Word-to-meaning connection: a comparison between auditory and visual modalities in the first and second language

Essa Batel
University of Arizona

Abstract

The present study examined the link between words and their semantic representations in two modalities, namely, the visual modality vs. the auditory modality. It compared the reaction times (RTs) of native vs. advanced-level non-native speakers of English on a word-picture matching task. The goal was to examine in which of the two modalities an L2 word can more quickly activate the target semantic representation. The participants were presented with one word at a time, each followed by either a matching or mismatched image. The task consisted in using a key on the keyboard to respond as quickly as possible to indicate whether or not the presented image matched the previously presented image. Word stimuli were presented either in auditory format (using a headset) or visually (a string of letters on the screen). The time taken from the word recognition phase to the matching image represents the time that it takes the perceptual system to activate the sematic representation of a given word. The results show that the link between an L1 word and its semantic representation is not significantly different in these two modalities. However, there is a statistically significant difference between the two modalities when a word is presented in the L2 (RT after the auditory modality was slower). The observed result might be due to the difficulty in matching an auditory-presented L2 word to its L2 phonological representation. This factor would cause a slower activation of the related semantic representation compared to the visually-presented words.

Keywords: Native language (L1), second language (L2), semantic representation, reaction times (RTs), auditory modality, visual modality

Introduction

The present study is an attempt to compare the strength of the auditory vs. visual links to the semantic representation in L1 vs. L2. In other words, it investigates whether the strength of the link between the semantic representation and a language differs based on the type of modality (visual vs. auditory) used to present these words. In the field of L2 psycholinguistics, there is ample evidence for the fact that L2 picture naming latencies tend to take a significantly longer time than those of the L1 (Ivanova & Costa, 2008; Kroll, Bobb, & Wodniecka, 2006). It is worth mentioning here that the term "visual modality" refers to the recognition of written words while the term "auditory modality" refers to the recognition of spoken words.

Many factors influence such results: among them are word frequency, age of L2 acquisition, the level of L2 proficiency, and among others (for a review, see Nicol, 2001). Hanulová, Davidson and Indefrey (2011) discussed different related studies that employed several mechanisms to investigate the process of bilingual word production. Hanulová et al. (2011) concluded that these studies, in general, suggest that slower L2 word naming might be due to the low frequency of L2 words, older age of L2 acquisition, and/or competition from L1 words that share some similarities with the L2 words. That is, adult L2 learners are expected to be slower when it comes to L2 word naming compared to those who acquired this L2 at a younger age. Additionally, cross-language homographs (i.e., words that share the same spelling across two languages, but with different meanings in each language, for example: *coin* means "corner" in French) would lead to a naming delay of L2 words when presented within an L2 context. Thus, comparing L2 speakers in these two modalities would help in understanding how L2 processing is modulated by the medium of input, besides other factors presented in the literature. Thus, the modality used for L2 input would be an additional factor for the slower L2 processing, in addition to the fact that comprehending an input in a second language is more difficult than that of an L1.

Furthermore, low frequency words (e.g., *dale* in English) result in longer reaction times (RTs) in naming tasks compared to high frequency words (e.g., *car* in English); this applies to both L1 and L2 words. However, L2 speakers take longer in such tasks compared to native speakers. Some studies suggest that this delay in L2 word recognition is due to a weak link between an L2 word and its semantic representation (i.e., the stored meaning in the cognitive system) compared to the link between an L1 word and its semantic representation (see Grosjean & Li, 2012 for a review). The stronger meaning-to-L1 word link compared to the weaker meaning-to-L2 word link is a factor for slower recognition of L2 words.

Another factor that contributes to the link between L1-L2 and semantic representations is a word's degree of concreteness. This point is well explained by the Distributed Feature Model (De Groot, 1992) which makes a distinction between concrete words (e.g., *car*, *pencil*, *shoes*) and abstract words (e.g., *loyalty*, *beauty*, *friendship*) in terms of their semantic representation across languages. According to this model, compared to words of abstract meanings, concrete words in L1 and L2 share more similar semantic features as their referents, in most cases, and are similar across different languages and cultures. Put simply, concrete words are translated more accurately than abstract words. Thus, with regard to the shared semantic representation, there is no reason to think that a concrete L1 word (e.g., *car*) would have a semantic representation that is different from its correspondent L2 word. This would lead to a more accurate selection of referents (i.e., pictures in this study) for the presented L2 words, as well as for L1 words. Thus, the link between the concrete L2 words is expected to be stronger than that between the abstract L2 words, a point to be considered when testing lexical items in general and across languages in particular.

