	-0.08	0.935	

Table B.48. Sentence-picture matching, Experiment C. Reaction times on Last Word.

Last Word					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	8.68	0.27	31.58	< 0.001	***
OVS	0.44	0.47	0.94	0.345	
Implausible	0.63	0.38	1.66	0.097	.
TwoGood	0.42	0.45	0.94	0.350	
log(Trial_order)	-0.36	0.06	-5.89	< 0.001	***
OVS:Implausible	0.16	0.67	0.24	0.808	
OVS:TwoGood	-1.10	0.81	-1.36	0.175	
OVS:log(Trial_order)	-0.17	0.16	-1.10	0.270	
Implausible:log(Trial_order)	-0.16	0.13	-1.27	0.206	
TwoGood:log(Trial_order)	-0.09	0.15	-0.57	0.566	
OVS:Implausible:log(Trial_order)	-0.04	0.22	-0.19	0.849	
OVS:TwoGood:log(Trial_order)	0.34	0.27	1.28	0.201	

Table B.49. Sentence-picture matching, Experiment C. Reaction times on Verb.

Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.53	0.17	37.75	< 0.001	***
OVS	-0.19	0.25	-0.77	0.444	
Implausible	0.24	0.20	1.17	0.244	
TwoGood	0.36	0.24	1.49	0.136	
log(Trial_order)	-0.20	0.03	-6.20	< 0.001	***
OVS:Implausible	0.24	0.36	0.66	0.512	
OVS:TwoGood	0.46	0.44	1.04	0.299	
OVS:log(Trial_order)	0.07	0.08	0.85	0.395	

Table B.49 (cont.)

Implausible:log(Trial_order)	-0.08	0.07	-1.10	0.272	
TwoGood:log(Trial_order)	-0.12	0.08	-1.48	0.139	
OVS:Implausible:log(Trial_order)	-0.10	0.12	-0.81	0.421	
OVS:TwoGood:log(Trial_order)	-0.14	0.14	-1.00	0.317	

Table B.50. Sentence-picture matching, Experiment C. Reaction times on Sentence.

Sentence

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.46	0.17	56.13	< 0.001	***
OVS	0.31	0.28	1.11	0.271	
Implausible	0.43	0.23	1.92	0.058	.
TwoGood	0.51	0.27	1.90	0.061	.
log(Trial_order)	-0.28	0.04	-7.77	< 0.001	***
OVS:Implausible	0.12	0.40	0.30	0.765	
OVS:TwoGood	-0.68	0.48	-1.41	0.163	
OVS:log(Trial_order)	-0.11	0.09	-1.21	0.231	
Implausible:log(Trial_order)	-0.12	0.08	-1.58	0.118	
TwoGood:log(Trial_order)	-0.13	0.09	-1.48	0.143	
OVS:Implausible:log(Trial_order)	-0.04	0.13	-0.30	0.766	
OVS:TwoGood:log(Trial_order)	0.21	0.16	1.34	0.184	

Table B.51. Sentence-picture matching, Experiment C. Reaction times on Accuracy.

Accuracy

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	2.79	0.60	4.63	< 0.001	***
OVS	-2.69	0.84	-3.21	< 0. 01	**

Table B.51. (cont.)

Implausible	-0.59	0.65	-0.91	0.365	
TwoGood	1.76	1.31	1.35	0.178	
log(Trial_order)	-0.01	0.12	-0.04	0.965	
OVS:Implausible	-1.09	1.21	-0.90	0.370	
OVS:TwoGood	-2.45	1.93	-1.27	0.204	
OVS:log(Trial_order)	-0.24	0.29	-0.82	0.415	
Implausible:log(Trial_order)	0.48	0.26	1.85	0.065	.
TwoGood:log(Trial_order)	-0.36	0.43	-0.84	0.403	
OVS:Implausible:log(Trial_order)	-0.17	0.44	-0.38	0.704	
OVS:TwoGood:log(Trial_order)	0.46	0.64	0.73	0.468	

Appendix C. Model Outputs for Part 4 Analyses

Appendix C contains the model outputs for the statistical analyses for Part 5 (eye-tracking experiments A and B). The results are organized by experimental task, and by experiment.

Eye-tracking learning phase

Table C.1. Learning phase, Experiment A. Predicted total time spent on Determiner 1.

Total Time on Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.26	0.13	48.33	< 0.001	***
OVS	1.20	0.17	7.17	< 0.001	***
good_agents:2	-0.10	0.17	-0.62	0.537	
log(presentation_order)	-0.13	0.03	-3.87	< 0.001	***
OVS:good_agents:2	-0.08	0.27	-0.31	0.754	
OVS:log(presentation_order)	-0.15	0.05	-2.98	< 0.01	**
good_agents:2:log(presentation_order)	0.03	0.05	0.66	0.508	
OVS:good_agents:2:log(presentation_order)	0.02	0.08	0.28	0.777	

Table C.2. Learning phase, Experiment A. Predicted total time spent on Noun 1.

Total Time on Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.44	0.13	58.08	< 0.001	***
OVS	0.48	0.16	2.94	< 0.01	**
good_agents:2	0.31	0.15	2.04	0.041	*
log(presentation_order)	-0.11	0.03	-3.79	< 0.001	***
OVS:good_agents:2	-0.23	0.25	-0.92	0.357	
OVS:log(presentation_order)	-0.16	0.05	-3.42	< 0.001	***
good_agents:2:log(presentation_order)	-0.09	0.04	-2.07	0.039	*
OVS:good_agents:2:log(presentation_order)	0.09	0.07	1.23	0.219	

Table C.3. Learning phase, Experiment A. Predicted total time spent on Verb.

