Exercise Design and Vocabulary Learning in Tutorial CALL: The Effects of Image Features and Combinations on Attention to Written Forms

by

Colleen Patchin

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved June 2019 by the
Graduate Supervisory Committee:

David Smith, Chair
Mark James
Andrew Ross

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Computer assisted language learning (CALL) has become increasingly common as a means of helping learners develop essential skills in a second or foreign language. However, while many CALL programs claim to be based on principles of second language acquisition (SLA) theory and research, evaluation of design and learning outcomes at the level of individual CALL exercises is lacking in the existing literature. The following proposed study will explore the design of computer-based vocabulary matching exercises using both written text and images and the effects of various design manipulations on learning outcomes. The study will use eye-tracking to investigate what users attend to on screen as they work through a series of exercises with different configurations of written words and images. It will ask whether manipulation of text and image features and combinations can have an effect on learners' attention to the various elements, and if so, whether differences in levels of attention results in higher or lower scores for measures of learning. Specifically, eye-tracking data will be compared to post-test scores for recall and recognition of target vocabulary items to look for a correlation between levels of attention to written forms in-task and post-test gains in scores for vocabulary learning.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF FIGURES

Figure Page

Introduction

Vocabulary acquisition is considered fundamental for second language learning and necessary for the development of reading, writing, listening, and speaking skills (Brown & Lee, 2015). At early stages, learners with limited to no knowledge of the target language (TL) require scaffolding of language and content in order for TL vocabulary to be processed and their meanings learned. Two examples of scaffolding, images and L1 translations, are often included in second language learning materials in order to assist early language learners in gaining vocabulary needed for the development of essential language skills. The inclusion of images is thought to aid in the activation of learners' existing conceptual knowledge and create connections between this knowledge and the TL (Clark & Paivio, 1991). Similarly, the use of L1 translations is meant to help beginning learners grasp the meanings of TL vocabulary quickly and easily through reference to known words in the L1 (Al-Seghayer, 2001; Yoshii, 2006). The use of images and translations together can therefore be very beneficial for language learners to access the meanings of unknown words through multiple modes and make stronger connections between these words and existing knowledge.

Following from these assumptions, many common tutorial CALL programs make use of both images and L1 translations in their various presentation and practice exercises as it is thought to facilitate language learning. One commonly observed exercise for vocabulary instruction found in many popular web and mobile based language learning programs is a matching exercise in which users match vocabulary items between the L1 and L2 (see App Review for examples). This type of exercise may be designed to include

only the written forms of vocabulary words to be matched, or it may feature images as well, accompanying either the L1 or L2 written forms. In the case in which images are paired with L2 written forms, making a match may be especially easy for users, and so special considerations must be made regarding image selection and combinations to ensure learners notice written forms and learn the target vocabulary. It is this version of the vocabulary matching exercise which the proposed study will focus on, and from which questions regarding attention, noticing, and vocabulary acquisition will be addressed.

This study will explore a number of questions raised by this dual-scaffolded matching exercise. The first question raised concerns the extent to which users attend to the images or the written text on screen in making decisions regarding matching translations. This question is relevant to ask since both text and images serve as sources of information for making a potential match, and if users draw on them in varying degrees under different design conditions, it could be revealing as to how these two types of information contribute to and interact in multimodal learning environments. The second question raised asks, if users do attend more to one source of information over the other in certain types of conditions, will those differences in attention have an effect on learning outcomes for target language items? Specifically, this study seeks to ask whether certain design conditions may draw users' attention more to images over written forms, and if in doing so, a lack of attention to forms during on-line practice may result in lower recognition and/or recall of target vocabulary following the task. This question is relevant to ask in order to spot potential conflicts in design which could interfere with learning

goals and to understand ways in which improvements could be made to better achieve learning outcomes.

This study will use eye-tracking to investigate where users fixate their gaze on the screen when working through a vocabulary matching exercise in which they are presented with TL vocabulary accompanied by images and asked to match an L1 word to the correct L2 equivalent. The study will include a pre- and post-assessment in which participants' vocabulary knowledge prior to and following the treatment task will be measured, and these scores will be compared to measures of attention during the treatment task. Discussion will focus on ways of improving CALL design based on the findings to better serve language learners' needs. The rest of the paper will proceed as follows: A review of the existing literature on second language vocabulary learning, a review of methods for measuring attention and noticing during language learning tasks, a discussion of the use of images and translations in second language materials, a discussion of CALL vocabulary exercises and a review of examples from popular programs, followed by a methods section and a discussion of predictions, potential applications, and possible future studies.

Second Language Learning: Definitions and Terms

Second language acquisition (SLA) refers to "the scholarly field of inquiry that investigates the human capacity to learn languages other than the first, during late childhood, adolescence or adulthood, and once the first language or languages have been acquired" (Ortega, 2014, pp. 1-2). SLA is a broad and continually growing field of study, within which exists a number of concepts, theories, and models which aim to describe the

various components and processes involved in learning a second language, as well as the distinctions and relations among them. In the following sections, several key conceptual and terminological distinctions of importance to the proposed study will be defined and explored with respect to the questions this study seeks to investigate.

The Distinction Between Second and Foreign Language Learning

The first distinction which is of relevance to the present proposal is that made between *second language* (SL) and *foreign language* (FL) learning. SL learning, also referred to as *naturalistic* language learning (Muñoz, 2008), usually refers to learning a language which is present in the local environment in which the learner is located, whether it be the primary language spoken by most residents as a first language, or an additional language used for wider "commercial, administrative, [or] educational" purposes (Broughton, Brumfit, Pincas, & Wilde, 2002, p. 7). FL learning on the other hand, describes learning a language which is not spoken in the learner's immediate environment and which does not play a role in their everyday activities (Broughton, Brumfit, Pincas, & Wilde, 2002). While SL learning typically takes place out of necessity for the learner to use the language for every day purposes, FL learning may take place for a number of different purposes, ranging from a simple interest in the culture, to in preparation for a future visit or move to the region, or for strictly academic purposes such as meeting a language requirement for a school program. FL learning often takes place in a remote classroom setting and is characterized as being "formal, planned, and systematic," while SL learning, although it may be supported by classroom instruction,

usually occurs through natural exposure and use and is characterized as being "informal and unstructured" (Mitchell, Myles & Marsden, 2019, p. 6).

Those maintain the distinction between SL and FL learning claim that "the differences in the amount and quality of the respective input of the two learning settings may have a significant influence on… the outcome of second language learning" (Muñoz, 2008, p. 578). Not all subscribe to this view however, with many preferring to use 'second language learning' to refer more broadly to "the learning of any language, to any level, provided only that the learning of the 'second language' takes place sometime later than the acquisition of the first language" (Mitchell, Myles & Marsden, 2019, p. 5). This more widely encompassing definition can be useful when making claims about language learning which need not account for differences in outcome which may be based on whether the language is learned primarily, partially, or not at all through explicit classroom instruction.

Despite the importance of this distinction in explaining certain types of observed differences in learning outcomes, in the remainder of this paper, the distinction between second and foreign language learning will not be maintained, and both types will be referred to under the umbrella term of 'second language learning,' as the differences between them are not considered significant for the purposes of this proposal. This is because the CALL applications which will be of interest can be used for both second and foreign language learners either inside or outside a classroom setting. In addition, since the proposed study will look at CALL at the level of the individual user's one-on-one interactions, it is considered unnecessary to make distinctions based on whether users are

learning primarily in remote or immersive environments since CALL technologies can be used under a variety of learning situations without changing the nature of these interactions.

The Distinction Between Learning and Acquisition

A second important distinction must be made between what are considered to be two types of language learning and knowledge; that of *learning* and that of *acquisition*. Learning can be described as "conscious knowledge obtained through grammar study," while acquisition refers to "incidentally developed ability to use an L2 fluently and naturally" (Ortega, 2014, pp. 136-137) which is "represented subconsciously in the brain" (Krashen, 1989, p. 440). Acquisition is characteristic of first language development, which is said to occur automatically for all normally developing children when exposed to language in their environment (Ellis, 2011). This process is thought to and take place rather rapidly in a period of about four years, by which time the "rudimentary aspects" of the child's primary language are thought to be acquired (Ingram, 1989, p. 2). However, while the first language acquisition process appears to occur automatically and with sufficient ease, later in life language acquisition can be more difficult, often requiring effort not needed in first language acquisition and which typically makes greater use of explicit, learned knowledge.

The distinction between learning and acquisition is significant because of the implications it has for the types of processes said to be involved in SL learning. While some claim that second language learning relies on fundamentally different processes from that of first language acquisition, others claim that they rely on what are essentially

the same processes, but perhaps with additional influencing factors. The idea that SL learning relies on different processes from first language acquisition has to do with the observation that the outcomes for SL learning contrast with the outcomes of first language acquisition in a number of ways. According to Bley-Vroman (2009), "Foreign language learning contrasts with native language development in two key respects: It is unreliable and it is nonconvergent" (p. 175), where *reliability* means, in terms of first language acquisition, that "children always succeed at language learning," and *convergence* means "children end up with systems that are so similar to those of others in the same speech community" (Pullum & Scholz, 2002 in Bley-Vroman, 2009, p.176).

This is in contrast to SL learners, whose language systems may be quite different from that of native monolingual speakers (Berko, 1958; Cook & Singleton, 2014; McLaughlin, 1990) and whose production often varies in many ways from not only adult native speakers of the language, but from other L2 speakers as well, and often contains inconsistencies that are not seen across native speakers (Kam & Newport, 2009). However, while this could be due to the fact that second language learning relies on different processes from that of first language acquisition, it could also have to do with a number of additional internal and external factors. This is because L2 learners follow a path of language development which is shaped based on unique encounters with the language along with influence from other languages they already know (Singleton, 1995). This is different from the development of a first (monolingual) language since children often receive similar types of exposure to the language in their native environment and develop without the influence of another already known language.

Theories of second language learning which propose similar learning mechanisms to that of first language acquisition account for the observed differences between first and second language speakers in a number of ways, such as possible later in life inefficiencies or weakening of working memory and inductive processing (Dekeyser, 2000), influences from an already fully developed first language (Ervin-Tripp, 1974), or reduced exposure to and practice with the TL (Muñoz, 2008). Ultimately, there are likely to be multiple factors involved in producing observed differences between first and second languages and being able to identify and understand them is important for helping learners to compensate for them in improving learning outcomes.

While not all researchers subscribe to the view that learning and acquisition are indeed separate processes (Zobl, 1995), it remains a significant distinction within SLA for many researchers and instructors. In the remainder of this proposal, the distinction between language learning and acquisition will not be maintained, and the more general term 'learning' will be used to cover both. Since the proposed study would cover a very brief period of learning, it is not thought that long-term acquisition processes can be accounted for under the period of study. For these reasons, it is believed that the use of the term 'second language learning' will be sufficient for this proposed study. The next section will review the various aspects and processes thought to be involved in second language development and how they each relate to the questions of interest for the proposed study.

The Role of Attention and Related Constructs in Second Language Learning

Some of the biggest questions about second language learning processes and the distinctions between them rest on understanding the role of attention in language learning. The construct of attention concerns why a learner may focus on a particular item in the input to the exclusion of others (Gu, 2017), and it plays an important role in "many aspects of SLA theory such as input, processing, development, variation, and instruction" (Al-Hejin, 2004, p. 1), as well as in "explain[ing] such diverse phenomena in second language acquisition (SLA) as variation in the accuracy, fluency and complexity of second language (L2) speech" (Robinson, Mackey, Gass, & Schmidt, 2012, p. 247).

Despite the common use of the word 'attention,' this term does not reflect a single concept, but rather a "variety of psychological phenomena" (Styles, 2006, p. 1), with "many mechanisms [that] have been proposed to explain how it affects different aspects of behavior and learning" (Robinson, Mackey, Gass, & Schmidt, 2012, p. 247). This concept, in all its complexity, has been variously described in the SLA literature under a number of terms including *attention*, *awareness*, *noticing*, and *detection* among others. The concept was most prominently first described by Schmidt (1990; 1995; 2001), and has been, and continues to be, qualified and elaborated on by SLA researchers and theorists.

According to Schmidt's *Noticing Hypothesis*, noticing is "the conscious registration of attended specific instances of language" (Schmidt, 2012, p. 32) which is "the necessary and sufficient condition for converting input to intake" (Schmidt, 1990, p. 129). Noticing, under this view, presupposes attention along with conscious awareness, and means the difference between features in the input being taken in for further

processing or being passed over. Specifically, it is the "hypothesis that input does not become intake for language learning unless it is noticed, that is, consciously registered" (Schmidt, 2010, p. 722). This assertion has been backed up by later researchers as well, commenting that "before something can serve as intake, it must be noticed" (Gass, 2017), and that "language learning can only take place if new linguistic structures are noticed in the input" (Godfroid & Schmidtke, 2013, p. 184). Noticing has also been described as a "mediating construct" composed of two "constituent processes," *attention* and *awareness* where it is thought that "attending to features is virtually the same as being aware of them" (Godfroid & Schmidtke, 2013, p. 184). Schmidt, by contrast, conflated the concepts of *noticing* and *attention*, commenting that noticing is "a low level of awareness… [which] is nearly isomorphic with attention and seems to be associated with all learning" (Schmidt, 1995, p. 1).

As with *noticing,* the concept of *awareness* has been defined both separately, as well as with respect to other related concepts. As its own construct, *awareness* has been defined as "the learner's knowledge or subjective experience that he/she is detecting a stimulus" (Gu, 2017, p. 27). With respect to *noticing,* it has been described as "conscious noticing," of features in the input, as well as in terms of knowledge representation such that if a learner is aware of a form, they have an "explicit representation of [it]" (Ellis, 2002, p. 29).

An additional mechanism, *detection,* has been included in many models to describe a "process that selects, or engages, a particular and specific bit of information," and which results in "the cognitive registration of sensory stimuli" (Tomlin & Villa,

1994, p. 192). Other authors have used this term differently, with Schmidt (1994) referring to *detection* as "the registration of the occurrence of a stimulus event in conscious awareness and subsequent storage in long term memory (p. 179). While some have considered *detection* as a key mechanism in language learning, others have noted that information which receives some low-level processing at early stages may be detected to a degree by the sensory system; however, this low-level detection is considered automatic and 'pre-attentive' in nature, and thus may not be sufficient to move information along to the level of conscious awareness (Ortega, 2014).

As can be seen from the review above, the various attention related constructs have been described throughout the SLA literature in ways that typically overlap, and much debate has risen up regarding the importance of each and the distinctions between them. The constructs are often treated together because they are "inherently connected, with one often entailing the other" (Gu, 2017, p. 26) and are all concerned in some way with describing different levels of, or variations in, conscious awareness during learning. When taken together, they can be thought of as different aspects or components of cognition which come into play during language learning or various levels on a continuum of conscious awareness which may be active from one point to the next. For the purposes of the present proposal, the terms *attention* and *noticing* will be used to describe specific mechanisms which play a role in guiding learners to select particular items in the input for further processing, and which may be necessary for those items to be retained for long-term learning. In the next section, some of the more specific

processes of language learning found in the SLA literature will be described with respect to the questions of interest in the proposed study.

Learning Process: Implicit and Explicit

Two other concepts of importance to second language learning, which deal with learners' conscious awareness of language and underly the distinction between learning and acquisition, are *explicit* and *implicit* learning. *Explicit* learning usually refers to that which can be gained from direct instruction, while *implicit* learning refers to "acquisition of linguistic patterns without explicit attention or instruction" (Brown, 2007, p. 66), such as is gained from passive exposure. Explicit learned knowledge is said to be consciously knowable and expressible but must also be recalled for use; implicit acquired knowledge on the other hand, is more similar to automatic behaviors which are carried out without conscious awareness and which often cannot be recalled or expressed (Berko, 1958). Implicit learning is thought to be subconscious and automatic and consists of abstractions about the underlying structure of the language derived from a number of instances of exposure. In contrast, explicit learning results from conscious selection of information in order to test working hypotheses about the language structure (Ellis, 2011).

Implicit learning processes are often associated with first language acquisition and native speakers are thought to "rely much less on explicit knowledge than on implicit knowledge" (Gass, & Selinker, 2001, p. 206) which can explain why speakers can intuitively know and use their first language without necessarily having an explicit, conscious understanding of the language. However, a common issue raised in SLA is whether or not second language learners can rely on implicit learning processes to the

same extent as children acquiring their first language (Birdsong, 1999; Johnson & Newport, 1989; Lenneberg, 1967). According to the *Fundamental Difference Hypothesis*, "adults can no longer rely on the innate mechanisms for implicit language acquisition and must, therefore, rely on… alternative, verbal-analytic problem-solving skills" (DeKeyser, 2000, p. 500). According to Ellis (2011), when it comes to adult SL acquisition, "what can be acquired implicitly from communicative contexts is typically quite limited in comparison to native speaker norms, and adult attainment of L2 accuracy usually requires additional resources of consciousness and explicit learning" (p. 35). In addition, it is though that "the processes involved in any language acquisition which takes place after the age of puberty will be qualitatively different from those involved in first language acquisition" (Snow & Hoefnagel-Höhle, 1978, p. 1114).