This paper, therefore, conducted an experiment that examines the L2-to-meaning link and whether it is of equal "connection" (i.e., compared to that of L1) in both visual and auditory modalities. In other words, do adult L2 learners respond to L2 words equally when they read them vs. when they hear them? A word-picture matching task was conducted to investigate this point, and some proposed interpretations are briefly discussed.

Literature Review

Although adult L2 learners are able to reach a very high proficient level in their L2 attainment, reaching a native-like level is still a challenge for the vast majority of L2 learners, even for advanced L2 speakers, who often show performance outside of the native range (e.g., Clahsen and Felser, 2006; Kroll and de Groot, 2005). Among these differences is the activation of semantic representations of a word in the L1 vs. the L2, and this is the reason why investigations of lexical and semantic representations of L2 words have been of great importance in the field of psycholinguistics for decades. For instance, Paradis (1978) proposed a model called the Three-Store Hypothesis in which he proposes that a bilingual speaker possesses two linguistic systems (one for each language) and a non-linguistic cognitive system where all semantic representations are stored and shared by these two separate linguistic systems.

With regards to the visual vs. auditory modalities, the Search Model (Forster, 1976) is one of the early models that made a distinction between the phonological (i.e. the auditory form of a word) and the orthographic representations (i.e. the visual form of a word) of lexical items. It proposes two separate sets of lexical representations: one for orthographic representation and another for phonological representation. Each one of the two representations is connected to "a master file" (Forster, 1976, p. 267) where information about the word's spelling, meaning, and part of speech is stored. Although this model does not confirm or deny an interaction between the two sets (i.e., the orthographic and the phonological representations), it suggests that each modality has a distinct trajectory to the semantic representation. Although the Search Model does not cover the difference between these two modalities in the bilingual speaker's cognitive system, it sheds light on the different trajectory of each modality. These two models (i.e., the Search Model and the Three-Store Hypothesis) are part of the early studies that made a distinction between a word's lexical representation and its semantic representation in the human cognitive system.

When it comes to the second language (L2) words, especially in early stages of L2 acquisition, a learner's first language (L1) plays a significant role in the acquisition process of the L2. For accessing the semantic representations of L2 words, it was thought that an L2 word's link to the semantic representation is mediated by the correspondent L1 word. That is, an L2 word activates its correspondent L1 word, which subsequently triggers the semantic representation. With higher L2 proficiency, L2 words create their own direct link to the semantic representation, independent of L1 words (for a discussion of the two mediation processes see Potter et al., 1984). This explanation (i.e., the dependence on correspondent L1 words at early L2 stage) has been debated by Kroll and Stewart (1994), who proposed the Revised Hierarchical Model (RHM) suggesting a different explanation. The question here is what accounts for the delay in the lexical recognition and the naming of L2 words compared to their correspondent L1 words by advanced-level l2 learners. The main point behind the RHM is whether an L1 word and its correspondent L2 word can equally activate the conceptual representation at the same pace. If not, what might be a possible explanation for this difference?

Kroll and Stewart (1994) compared the links to the conceptual representation between an L1 and an L2. Native vs. nonnative speaker participants were asked to identify images by name, one at a time, both with L1 and L2 words. Then participants were asked to translate words from L1 to L2 and vice versa. The results showed that picture naming with the L2 was slower than with the L1. In addition, translation from L2 to L1 was faster than translation from L1 to L2. The study concluded that the link between an image, or the conceptual representation, was stronger with the L1 than with the L2. Kroll and Stewart attributed the delay in activating the target L2 word after seeing the presented image to the weaker link between conceptual representation and L2.