Total Time on Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.48	0.13	59.61	< 0.001	***
OVS	0.72	0.17	4.34	< 0.001	***
good_agents:2	-0.02	0.16	-0.10	0.921	
log(presentation_order)	-0.19	0.03	-6.30	< 0.001	***
OVS:good_agents:2	-0.20	0.26	-0.76	0.447	
OVS:log(presentation_order)	-0.16	0.05	-3.43	< 0.001	***
good_agents:2:log(presentation_order)	-0.03	0.04	-0.66	0.509	
OVS:good_agents:2:log(presentation_order)	0.05	0.07	0.74	0.459	

Table C.4. Learning phase, Experiment A. Predicted total time spent on Determiner 2.

Total Time on Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.87	0.12	56.86	< 0.001	***
OVS	-0.17	0.19	-0.92	0.361	
good_agents:2	0.11	0.16	0.68	0.500	
log(presentation_order)	-0.17	0.03	-5.52	< 0.001	***
OVS:good_agents:2	0.09	0.30	0.31	0.756	
OVS:log(presentation_order)	-0.10	0.06	-1.77	0.077	.
good_agents:2:log(presentation_order)	-0.02	0.05	-0.47	0.637	
OVS:good_agents:2:log(presentation_order)	-0.01	0.09	-0.07	0.945	

Table C.5. Learning phase, Experiment A. Predicted total time spent on Noun 2.

Total Time on Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.87	0.12	56.86	< 0.001	***
OVS	-0.17	0.19	-0.92	0.361	
good_agents:2	0.11	0.16	0.68	0.500	
log(presentation_order)	-0.17	0.03	-5.52	< 0.001	***
OVS:good_agents:2	0.09	0.30	0.31	0.756	
OVS:log(presentation_order)	-0.10	0.06	-1.77	0.077	.
good_agents:2:log(presentation_order)	-0.02	0.05	-0.47	0.637	
OVS:good_agents:2:log(presentation_order)	-0.01	0.09	-0.07	0.945	

Table C.6. Learning phase, Experiment A. Predicted total time spent on Picture.

Total Time on Picture					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.64	0.17	39.45	< 0.001	***
OVS	0.27	0.21	1.31	0.191	
good_agents:2	0.52	0.19	2.73	< 0.01	**
log(presentation_order)	-0.14	0.04	-3.50	< 0.001	***
OVS:good_agents:2	0.20	0.33	0.62	0.534	
OVS:log(presentation_order)	-0.05	0.06	-0.74	0.459	
good_agents:2:log(presentation_order)	-0.11	0.06	-1.97	0.050	*
OVS:good_agents:2:log(presentation_order)	-0.05	0.10	-0.53	0.595	

Table C.7. Learning phase, Experiment A. Predicted total time spent on Trial.

Total Time on Trial					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.37	0.10	96.63	< 0.001	***
OVS	0.41	0.10	4.00	< 0.001	***
good_agents:2	0.28	0.10	2.76	< 0.01	**
log(presentation_order)	-0.14	0.02	-6.92	< 0.001	***
OVS:good_agents:2	-0.17	0.16	-1.01	0.315	
OVS:log(presentation_order)	-0.13	0.03	-4.18	< 0.001	***
good_agents:2:log(presentation_order)	-0.09	0.03	-2.98	< 0.01	**
OVS:good_agents:2:log(presentation_order)	0.07	0.05	1.35	0.178	

Table C.8. Learning phase, Experiment A. Predicted GPT on Determiner 1.

GPT on Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.91	0.11	56.15	< 0.001	***
OVS	0.47	0.15	3.10	< 0.01	**
good_agents:2	0.15	0.15	0.96	0.337	
log(presentation_order)	0.00	0.03	-0.05	0.963	
OVS:good_agents:2	-0.11	0.25	-0.45	0.654	
OVS:log(presentation_order)	-0.09	0.05	-1.90	0.058	.
good_agents:2:log(presentation_order)	-0.07	0.05	-1.50	0.133	
OVS:good_agents:2:log(presentation_order)	0.05	0.07	0.69	0.490	

Table C.9. Learning phase, Experiment A. Predicted GPT on Noun 1.

GPT on Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.30	0.13	57.54	< 0.001	***
OVS	-0.32	0.18	-1.83	0.068	.
good_agents:2	-0.10	0.16	-0.63	0.532	
log(presentation_order)	-0.15	0.03	-4.58	< 0.001	***
OVS:good_agents:2	0.15	0.27	0.56	0.575	
OVS:log(presentation_order)	0.06	0.05	1.21	0.226	
good_agents:2:log(presentation_order)	0.03	0.05	0.52	0.606	
OVS:good_agents:2:log(presentation_order)	-0.04	0.08	-0.53	0.593	

Table C.10. Learning phase, Experiment A. Predicted GPT on Verb.

GPT on Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.07	0.14	51.73	< 0.001	***
OVS	0.33	0.20	1.68	0.094	.
good_agents:2	-0.08	0.19	-0.42	0.678	
log(presentation_order)	-0.13	0.04	-3.61	< 0.001	***
OVS:good_agents:2	-0.70	0.31	-2.24	0.025	*
OVS:log(presentation_order)	-0.07	0.06	-1.15	0.252	
good_agents:2:log(presentation_order)	-0.01	0.05	-0.17	0.862	
OVS:good_agents:2:log(presentation_order)	0.18	0.09	2.01	0.045	*

Table C.11. Learning phase, Experiment A. Predicted GPT on Determiner 2.