Despite these limitations on SL learning through implicit mechanisms, it has been argued that learners "can use explicit learning mechanisms to bypass the increasingly inefficient implicit mechanisms" (Dekeyser, 2000, p. 518), and while there is evidence for implicit processes at work in second language learning, there are also indications that conscious understanding is helpful for many learners with arguments that explicit instruction is necessary for SL learners since "[they] often fail to discover grammatical patterns on their own in the absence of formal instruction" (Brooks & Kempe, 2013, p. 281). The implication for SL instruction is that, while acquisition may be the goal, it may nonetheless be necessary to provide second language learners with explicit instruction if they are indeed unable to acquire language by the same processes as in a first language. However, classroom instruction is not always effective in achieving this aim due to

limitations in providing learners with necessary exposure to and practice with the target language (Asher, 1969; Laufer, 2003; Muñoz, 2008; Singleton, 1995). For this reason, language instructors must find ways to help learners move explicitly learned language knowledge into acquired knowledge. If learned language can be automatized through continued practice, learners could eventually achieve fluent, automatic skill at a level similar to that of an acquired first language. However, there is disagreement over whether and how this may be accomplished.

The essence of this debate, known as the *Interface Question*, "concern[s] the differences between implicit and explicit language knowledge and the ways in which they interact" (Ellis, 2011, p. 35). Specifically, it asks whether or not explicitly learned language rules can be transferred into acquired knowledge and applied automatically without conscious reference back to that knowledge. Those in SLA and related fields who view language learning as a type of skill acquisition consider the process of *automatization* to be at work in moving explicit knowledge to automatic, fluent performance of behavior. *Automatization* is the process which leads to *automaticity*, which refers to "perform[ing] a complex series of tasks very quickly and efficiently, without having to think about the various components and subcomponents of action involved" (DeKeyser, 2001, p. 125). According to Gass, Behney and Plonsky (2013), it is "because of the limited capacity that humans have available for processing, the more that can be handled routinely-that is, automatically-the more attentional resources are available for new information (p. 256), so as learners 'routinize' language skills, this

leads to “greater ease with which they can be put to use” (Gass, Behney & Plonsky, 2013, p. 255).

While many researchers argue that explicit, learned knowledge can become automatized in a way similar to implicitly acquired knowledge, others claim that there exists a “fundamental difference” between the two types of knowledge, and there is no cross-over between them that would allow for transfer. This view, referred to as the *noninterface position*, claims that implicit acquisition does not require explicit knowledge, and knowledge gained from explicit instruction “does not affect the acquisition of implicit knowledge” (Laufer & Hulstijn, 2001, p. 3). In other words, “[implicit] knowledge of an L2 cannot come into existence as the direct result of [explicit] knowledge, nor can it indirectly be influenced by [explicit] knowledge of that language” (Hulstijn & De Graaff, 1994, p. 98). If it is the case that such transfer is possible, SL instructors can make use explicit presentation and practice exercises, which may be aided by CALL technologies, to help learners automatize learned language. In addition, the use of CALL technologies which can provide learners access to language content that they may not otherwise encounter in their physical environment, overcoming some of the obstacles of remote, classroom-based instruction.

Learning Processes: Intentional and Incidental Learning

A final related distinction which needs to be made is that between two other types of learning processes: *intentional* and *incidental*. *Intentional* learning refers to “the deliberateness on the part of learners to attend to the stimulus” (Gu, 2017, p. 27), while *incidental* learning refers to “learning without intention, while doing something else

(Ortega, 2014, p. 94), such as learning words by reading in context and inferring meaning without looking it up or acquiring grammatical structures while focusing on comprehension rather than form. Whether learning with intention is better overall for learning, or merely facilitative is debated; however, "the superiority of intentional over incidental learning," if it is indeed superior, "depends upon [many] factors" (Eagle & Leiter, 1964, p. 58). For example, it has been shown across studies that "while intentional learning [is] more effective for recall, incidental learning [is] more effective for recognition" (Eagle & Leiter,1964, p. 62).

At the same time, lack of intention does not necessarily mean attention cannot be drawn to certain items in the input, as novel items are thought to automatically draw some level of attention (Laufer, 2003). Some have considered incidental learning to be akin to "guessing" or "contextual guessing" (Kelly, 1990, p. 200), by which means learners are able to figure out the meaning of unknown items in the input. If it is the case that novel items can be learned by guessing from context without intention to learn those items, then benefits for this strategy under certain conditions can be seen as well.

Finally, not all researchers agree in the dichotomy between intentional and incidental learning, or that one is preferable to the other as a means of language learning, but rather view the two as "complementary activities, each one enhancing the learning that comes from the other" (Nation, 2001, p. 238), and so it cannot necessarily be said that one strategy is preferable to the other overall, as it may depend on task demands and immediate goals.

The distinction between these two concepts is important because of its implications for language instruction in terms of approach in either classroom or CALL based settings. The distinction also has implications for SLA and CALL research since the respective processes can be used to account for different types of observed outcomes. For example, in the computer mediated vocabulary exercise which constitutes the proposed study, learners are engaged in an intentional learning task since they are completing the task with a clear objective to learn the target vocabulary items. However, they are also being exposed to items in the input (e.g. distractors) which are not explicit targets for learning, and to which they may not give direct, intentional attention. Therefore, learning outcomes indicated on post-task assessments need to be evaluated for whether they occurred through intentional learning of the target items or incidental learning of distractors.

Measuring participants' attention to various items in the display as they work through the task can help to make an inference one way or the other since it can be observed whether participants attended to the items that they are shown to have made learning gains on, and whether or not those items were explicit targets. If participants are shown to have attended to distractor items without being explicitly instructed, this could be an indication of incidental learning of items in the surrounding input. Thus, the proposed study will interpret findings on language learning considering implicit, explicit, intentional, and incidental learning as well as the possible connection, or lack thereof, between them. In the following section, theory and research regarding specifically how vocabulary is learned in another language will be reviewed and discussed.

How Vocabulary is Learned in a Second or Foreign Language

While the learning processes overviewed in the previous section tend to describe language learning as it relates to grammar, vocabulary learning is said to rely on these same foundational processes and is thought to be "acquired in fundamentally the same way the rest of language is acquired" (Krashen, 1989, p. 440). Some have argued that vocabulary acquisition occurs primarily through incidental processes which take place mainly through extensive reading in the target language (Krashen, 1981; Schouten-van Parreren, 1989; Sternberg, 1987), and which thus minimize the role of vocabulary focused instruction. Others have claimed however, that while incidental vocabulary learning may play a major role in first language acquisition for speakers who receive literacy instruction throughout school, the process may not be as straightforward for second language learning (Laufer, 2003). Incidental vocabulary learning has been identified under a number of conditions including reading a text in the target language which contains a number of unknown target items (Godfroid, Boers & Housen, 2013), and while engaging in reading and listening comprehension tasks (Yoshii & Flaitz, 2002). However, while such learning may be possible under certain conditions, it is not necessarily the case that incidental learning is the most common means by which L2 learners acquire vocabulary.

While incidental learning seems to responsible for a good deal of second and foreign language vocabulary acquisition, questions regarding the mechanisms, existing vocabulary requirements, amount of exposure needed, and usefulness of input modification still remain (Huckin & Coady, 1999). In addition, the process of incidental

vocabulary learning through reading takes place at a rather slow rate (Hulstijn, 1992), and it is thought that numerous exposures are required for L2 learners to acquire a word (Goulden, Nation & Read, 1990). It has also been noted that incidental learning relies in part on learners' noticing of unknown vocabulary in the sense that they are aware they do not know a word, which second language learners may not always do with accuracy, perhaps believing they know the meaning of a word which they do not, or guessing incorrectly at the meaning of a word, which would have a negative effect on retention (Laufer, 2003).

While supporters of uninstructed vocabulary learning through reading argue that it leads learners to guess the meaning of unknown words from context, which would arguably aid in retention (Schouten-van Parreren, 1989), this has not in fact been shown to be the case in many empirical studies (Hulstijn, 1992; Mondria, & Boer, 1991), and use of this strategy seems to require a large existing vocabulary base on which to draw for making guesses- a base which first language learners, but not early second language learners, would have. In fact, it is thought that around "95% coverage of text is needed to gain adequate comprehension and to guess unknown words from context" (Hirsh & Nation, 1992, p. 690). This means readers will know 95% of the individual words (tokens) in a text which will result in around one unknown word occurring every two lines and translates to a vocabulary of around 2,600 words for English novel reading (Hirsh & Nation, 1992), and nearly 15,000 words for a more diverse range of texts (Nation & Waring, 1997).

Therefore, under this view, reading alone would not be sufficient for L2 vocabulary acquisition because in order for second language learners to be able to acquire vocabulary from passive reading, they would first need to have a substantially high vocabulary of the most common words in a language. Intentional learning aided by explicit, vocabulary focused instruction can be useful in getting learners to this point, and these goals can be aided in part by CALL technologies which allow for increased exposure and repetition of practice with forms. The proposed study does not take one view over the other but considers that learners likely rely on both implicit and explicit processes in language learning, and so there is likely to be some benefit to explicit vocabulary instruction, even if it is not, in and of itself sufficient. The CALL vocabulary exercises which make up the proposed study, being tutorial in nature, rest on the assumption that such instruction can be useful for learners but considers also that inductive/implicit processes may also be at play for learners even when engaged in an explicit/intentional task as used in the study.

Scaffolding for Vocabulary Learning

Given the high demands of vocabulary learning and the challenges this can present for SL learners, it has been argued across much of the SLA literature that it is helpful, if not necessary, for early language learners to have access to scaffolded materials for learning vocabulary. Scaffolding can take many different forms depending on the skill, materials, medium of delivery, and many other factors, and may include "modeling, using simplified language, visuals, demonstrations, graphic organizers, and cooperative learning activities" (Ovando & Combs, 2018, p. 331). The proposed study

will focus on the use of images and L1 translations as a form of scaffolding for SL vocabulary learning, particularly with respect to the multimodal integration of both in computer mediated modes of delivery and will focus on a number of potential issues raised by this integration.

Images as Scaffolding for L2 Vocabulary Learning

In the absence of physical objects or real-world environments, which is often the case in SL, and particularly FL learning, pictures of objects, people, and places are typically used as stand-ins to prime target concepts. In such cases, learners are presented with pictures and the names of represented objects in the target language (TL), with or without L1 translations. The inclusion of visual representations alongside target language items is thought to help learners "associate new words to a meaningful context to which they apply" (Brown & Lee, 2015, p. 481) and to provide cues or 'assistance to memory' of the content primarily carried by the written language (Wysocki, 2001, p.137). This could be possible because it is believed that visuals "may be more useful for transmitting large amounts of certain kinds of information… [and] more effective means of processing" (Kress, 1998, p. 55) and are said to offer "more robust nonverbal information than that presented in the static pictures… and allow[s]… their nonverbal processing system to support their verbal processing" (Silverman & Hines, 2009, p. 306).

In addition, it is thought that multilingual speakers, despite differences in lexis and grammar, have once common conceptual system (Hulstijn, 2001) or "two separate verbal systems (L1 and L2) and a common imagery system" (Al-Seghayer, 2001, p. 205), which suggests that presenting target language items in both visual and verbal formats

should “not only link the two verbal systems, but that this storage in the second verbal system would also have an additive effect on learning” (Al-Seghayer, 2001, p. 205). According to the *dual coding theory* of language processing, when words are associated with “appropriate nonverbal referents (pictures, objects, events, and emotions),” connections between the “verbal and visual systems” will be strengthened, which should lead to “better recall and appropriate use” of words than if learned in only a verbal or visual mode alone (Yoshii, 2006, p. 87). Thus, the use of visual aids can be helpful in vocabulary learning by priming the concepts associated with the target words and helping learners to associate the meaning of the L2 word with its form.

L1 Translations as Scaffolding for L2 Vocabulary Learning

The use of L1 translations as a form of scaffolding for language learners has been given support from the SLA literature in a number of ways. It has been commented that “effective ways to encourage noticing include giving a definition or an L1 translation” (Allum, 2004, p. 489) and that “for many learners learning is faster if the meaning of the word is conveyed by a first language translation” (Nation, 2013, p. 105). The inclusion of L1 translations for vocabulary learning may be useful especially for “low level learners [who] may not be able to understand L2 definitions or synonyms” (Nakata, 2011, p. 19). For this reason, many tutorial CALL programs include translations in early level vocabulary exercises to assist learners who do not have the vocabulary knowledge in the L2 needed for matching based on L2 synonyms or definitions.

The inclusion of L1 forms along with L2 targets may have either facilitative or interfering effects, depending on the specific relationship between the word pair. Formal

similarity between words is a useful source of information for language learners to make guesses regarding the meaning of an unknown word, either because it is similar to a related known word in the target language, or because it is similar to a word in another known language (Kelly, 1990). This strategy of guessing based on formal similarity is not only often taught as an explicit strategy to learners, but it is also thought to be commonly used by learners on their own as "foreign language learners… rely heavily on it when reading languages related to known ones" (Kelly, 1990, p. 200). It thus seems fair to expect that, during a language task, learners will attend to forms which appear similar to known words, whether correctly or not, since formal similarity is a potential clue to word relatedness.

Formal similarities can cause interference effects with learning when either similarities are misleading as to the relationship between words (e.g. false cognates), or when additional sources of information in the input provide contrary evidence to the meaning. For example, a conflict may occur when a user is asked to select an L2 item which does not have formal similarity to its L1 counterpart, while distractor L2 items do have formal similarity with their L1 counterparts. In this case, knowledge of the meaning of distractors given by formal similarities with known L1 words may interfere with processing and memory for unknown L2 targets. In the proposed study, a number of orthographically similar word pairs between the target Spanish language and the English source language will be included in order to look for possible differences in attention between these pairs and those which do not have formal similarity, as well as the types of facilitative and interference effects between word pairs which may occur.

Multimodality in Language Learning Materials

The use of both written text and images in language learning materials raises a number of considerations for potential learning outcomes, given the complex ways verbal and visual elements combine to convey meaning in a multimodal text. While the theory supporting the use of visuals for L2 learning is a good starting point for providing scaffolding needed to acquire early vocabulary, there is much more to consider in trying to utilize the most effective materials and combinations of text and image. While the use of images in tutorial CALL programs is useful for activating users' existing knowledge for the content in a particular activity or lesson, the way in which these images are paired with one another and with the written words which are the target items to be learned can make a difference as to whether users are relying on images, forms, or both in processing content and making decisions regarding responses.

Many current language materials use text and image combinations which could be ineffective or even counterproductive in leading the learner to the intended meaning. However, it is not always clear whether the arrangement in a particular display draws users to attend to certain elements over others. For this reason, it is important that further research is carried out on multimodal SL materials in order to answer these questions and to improve the use of text and images in SL learning materials to be more effective for vocabulary acquisition and overall language learning.

One hypothesis of the proposed study is that users will attend first and more strongly to images over written text. It will be argued that while highly representative images may be useful for priming learners' content knowledge and may aid in recall of

content, these types of images, because of their high representativeness, may distract learners from attending to the written forms of the target items which may negatively affect their recall of the meaning of the forms. To do this, the study will need to measure participants attention to visual and written elements on screen during a vocabulary learning task. In the following section, common methods for studying attention and noticing in computer-mediated learning tasks will be discussed with respect to the proposed study.

Measuring Attention and Noticing During Language Learning Tasks

A number of different methods exist for measuring the various constructs associated with attention and noticing during language learning tasks, and these are typically distinguished based on whether they measure *online* (concurrent) or *offline* (non-concurrent) processing. The earliest methods used were two offline measures- *recall*, "in which participants retrieve from memory previously heard or seen words, sentences, or texts," and *recognition*, "in which participants indicate whether or not they had previously heard or seen words, sentences, or texts" (Segalowitz & Trofimovich, 2012, p. 185). Examples of task typically used for measuring offline recall and recognition are post-task stimulated recall, interviews, or questionnaires. Later on, online measures were adopted which allowed researches to get an idea of what users were noticing during a task. The most common of this type of task was the think-aloud in which participants report their thoughts as they work through a task (Leow, 2001). For some, online measures are preferable because it is thought that offline measures such as recall and recognition are "slow and consciously controlled language processing

influenced by participants' test-taking strategies" while online measures look at automatic processing which "[is] not subject to conscious control and… proceed[s] without much attention or awareness" and therefore "reflect[s] how languages are organized and used" (Segalowitz & Trofimovich, 2012, p. 186).