Therefore, based on the RHM, Kroll and Stewart suggested the link between the conceptual representation and the target word (be it in the L1 or the L2) is a reason behind the slower L2 word retrieval. However, this research only addresses the concept-to-word direction using the picture naming task. That is, naming a picture explains the direction from the conceptual representation (i.e., an image) to the lexical representation (i.e., a word). Therefore, one key question this paper investigated is the word-to-concept direction to establish whether the activation of the conceptual representation is faster after recognition of an L1 word vs. an L2 word. If so, the question is whether the modality (visual vs. auditory) influences the access of the semantic representation in the L1 vs. the L2.

One reason behind this modality comparison is that late L2 speakers, like L1 speakers, need to establish new orthographic and phonological representations for L2 words. With an existing L1 word in place, an interaction between L1 and L2 properties is likely to occur. For this goal, the Bilingual Interactive Activation Plus (BIA+) model proposed by Dijkstra and Heuven (2002) argued for the interaction between phonological, orthographic, and semantic representations within a language and across the bilingual speaker's two languages. Thus, the activation of one of these representations in either language will subsequently lead to the activation of the other two representations within the same language and the activation of the same representations of the correspondent word in the other language. That is, activating the lexical representation of a word in an L2 will subsequently activate the phonological and the semantic representation of this word within the L2 and will also activate the lexical, the phonological, and the semantic representations of the correspondent L1 word. In addition, competition for selection between the L1 and the L2 words increases when both words have cross-linguistic phonological and/or orthographical similarities (e.g., cognates or homographs). For instance, when a Spanish native speaker whose L2 is English reads the word *transportation*, this will subsequently activate the Spanish correspondent word as well (i.e., *transporte*). This competition continues until either a lexical, phonological uniqueness point, or a contextual cue appears that determines to which of the two languages this word belongs. Therefore, this co-activation implies an interaction process between these two languages (for further details, see Dijkstra and Heuven, 2002).

After a word is recognized at the lexical level (i.e., the lexical form is identified to determine to which of the two languages it belongs), the search for the semantic representation of this word starts (i.e., the search for its meaning) (for a review, see Forster, 1976). Thus, the semantic representation, based on the RHM model and afore-mentioned models, is a store where all meanings of words, either from one language or from different languages, are saved. Therefore, these models clearly concluded that this semantic representation is a shared store between the two languages of a bilingual (see the RHM & the Three-Store models described above). Since there is a distinction between the lexical level and the semantic level, accessing the lexical representation (i.e., the stored form) of a word and accessing the semantic representation of the same word (i.e., its stored meaning) are two different steps occurring in a rapid, but sequential way.

The Bilingual Language Interaction Network for Comprehension of Speech (BLINCS) (Shook & Marian, 2013) model explains how spoken words are recognized across languages. This model includes four key representations of a word – phonological, phono-lexical, ortho-lexical, and semantic. Each of these language aspects is represented as a separate "self-organizing map" (Shook & Marian, 2013, p. 305). It suggests that each language is a separate "island" on the representation map, and words of similar sounds or meaning in both languages are close to one another on this representation map. Thus, this model suggests a bidirectional connection between similar words from both languages and thus a faster activation of the semantic representation of these similar cross-language words. It implies that, although the semantic representation is shared by all languages of a bilingual speaker, the organization of the semantic representation is actually affected by the similarities within the lexical representation (e.g., cognates). This could be the case when comparing words with cross-linguistic similarities to those words without cross-linguistic similarities. For instance, comparing the recognition of the English word *bicycle* and the Spanish word *bicicleta* vs. English word *key* and the Spanish word *llave*.

Among the cross-linguistic differences that may lead to the processing difficulty of L2 words is the orthographic system differences between the L1 and the L2. For instance, many languages around the world have different spelling characters or scripts from one another (e.g., Chinese vs. Arabic) while other languages, although distinct from one another, share similar alphabets (e.g., English and Italian). Yet, shared pronunciations of words or similar phonological features could exist across languages, even when these words have different meaning(s) in each language. For example, Hoshino and Kroll (2006) studied cross-language activation in bilinguals whose L1 (i.e., Japanese) does not share the same writing system with their L2 (i.e., English). The authors wanted to see whether there was an effect of cross-language phonological similarities between words of different orthographic systems. They found that the phonologically cognate words (objects that have a similar pronunciation in both L1 and L2) activate one another in both languages although these two words are written with different orthographic systems. For instance, a picture's name that is a phonologically cognate in both languages (e.g., シャツ "*sha.tsu*" in Japanese [meaning *shirt*] and *"shirt"* in English) is named faster in the L2 than a non-cognate word. It could, therefore, be inferred that, despite the different orthographic systems, phonologically cross-language cognates form a stronger link to the semantic representation compared to the words that do not have such similarities.