GPT on Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.23	0.13	48.08	< 0.001	***
OVS	-0.09	0.22	-0.40	0.692	
good_agents:2	-0.02	0.19	-0.11	0.914	
log(presentation_order)	-0.07	0.04	-1.88	0.061	.
OVS:good_agents:2	0.10	0.36	0.27	0.789	
OVS:log(presentation_order)	-0.04	0.07	-0.58	0.563	
good_agents:2:log(presentation_order)	0.00	0.06	0.05	0.962	
OVS:good_agents:2:log(presentation_order)	0.01	0.11	0.06	0.950	

Table C.12. Learning phase, Experiment A. Predicted GPT on Noun 2.

GPT on Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.61	0.14	48.65	< 0.001	***
OVS	0.57	0.20	2.87	< 0.01	**
good_agents:2	0.43	0.19	2.31	0.022	*
log(presentation_order)	-0.05	0.04	-1.38	0.169	
OVS:good_agents:2	-0.45	0.31	-1.42	0.155	
OVS:log(presentation_order)	-0.09	0.06	-1.55	0.121	
good_agents:2:log(presentation_order)	-0.09	0.06	-1.67	0.095	.
OVS:good_agents:2:log(presentation_order)	0.13	0.09	1.36	0.175	

Table C.13. Learning phase, Experiment A. Predicted regressions in Picture.

Regressions_IN_picture					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.33	0.15	2.15	0.032	*
OVS	0.29	0.16	1.82	0.068	.
good_agents:2	0.37	0.14	2.55	0.011	*
log(presentation_order)	-0.15	0.03	-4.80	< 0.001	***
OVS:good_agents:2	0.27	0.25	1.09	0.276	
OVS:log(presentation_order)	-0.05	0.05	-0.93	0.350	
good_agents:2:log(presentation_order)	-0.07	0.05	-1.55	0.122	
OVS:good_agents:2:log(presentation_order)	-0.07	0.08	-0.88	0.377	

Table C.14. Learning phase, Experiment A. Predicted regressions in Determiner 1.

Regressions IN Determiner 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.49	0.13	3.82	< 0.001	***
OVS	0.82	0.14	5.70	< 0.001	***
good_agents:2	-0.25	0.17	-1.49	0.137	
log(presentation_order)	-0.19	0.03	-5.74	0.000	***
OVS:good_agents:2	-0.02	0.26	-0.07	0.944	
OVS:log(presentation_order)	-0.14	0.05	-3.06	< 0.01	**
good_agents:2:log(presentation_order)	0.06	0.06	1.09	0.274	
OVS:good_agents:2:log(presentation_order)	0.02	0.08	0.28	0.777	

Table C.15. Learning phase, Experiment A. Predicted regressions in Noun 1.

Regressions IN Noun 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.14	0.15	0.99	0.322	
OVS	0.87	0.16	5.47	< 0.001	***
good_agents:2	0.39	0.16	2.50	0.013	*
log(presentation_order)	-0.10	0.03	-2.89	< 0.01	**
OVS:good_agents:2	-0.44	0.26	-1.68	0.093	.
OVS:log(presentation_order)	-0.24	0.05	-4.71	< 0.001	***
good_agents:2:log(presentation_order)	-0.11	0.05	-2.19	0.028	*
OVS:good_agents:2:log(presentation_order)	0.13	0.08	1.59	0.112	

Table C.16. Learning phase, Experiment A. Predicted regressions in Verb.

Regressions IN Verb					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.18	0.17	-1.12	0.264	
OVS	0.83	0.19	4.39	< 0.001	***
good_agents:2	0.41	0.19	2.19	0.029	*
log(presentation_order)	-0.17	0.04	-4.10	< 0.001	***
OVS:good_agents:2	0.18	0.30	0.61	0.545	
OVS:log(presentation_order)	-0.16	0.06	-2.62	< 0.01	**
good_agents:2:log(presentation_order)	-0.09	0.06	-1.55	0.122	
OVS:good_agents:2:log(presentation_order)	-0.08	0.10	-0.83	0.405	

Table C.17. Learning phase, Experiment A. Predicted regressions in Determiner 2.

Regressions IN Determiner 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.47	0.13	3.74	< 0.001	***
OVS	0.64	0.18	3.47	< 0.001	***
good_agents:2	0.15	0.16	0.96	0.339	
log(presentation_order)	-0.25	0.03	-7.35	< 0.001	***
OVS:good_agents:2	0.16	0.32	0.50	0.615	
OVS:log(presentation_order)	-0.24	0.06	-3.85	< 0.001	***
good_agents:2:log(presentation_order)	-0.05	0.05	-0.90	0.371	
OVS:good_agents:2:log(presentation_order)	0.00	0.11	0.02	0.985	

Table C.18. Learning phase, Experiment A. Predicted regressions in Noun 2.

Regressions IN Noun 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.02	0.15	-0.16	0.872	
OVS	0.64	0.19	3.43	< 0.001	***
good_agents:2	0.18	0.19	0.95	0.344	
log(presentation_order)	-0.17	0.04	-4.49	< 0.001	***
OVS:good_agents:2	0.48	0.31	1.57	0.116	
OVS:log(presentation_order)	-0.16	0.06	-2.64	< 0.01	**
good_agents:2:log(presentation_order)	-0.11	0.06	-1.85	0.065	.
OVS:good_agents:2:log(presentation_order)	-0.10	0.10	-1.06	0.289	

Table C.19. Learning phase, Experiment A. Predicted regressions out of Determiner 1.

Regressions OUT of Determiner 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.01	0.15	-0.08	0.933	
OVS	0.83	0.19	4.45	< 0.001	***
good_agents:2	-0.27	0.22	-1.23	0.218	
log(presentation_order)	-0.14	0.04	-3.35	< 0.001	***
OVS:good_agents:2	-0.05	0.33	-0.15	0.879	
OVS:log(presentation_order)	-0.18	0.06	-2.95	< 0.01	**
good_agents:2:log(presentation_order)	0.06	0.07	0.85	0.398	
OVS:good_agents:2:log(presentation_order)	0.02	0.10	0.17	0.866	

Table C.20. Learning phase, Experiment A. Predicted regressions out of Noun 1.