What all three of these measures have in common is that they are types of verbal reports and are thus *introspective* methods which seek to identify the cognitive processes employed by learners during a task through their own self-reporting (Bowles, 2010).While there has been debate regarding the effectiveness of introspective methods in revealing information about cognitive processes (Selinker, 1972), it is considered by many to be a useful, albeit imperfect, method of eliciting data about learner processing. Others however consider only directly observable behavior data to be useful for making inferences regarding language acquisition (Selinker, 1972), and that introspective reports are too problematic to be useful.

Despite the imperfect nature of introspective data elicitation techniques, they may be necessary in order to gain insight into the full picture of second language learning since productive language is only one aspect of knowing a language, and comprehension cannot be directly observed. If only observed behavior is used, researchers must make inferences about learners' processes which may not be accurate to what learners' actually do and may miss key aspects of the learning process (Bowles, 2010). To overcome the limitations of introspective methods alone, the method of eye-tracking can be used to record and measure participants' behavior in terms of eye movements and fixations to

items of interest. Used in conjunction, these two types of methods can be useful in building a bigger picture of language learning as it occurs during task execution.

Eye-tracking as a Method of Measuring Attention

One useful method of measuring attention in computer-based activities is eye tracking, which measures gaze and fixation patterns as the user interacts with content on screen. Typically, the eyes move over a scene in visual space in short, rapid movements known as *saccades*, which occur about 3-4 times per second (Bojko, 2013). These eye movements can be measured in terms of *gaze points*, which are "the instantaneous spatial locations of the visual axis landing on the stimulus" ("Types of eye movement," para. 2). From time to time, gaze will rest on a location within the visual field for a period of time, which is referred to as a *fixation* (Bojko, 2013). Fixations are "period[s] of time when the focus of the participant's gaze is relatively still on an area" ("Eye tracking metrics," para. 2).

Fixations are a common measure of interest in eye tracking research and are considered to be revealing of attentional processing since, because "vision is suppressed during a saccade" (Rayner, 2009, p. 1458), visual information can only be processed "when the eyes are relatively motionless and are focused on something" (Bojko, 2013, p. 12). Since fixation is needed to process an object and interpret its features, eye fixations can reveal something about what features individuals are trying to process in a visual display. They can also be indicative of which items are of more or less interest to an individual since "the more complicated, confusing or interesting those features are the

longer we need to process them and, consequently, more time is spent fixating on them" ("What do we study," para. 1).

While fixations are the primary measurement unit of interest to eye tracking studies, fixations themselves are not directly measured by eye-tracking but are composed of clusters of gaze points which are close in time and space ("Types of eye movement"). Gaze points are a feature of the eye-tracking hardware, and the frequency at which a gaze is recorded depends on the recording frequency of the device. The duration of fixations can range "from 60 ms during reading to several hundred milliseconds when examining a photograph or image" ("Eye tracking metrics," para. 2).

Eye tracking as a technology is useful in identifying *what* elements in a display an individual is looking at, but it does not, in and of itself, identify *why* they look at these elements. There are a number of reasons as to why eye movements may be directed to a particular element on a screen, including bottom-up stimulus driven movements based on features of the element itself such as color, shape, or size contrasts (Bojko, 2013), as well as top-down cognitive driven selection based on features of the task, participant goals, semantic interpretation, and existing knowledge (Nyström & Holmqvist, 2008; Smith & Colunga, 2012). Fixations on particular elements in a display indicates both what is interesting and what is salient to a participant, and increased fixation to a particular area "could be associated with motivation and top-down attention as respondents refrain from looking at other stimuli in the visual periphery that could be equally interesting" (Farnsworth, 2018, para. 18). Revisits to an area can indicate either that something was particularly interesting or confusing to an individual (Farnsworth, 2018). Now that the

important concepts associated with SL learning and how to measure the variables of interest for this study have been addressed, the more specific topic of CALL and its importance to the proposed study will be discussed.

Computer Assisted Language Learning: History and Classifications

"[In] the 21st century, everyday language use is so tied to technology that learning language through technology has become a fact of life" (Chapelle, 2001, p. 1).

In the modern language learning environment, computer assisted language learning (CALL) technologies are often incorporated into language classrooms and curricula to facilitate acquisition through increased exposure to and practice with the target language. Computer assisted language learning may be defined as "any process in which a learner uses a computer and, as a result, improves his or her language" (Beatty, 2013, p. 7). CALL arose as the various affordances of computer technologies for language learning and the means of capitalizing on them began to be recognized, leading to the development of a range of programs and tools. Its early development saw the rise of a number of different types of applications all meant to assist in some way with language learning, but which did so through a range of structures, features, designs, and modes of delivery (Beatty, 2013), and which rested on different technologies, theories, and pedagogies (Stockwell, 2012).

One of the first to develop a classification for types of CALL programs was Levy (1997) who made a distinction at that time between tutor and tool CALL applications. Defined under this framework, tutor software "evaluates the learner, controls the learning

process and temporarily substitutes for the human teacher," while tool software "increases the efficiency or effectiveness of actions related to learning" (Hubbard & Siskin, 2004, p. 449). The key distinction here is that while tool-oriented programs serve to assist or support the learner in some aspect of language learning, they do not control the delivery of information or evaluate learning in any way. Rather, the learner guides themselves through the information provided by such tools and determines in what ways that information can be made of use to them. In contrast, tutor-oriented programs do control the delivery of information to learners, in more or less structured ways, and offer feedback and evaluation of user input.

Another common classification for CALL programs which has arisen more recently with the increasing development of social uses for computer technologies is that of tutorial and social CALL. This distinction has been motivated by the development of programs featuring communicative components which aim to facilitate language learning through interaction with other speakers. Within this classification, tutorial CALL is generally used to describe programs which are characterized by "one-to-one interactions where the computer evaluates the student input and then presents the new material accordingly" (Levy, 1997, p. 178), while social CALL describes "using a computer or mobile device to connect with other people through email, blogs, texts, instant-messaging, and social networks" (Blake, 2013). However, rather than being of one clearly defined type, social CALL is wide ranging and can be better described in terms of the *affordances* an app or program has for social interaction (Smith, 2016).

While the proposed study will not look at social elements of CALL programs or tool-based applications, it is worth noting these features as part of the larger landscape which currently makes up CALL software. The main focus of this study will be tutorial CALL programs which include structured delivery of content along with feedback and tracking of learner progress and which contain the features which will allow the research questions of interest to be explored.

Tutorial CALL: Its Characterization and Features

Tutorial CALL programs are characterized by the presence of a wide range of exercises and tasks to cover language skills such as reading, writing, listening and speaking, as well as the fundamental vocabulary and grammar knowledge underlying these skills. While the different skills focused on by each program can vary widely, altogether, tutorial CALL covers all of these areas, and most programs contain components for many of these skills together (see App Review for examples). Besides the skills targeted by these different programs, they also vary in a number of ways in terms of the visual interface, layout, navigation, and other usability features. Most importantly, they also vary in the way their exercises and activities are designed, the structure of levels and tracking of progress, as well as how possible game-based elements may be worked into the overall program. This variety of design and structure across tutorial CALL needs to be considered in any question which concerns the effects of CALL on language learning.

Since most tutorial CALL programs today feature multimodal elements including both written text and images, differences in the design of exercises and the ways text and

image are incorporated in providing information to users are especially worthy of exploration to gain insight into the ways different design features can impact learning outcomes for learners. The design of CALL courses and exercises is a recurring topic of discussion in field related publications and discourse (Levy & Stockwell, 2013), and the question of design at the level of the exercise is what is of interest to the present study. While most commercial software claims to be based on SLA research, very few appear to actually apply SLA principles, and oftentimes, the rationale for the specific design and structure of vocabulary exercise is unclear for a given program.

At present, there exists no currently known research which comprehensively explores the ways in which these variations may affect in-task response and later recognition and recall. Investigations of this type are important for SLA and CALL researchers to understand the many types of structures and combinations which exist and to explore whether certain types may be more or less beneficial for achieving learning outcomes. It is this design at the level of the individual exercise which will be the focus of the proposed study, as it aims to discover if variations in the type and arrangement of visual and written elements on the screen can have an effect on attention to and processing of these elements, and whether variation in attention and processing could lead to differences in recall and recognition of target items post-task. Specifically, this study will look at multiple choice vocabulary exercises in tutorial CALL which use both written text and image to guide users to making a match between L1 and L2 target vocabulary, and which most clearly represent the type of multimodal interactional phenomena with which this study is interested.

Vocabulary Exercises in Tutorial CALL

While there is no set standard for the structure of vocabulary exercises across tutorial CALL programs, a typical design consists of a two-part presentation-practice format which can be seen as a simplified version of the classroom-based presentation-practice-production (PPP) model. In the standard model, students learn target language items by moving through a series of stages in which instructor control is reduced and student autonomy is increased from controlled presentation and practice, to free production (Brown & Lee, 2015; Richards & Renandya, 2002; Richards & Rodgers, 2014). In the CALL version of this model, instruction is given in two stages; first, through a presentation mode in which "learners familiarise themselves with the target words," followed by a mode in which "they practise retrieval of previously met words… [having to] recall or recognise the L2 word form or its meaning" (Nakata, 2011, p. 21). Presentation modes in CALL programs often include the use of flashcards in which target items are shown without requiring any input response from the user, and which may incorporate either L1 translations, visual depictions, or some combination of both to signal the meaning of the L2 item (see App Review for examples). Practice modes typically consist of sets of exercises which may include any number of matching, T/F, multiple choice, or fill-in-the blank activities and which do require some type of response on the part of the user to indicate their recognition or recall of the target items. The number and types of exercises included in practice modes, as well as the way they are incorporated throughout the courseware in terms of activity sets and lessons varies for

each individual program, but all have the underlying goal to promote practice in recall and recognition of the meaning and/or use of the target L2 items.

Another type of presentation-practice format exists which, rather than separating the two modes, mixes them together in a highly scaffolded design which requires users to respond in some way to the newly presented items as they are first being introduced. In this design type, users are typically asked to determine the meaning of a novel target language item by selecting its corresponding L1 translation, indicated to the user by an accompanying image depicting the underlying concept shared by the two words (see App Review for examples). In later stages, users may again be asked to select the meaning of the target L2 word with either images or L1 translations removed, which could be seen as an additional level of now unassisted practice. In both versions, it is thought that by associating the target items with known L1 equivalents and concepts, learners will be able to make connections with existing knowledge that will allow for deeper learning, and therefore, better retention. The theory and research underlying these assumptions is further described below.

The Paired Associate Paradigm in Vocabulary CALL Tasks

Vocabulary exercises in many tutorial CALL programs operate on a learning principle known as paired associate learning (PAL) which has been used in psychological research since the1950s (Bower, 1961; Glenberg, 1976; Kleinsmith, & Kaplan, 1963; Paivio, 1965; Russell & Storms, 1955). In this paradigm, "subjects may be asked to learn several pairs of words or figures and then recall the pair to each stimulus presented as a cue" (Honda, Barrett, Yoshimura, Ikeda, Nagamine & Shibasaki, 1996, p.407). Language

learning studies which make use of this paradigm have looked both at how participants learn to associate known words in an L1, as well as novel L2 words with known L1 words (Steinel, Hulstijn & Steinel, 2007). The assumption underlying PAL is that it "involves establishing connections between a given stimulus and a response in memory" (Georgiou, Liu, & Xu, 2017, p. 82). If L2 vocabulary learning "is viewed as learning the L2 equivalents for L1 words, then it is essentially a paired-associate task in which the L1 word serves as a stimulus and the L2 word serves as a response" (Schneider, Healy, & Bourne, 2002, p. 419). This view of L2 vocabulary learning can be seen in CALL exercises in which "learners are asked to associate the L2 word form with its meaning, usually in the form of a first language (L1) translation, L2 synonym, or L2 definition" (Nakata, 2011, p. 17). In these types of exercises, the idea is that users will learn the meaning of the novel L2 words by association with L1 words whose meanings are known. While the assumptions underlying the principle of paired associate learning will not be examined here, what will be explored is how CALL vocabulary exercises which rest on this principle in their design may be more or less effective in achieving their desired learning outcomes based on how target associates are displayed on screen.

The type of presentation of interest for the present study is that which asks the learner to match a known L1 word to a target L2 translation and which include visual aids as support in the form of images representing the target concept to accompany L2 written vocabulary options. Despite the underlying similarity for these types of tasks, there is a great degree of variation amongst the details of their designs. One way this variation occurs is in the types of images chosen to accompany written vocabulary items;

specifically, some images in a given display may be more or less representative of the concepts associated with their L2 paired words (see App Review for examples). This variation depends in part on that nature of the underlying concept and how easy it is to depict the meaning of the word visually, but also on the language level of the particular exercise. Most early vocabulary that learners are introduced to in an L2 are common, concrete items that are easier to represent, so most basic level vocabulary exercises feature images which clearly depict the intended meaning. The more abstract and idiomatic vocabulary found at higher levels can be more difficult to depict visually, and so these images may be less reliable in conveying the target meaning.

One of the variations this study will look at concerns how representative the images used in a particular display are of their respective concepts. The reason this is important to address is because the purpose of including images is to serve as a cue to the meaning of an unknown word, and so it needs to be considered how likely a user is to understand the meaning of the target word based on the image of the concept associated with the L2 item. Another variation this study will consider is how similar L2 options are in terms of their meaning and their accompanying images. This has implications for how clear it may be to users which image is a match for the L1 cue word since related concepts or images which could represent more than one concept in a display could create interference for users' interpretations (see Methods for further description). A final variation which also has to do with the distractors chosen in a display has to do with formal similarities between L1 and L2 written forms which can either facilitate or interfere with making a match to the desired target. For example, if the L1 cue and L2

target are cognates, and users recognize this, an easy match can be made. At the same time, if words are false cognates, if they are superficially similar without actually being related, or if distractors are similar in form to the L1 cue, then more time must be spent in processing information coming from written text and images to make a match (see Methods for further description).

The translation matching exercise which is of interest in this study offers support in the form of both images and written text in the L1. In this task, learners must select the meaning of the target L2 word by matching it with the corresponding word in the L1 with the equivalent meaning. The target meaning is doubly indicated by the provision of both the image and the written text, so learners are able to make a match drawing on information from either source. One challenge presented by this type of exercises is that it aims, ideally, to focus learners' attention on both meaning and form. By including both images representative of the words' meanings, as well as L1 translations, it is hoped that learners will understand the meaning of the L2 word. However, learners must also associate that meaning with the word form in order to later recognize and recall the word for its appropriate meaning. The problem here is that it may be too much to ask of learners to focus on both meaning and form in a given task due to "limited cognitive resources" (Nakata, 2011, pp. 21). Thus, it is important to evaluate whether these types of exercises are effective in achieving the learning goal of users being able to retain the target vocabulary beyond the immediate task. In order for this to be the case, users must be able to recognize and recall the L2 vocabulary in their written/verbal form without the assistance of images. Whether or not this goal is achieved from these exercises is the

question this study will aim to address. Across all variations of this design, low level design features, including the types of images chosen and the way they are paired with distractor words and images across trials, could affect what elements users attend to in-task and may have implications for which items, out of all target items in a set, are retained post-task. It is these features and their possible influence within the specific format outlined above which will be explored in the proposed study. In the following section, examples from various tutorial CALL programs will be provided to illustrate the design features described above.

App Review

In this section, I will review a number of examples from popular language learning programs to illustrate the different kinds of vocabulary matching tasks which can be found, the different types of images used to accompany target vocabulary, and the different combinations of text and images which occur. Using the theory described in the previous section, I will discuss the varying degrees of representativeness of images, distinctiveness from distractor images, and the similarity between L1 and L2 forms for targets and distractors across some of these examples. I will then establish the problem that appears to exist with the design of these tasks and the effect it may have on learning outcomes. Finally, I will frame the research questions with respect to the problem and discuss the importance of exploring these questions in order to address the problem and improve design.

Split Presentation-Practice and Practice only Exercises

As described in the previous section, there are a number of ways vocabulary exercises may be designed and structured across CALL programs. One commonly seen format makes use of an initial presentation of target vocabulary through a flashcard-like display, followed by a series of exercises to practice this vocabulary. Variations of this format may mix presentation and practice modes, requiring users to give a response to indicate the meaning of L2 items upon initial presentation. A number of examples are provided below, sampled from different web-based software and covering multiple languages, to illustrate the many ways this exercise can appear across tutorial CALL programs and to describe the features which are of interest for the proposed study.

Fig. 1: Busuu, Spanish, Beginner A1

In the first set of examples, taken from the web version of Busuu, users are first shown target vocabulary on a flashcard which provides both the written and aural form of the L2 word, accompanied by an English translation and an image depicting the associated concept (fig. 1). In this particular program, lessons are broken down into a number of practice sub-sets which each feature 3-4 vocabulary items. Practice exercises

are interspersed throughout the set between presentation of novel items, so a straightforward presentation-practice format is not followed; however, presentation and practice modes are separated in the sense that presentation flashcards do not require a response from the user, while later exercises ask users to recognize or recall vocabulary items through various multiple choice, matching, and fill in the blank exercises (figs. 2, 3, 4).