In addition to the interaction between the words of phonological similarities across languages, words presented in an auditory manner are also affected by many other factors that do not exist in the visually-presented word. These factors include the speaker's accent and speed of articulation, to name a few. The Event-Related Potentials (ERP) study (Hatzidaki, Baus & Costa, 2015) found that words in auditory modality are processed according to the way they are said. The participants in this study, who are Spanish monolinguals, listened to and read emotional words and negative-meaning words. Their results showed that emotional and negative-meaning words revealed greater amplitude than neutral words. What could be taken from this study is that there is a reason to speculate that more processing is required when a word is presented in an auditory manner than when presented visually; a written word does not usually give clues about a speaker's accent and emotions compared to words presented orally. In addition, the speed of articulation is a crucial factor in the recognition process of a word presented in an auditory manner. However, a reader may or may not take longer to move his/her eyes across the visually-presented word, leading to a faster (or possibly slower) recognition of the presented word.

For modality comparison in general (i.e., aside from the lexical stimuli), it was found that participants' RTs to auditory stimuli (e.g. a beep) were faster compared to the RTs to visual stimuli (e.g., a red circle) (See Jain et al., 2015 for a review). However, for the recognition of lexical stimuli in the L1 domain, Shelton and Kumar (2010) observed faster RTs for auditory stimuli compared with visual stimuli. On the other hand, Yagi et al. (1999) concluded that RT to auditory lexical stimuli is slower than visual to lexical stimuli.

It is worth noting that these fore-mentioned studies did not compare the semantic link of words in different modalities, yet they tested the form recognition in different modalities. Therefore, although these studies were conducted for different purposes than L1 vs. L2 comparison, their results revealed opposing results, which presents a clue that modality matters when it comes to word recognition in general.

However, when it comes to cross-language comparisons, Ibrahim's (2008) lexical decision study found that bilinguals' RTs to Hebrew words (L2) were faster when the words were presented visually. However, the reaction times to the Literary Arabic words (L1) were faster when they were presented orally. Although Ibrahim's (2008) study did not cover the link between a word and its semantic representation in these two modalities, it found a difference in the recognition process between the visual vs. the auditory modality in L1 vs. L2. With regards to word guessing in the L1 and the L2, the Bilingual Interactive Model of Lexical Access (BIMOLA) (Grosjean, 1998) studied the language perception phenomena by recruiting bilinguals who were presented with a portion of an L1 vs. an L2 word in the auditory modality. Participants were then asked to guess the target word. As expected, participants' performance in guessing the L1 target word was better than that of the target L2 word. Although guessing a target word from a presented part of it implies an activation of the semantic representation of this word, it is hard to distinguish between the guessing process time of the target form and the time course to the semantic representation. That is, it is challenging to establish whether the delay in naming the target L2 words comes from the difficulty of guessing the L2 form or the weak link between the L2 word and its semantic representation. Generally speaking, it can be inferred from the above literature review that the phono-lexical representation and the ortho-lexical representation of L1 words have stronger links to the semantic representation than those of L2 words. Thus, there is a valid reason here to compare these two modalities within each language and between L1 and L2 with regards to the link to the semantic representation.

The Present Study

The present study conducted an experiment to test word-meaning matching processes in both L1 and L2. In addition, a within-language comparison is performed to compare visually-presented and auditory-presented words with regards to the picture matching process. Therefore, this study's research questions are:

1. Is the link of a visually-presented L2 word to the semantic representation stronger than that of the auditory-presented L2 word, or are both links of equal strength?
2. Does L1 differ from L2 with regards to these different modalities?

That is, the two modalities were not compared in terms of whether they could accurately activate the correct semantic representation, yet the comparison is related to the time course taken by a participant to activate the correct semantic representation after a visually-presented word vs. an auditory-presented word.