Regressions out of Noun 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.82	0.12	7.06	< 0.001	***
OVS	0.48	0.13	3.66	< 0.001	***
good_agents:2	0.17	0.13	1.37	0.170	
log(presentation_order)	-0.20	0.03	-7.54	< 0.001	***
OVS:good_agents:2	-0.12	0.22	-0.54	0.590	
OVS:log(presentation_order)	-0.12	0.04	-2.77	< 0.01	**
good_agents:2:log(presentation_order)	-0.05	0.04	-1.25	0.213	
OVS:good_agents:2:log(presentation_order)	0.05	0.07	0.80	0.424	

Table C.21. Learning phase, Experiment A. Predicted regressions out of Verb.

Regressions out of Verb					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.05	0.15	0.31	0.757	
OVS	1.05	0.16	6.62	< 0.001	***
good_agents:2	0.31	0.17	1.82	0.069	.
log(presentation_order)	-0.11	0.04	-2.97	< 0.01	**
OVS:good_agents:2	-0.54	0.26	-2.05	0.041	*
OVS:log(presentation_order)	-0.20	0.05	-3.89	< 0.001	***
good_agents:2:log(presentation_order)	-0.08	0.05	-1.44	0.151	
OVS:good_agents:2:log(presentation_order)	0.12	0.08	1.46	0.144	

Table C.22. Learning phase, Experiment A. Predicted regressions out of Determiner 2.

Regressions out of Determiner 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.12	0.15	0.83	0.409	
OVS	0.40	0.22	1.80	0.072	.
good_agents:2	0.27	0.18	1.55	0.121	
log(presentation_order)	-0.21	0.04	-5.60	< 0.001	***
OVS:good_agents:2	0.32	0.36	0.88	0.381	
OVS:log(presentation_order)	-0.23	0.08	-3.01	< 0.01	**
good_agents:2:log(presentation_order)	-0.09	0.06	-1.61	0.108	
OVS:good_agents:2:log(presentation_order)	0.02	0.12	0.19	0.853	

Table C.23. Learning phase, Experiment A. Predicted regressions out of Noun 2.

Regressions out of Noun 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.69	0.13	5.50	< 0.001	***
OVS	0.70	0.13	5.26	< 0.001	***
good_agents:2	0.18	0.13	1.37	0.169	
log(presentation_order)	-0.22	0.03	-7.66	< 0.001	***
OVS:good_agents:2	0.30	0.22	1.38	0.168	
OVS:log(presentation_order)	-0.17	0.04	-3.88	< 0.001	***
good_agents:2:log(presentation_order)	-0.03	0.04	-0.66	0.510	
OVS:good_agents:2:log(presentation_order)	-0.09	0.07	-1.27	0.206	

Table C.24. Learning phase, Experiment A. Predicted probability of skipping Ending 1.

Probability of skipping ending1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.11	0.34	3.22	< 0.01	**
OVS	-0.30	0.52	-0.58	0.565	
good_agents:2	-0.12	0.49	-0.26	0.797	
log(presentation_order)	-0.24	0.09	-2.55	0.011	*
OVS:good_agents:2	0.37	0.81	0.45	0.650	
OVS:log(presentation_order)	0.28	0.15	1.88	0.060	.
good_agents:2:log(presentation_order)	0.03	0.14	0.23	0.817	
OVS:good_agents:2:log(presentation_order)	-0.17	0.23	-0.73	0.463	

Table C.25. Learning phase, Experiment A. Predicted probability of skipping Ending 2.

Probability of skipping ending2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.36	0.36	1.00	0.318	
OVS	-0.99	0.55	-1.82	0.069	.
good_agents:2	-0.93	0.52	-1.78	0.076	.
log(presentation_order)	-0.03	0.10	-0.29	0.770	
OVS:good_agents:2	0.50	0.86	0.58	0.563	
OVS:log(presentation_order)	0.07	0.15	0.47	0.640	
good_agents:2:log(presentation_order)	0.22	0.14	1.53	0.126	
OVS:good_agents:2:log(presentation_order)	-0.07	0.24	-0.30	0.761	

Table C.26. Learning phase, Experiment B. Predicted total time on Determiner 1.

Total time Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.03	0.11	53.33	< 0.001	***
OVS	1.18	0.15	7.90	< 0.001	***
good_agents:2	0.09	0.15	0.61	0.545	
log(presentation_order)	-0.10	0.03	-3.39	< 0.001	***
OVS:good_agents:2	-0.39	0.24	-1.65	0.100	
OVS:log(presentation_order)	-0.14	0.05	-3.12	< 0.01	**
good_agents:2:log(presentation_order)	-0.03	0.05	-0.73	0.463	
OVS:good_agents:2:log(presentation_order)	0.12	0.07	1.72	0.086	.

Table C.27. Learning phase, Experiment B. Predicted total time on Noun 1.

Total Time on Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.08	0.11	64.59	< 0.001	***
OVS	0.55	0.14	3.85	< 0.001	***
good_agents:2	0.25	0.13	1.88	0.060	.
log(presentation_order)	-0.06	0.03	-2.19	0.029	*
OVS:good_agents:2	-0.56	0.23	-2.48	0.013	*
OVS:log(presentation_order)	-0.19	0.04	-4.47	< 0.001	***
good_agents:2:log(presentation_order)	-0.07	0.04	-1.73	0.084	.
OVS:good_agents:2:log(presentation_order)	0.17	0.07	2.56	0.011	*

Table C.28. Learning phase, Experiment B. Predicted total time on Verb.