Fig. 2: Busuu, Spanish, Beginner A1

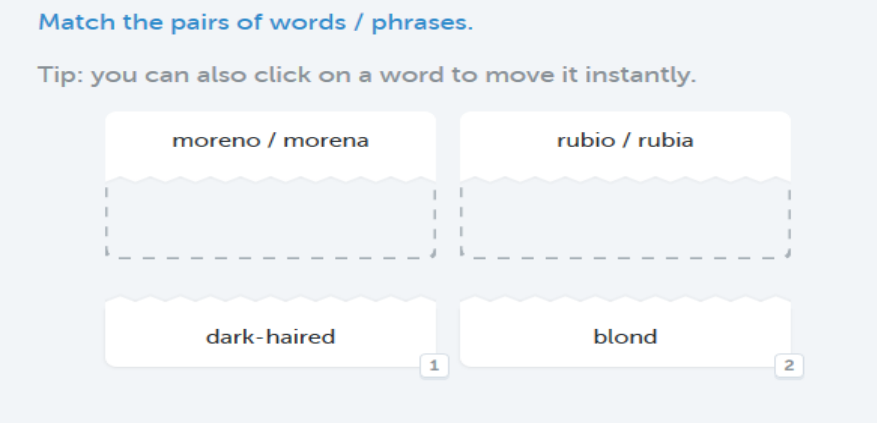


Fig. 3: Busuu, Spanish, Beginner A1

Fig. 4: Busuu, Spanish, Beginner A1

In the next set of examples, taken from the web version of Babbel, users are again presented with the target vocabulary, along with both L1 translations and images, and are simply asked to learn them without making any sort of response (fig. 5). One difference with this example is that in the later multiple-choice recognition task (fig. 6), users must select the L1 translations for the L2 targets, rather than the other way around. So here, even though there is again the use of both translations and images, the images are aligned with the L1 forms, and so the user cannot guess the meaning of the L2 word from the image; rather, they must be able to recognize either the L1 word or the image that the L2 target was associated with in the earlier presentation stage to make a correct match.

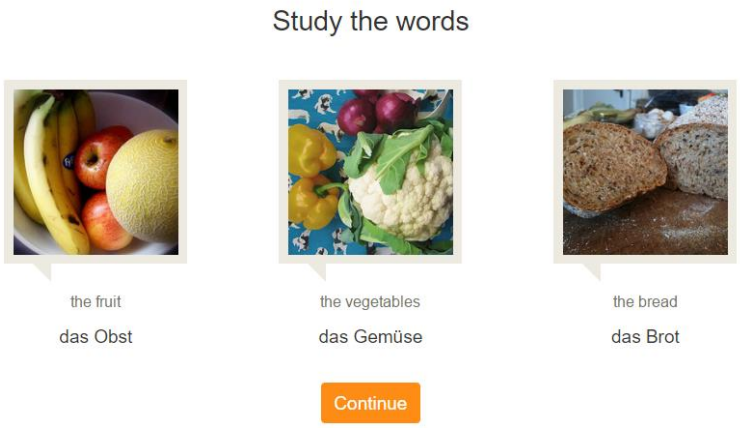


Fig. 5: Babbel, German, Level One

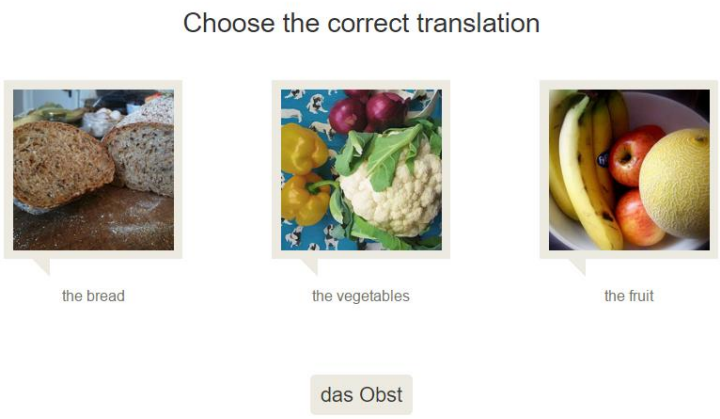


Fig. 6: Babbel, German, Level One

Mixed Presentation-Practice Exercises

As mentioned earlier, the type of vocabulary exercise which will be the focus of this study is one in which presentation and practice modes are combined, users are required to give a response to target vocabulary as they are introduced by indicating the L2 match for an L1 cue word, and which include both L1 translations and visuals to help

guide the user to the correct match. This type of exercise and its features of interest are described through the examples below.

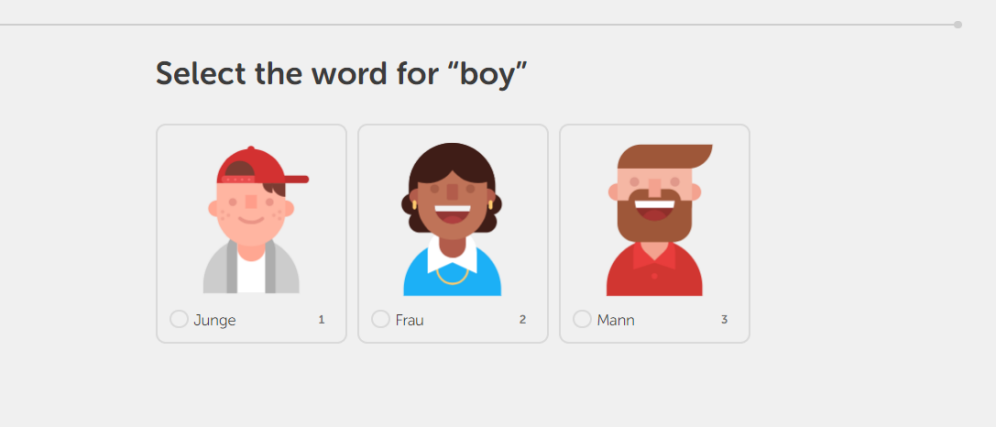


Fig. 7: Duolingo, German, Basics 1, Level Zero

In the example above (fig. 7), taken from the web-based version of Duolingo, the user is being asked to select the German translation for the English word 'boy.' All three L2 options are presented in written form and are accompanied by a graphic depicting the concept associated with that word. This exercise is the first the user will be presented with upon beginning the course; therefore, the user has not already been presented with this vocabulary or had a chance to practice them in the program prior to this point. For this reason, it is considered to represent an example of a mixed presentation-practice design since the user is practicing making a correct match upon first being introduced to the words. The use of images seen in this display is therefore necessary for users to make a choice which is based on an understanding of the meaning of the word, without which, users would be guessing randomly as to the correct translation, assuming they had no prior knowledge of any of the L2 words in the set.

With respect to the question of representativeness, all three images seen in this display would be considered representative of their associated concepts, in that users would likely be able to name what was depicted in the image as being something close to their written labels (see Methods for further description). In addition, it is considered unlikely that the user would be confused by the distractor items or consider them to be possible responses due to the distinctiveness of the three concepts and their images from one another. The images are considered distinct in that each image would be said to depict only its associated word, and not that of the other options (e.g. the pictures accompanying 'Frau' and 'Mann' do not depict the concept of *boy*). So, given the English cue word 'boy,' it seems likely that users would see the picture accompanying the German word 'Junge' as most clearly depicting the concept of *boy*, and so it would be thought that the majority of users would select this option as the matching translation for 'boy.'

Finally, it is important to point out the similarity in written form between the L2 option 'Mann' and its L1 equivalent 'man.' While the similarity alone does not tell whether the words are related (i.e. have the same meaning), as they could be false cognates, in conjunction with the image connected to it, it is safe to assume the two words are related, and it is thought that most users would see this connection. This is important because it would allow the user to eliminate 'Mann' as an option for the translation of 'boy,' making the selection easier. While in this particular display, the formal similarity is less relevant since the representativeness and distinctiveness of the images is sufficient for a match, in other examples in which visual information is less

helpful, formal similarity can become more important for either selecting or ruling out certain options.

The question which rises from the example in figure one and others like it is that since the match for the correct L2 translation can be easily made by reference to the images alone, will users attend to the written forms associated with the images, or overlook them because they are not used as sources of information in selecting a response? This is important because in later questions, users will be asked to choose the appropriate translation from the written forms alone without accompanying images, as well as to provide the written L2 translation for the L1 form without multiple choice options (figs. 8 & 9). If the user has not attended to the written forms in the presentation exercise but made their choice based on images alone, then it may be difficult for them to later recognize and recall the correct forms in the absence of images. Of course, it may be the case that users still would attend to the written words even if not needed, but since they do not need to, it is also possible they would simply overlook them.

It is also considered possible that even if a user were to look at the written forms, knowing that learning the L2 words was the goal, being able to rely on the images to make a fast, instinctual match could mean the user will attend less to the forms than they would if attention were needed to make the decision regarding a match; for example, by trying to remember the meaning of the L2 word after the fact, rather than during the decision making process, the latter of which is thought to possibly result in stronger memory of the target item. This would be similar to the process involved in learning via flashcards in a presentation only mode in which users try and hold the associated items in

memory for use during later practice. This outcome rests of course on the assumption that users will decide on the correct match by relying on the images since they will be initially drawn to them, and being sufficient for a match, the search task itself will end. If users do attend to both images and written forms in deciding on a match, regardless of the features of the images and written forms, a reduced memory for the target items may not necessarily be the case.

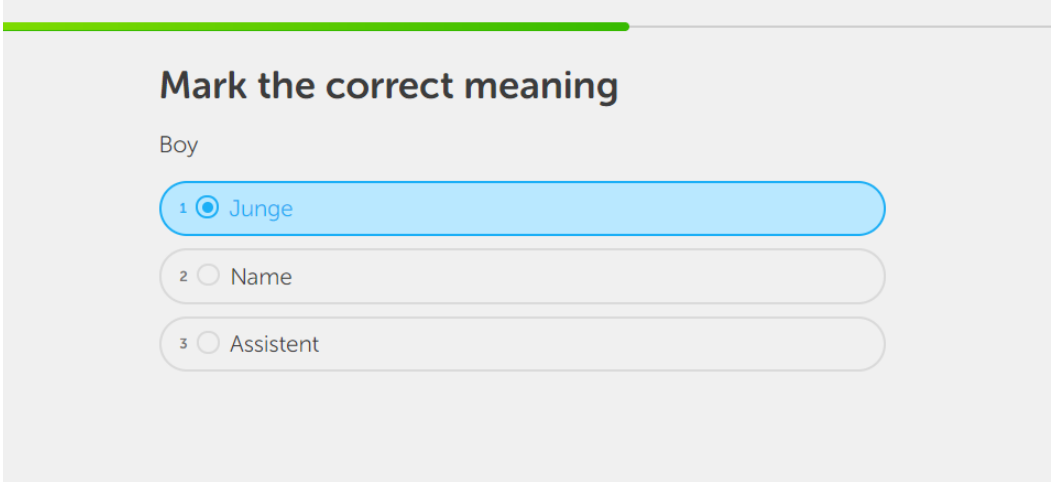


Fig. 8: Duolingo, German, Basics 1, Level Zero

Fig. 9: Duolingo, German, Basics 1, Level Zero

In the examples above (figs. 10 & 11), the user is being tested on their recognition and recall of the written forms of the target vocabulary. Continued practice through multiple trials of presentation/practice and recognition/recall of this sort is meant to consolidate the vocabulary as they are being learned and reviewed. The mixing of options with vocabulary from different presentation trials (fig. 9) may be meant to eliminate the likelihood that users will select the correct response based on memory of the earlier presentation task. Whether or not this is the case will not be explored here, but what is of interest to the present study is whether on a follow up exercise such as this, participants would have higher or lower percentages of correct responses based on features of the presentation exercise the target had been seen in earlier; specifically, whether features of image representativeness and distinctiveness as well as formal similarity between L1 and L2 words made a difference on later recall and recognition. In addition, if differences are observed across display types, could it have something to do with which elements on the screen users attended to during in-task decision making?

Whether or not users attend to written forms in the absence of necessity is one question which will be explored in the proposed study, as well as whether or not increased attention to written forms may result in better recall and/or recognition in follow-up exercises, and whether or not there are certain design conditions under which attention can be drawn to written forms out of necessity in making a match, given such attention improves later recall and/or recognition. Comparison of learning outcomes (e.g. recall and recognition scores) along with eye-tracking data (e.g. fixation measures) can be used to draw conclusions regarding these questions and answering these questions could

help to determine whether altering certain of these features to draw more attention to the written forms *during in-task decision making* could make a difference for outcomes on follow up recognition and recall tasks. The effects of these differences can be studied my manipulating task features to see if learning outcomes change based on the predictions above. The proposed study will test for this by manipulating image and text features to affect attention levels to images and text and to explore if differences in outcomes on follow-up exercises occur. Some additional examples, provided below, will further illustrate the type of exercise and design features of interest.

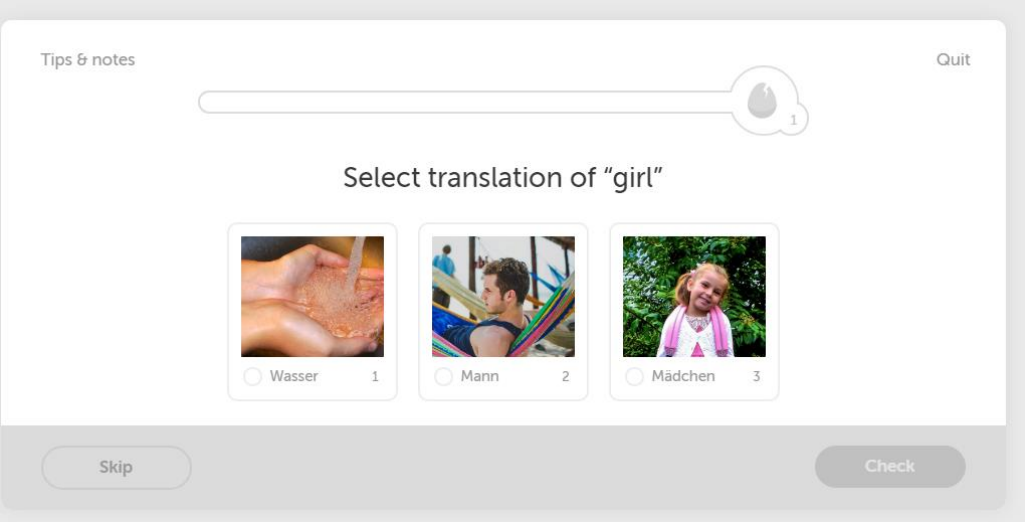


Fig. 10: Duolingo, German, Basics 1, Level One

In the next example (fig. 10), also from Duolingo, there are once again three L2 options for the user to select from as a match for 'girl.' The images here are less representative but are distinct in this particular display. In this example, two of the L2 options have written forms which resemble their L1 counterparts (water-Wasser, man-Mann), and so could be eliminated as options for the match to 'girl' on this basis alone.

Because the images in the display are distinct as to what they represent, and the images for ‘Wasser,’ and ‘Mann’ could not be said to depict the concept of *girl*, the images serve as sufficient cues for making a choice without needing to consider the written forms at all. Because of the high representativeness and distinctiveness of the picture accompanying ‘Mädchen’ as depicting *girl*, the elimination of cognates may or may not be relevant in this particular display. It is difficult to know whether or not users rely on these formal similarities in making a decision regarding a match where the images alone are sufficient, or if the even attend to the written forms after noticing the match. The use of eye-tracking to monitor what users attend to on screen when working through these types of tasks can be useful as a starting place for exploring some of these questions.

Once again, in a later exercise, users must give the L2 translation of the L1 cue word by recall without multiple choice options (fig. 11). While a recall exercise such as this would be expected to be challenging in any case since users will not have had enough exposure to the target to have retained it after only one trial, it would be especially difficult if users have not attended to the written form of the word during the presentation exercise. Whether or not learners can produce the written form partially may indicate whether or not attention had been given, even though not perfectly retained. Data from eye-tracking could be used in conjunction with recall response data to evaluate this hypothesis.