Participants:

Two groups of participants were recruited in this experiment: the control and the experimental groups. The experimental group consisted of 18 adult L2-English learners. They were five females and 13 males, and all were students at the University of Arizona with ages ranging from 25 to 38 (mean age = 31.5) at the time of this experiment. They had lived in the U.S. from three to five years and had spent from three to five years (mean = 4 years) studying in English-medium classes. They all reported that they had no vision and/or hearing difficulties. Arabic is the native language of 15 participants, Farsi is the L1 of two participants and Taiwanese is the L1 of one participant. The control group also consisted of 18 adult participants whose native language is English. Twelve of them were males and six were females, and all of them were students at the University of Arizona at the time of this study. Their ages ranged from 23 to 33 years old (mean = 28) at the time of this experiment. They all reported that they had no vision and/or hearing difficulties.

Materials and Design:

The materials for this experiment consisted of 34 images of different household objects, fruits, vegetables, kitchen tools, and animals, as well as 34 English word stimuli (words length ranged from 3-8 letters of different frequencies, mean Freq_HAL = 20,891.441, mean_log_Freq_HAL = 8.775). Seventeen of these critical words were visually-presented (written on the screen) and the other 17 were auditory-presented (listened to through a headset). The experiment was carried out using DMDX, a computer-based software created by Forster and Forster (2003).

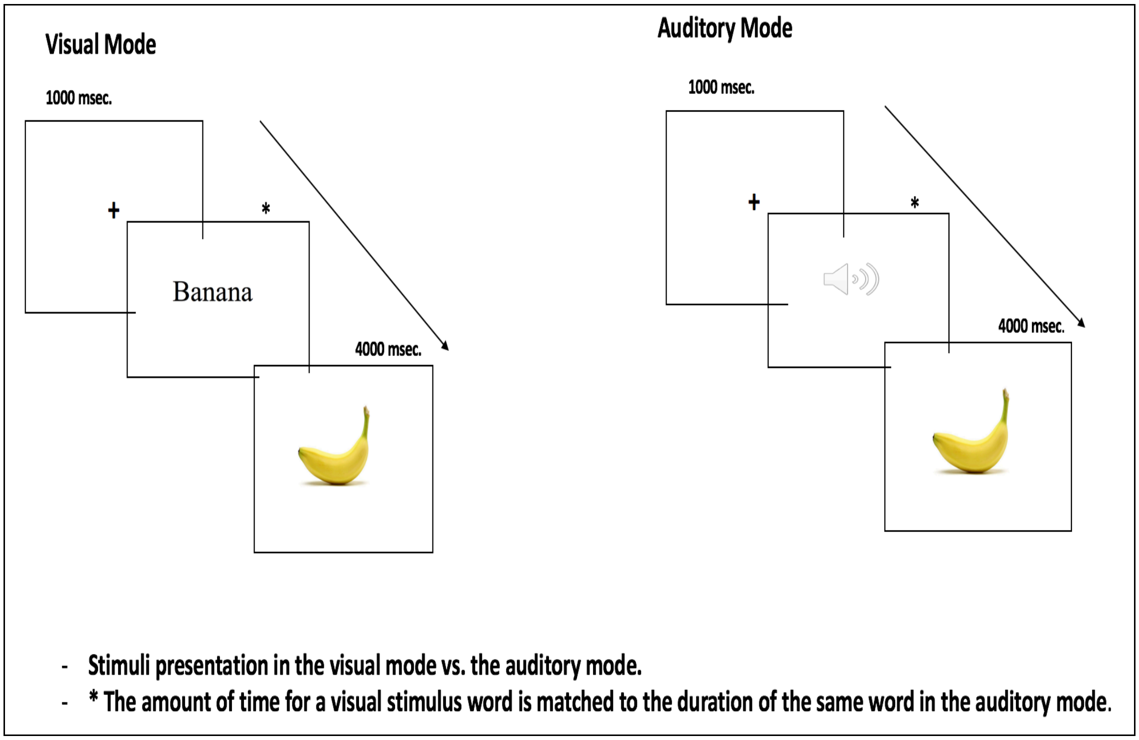
A female native speaker of English volunteered to record the auditory stimuli for this experiment. The recording of the auditory stimuli was done in a sound booth located in the Douglass Phonetics Lab at the University of Arizona. In addition to the 34 items, a practice trial that preceded the experiment consisted of three different items on each modality administrated right before the experiment trials. Finally, the dimension of each image was 5x5 cm and the words were presented in 30 point Times New Roman font. In addition, clear sound headphones were used for the auditory-presented words. In order to find out whether these pictures are easy to identify, 7 randomly-selected students at the University of Arizona (4 of them were native speakers of English and 3 of them were English L2 late learners) were asked to name each experiment picture in order to verify how easy or difficult the recognition of each picture was. Pictures that were not easy to identify were replaced by others.