Total Time on Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.88	0.11	63.53	< 0.001	***
OVS	0.76	0.15	5.19	< 0.001	***
good_agents:2	0.14	0.14	1.00	0.318	
log(presentation_order)	-0.08	0.03	-2.88	< 0.01	**
OVS:good_agents:2	-0.78	0.23	-3.36	< 0.001	***
OVS:log(presentation_order)	-0.19	0.04	-4.37	< 0.001	***
good_agents:2:log(presentation_order)	-0.08	0.04	-1.96	0.050	*
OVS:good_agents:2:log(presentation_order)	0.22	0.07	3.29	< 0.01	**

Table C.29. Learning phase, Experiment B. Predicted total time on Determiner 2.

Total Time on Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.51	0.10	64.25	< 0.001	***
OVS	-0.48	0.16	-2.95	< 0.01	**
good_agents:2	0.14	0.13	1.01	0.314	
log(presentation_order)	-0.14	0.03	-5.31	< 0.001	***
OVS:good_agents:2	-0.21	0.27	-0.79	0.427	
OVS:log(presentation_order)	0.01	0.05	0.25	0.803	
good_agents:2:log(presentation_order)	-0.03	0.04	-0.71	0.478	
OVS:good_agents:2:log(presentation_order)	0.05	0.08	0.60	0.552	

Table C.30. Learning phase, Experiment B. Predicted total time on Noun 2.

Total Time on Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.00	0.11	65.21	< 0.001	***
OVS	0.43	0.15	2.93	< 0.01	**
good_agents:2	0.21	0.14	1.50	0.133	
log(presentation_order)	-0.17	0.03	-6.08	< 0.001	***
OVS:good_agents:2	-0.15	0.23	-0.66	0.508	
OVS:log(presentation_order)	-0.05	0.04	-1.17	0.244	
good_agents:2:log(presentation_order)	-0.04	0.04	-1.05	0.295	
OVS:good_agents:2:log(presentation_order)	0.04	0.07	0.65	0.514	

Table C.31. Learning phase, Experiment B. Predicted total time on Trial.

Total Time on Trial					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	9.17	0.10	94.89	< 0.001	***
OVS	0.48	0.11	4.45	< 0.001	***
good_agents:2	0.19	0.11	1.76	0.080	.
log(presentation_order)	-0.11	0.02	-4.86	< 0.001	***
OVS:good_agents:2	-0.33	0.17	-1.91	0.056	.
OVS:log(presentation_order)	-0.14	0.03	-4.39	< 0.001	***
good_agents:2:log(presentation_order)	-0.05	0.03	-1.65	0.100	.
OVS:good_agents:2:log(presentation_order)	0.09	0.05	1.78	0.075	.

Table C.32. Learning phase, Experiment B. Predicted total time on English Sentence.

Total Time on English Sentence					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.14	0.18	40.61	< 0.001	***
OVS	0.94	0.20	4.78	< 0.001	***
good_agents:2	0.53	0.19	2.86	< 0.01	**
log(presentation_order)	-0.10	0.04	-2.59	< 0.01	**
OVS:good_agents:2	-0.58	0.32	-1.82	0.069	.
OVS:log(presentation_order)	-0.24	0.06	-4.01	< 0.001	***
good_agents:2:log(presentation_order)	-0.13	0.06	-2.34	0.019	*
OVS:good_agents:2:log(presentation_order)	0.18	0.10	1.86	0.063	.

Table C.33. Learning phase, Experiment B. Predicted GPT on Determiner 1.

GPT on Determiner 1	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.02	0.11	54.79	< 0.001	***
OVS	0.47	0.16	2.94	< 0.01	**
good_agents:2	-0.14	0.16	-0.86	0.389	
log(presentation_order)	-0.07	0.03	-2.03	0.043	*
OVS:good_agents:2	-0.16	0.26	-0.62	0.533	
OVS:log(presentation_order)	-0.08	0.05	-1.56	0.119	
good_agents:2:log(presentation_order)	0.01	0.05	0.23	0.816	
OVS:good_agents:2:log(presentation_order)	0.07	0.08	0.92	0.360	

Table C.34. Learning phase, Experiment B. Predicted GPT on Noun 1.

GPT on Noun 1	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.03	0.11	61.54	< 0.001	***
OVS	-0.14	0.17	-0.84	0.399	
good_agents:2	-0.09	0.16	-0.59	0.556	
log(presentation_order)	-0.11	0.03	-3.56	< 0.001	***
OVS:good_agents:2	0.15	0.27	0.55	0.581	
OVS:log(presentation_order)	-0.01	0.05	-0.17	0.864	
good_agents:2:log(presentation_order)	0.02	0.05	0.49	0.623	
OVS:good_agents:2:log(presentation_order)	-0.02	0.08	-0.26	0.794	

Table C.35. Learning phase, Experiment B. Predicted GPT on Verb.

GPT on Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.70	0.12	54.00	< 0.001	***
OVS	0.22	0.19	1.18	0.238	
good_agents:2	-0.11	0.18	-0.60	0.552	
log(presentation_order)	-0.08	0.04	-2.30	0.022	*
OVS:good_agents:2	-0.72	0.30	-2.43	0.015	*
OVS:log(presentation_order)	-0.06	0.06	-1.00	0.316	
good_agents:2:log(presentation_order)	-0.01	0.05	-0.17	0.863	
OVS:good_agents:2:log(presentation_order)	0.21	0.09	2.39	0.017	*

Table C.36. Learning phase, Experiment B. Predicted GPT on Determiner 2.