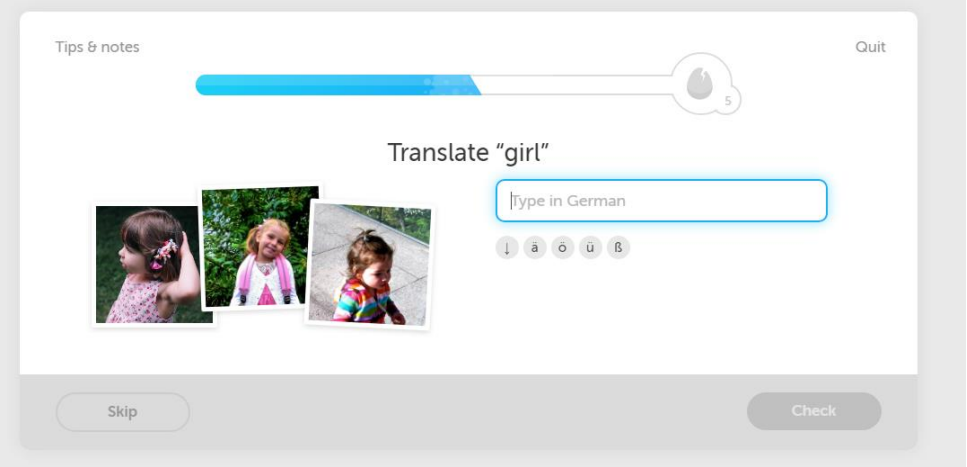


Fig. 11: Duolingo, German, Basics, Level One

Two additional examples from Duolingo demonstrate how the arrangement of formally similar items amongst items in a display can make a difference on which items may be attended to, selected, and retained. In the first (fig. 12), there is formal similarity between not only the L1 cue word and the L2 target, but between the distractor items and their L1 translations as well. All three images are also highly representative of the concepts they depict and are distinct from one another with respect to their associated concepts. In the second example (fig. 13), while there is similarity between the L1 cue word and the L2 target, there is not similarity between the distractors and their L1 forms. The question of interest here is whether, given such similarities, users will attend more to written forms than images, or more equally to both, in making their selection than they would if there were no formal similarities and the images were sufficient for making a choice.

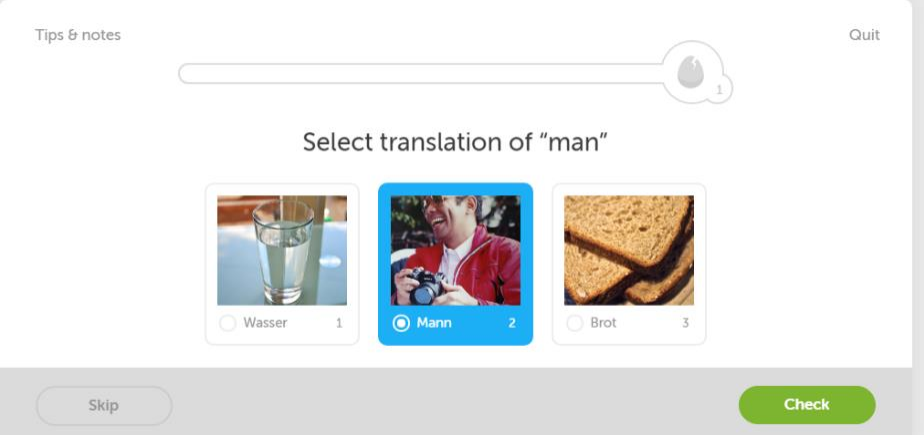


Fig. 12: Duolingo, German, Basics, Level One

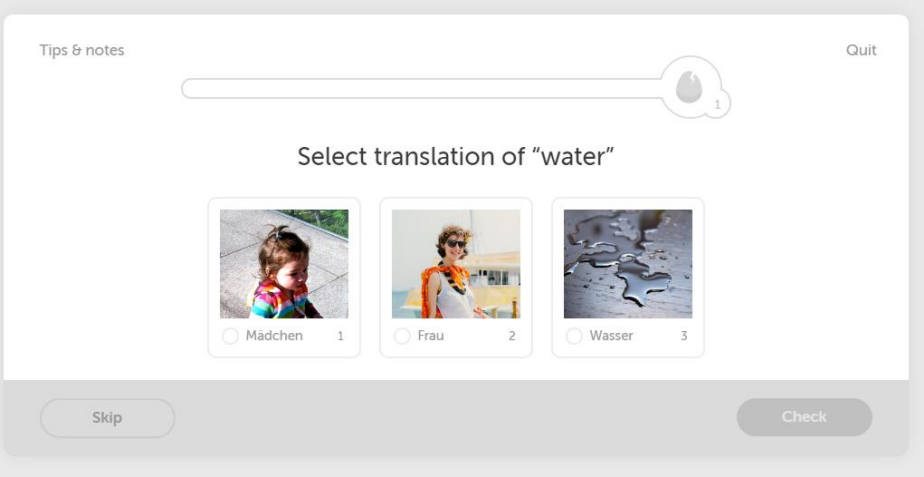


Fig. 13: Duolingo, German, Basics, Level One

While the examples above do not of course represent the entire range of vocabulary tasks found across CALL programs, they do illustrate a sample of the type of exercise which is of interest to the present study, in which users are presented with the form of target vocabulary (written and/or aural) along with images used to indicate meaning. Many of these exercises include L1 written translations as well, but this is not always the case. The next sections will detail the design and procedure for the proposed

study along with predictions of outcomes and methods for measuring and interpreting the results.

Methods

Participants

Participants will be undergraduate students in their first semester of study in a Spanish foreign language course at a public American university. Participants will be recruited via email by first contacting the course instructor to request permission to visit the class, followed by a classroom visit to request participation from students. Participants will be contacted from two classrooms initially. The classes from which participants will be recruited will be selected based on the first two instructors to reply positively to the request for a visit. The researcher will visit the classes at a time which is suitable for the instructors and will describe to the class the research being done. They will also go over the consent forms outlining details of participation, and the forms will be left with the instructor to hand out to the students. If enough students have not expressed interest from the first two classes, additional classes will be contacted by the same process. Ten participants will be selected for this study. Participants will be selected based on the first ten students to respond positively to the request to participate.

Consent forms will include wording that informs students they must be eighteen to participate. This requirement is to avoid the need of obtaining additional consent for persons under the age of eighteen. Consent forms will be in English. All participants will have a high degree of proficiency in English as determined by the instructor. Since no

personal information will be collected from participants, consent will be given through verbal agreement, as indicated in the written consent form. The consent forms will also inform students how to contact the research team to express interest in the study, which will be via email or phone. Participant responses will be kept anonymous through a fake log in name created by each participant to enter the quiz, which will be based on their birth month and favorite color. Participants will be instructed in the consent form as to how to create their log in information and informed that their responses will be anonymous.

First semester students were selected as the participant group due to the nature of the vocabulary items being presented to participants. Target vocabulary used in the treatment task are basic and relatively common words which learners are expected to be exposed to early on in a foreign language course. By selecting first semester students, it decreases the likelihood that participants will already be familiar with the target items. A pre-task assessment will be used to ensure this is the case and to eliminate from the data participant responses for items on which they are already familiar prior to the treatment task. The assessment will include all of the target items in the study as well as twice as many distractor items. The number of participants was set at ten since this is an early exploratory study meant to gather data and generate some initial findings. Later, follow-up studies would aim to include larger numbers of participants drawn from a greater number of classrooms or over a number of different institutions. Different age groups and language backgrounds would also need to be included to obtain a wider set of data and more generalizable results.

Vocabulary and Images

Target vocabulary items were chosen based around the theme of transportation and travel and selected to include both concrete and abstract concepts. Due to the specific requirements of each condition, not all items appear as targets in the treatment task; however, all items will be presented multiple times across the various conditions as distractors. For this reason, all items will be included in the post-test, and used to address questions regarding incidental and implicit learning processes.

A Google image search was used to select the images accompanying each vocabulary item and were chosen based on the apparent representativeness of each image for the associated concept as well as whether or not the image could also be used to represent other concepts on the vocabulary list (i.e. to be useful as a distractor image). Independent raters will be used to confirm the representativeness and distinctiveness of the images selected by the researcher, which will be carried out in two parts. First, to assess representativeness, raters will be presented with each image and asked simply, "what is this a picture of?" For this part, raters will be instructed that each image represents only one object or concept, and to provide a single response for what is depicted. If raters are easily able to name the target concept from the picture using its associated English vocabulary word from the study set, images will be considered representative. If they are not able to easily name the concept using the associated vocabulary item, the image will be considered unrepresentative. If images which had been selected by the researcher as being either representative or unrepresentative are

shown not to be so based on rater responses, new images will be selected for those items and re-checked.

For the second part, to determine distinctiveness, a different group of raters will be shown the same set of images and asked to name every object or concept they think the image could represent, or that they see in the picture. The responses from all raters across images will be compared to ensure that the concepts for which those images are meant to serve as distractors are considered likely to be represented by that image. If images which had been selected by the researcher as being either representative or unrepresentative of concepts for which they are meant to serve as distractors are shown not to be so based on rater responses, new images will be selected for those items and re-checked.

Five raters will be selected for each group, for a total of ten, who will be recruited from the researcher's professional and personal contacts. Recruited individuals will be given a written document describing the details of participation and informed that their verbal agreement will count as consent. Data will be collected through an online form sent to participants to submit their responses. Data from this portion will not be included in the final report on the findings of the study, nor any associated publications. It will only be noted that raters were used to confirm judgements of representativeness and distinctiveness, and in brief how this was done. The confirmation of the association between images and vocabulary from the study set will only be carried out on the English vocabulary items. This will be done primarily to reduce the number of raters needed, which would be double if tested for both English and Spanish sets, and because the

participants in the main study will be English speakers who are not expected to know the associated Spanish vocabulary for the images, and thus will be drawing on knowledge of English vocabulary to make an association between words and depicted concepts.

Design

Participants will complete all tasks via a computer mediated quiz which will be created, stored, and accessed through an online quiz generator program. Participants will first receive a pre-test to assess their existing knowledge of the vocabulary items in the treatment task. The structure of this task will be a fill-in-the blank exercise which will ask participants to provide translations for the target items in two different conditions: English to Spanish and Spanish to English. Items will be presented to participants one at a time and will include a total of eighteen items in each set for a total of 36 items (see Appendix I). The purpose of the pre-test will be to screen for participants' knowledge of the target Spanish vocabulary prior to the treatment task. This will serve to filter data from the treatment task such that only items for which participants do not have prior existing knowledge will be tracked on the post-test for measures of learning.

In the treatment task, participants will work through a series of vocabulary matching exercises while being monitored through an eye-tracking device. Participants will be asked to select the Spanish translation for a set of English words presented in panels of four items at a time. Each English word will be presented in written form, and each translation option will be presented as a written form with an accompanying image. There will be a total of twelve experimental conditions which vary in terms of the representativeness of the images, their distinctiveness from distractor images, and the

similarity in form between the English and Spanish translations. Participants will see a total of thirty-six screens, three for each condition, which will change in only the positions of items on the panel. Each screen will count as one trial in the treatment task. The purpose of the treatment task will be to track participants' attention to the words and images on the screen as they make decisions regarding the matching translation across trials.

Following the treatment task, participants will take a post-test consisting of two parts. This test will be taken first immediately after the main task and again three days later. The design of the first part of the test will be the same as the pre-test, except items will be scrambled so as to be presented in a different order. The purpose of this part of the post-test will be to assess participants' recall knowledge of the target Spanish words following the experimental training. The design of the second part will be similar to the treatment task in that participants will have to match the Spanish translations for the English vocabulary items in the set; however, in the post-test, all supporting images will be removed, and participants will have to choose the correct translation based on the written forms alone. Participants will see a total of fifty-four screens, three for each of the target items, which will vary in the positions of the options on the panel. The purpose of this part of the post-test will be to assess participants' recognition knowledge of the target Spanish words following the treatment task.

In the treatment task, there will be a total of eighteen translation pairs (thirty-six items) appearing in various combinations in the different tasks; however, due to the constraints of the different conditions, not all of the written forms will be presented the

same number of times. There are six pairs that have similarity in written form between the two languages, and these items appear only once or twice across conditions. The other twelve pairs do not share similarity in form and appear three to four times across conditions. Findings for the learning outcomes of each of these two groups will be compared separately with respect to the number of times participants are presented with them (see Analysis section).

Variables of Interest

In order to address the research questions posed in this study, three variables will be considered with respect to the visual and written stimuli; the *representativeness* of the image in depicting the associated concept, the *distinctiveness* of the image from distractor items, and the similarity in written form between Spanish and English translations. *Representativeness,* for the purposes of this study, is defined by the likelihood of identifying the associated concept from the image alone without written text or other contrasting images. An image is considered *representative* if its associated concept is likely to be recognized and/or named easily as depicting that concept. *Distinctiveness* is defined as the clear separation of the images in a display as depicting their own concepts such that they could not be confused for one another. An image is considered *distinct* if other items in the panel are not likely to be considered representative of the same concept. *Formal similarity* is defined as the orthographic similarity in written form between target L2 items and their known L1 counterparts. An item is considered to have *formal similarity* if its written form shares a similar spelling pattern with that of its translation equivalent, such that the two words appear to be related.

Conditions and Trials

The experimental conditions will be set up to vary the combinations of variables in four ways. In condition A, images are defined as being *representative* and *distinct*, meaning that it is supposed to be the case that the concept being depicted in each image is clear from the image alone, and distractor items are not thought to be representative of the same concepts. In condition A1, there is no *formal similarity* between the L1 form and target L2 form or between any of the distractors and their English equivalents. In condition A2, there is *formal similarity* between the L1 form and target L2 form, but not with distractor forms and their English equivalents. In condition A3, there is no *formal similarity* between the L1 form and target L2 form, but there is similarity between one of the distractor items and its English equivalent.

For all conditions, the variable of formal similarity was manipulated in the same way, such that for all conditions X1, there is no formal similarity between any of the items in the panel, for all conditions X2, there is similarity with the target but none of the distractors, and for all conditions X3, there is similarity with a distractor but not with the target. In condition B, target images are defined as being *representative* but *indistinct* in that distractor images are also considered to be potentially representative of the same concept as the target since they all depict similar features. In condition C, target images are defined as being *unrepresentative* of their concepts because they refer to what are considered abstract concepts which cannot be easily represented. These images are also defined as being *indistinct* in that distractor images could also potentially refer to the same concept. Finally, in condition D, target images are defined as being

unrepresentative of their concepts in that they again have abstract concepts as their referent but are also *distinct* in that distractor images are not considered to be representative of the same concept.

Across all conditions, it is predicted that participants who attend more to images than written forms when looking to make a translation match will have lower scores on learning outcomes for target forms in post-task assessments than participants who attend more to written forms than images, or more equally to images and written forms. It is predicted that participants will have the lowest recognition and recall scores on items which are tested under condition A, in which accompanying images are highly representative and distinct so as to not require drawing on information from written forms. The exception being for items which have formal similarity to their English translations, which participants are expected to give more attention to and have higher post-test scores on regardless of the conditions in which they are presented.

It is also predicted that in all conditions in which there is a written form which has similarity to its English translation, whether the target or a distractor, participants will attend more to written forms than in conditions where no items have written formal similarity because it is thought that once participants have been cued to attend to written forms by the similar item, they will then attend to the written forms of the other items as well. It is further predicted that participants will tend to attend to images first and primarily as long as images are distinct from one another (conditions A and D), and that they will not attend to written forms unless either cued by a similar form in the panel, or images alone are not sufficient to indicate a match, which would occur in cases in which

images for targets and distractors are not distinct (conditions B and C). Even in cases in which images are unrepresentative (conditions C and D), it is predicted that participants will still attend mainly to images when they are distinct and only one image is likely to represent the target concept (condition D). Finally, it is predicted that the only case in which participants will tend to select a particular distractor is if it is similar in form to its English equivalent and the images for items in the set are unrepresentative and indistinct such that the concept associated with the distractor could be said to match the English translation (condition C3). A detailed description of each condition and its predictions can be found in the Appendix I.

Equipment

This study will use a Mirametrix S2 desktop eye-tracker. The device is a portable USB device which is placed at the bottom of the computer screen and is relatively unobtrusive. The hardware has a sampling rate of 60 Hz and a point-of-gaze measurement accuracy of 0.5 – 1 degree range. It uses a 9 point calibration which completes in about 15 seconds and holds for a long period of testing. The Viewer software records a video during participant interaction and provides an overlay marking participants' gaze behaviors. The accompanying API interface can integrate the eye-tracking data with other XML and CSV compatible applications for export.

Measures and Metrics

In order to address the research questions, this study will consider the following eye-tracking metrics: time to first fixation, fixation duration, and fixation count. The

reason for including time to first fixation is to determine whether participants' first item of attention in a given trial is that of an image or a written word. Answering this question is necessary to test the prediction that participants will largely first attend to images, which would indicate a bias towards images over text in these types of tasks. Fixation duration and fixation count were included to determine for how long and how often participants attend to particular images and written words in each trial and to summarize whether, for each trial, participants attended longer or more often to images or written words in making their selection. Answering these questions is necessary to test the predictions that participants will attend less often and for shorter durations to written words under conditions where images are representative and distinct and there is no formal similarity but will attend more often and for longer durations to written words when images are unrepresentative and indistinct or there is formal similarity.