Procedure:

After the experimental procedure was explained to the participants and signatures of consent were obtained from all participants, each was seated in front of the computer screen to start with the practice session that consisted of six items before the actual experimental items started. The procedure (see Fig. 1) went as follows for the visual block: the trial started with a black cross sign (+) that appeared in the center of a white screen as a fixation point which remained on the screen for 1000 msec. This was followed by a visual stimulus English word which was also centered on the screen. Right after the offset of the stimulus word, the word disappeared and was replaced by a picture at the center of the screen. The participant's job was to press the RIGHT ARROW key on the keyboard if the picture matched the previously presented word or the LEFT ARROW key if the picture mismatched that word. Each picture remained on the screen until the participant pressed the match or mismatch key or for 4000 msec before it timed out and a new trial started. The auditory block was presented with the same procedure of the visual block except that the participants listened to the stimuli instead of reading them.

Again, the picture appeared immediately after the offset of the stimulus word. These two blocks were counterbalanced; half of each language group started the experiment with the visual block and the other half started with the auditory block, and vice versa for the other half of the participant group. The clock started counting from the picture's onset time until a response was given or for 4 seconds if no response was provided. Participants' reaction times (RTs) were recorded as data of analysis.

Regarding the time each stimulus word (the visual vs. the auditory) remained on the screen, the duration of each visually-presented word was matched to the duration of the same word in the auditory modality. For example, the duration of the stimulus word *Tomato* is 557 msec (milliseconds) when presented in the auditory modality from the word onset time to the word offset time (i.e., it took the speaker 557 msec to pronounce the whole word). Therefore, when the same word was visually presented, it remained on the screen for 557 msec before it disappeared, and a picture was presented where the participant had to decide whether this picture matched or mismatched this stimulus word. It is worth mentioning that no single word was repeated for the same participant as the experiment stimuli were counterbalanced across modalities within each language group (for a review of the counterbalance design in research, see Gravetter and Wallnau, 2016, p. 368)



- Stimuli presentation in the visual mode vs. the auditory mode.
- * The amount of time for a visual stimulus word is matched to the duration of the same word in the auditory mode.

Figure 1: *Visual vs. Auditory stimuli presentation*

Analysis

The reaction times of the participants were analyzed using a two-factor mixed design ANOVA, with Language (L1 and L2) as a between-subjects factor and Modality (visual and auditory) as a within-subject factor. Using RStudio (RStudio Team, 2015) for analysis, the results showed that the main effect of Language was statistically significant (F (1, 34) = 8.528, $p < 0.01$), in that native speakers' reaction times, as expected, were faster than those of L2 speakers, but the main effect of Modality was not statistically significant (F (1, 34) = 1.418, $p > 0.05$) in that the difference between the mean of the visual block and the mean of the auditory block were not significantly different from one another statistically. In addition, the Language by Modality interaction was statistically significant, (F (1, 34) = 5.029, $p < 0.04$).