GPT on Determiner 2				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.74	0.13	44.54	< 0.001
OVS	-0.10	0.22	-0.46	0.646
good_agents:2	0.28	0.19	1.52	0.131
log(presentation_order)	0.02	0.04	0.47	0.639
OVS:good_agents:2	-0.19	0.37	-0.52	0.604
OVS:log(presentation_order)	0.04	0.07	0.55	0.585
good_agents:2:log(presentation_order)	-0.05	0.06	-0.81	0.417
OVS:good_agents:2:log(presentation_order)	0.02	0.11	0.21	0.832

Table C.37. Learning phase, Experiment B. Predicted GPT on Noun 2.

GPT on Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.50	0.13	49.67	< 0.001	***
OVS	0.39	0.20	1.99	0.048	*
good_agents:2	0.13	0.18	0.70	0.486	
log(presentation_order)	-0.05	0.04	-1.38	0.168	
OVS:good_agents:2	-0.29	0.31	-0.92	0.356	
OVS:log(presentation_order)	-0.03	0.06	-0.51	0.608	
good_agents:2:log(presentation_order)	0.00	0.06	-0.05	0.959	
OVS:good_agents:2:log(presentation_order)	0.06	0.09	0.61	0.546	

Table C.38. Learning phase, Experiment B. Predicted regressions in English Sentence.

Regression IN English sentence					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.99	0.13	7.57	< 0.001	***
OVS	0.71	0.10	7.08	< 0.001	***
good_agents:2	0.45	0.10	4.66	< 0.001	***
log(presentation_order)	-0.10	0.02	-4.50	< 0.001	***
OVS:good_agents:2	-0.63	0.16	-3.85	< 0.001	***
OVS:log(presentation_order)	-0.18	0.03	-5.71	< 0.001	***
good_agents:2:log(presentation_order)	-0.12	0.03	-3.78	< 0.001	***
OVS:good_agents:2:log(presentation_order)	0.19	0.05	3.77	< 0.001	***

Table C.39. Learning phase, Experiment B. Predicted regressions in Determiner 1.

Regression IN Determiner 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.18	0.15	1.21	0.225	
OVS	0.85	0.16	5.29	< 0.001	***
good_agents:2	-0.35	0.20	-1.74	0.081	.
log(presentation_order)	-0.19	0.04	-4.86	< 0.001	***
OVS:good_agents:2	-0.23	0.29	-0.81	0.416	
OVS:log(presentation_order)	-0.11	0.05	-2.17	0.030	*
good_agents:2:log(presentation_order)	0.10	0.07	1.47	0.141	
OVS:good_agents:2:log(presentation_order)	0.08	0.09	0.85	0.396	

Table C.40. Learning phase, Experiment B. Predicted regressions in Noun 1.

Regression IN Noun 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.89	0.19	-4.72	< 0.001	***
OVS	1.53	0.21	7.38	< 0.001	***
good_agents:2	0.75	0.21	3.53	< 0.001	***
log(presentation_order)	0.13	0.05	2.79	< 0.01	**
OVS:good_agents:2	-1.19	0.32	-3.68	< 0.001	***
OVS:log(presentation_order)	-0.43	0.06	-6.71	< 0.001	***
good_agents:2:log(presentation_order)	-0.24	0.07	-3.72	< 0.001	***
OVS:good_agents:2:log(presentation_order)	0.37	0.10	3.80	< 0.001	***

Table C.41. Learning phase, Experiment B. Predicted regressions in Verb.

Regression IN Verb					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-1.25	0.22	-5.67	< 0.001	***
OVS	1.14	0.26	4.39	< 0.001	***
good_agents:2	0.70	0.27	2.63	< 0.01	**
log(presentation_order)	0.06	0.06	1.04	0.301	
OVS:good_agents:2	-0.68	0.40	-1.70	0.090	.
OVS:log(presentation_order)	-0.22	0.08	-2.81	< 0.01	**
good_agents:2:log(presentation_order)	-0.21	0.08	-2.50	0.012	*
OVS:good_agents:2:log(presentation_order)	0.17	0.12	1.39	0.164	

Table C.42. Learning phase, Experiment B. Predicted regressions in Determiner 2.

Regression IN Det2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.13	0.14	0.95	0.342	
OVS	-0.07	0.23	-0.32	0.746	
good_agents:2	-0.06	0.19	-0.35	0.728	
log(presentation_order)	-0.19	0.04	-4.99	< 0.001	***
OVS:good_agents:2	-0.30	0.42	-0.71	0.479	
OVS:log(presentation_order)	-0.01	0.08	-0.17	0.867	
good_agents:2:log(presentation_order)	-0.03	0.06	-0.43	0.664	
OVS:good_agents:2:log(presentation_order)	0.12	0.13	0.88	0.377	

Table C.43. Learning phase, Experiment B. Predicted regressions in Noun 2.

Regression IN Noun 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.25	0.16	-1.58	0.115	
OVS	0.17	0.23	0.75	0.456	
good_agents:2	0.20	0.21	0.96	0.339	
log(presentation_order)	-0.25	0.05	-5.50	< 0.001	***
OVS:good_agents:2	0.50	0.37	1.35	0.176	
OVS:log(presentation_order)	0.02	0.07	0.28	0.783	
good_agents:2:log(presentation_order)	-0.05	0.07	-0.69	0.493	
OVS:good_agents:2:log(presentation_order)	-0.18	0.12	-1.54	0.125	

Table C.44. Learning phase, Experiment B. Predicted regressions out of Determiner 1.

Regression OUT of Determiner 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.11	0.14	0.77	0.444	
OVS	0.85	0.17	4.94	< 0.001	***
good_agents:2	-0.15	0.21	-0.75	0.453	
log(presentation_order)	-0.16	0.04	-4.12	< 0.001	***
OVS:good_agents:2	-0.20	0.30	-0.65	0.517	
OVS:log(presentation_order)	-0.16	0.06	-2.82	< 0.01	**
good_agents:2:log(presentation_order)	0.00	0.07	0.02	0.981	
OVS:good_agents:2:log(presentation_order)	0.10	0.10	1.02	0.310	

Table C.45. Learning phase, Experiment B. Predicted regressions out of Noun 1.