Fixation duration and fixation count are also needed to determine which items of image and text may be easier or more difficult for participants to process in each trial. Answering this question is necessary to test the predictions that under conditions in which images are unrepresentative or indistinct, participants will attend longer and more often to both images and written words because items in these trials will require more consideration in order to make a match. These measures are also needed to determine if certain items of image or text in each trial are more interesting for participants, which is related to the prediction that written forms will be attended to longer and more often when there is formal similarity with the L1 translation, since the recognition of similarity

is thought to attract users' attention. This increased attention in response to familiarity may signal higher interest to that item for the participant.

Areas of Interest (AOIs) are defined for each image and written word that appears on the screen during a given trial. AOIs were selected which corresponded to image features which are considered objects of possible attention for participants, and which may serve as cues for a possible match. Attention to particular AOIs in a given trial will be compared to participant responses to make inferences regarding which features of the image were drawn on for making a choice. AOIs for written words were not broken down to smaller units in order to avoid overlapping AOIs; however, doing so may be beneficial for exploring whether participants attend to specific spelling patterns within words which are similar in either the treatment task, or post-task assessments.

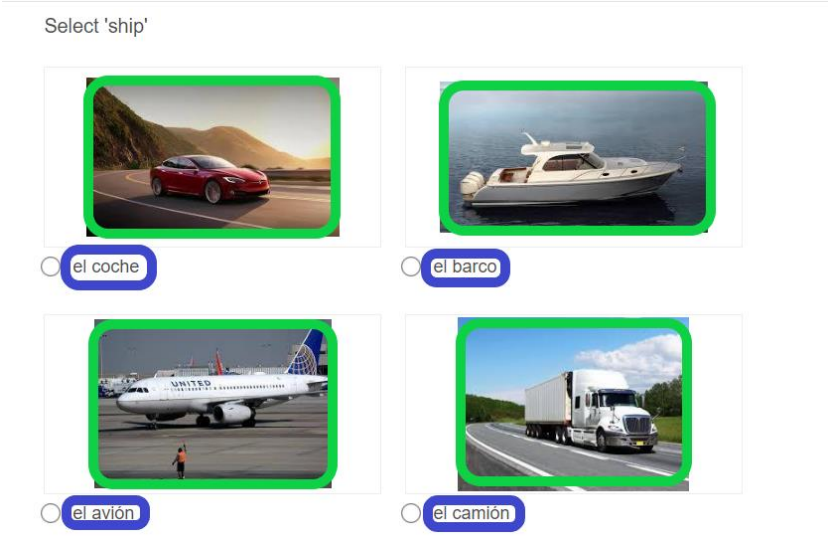


Fig. 14: Example AOIs for a single trial.

Times of Interest (TOIs) were set for the duration of a given trial as the time between when the panel appears on the screen to when the participant clicks the mouse

for a response. This TOI was set because all attentional behavior within the time of exposure to the panel on screen is considered relevant for participant response.

Procedure

Following their expression of interest and selection, participants will be instructed as to the time and place where to appear for the experiment. The experiment will take place at a location on the campus at which the students attend as made available by the university. Participants will be tested individually in a room by themselves, and the researcher will not be in the room while the participants complete the tasks. Participants will be seated in front of a desktop computer, on which the remote eye-tracking device will be located. They will be seated at a distance from the screen to their preference which is comfortable to them. The room will be lit to natural lighting conditions. Due to time constraints based on the length of the test, participants will be tested over two days, five each day. Participants will first take the pre-test, followed right after with the treatment task. After the treatment task, there will be a 15-minute break, and then participants will take the first post-test. Participants will be instructed as to when and where to return for the second post-test, which will take place 3 days later.

Participants will receive feedback following each response as to whether it was incorrect or not but will not be presented with a final grade or score upon completion. Scores will only be accessible to the research team, and the only scores that will be tracked are those for correct and incorrect responses. None of the tests will be timed, and participants may take as long as they need to answer each question and complete the entire exercise. Reaction times will not be recorded for the purposes of this study.

Participants will be informed prior to beginning that they will not be presented with a final score or review of their performance upon completion.

Analysis

Visual Analysis

To visualize the eye-tracking data, both heat maps and gaze path plots will be used. Heat maps, in showing the general distribution of gaze points over a display, are useful for getting a view of the overall amount of attention given to a particular AOI. This allows for a fast interpretation regarding whether participants attend more to images or written words within a given trial since these will constitute the AOIs in each display. Gaze path plots on the other hand, show more detailed information regarding for how long, how many times, and in what order participants look at different locations on the screen. This type of visualization is more detailed than a heat map since it includes data for each individual metric, and so is useful for answering more specific questions regarding time and order of fixations. At the same time, the more summary information provided by heat maps is beneficial since the overall attention to each location would otherwise need to be calculated from each of the individual metrics. For these reasons, both visualizations will be used in conjunction in order to answer both more general and more specific questions regarding participants' gaze behavior.

Using heat map visualizations, fixation measures for images and written forms in each trial will be compared and labeled either 1, for trials in which more attention was given to images overall, or 2, for trials in which more attention was given to written

forms overall. Labels for trials of the same condition will be compared to see if there is a difference in attention to words or images across trials, and each condition will be labeled overall based on which number, 1 or 2, had the highest frequency across the tree trials. Interpretations will be made regarding the design of each condition and the findings for overall attention to images or written forms in each based on whether the findings align with predictions or not. This portion of the analysis will answer questions regarding whether, under certain conditions, participants attend more to images or written forms when making decisions regarding a translation match. Information from the gaze path plots will be used to answer questions about whether participants attend first, or more often (i.e. return more) to images or forms across conditions, and interpretations of these findings will be included in the discussion.

Statistical Analysis

The purpose of this portion of the analysis will be to compare scores for time spent fixating on images and written forms associated with each vocabulary item to recall and recognition scores for those items on immediate and delayed post-tests. Participant scores will be collected from the online quiz and put into a spreadsheet. SPSS will be used to perform statistical analyses on the data. This portion of the analysis will consist of three stages. First, participant responses on each trial will be summed and assigned a label A, if the majority of participants got the response correct, or B if the majority of participants got the response wrong. Labels for trials of the same condition will be compared to see if there is a difference in correct responses across trials, and each

condition will be labeled overall based on whether participants mostly got the responses correct or incorrect across the tree trials.

Next, a comparison will be made between findings on levels of attention to images and written text, along with tendencies towards the correct response, for each condition. Interpretations will be made regarding the design of each condition and whether or not findings for attention to words or images and responses for that condition align with predictions or not. Since the experimental task consists of an initial presentation of the target vocabulary, and participants will be pre-assessed to ensure they do not already know the correct responses, participants should have only a twenty-five percent chance of choosing the correct response. Therefore, an indication of a higher number of correct responses, A labeled conditions, will be considered to demonstrate possible facilitation towards the correct response based on design features of that particular display, and measures of attention to particular elements in the display will be used to support or modify that conclusion.

Finally, a correlational analysis will be run to identify if there exists a relationship between fixation scores and post-test learning scores. To prepare the data, each vocabulary item will be given score based on whether, across all trials in which it appears, either its written form or associated image was given more attention overall, summed across all participants. Items which received more attention to their images overall will be labeled with a 3, and those which received more attention to their form overall will be labeled with a 4. Post-test results will be scored as either correct or incorrect for each item and summed across all participants. Scores for items on which

participants had prior knowledge based on the pre-test will not be included in this data. Each item will be given a score based on whether most participants got the answer correct or incorrect on the post-test (this analysis will be run separately on both the immediate and delayed recall task). If most participants got the answer correct, that item will be labeled C, if most did not get it correct, it will be labeled D. Based on the hypotheses of this study, the prediction is that a correlation will be seen between items labeled 4 (more attention to form) and those labeled C (most participants chose correct on post-test), as well as those labeled 3 (more attention to images) and those labeled D (most participants chose incorrect on post-test). Interpretations will be made regarding the correlations between labeled items, and whether they align with predictions or not.

Conclusion

While the use of CALL applications has become increasingly common in second language instruction as a means of overcoming many of the challenges learners face in receiving adequate exposure to and practice with the target language necessary to improve learning towards their goals, a number of issues exist regarding the ways in which particular design features of CALL activities may affect users' short- and long-term retention of target language structures. In order for tutorial CALL to advance as a technology for improving language learners' ability to acquire essential grammar and vocabulary needed for further language development, it must thoroughly and critically analyze the ways in which its designs facilitate or impede learning objectives by drawing on principles of SLA as a source of evaluation. At the same time, SLA theorists and researchers must continue to explore means of testing CALL activities and their effects

on language learning in order to better inform improvements within CALL and practices in second language instruction. The above proposed study aims to serve as a starting place in the direction towards increasing critical evaluation, research, and development at the cross-section between CALL, SLA, and second language instruction.

References

Al-Hejin, B. (2004). Attention and awareness: Evidence from cognitive and second language acquisition research. *Working Papers in TESOL & Applied Linguistics, 4*(1).

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology, 5*(1), 202-32.

Asher, J. J. (1969). The total physical response approach to second language learning. *The modern language journal, 53*(1), 3-17.

Beatty, K. (2013). *Teaching & researching: Computer-assisted language learning*. Routledge.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*(2-3), 150-177.

Birdsong, D. (Ed.). (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Blake, R. J. (2013). *Brave new digital classroom: Technology and foreign language learning*. Georgetown University Press.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*(2), 175-198.

Bojko, A. (2013). *Eye tracking the user experience: A practical guide to research*. Rosenfeld Media.

Bower, G. H. (1961). Application of a model to paired-associate learning. *Psychometrika, 26*(3), 255-280.

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

Brooks, P. J., & Kempe, V. (2013). Individual differences in adult foreign language learning: The mediating effect of metalinguistic awareness. *Memory & cognition*, *41*(2), 281-296.

Broughton, G., Brumfit, C., Pincas, A., & Wilde, R. D. (2002). *Teaching English as a foreign language*. Routledge.

Brown, H. (2007). *Principles of language learning and teaching*. Pearson Longman.

Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy*. Pearson Education Inc.

Chapelle, C. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press.

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

Cook, V., & Singleton, D. (2014). *Key topics in second language acquisition* (Vol. 10). Multilingual Matters.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in second language acquisition*, *22*(4), 499-533.

DeKeyser, R. M. (2001). Automaticity and automatization. *Cognition and second language acquisition*. Cambridge University Press.

Eagle, M., & Leiter, E. (1964). Recall and recognition in intentional and incidental learning. *Journal of experimental psychology*, *68*(1), 58.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition, 24*(2), 143-188.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and Bilingualism*. Georgetown University Press.

Ervin-Tripp, S. M. (1974). Is second language learning like the first. *TESOL Quarterly*, 111-127.

Eye tracking metrics. (n.d.). *Tobii Pro*. Retrieved from https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/interpret/eye-tracking-metrics/

Farnsworth, B. (2018, August 14). 10 Most Used Eye Tracking Metrics and Terms. *Imotions*. Retrieved from https://imotions.com/blog/7-terms-metrics-eye-tracking/

Gass, S. M. (2017). *Input, interaction, and the second language learner*. Routledge.

Gass, S. M., Behney, J., & Plonsky, L. (2013). Looking at inter-language processing. *Second language acquisition: An introductory course*. Routledge.

Gass, S., & Selinker, L. (2001). *Second language acquisition: An introductory course*. L. Erlbaum Associates.

Georgiou, Liu, & Xu. (2017). Examining the direct and indirect effects of visual–verbal paired associate learning on Chinese word reading. *Journal of Experimental Child Psychology, 160*, 81-91.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*(1), 1-16.

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eyetracking. *Studies in Second Language Acquisition, 35*(3), 483-517.

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. *Noticing and second language acquisition: Studies in honor of Richard Schmidt*, 183-205.

Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be*? Applied linguistics, 11*(4), 341-363.

Gu, S. (2017). *Interaction Process and Chinese EFL Learners' Proficiency Development*. Springer.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a foreign language, 8*, 689-689.

Honda, M., Barrett, G., Yoshimura, N., Ikeda, A., Nagamine, T., & Shibasaki, H. (1996). Event-related potentials during paired associate memory paradigm. *Electroencephalography and Clinical Neurophysiology*, *100*(5), 407-421.

Hubbard, P., & Siskin, C. B. (2004). Another look at tutorial CALL. *ReCALL, 16*(2), 448-461.

Huckin, T., & Coady, J. (1999). Incidental vocabulary acquisition in a second language: A review. *Studies in second language acquisition, 21*(2), 181-193.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. *Vocabulary and applied linguistics*. Palgrave Macmillan.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. *Cognition and Second Language Instruction*. Cambridge University Press.

Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *Aila Review, 11*(97-112).

Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge University Press.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology, 21*(1), 60-99.

Kam, C. L. H., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology, 59*(1), 30-66.

Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System, 18*(2), 199-207.

Kleinsmith, L. J., & Kaplan, S. (1963). Paired-associate learning as a function of arousal and interpolated interval. *Journal of experimental psychology, 65*(2), 190.

Krashen, S. (1981). *Second language acquisition and second language learning* (1st ed., Language teaching methodology series). Pergamon Press.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440-464.

Kress, G. (1998). Visual and verbal modes of representation in electronically mediated communication: The potentials of new forms of text. *Page to screen: Taking literacy into the electronic era*, *53*.

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review, 59*(4), 567-587.

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied linguistics, 22*(1), 1-26.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice, 2*(12), 59-67.

Leow, R. (2001). Attention, awareness, and foreign language behavior. *Language Learning, 51*(Supplement 1), 113-155.

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Oxford University Press.

Levy, M., & Stockwell, G. (2013). *CALL dimensions: Options and issues in computer-assisted language learning*. Routledge.

McLaughlin, B. (1990). Restructuring. *Applied linguistics, 11*(2), 113-128.

Mitchell, R., Myles, F., & Marsden, E. (2019). *Second Language Learning Theories*. Routledge.

Mondria, J. A., & Boer, M. W. D. (1991). The effects of contextual richness on the guessability and the retention of words in a foreign language. *Applied linguistics, 12*(3), 249-267.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning, 24*(1), 17-38.

Nation, I. (2001). *Learning vocabulary in another language* (Cambridge applied linguistics series). Cambridge University Press.

Nation, I. P. (2013). *Learning vocabulary in another language* (Second ed., Cambridge applied linguistics series).

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy, 14*, 6-19.

Nyström, M., & Holmqvist, K. (2008). Semantic override of low-level features in image viewing—both initially and overall. *Journal of Eye Movement Research, 2*(2), 1-11.

Ortega, L. (2014). *Understanding second language acquisition*. Routledge.

Ovando, C. J., & Combs, M. C. (2018). *Bilingual and ESL classrooms: Teaching in multicultural contexts*. Rowman & Littlefield.

Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior, 4*(1), 32-38.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The linguistic review, 18*(1-2), 9-50.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457-1506.

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge University Press.

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

Russell, W. A., & Storms, L. H. (1955). Implicit verbal chaining in paired-associate learning. *Journal of Experimental Psychology, 49*(4), 287.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129–158.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. *Implicit and explicit learning of languages, 22*, 165-209.

Schmidt, R. W. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning* (pp. 1–63). University of Hawai‘i, National Foreign Language Resource Center.

Schmidt, R. (2001). Attention. *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. *Proceedings of CLaSIC 2010, Singapore, December 2-4* (pp. 721-737). National University of Singapore, Centre for Language Studies.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on individual characteristics and foreign language education* (pp. 27–50). Mouton de Gruyter.

Schneider, Healy, & Bourne. (2002). What Is Learned under Difficult Conditions Is Hard to Forget: Contextual Interference Effects in Foreign Vocabulary Acquisition, Retention, and Transfer. *Journal of Memory and Language, 46*(2), 419-440.

Schouten-van Parreren, C. (1989). Vocabulary learning through reading: Which conditions should be met when presenting words in texts. *AILA review*, *6*(1), 75-85.

Segalowitz, N., & Trofimovich, P. (2012). Second language processing. *The Routledge handbook of second language acquisition*, 179-192.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*(1-4), 209-232.

Silverman, R., & Hines, S. (2009). The effects of multimedia-enhanced instruction on the vocabulary of English-language learners and non-English-language learners in pre-kindergarten through second grade. *Journal of educational psychology, 101*(2), 305.

Singleton, D. (1995). Introduction: A critical look at the critical period hypothesis in second language acquisition research. *The age factor in second language acquisition*. Multilingual Matters.

Smith, L. B., & Colunga, E. (2012). Developing categories and concepts. *The Cambridge handbook of psycholinguistics*, 283-308.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child development*, 1114-1128.

Stadthagen-Gonzalez, H., Damian, M. F., Pérez, M. A., Bowers, J. S., & Marín, J. (2009). Name–picture verification as a control measure for object naming: A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology, 62*(8), 1581-1597.

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition, 29*(3), 449-484.

Sternberg, R. J. (1987). Most vocabulary is learned from context. *The nature of vocabulary acquisition*. Psychology Press.

Stockwell, G. (Ed.). (2012). *Computer-assisted language learning: Diversity in research and practice*. Cambridge University Press.