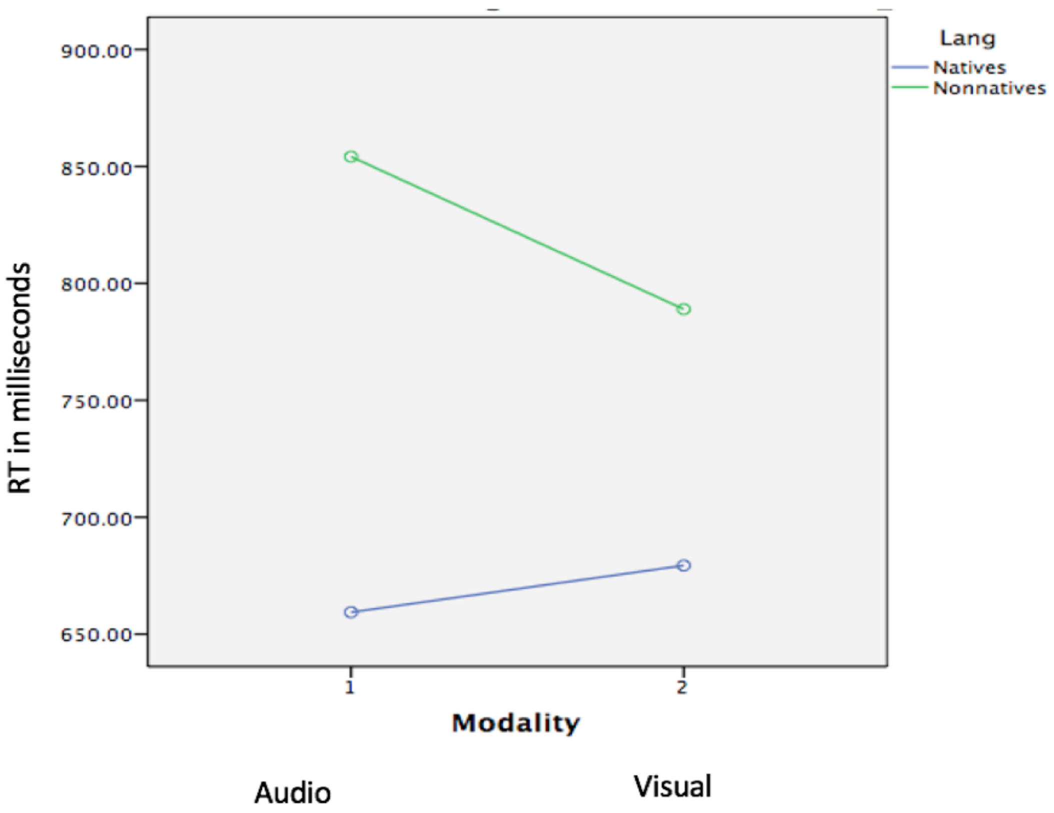


Figure 2: *Reaction times in L1 vs. L2 in milliseconds*

Because the main effect of the Language factor was statistically significant, the simple effect of Language was tested for the L1 and L2 separately with Bonferroni correction for two post hoc tests (i.e., alpha = 0.05/2 = 0.025) due to the possibility of an increased Type I error.

The analysis showed that L2 speakers' RTs on the auditory modality (mean RT = 854.2) were slower than those on the visual modality (mean RT = 789.0), and they are significantly different from one another (F (1,17) = 7.051, $p < 0.016$). On the other hand, native speakers' RTs on the auditory modality (mean RT = 659.4) were a little bit faster than those on the visual modality (mean RT = 679.3), yet they are not significantly different from one another (F (1, 17) = 0.475, p = 0.5).

Discussion and Conclusion

The above results showed that the semantic representations of L2 words were activated with statistically different paces in visual modality vs. auditory modality – an effect that was not observed in the L1. As expected, L1 speakers surpassed their L2 counterparts in the RTs in the word-picture matching task. This is an indication of a faster activation of the L1 semantic representation as a result of seeing or hearing an L1 word. In other words, advanced L2 adult speakers of English showed a significantly longer reaction time to match a target picture to an auditory presented word compared to time taken to match a target picture to a visually presented word. On the other hand, this difference was not observed between these two modalities when they were processed by L1 speakers. In line with RHM model (1994), that L2 words' link to the semantic representation is weaker than that link of L1 words; the present study observed the same results when it came to L1 vs. L2. However, the link of the visual vs. the auditory modalities to the semantic representation is of different strength in the L2, but not in the L1. Therefore, it is obvious that the phono-lexical representation of L2 words takes longer to activate the target semantic representation compared to the ortho-lexical representation. Therefore, if the figure of RHM model is borrowed to just explain the observed result in this paper, it would somehow look like Fig. 2 below.