Regression OUT of Noun 1					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.56	0.13	4.40	< 0.001	***
OVS	0.80	0.13	6.30	< 0.001	***
good_agents:2	0.20	0.13	1.55	0.120	
log(presentation_order)	-0.11	0.03	-4.13	< 0.001	***
OVS:good_agents:2	-0.63	0.21	-2.95	< 0.01	**
OVS:log(presentation_order)	-0.19	0.04	-4.79	< 0.001	***
good_agents:2:log(presentation_order)	-0.04	0.04	-1.04	0.299	
OVS:good_agents:2:log(presentation_order)	0.18	0.07	2.80	0.005	**

Table C.46. Learning phase, Experiment B. Predicted regressions out of Verb.

Regression OUT of Verb					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.25	0.15	-1.65	0.100	.
OVS	1.22	0.17	7.26	< 0.001	***
good_agents:2	0.38	0.18	2.06	0.039	*
log(presentation_order)	0.00	0.04	0.06	0.951	
OVS:good_agents:2	-0.97	0.28	-3.46	< 0.001	***
OVS:log(presentation_order)	-0.29	0.05	-5.56	< 0.001	***
good_agents:2:log(presentation_order)	-0.15	0.06	-2.72	< 0.01	**
OVS:good_agents:2:log(presentation_order)	0.32	0.09	3.73	< 0.001	***

Table C.47. Learning phase, Experiment B. Predicted regressions out of Determiner 2.

Regression OUT of Det 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.21	0.16	-1.36	0.173	
OVS	-0.60	0.28	-2.13	0.033	*
good_agents:2	0.40	0.19	2.09	0.036	*
log(presentation_order)	-0.15	0.04	-3.45	< 0.001	***
OVS:good_agents:2	-0.30	0.47	-0.65	0.519	
OVS:log(presentation_order)	0.16	0.09	1.76	0.079	.
good_agents:2:log(presentation_order)	-0.08	0.06	-1.34	0.179	
OVS:good_agents:2:log(presentation_order)	0.05	0.14	0.36	0.721	

Table C.48. Learning phase, Experiment B. Predicted regressions out of Noun 2.

Regression OUT of Noun 2					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	0.39	0.14	2.89	< 0.01	**
OVS	0.66	0.14	4.64	< 0.001	***
good_agents:2	0.20	0.14	1.40	0.163	
log(presentation_order)	-0.14	0.03	-4.48	< 0.001	***
OVS:good_agents:2	-0.37	0.24	-1.56	0.119	
OVS:log(presentation_order)	-0.15	0.05	-3.30	< 0.001	***
good_agents:2:log(presentation_order)	-0.04	0.05	-1.00	0.319	
OVS:good_agents:2:log(presentation_order)	0.11	0.07	1.56	0.120	

Table C.49. Learning phase, Experiment B. Predicted probability of skipping Ending 1.

Probability of skipping ending 1				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.67	0.36	1.88	0.0595 .
OVS	0.21	0.53	0.39	0.699
good_agents:2	0.34	0.50	0.68	0.498
log(presentation_order)	-0.14	0.09	-1.51	0.131
OVS:good_agents:2	0.23	0.85	0.28	0.783
OVS:log(presentation_order)	0.13	0.15	0.89	0.376
good_agents:2:log(presentation_order)	-0.03	0.14	-0.21	0.830
OVS:good_agents:2:log(presentation_order)	-0.12	0.24	-0.51	0.612

Table C.50. Learning phase, Experiment B. Predicted probability of skipping Ending 2.

Probability of skipping ending 2				
	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.08	0.36	0.24	0.814
OVS	-0.35	0.53	-0.66	0.509
good_agents:2	0.03	0.51	0.07	0.946
log(presentation_order)	0.10	0.10	1.07	0.283
OVS:good_agents:2	-0.38	0.83	-0.46	0.646
OVS:log(presentation_order)	-0.06	0.15	-0.44	0.662
good_agents:2:log(presentation_order)	-0.05	0.14	-0.34	0.736
OVS:good_agents:2:log(presentation_order)	0.12	0.23	0.53	0.600

Visual world

Table C.51. Visual World, Experiments A & B. Predicted TT on Determiner 1.

Determiner 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.56	0.11	51.86	< 0.001	***
OVS	0.62	0.22	2.86	< 0.01	**
TwoGood	-0.10	0.24	-0.42	0.678	
Implausible	0.19	0.22	0.85	0.401	
log(presentation_order)	-0.04	0.03	-1.42	0.158	
ExperimentB	0.02	0.08	0.31	0.758	
OVS:TwoGood	0.25	0.39	0.63	0.529	
OVS:Implausible	0.20	0.36	0.55	0.586	
OVS:log(presentation_order)	0.01	0.07	0.15	0.882	
TwoGood:log(presentation_order)	0.04	0.08	0.55	0.582	
Implausible:log(presentation_order)	-0.06	0.08	-0.73	0.467	
OVS:TwoGood:log(presentation_order)	-0.11	0.13	-0.87	0.387	
OVS:Implausible:log(presentation_order)	-0.09	0.12	-0.80	0.429	

Table C.52. Visual World, Experiments A & B. Predicted TT on Noun 1.

Noun 1					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	7.31	0.10	73.99	< 0.001	***
OVS	-0.19	0.19	-1.01	0.317	
TwoGood	0.11	0.18	0.63	0.528	
Implausible	0.35	0.17	2.05	0.046	*
log(presentation_order)	-0.17	0.02	-7.28	< 0.001	***
ExperimentB	0.04	0.09	0.38	0.703	

Table C.52. (cont.)