Styles, E. (2006). *The psychology of attention*. Psychology Press.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in second language acquisition, 16*(2), 183-203.

Types of eye movement. (n.d.). *Tobii Pro*. Retrieved from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/types-of-eye-movements/

What do we study when we use eye tracking data?. (n.d.). *Tobii Pro*. Retrieved from https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/what-do-we-study-when-we-use-eye-tracking-data/

Wysocki, A. F. (2001). Impossibly distinct: On form/content and word/image in two pieces of computer-based interactive multimedia. *Computers and Composition, 18*(2), 137-162.

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

Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO journal*, 33-58.

Zobl, H. (1995). Converging evidence for the ‘acquisition-learning ‘distinction. *Applied Linguistics, 16*(1), 35-56.

APPENDIX A

TASKS AND CONDITIONS

Pre-Test

Participants will be asked to provide translations from English to Spanish and Spanish to English for the vocabulary items presented in the experimental conditions. Participants will be presented with each item individually for a total of 36 items (18 in each language). The purpose of the pre-test will be to assess participants' existing knowledge of the target Spanish words that appear in the main task.

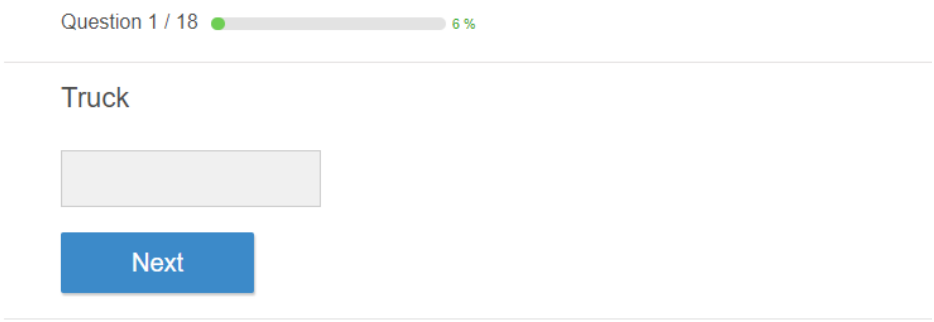


Fig 1. Example item from the English to Spanish translation set.

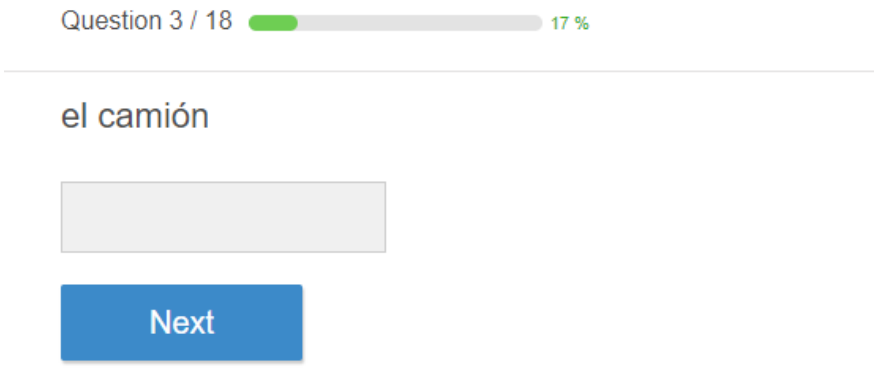
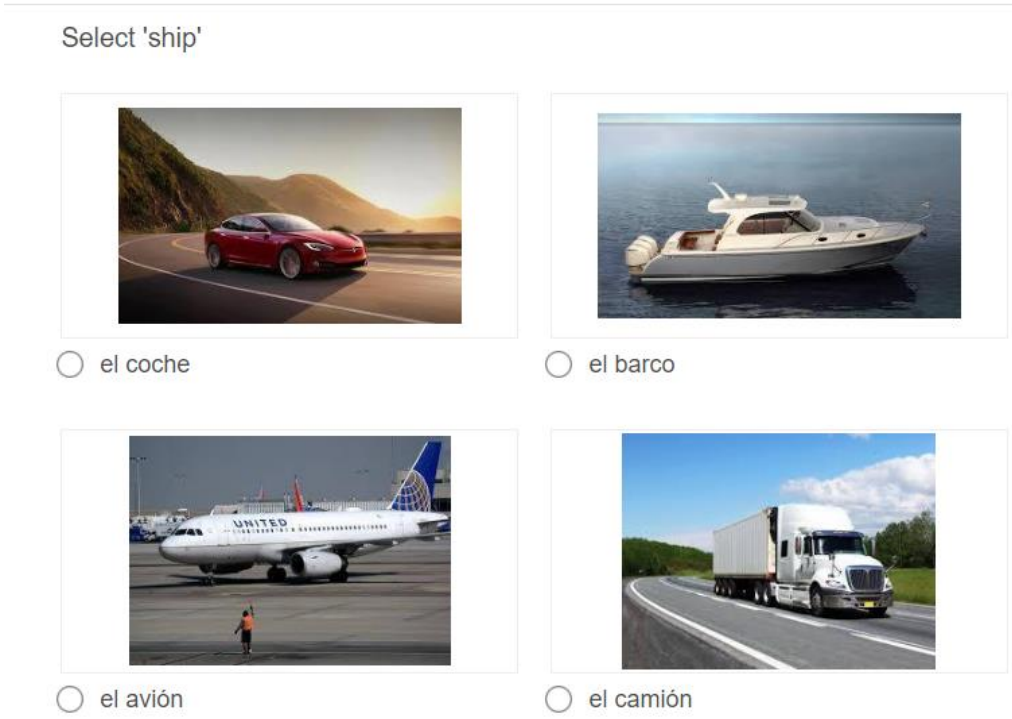


Fig 2. Example item from the Spanish to English translation set.

Treatment Task

Condition A1



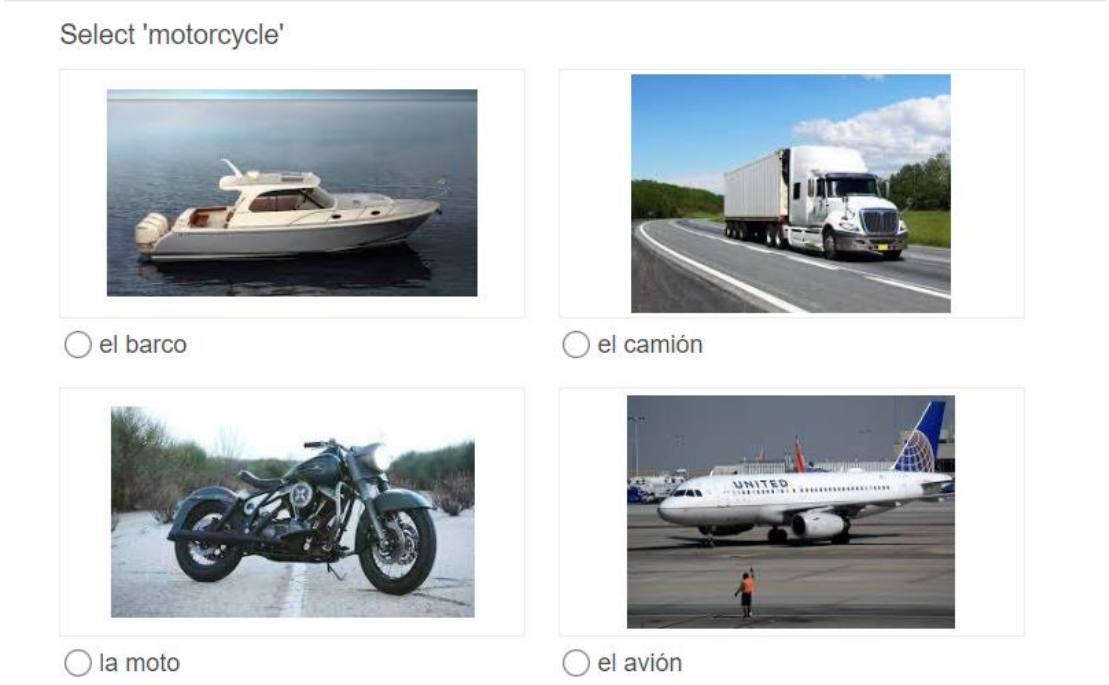
Description

The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted and *distinct* in that distractor items are not also representative of the concept. Distractor items are representative of their respective concepts and distinct from other items. There is no *formal similarity* between the L1 form and target L2 form or between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target ‘el barco’ because of the high representativeness and distinctiveness of the image accompanying the target word which leads the participant to quickly identify a likely match. It is also predicted that participants will attend more to the images than written forms since attention to forms is not needed to make a match, and there is no formal similarity between any of the L2 forms and their English equivalents which could provide additional information regarding a potential match.

Condition A2



Description

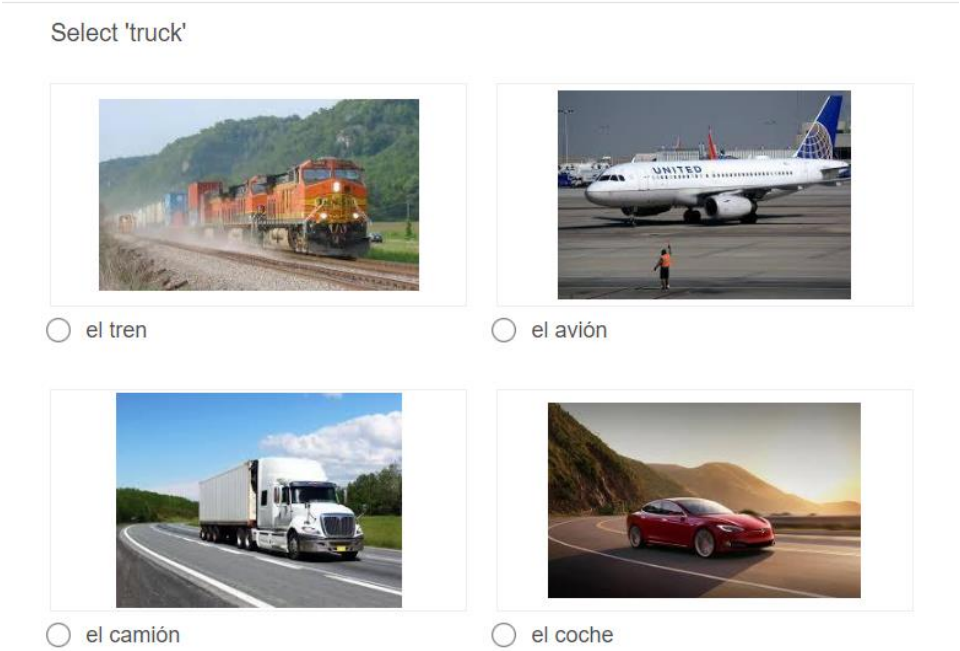
The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted and *distinct* in that distractor items are not

also representative of the concept. Distractor items are representative of their respective concepts and distinct from other items. There is *formal similarity* between the L1 form and target L2 form, but not with distractor forms. There is no formal similarity between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target ‘la moto’ because of the high representativeness and distinctiveness of the image accompanying the target word, as well as the similarity in form to its English translation which further indicates this option as a likely match. It is also predicted that participants will attend more equally to images and written forms because of the similarity between one of the translation options and its English equivalent.

Condition A3



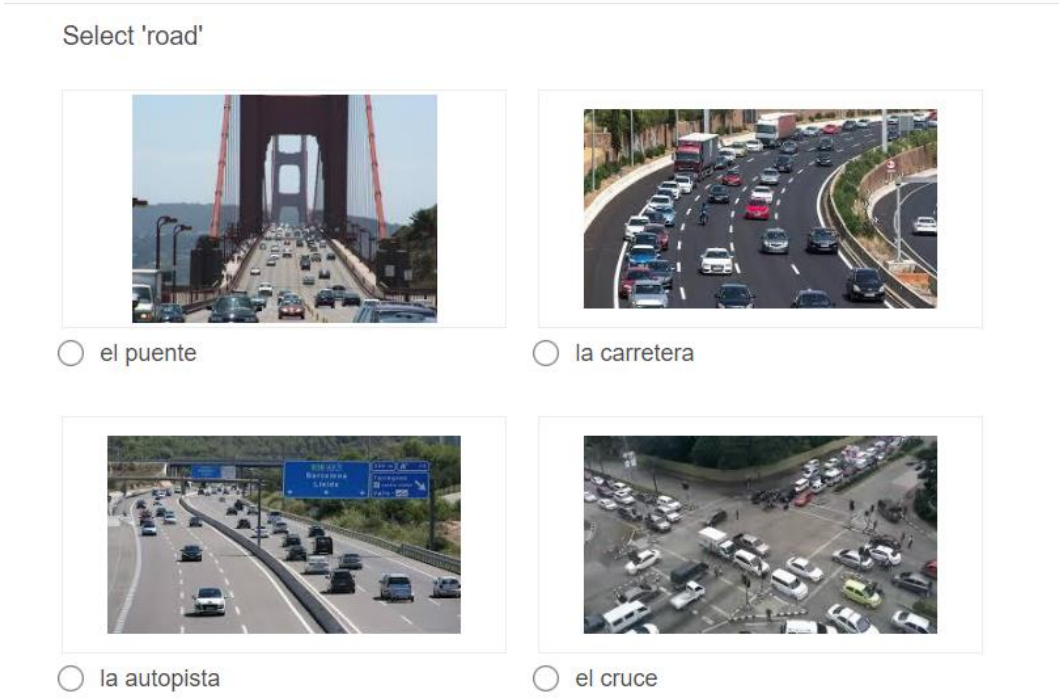
Description

The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted and *distinct* in that distractor items are not also representative of the concept. Distractor items are representative of their respective concepts and distinct from other items. There is no *formal similarity* between the L1 form and target L2 form, but there is similarity between one of the distractor items and its English equivalent.

Predictions

Participants will tend to select the target ‘el camión’ because of the high representativeness and distinctiveness of the image accompanying the target word which leads the participant to quickly identify a likely match. It is also predicted that participants will attend more equally to images and written forms due to the similarity between one of the items and its English equivalent. While the formal similarity occurs as a feature of one of the distractor items, this is not expected to have an impact on the selection made, due to the high representativeness of the images accompanying each option and because the concept cued by the formally similar ‘el tren’ is not that of a truck.

Condition B1



Description

The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted but *indistinct* in that distractor images are also representative of the concept since a road is featured in each image. Each distractor image is also representative of its own concept, a bridge, a highway, and an intersection respectively, although not highly representative since other items (e.g. cars, road signs) are also featured in each image. There is no *formal similarity* between the L1 form and target L2 form or between any of the distractors and their English equivalents.

Predictions

Participants will select the target ‘la carretera,’ at chance because neither the image or the L2 form is sufficient to indicate a clear match. It is predicted that participants will attend to the different images in an effort to determine which once best depicts the target concept. It is also predicted that participants will attend more equally to written forms and images in an attempt to make a match even though there is no formal similarity between L1 and L2 words to cue a relationship because, since the images are ambiguous with reference to the target concept, participants will need to look to written forms as a possible source of additional information.

Condition B2

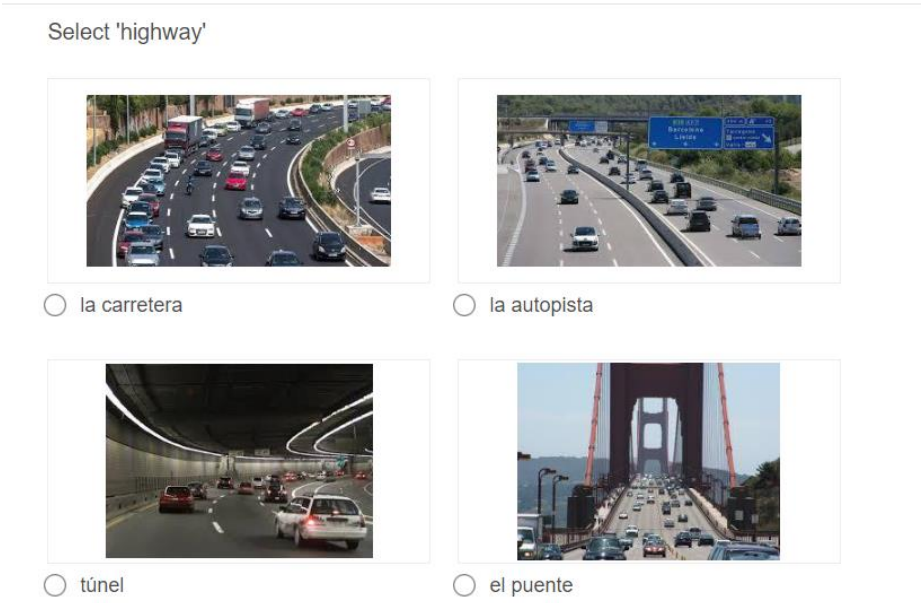
Description

The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted but *indistinct* in that distractor images are also representative of the concept since traffic is a feature of each image. Each distractor image is also representative of its own concept, a road, a highway, and an intersection respectively, although not highly representative since other items (e.g. cars, roads, signs) are also featured in each image. There is *formal similarity* between the L1 form and target L2 form, but not with distractor forms. There is no formal similarity between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target 'el tráfico' because the similarity between the L1 and L2 forms indicates it as a possible correct match, and there is no additional feedback provided by the images to indicate otherwise (i.e. the corresponding image does depict traffic). It is also predicted that participants will attend more equally to images and written forms because the images are ambiguous with reference to the target concept and the similarity between one of the translation options and its English equivalent will drive attention to written forms.

Condition B3



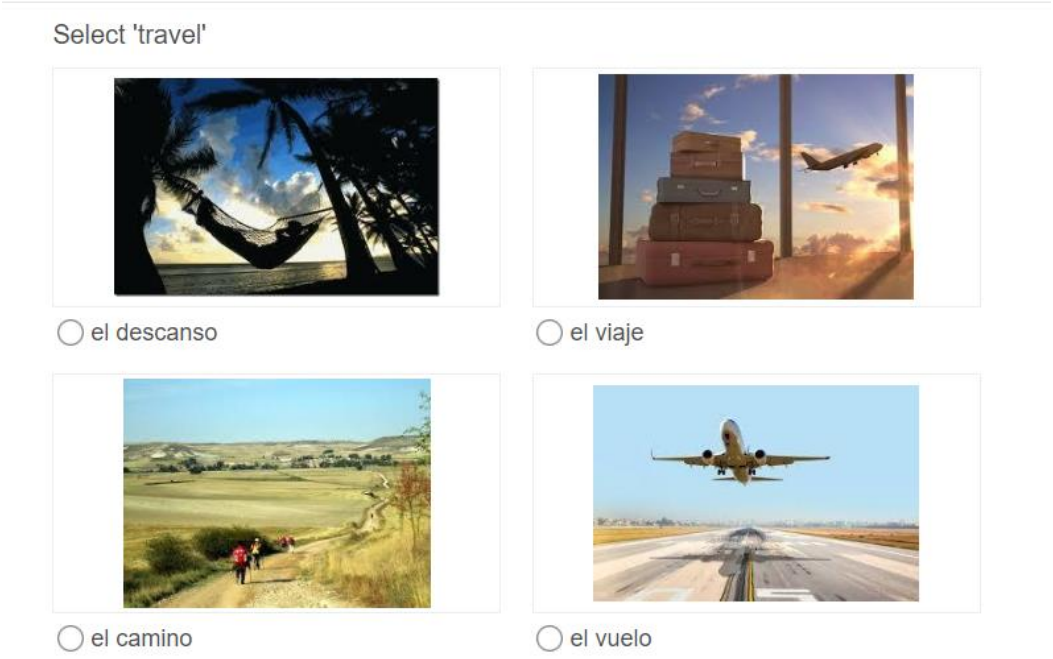
Description

The image for the target word is *representative* of the concept in that it is clear from the image alone what concept is depicted but *indistinct* in that distractor images are also representative of the concept since a highway is featured in each image. Each distractor image is also representative of its own concept, a road, a tunnel, and a bridge respectively, although not highly representative since other items (e.g. cars, road signs) are also featured in each image. There is no *formal similarity* between the L1 form and target L2 form, but there is formal similarity between one of the distractor items and its English equivalent.

Predictions

Participants will select the target ‘la autopista,’ at chance because neither the image or the L2 form is sufficient to indicate a clear match. It is also predicted that participants will attend more equally to written forms and images in an attempt to make a match since the images are ambiguous with reference to the target concept and the similarity between one of the translation options and its English equivalents will drive attention to forms. While the formal similarity occurs as a feature of one of the distractor items, this is not expected to have an impact on the selection made since the concept cued by the formally similar ‘túnel’ is not that of a highway.

Condition C1



Description

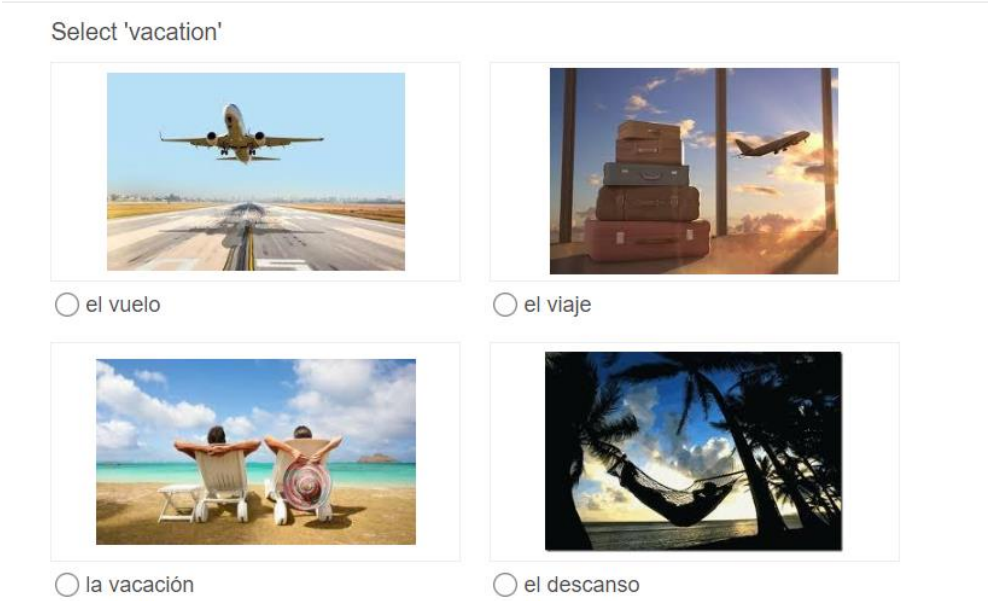
The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) and

indistinct in that distractor images could also potentially refer to travel. There is no *formal similarity* between the L1 form and target L2 form or between any of the distractors and their English equivalents.

Predictions

Participants will select the target ‘el viaje,’ at chance because neither the image or the L2 form is sufficient to indicate a clear match. It is predicted that participants will attend to the different images in an effort to determine which once best depicts the target concept. It is also predicted that participants will attend more equally to written forms and images in an attempt to make a match even though there is no formal similarity between L1 and L2 words to cue a relationship because, since the images are ambiguous with reference to the target concept, participants will need to look to written forms as a possible source of additional information.

Condition C2



Description

The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) and *indistinct* in that distractor images could also potentially refer to vacation. There is *formal similarity* between the L1 form and target L2 form, but no similarity between the L1 form and distractor L2 forms. There is no formal similarity between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target ‘la vacación’ because the similarity between the L1 and L2 forms indicates it as a possible match and there is no additional feedback provided by the images to indicate otherwise (i.e. the accompanying image does

depict what could be interpreted as a vacation). It is also predicted that participants will attend more equally to images and written forms because the images are ambiguous with reference to the target concept and the similarity between one of the translation options and its English equivalent will drive attention to forms.

Condition C3

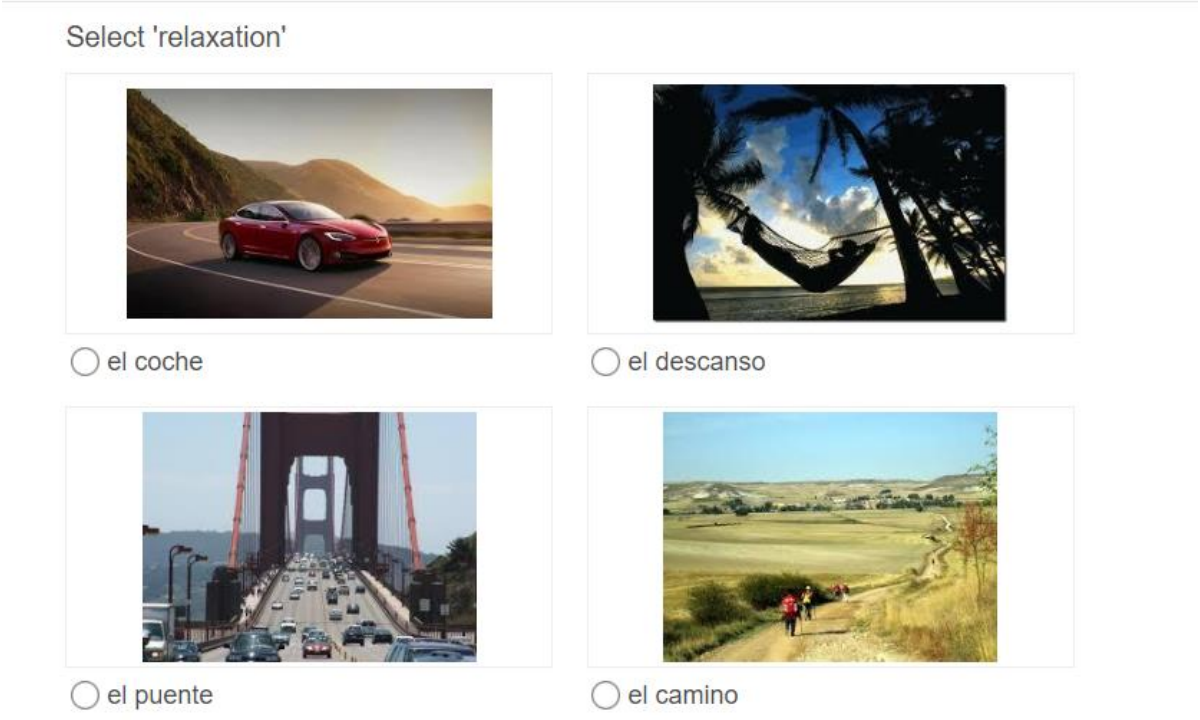
Description

The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) and *indistinct* in that distractor images could also potentially refer to a journey. There is no *formal similarity* between the L1 form and target L2 form, but there is formal similarity between one of the distractor items and its English equivalent.

Predictions

Participants will select the target ‘el camino,’ at chance because neither the image or the L2 form is sufficient to make a clear match. It is also predicted that participants will attend more equally to written forms and images in an attempt to make a match since the images are ambiguous with reference to the target concept and the similarity between one of the translation options and its English equivalent will drive attention to forms. Participants may tend to select the distractor item ‘la excursión’ due to its similarity to its English equivalent and the possibility of the corresponding image representing the concept of a journey.

Condition D1



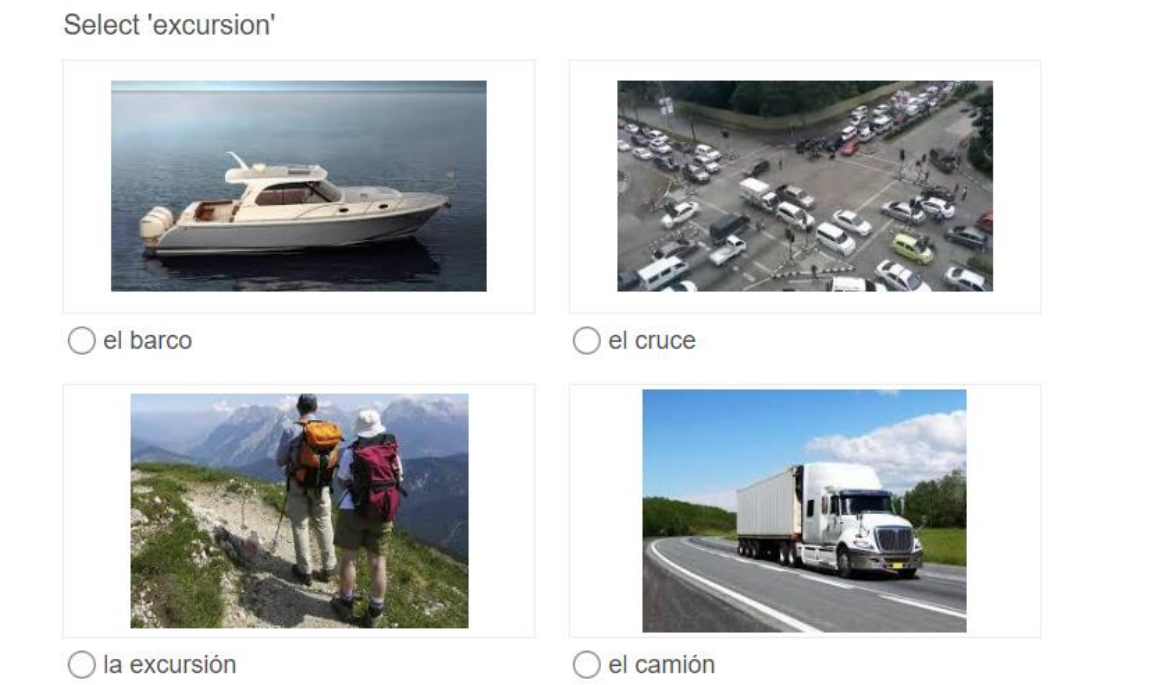
Description

The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) but *distinct* in that distractor images are not representative of the target concept. There is no *formal similarity* between the L1 form and target L2 form or between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target 'el descanso' because of the distinctiveness of the image as possibly representing the concept of 'relaxation.' It is also predicted that participants will attend more to the images than written forms since only one item is a possible representation of the concept and attention to forms is not helpful in making a match since there is no similarity between any of the L2 forms and their English equivalents.

Condition D2



Description

The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) but *distinct* in that distractor images are not representative of the target concept. There is *formal similarity* between the L1 form and target L2 form, but not with distractor forms. There is no formal similarity between any of the distractors and their English equivalents.

Predictions

Participants will tend to select the target 'la excursión' because of similarity in form to its English translation and the distinctiveness of the corresponding image as possibly representing the target concept. It is also predicted that participants will attend

more equally to images and forms because of the similarity between one of the translation options and its English equivalent.

Condition D3

Description

The image for the target word is *unrepresentative* of the concept in that it is not clear from the image alone what concept is depicted (it is an abstract concept) but *distinct* in that distractor images are not representative of the target concept. There is no *formal similarity* between the L1 form and target L2 form, but there is similarity between one of the distractor items and its English equivalent.

Predictions

Participants will tend to select the target ‘el vuelo’ because of the distinctiveness of the image as possibly representing the concept ‘flight.’ It is also predicted that participants will attend more equally to images and written forms because of the similarity between one of the translation options and its English equivalent. While the formal similarity occurs as a feature of one of the distractor items, this is not expected to have an impact on the selection made due to the high distinctiveness of the target image in representing the target concept and because the concept cued by the formally similar ‘el tren’ is not that of a flight.

Post-Test A

Participants will be asked to provide translations from English to Spanish and Spanish to English for the target vocabulary items in the experimental conditions. Participants will be presented with each item individually for a total of 36 items (18 in each language). The purpose of the post-test will be to assess participants' recall knowledge of the target L2 words following the experimental training. Post-Test A will be presented in the same way and with the same items as the Pre-Test.

Post-Test B

Participants will be asked to select the Spanish translation for the English words from the experimental conditions. Participants will see a total of 54 screens, 3 for each of the target items, which will vary in the positions of the options on the panel. The purpose of the post-test will be to assess participants' recognition knowledge of the target L2 words following the experimental training.

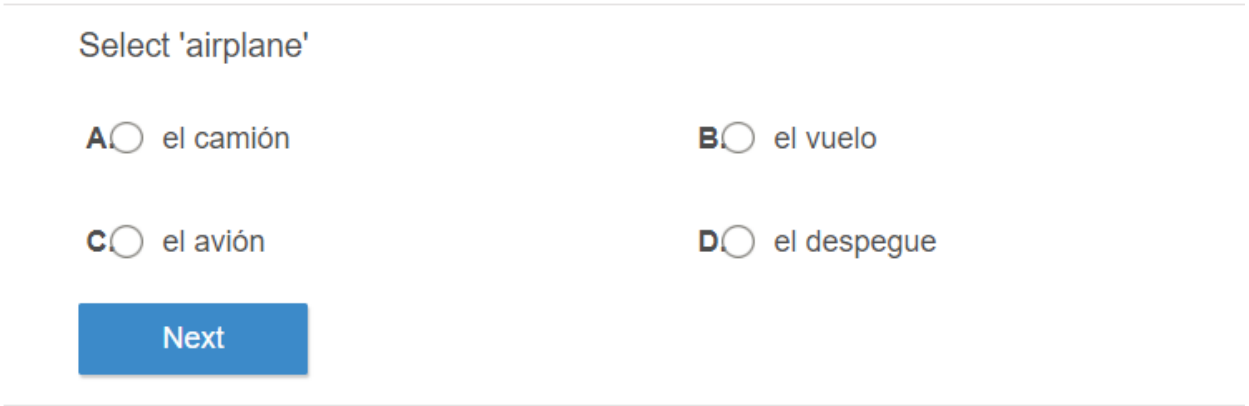


Fig 3. Example item from the translation matching post-test.