OVS:TwoGood	-0.30	0.32	-0.93	0.354	
OVS:Implausible	-0.05	0.30	-0.16	0.877	
OVS:log(presentation_order)	0.01	0.06	0.20	0.842	
TwoGood:log(presentation_order)	0.00	0.06	-0.05	0.957	
Implausible:log(presentation_order)	-0.10	0.06	-1.74	0.088	.
OVS:TwoGood:log(presentation_order)	0.08	0.11	0.75	0.457	
OVS:Implausible:log(presentation_order)	0.01	0.10	0.12	0.905	

Table C.53. Visual World, Experiments A & B. Predicted TT on Verb.

Verb					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.92	0.12	59.18	< 0.001	***
OVS	-0.09	0.21	-0.43	0.669	
TwoGood	0.27	0.21	1.30	0.197	
Implausible	0.31	0.20	1.54	0.132	
log(presentation_order)	-0.16	0.03	-5.87	< 0.001	***
ExperimentB	0.03	0.11	0.29	0.771	
OVS:TwoGood	0.27	0.37	0.73	0.471	
OVS:Implausible	0.09	0.35	0.25	0.803	
OVS:log(presentation_order)	0.03	0.07	0.38	0.703	
TwoGood:log(presentation_order)	-0.14	0.07	-2.03	0.047	*
Implausible:log(presentation_order)	-0.06	0.07	-0.88	0.387	
OVS:TwoGood:log(presentation_order)	-0.10	0.12	-0.81	0.422	
OVS:Implausible:log(presentation_order)	-0.06	0.11	-0.49	0.627	

Table C.54. Visual World, Experiments A & B. Predicted TT on Determiner 2.

Determiner 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.26	0.12	51.39	< 0.001	***
OVS	-0.35	0.30	-1.17	0.246	
TwoGood	0.09	0.27	0.32	0.747	
Implausible	0.41	0.27	1.51	0.138	
log(presentation_order)	-0.14	0.04	-3.92	< 0.001	***
ExperimentB	-0.04	0.08	-0.51	0.609	
OVS:TwoGood	-0.89	0.52	-1.72	0.089	.
OVS:Implausible	-0.56	0.49	-1.15	0.253	
OVS:log(presentation_order)	0.01	0.10	0.14	0.887	
TwoGood:log(presentation_order)	0.02	0.09	0.25	0.802	
Implausible:log(presentation_order)	-0.10	0.09	-1.08	0.284	
OVS:TwoGood:log(presentation_order)	0.22	0.17	1.29	0.202	
OVS:Implausible:log(presentation_order)	0.11	0.16	0.67	0.504	

Table C.55. Visual World, Experiments A & B. Predicted TT on Noun 2.

Noun 2					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.56	0.10	63.49	< 0.001	***
OVS	0.38	0.21	1.85	0.069	.
TwoGood	0.11	0.21	0.51	0.611	
Implausible	0.24	0.20	1.20	0.236	
log(presentation_order)	-0.13	0.03	-4.63	< 0.001	***
ExperimentB	0.01	0.08	0.18	0.861	
OVS:TwoGood	-0.22	0.37	-0.61	0.546	

Table C.55. (cont.)

OVS:Implausible	0.05	0.34	0.16	0.878	
OVS:log(presentation_order)	0.00	0.07	-0.05	0.960	
TwoGood:log(presentation_order)	0.02	0.07	0.25	0.803	
Implausible:log(presentation_order)	-0.05	0.07	-0.72	0.474	
OVS:TwoGood:log(presentation_order)	0.04	0.12	0.33	0.746	
OVS:Implausible:log(presentation_order)	-0.02	0.11	-0.18	0.859	

Table C.56. Visual World, Experiments A & B. Predicted TT on the Correct Picture.

Correct picture					
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	6.98	0.08	92.48	< 0.001	***
OVS	-0.14	0.13	-1.07	0.284	
TwoGood	0.17	0.13	1.37	0.171	
Implausible	0.06	0.11	0.60	0.549	
log(presentation_order)	-0.09	0.02	-5.34	< 0.001	***
ExperimentB	-0.05	0.08	-0.60	0.551	
OVS:TwoGood	-0.19	0.23	-0.86	0.391	
OVS:Implausible	0.28	0.19	1.49	0.136	
OVS:log(presentation_order)	0.05	0.04	1.06	0.287	
TwoGood:log(presentation_order)	-0.03	0.04	-0.67	0.500	
Implausible:log(presentation_order)	-0.02	0.04	-0.47	0.637	
OVS:TwoGood:log(presentation_order)	0.04	0.07	0.59	0.555	
OVS:Implausible:log(presentation_order)	-0.09	0.06	-1.51	0.132	

Table C.57. Visual World, Experiments A & B. Predicted Accuracy.

Accuracy					
	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	3.28	0.39	8.44	< 0.001	***
ExperimentB	-0.39	0.53	-0.73	0.467	
OVS	-0.57	0.26	-2.16	0.030	*
TwoGood	-0.54	0.26	-2.05	0.041	*
Implausible	-1.07	0.26	-4.17	< 0.001	***
ExperimentB:OVS	-0.11	0.35	-0.32	0.753	
ExperimentB:TwoGood	-0.21	0.35	-0.61	0.542	
ExperimentB:Implausible	-0.20	0.34	-0.58	0.559	
OVS:TwoGood	0.22	0.36	0.62	0.539	
OVS:Implausible	0.28	0.35	0.79	0.427	
ExperimentB:OVS:TwoGood	-0.22	0.47	-0.47	0.635	
ExperimentB:OVS:Implausible	0.22	0.46	0.48	0.631	