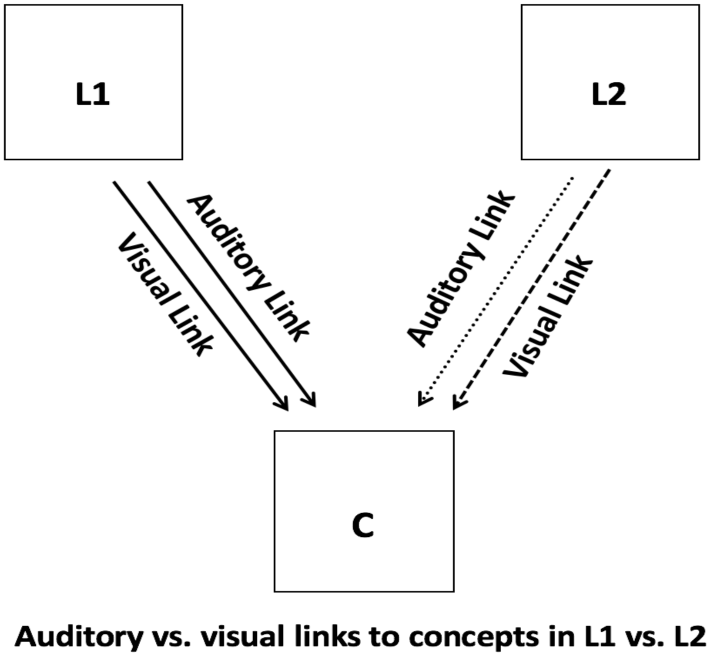


Figure 3: *Auditory vs. Visual links in L1 vs. L2*

One possible reason for the L2 visual modality vs. auditory modality being processed differently is that the L1 and the L2 of these L2 participants do not share cross-language similar orthographic or alphabet systems (i.e., English spelling system is different from Arabic spelling system). Thus, the Latin letters of English words (L2) is not going to be confused with an Arabic one due to the different orthographic (or spelling) systems. That is no reason to think that an Arabic native speaker would confuse a written English word with an Arabic word. However, this is not the case when a word is presented auditory. The lexical competition between an L2 word and a similar-sounding L1 word is greater when L2 words are presented in an auditory. This auditory competition across languages is supported by Schulpen et al. (2003), who found that Dutch listeners, who are learning English as an L2, not only activate the English meaning of word *leaf,* but also activate the Dutch meaning of a similar-sounding word *lief* (meaning 'sweet'). In other words, it is possible that the phonological sequence within some L2 words might be shared by some L1 words, which might result in a cross-language phonological co-activation – an interaction that does not likely to occur between the written L1 and L2 words at the initial processing of word recognition starts. This is in line with the previously mentioned study of Hoshino and Kroll (2006), who observed a cross-language co-activation of phonologically similar words despite the different spelling systems between English and Japanese (e.g., シャツ "*sha.tsu"* in Japanese [meaning *shirt*] and *"shirt"* in English). Another possible reason for this result is that spoken words, either in L1 or L2, are more likely to be pronounced differently by different people. These different pronunciations would be the cause of the delayed RTs in the auditory L2 words since L2 participants cannot store all possible pronunciations of each L2 word. Although these participants have been in the U.S. for a good number of years (mean = 4 years) and have been attending academic classes on a regular basis, their interaction with L2 textbooks is more frequent that their social interaction with native speakers. Thus, the difference in the amount of exposure to L2 words in each modality (listening vs. reading) would lead to the less familiarity of the L2 word in auditory form compared to that in the visual form.

Although with auditory word recognition in general, a listener may take a longer (or shorter) time to recognize a word due to the slow (or fast) speed of articulation of the speaker; it is obvious that this result cannot be attributed to the pace of articulation factor. It is mainly because each target picture is presented after the offset of its related stimulus word (in both the auditory and the visual blocks). However, if these results are due to the pace of articulation, this would have been observed for L1 speakers who showed no statistical difference between the two modalities. In addition, the results may not be attributed to the difficulty in recognizing the auditory-presented words (incorrect or “odd” pronunciation of words) since the results showed low error rates in both modalities across the two language groups (i.e., accepting a mismatch or rejecting a correct match between the stimuli and the picture).

It is worth mentioning here that one of the limitations of this study is that the majority of L2 participants were native Arabic speakers. A future study should include a more linguistically-diverse group of L2 speakers. Another limitation is that the native languages of these L2 participants do not share the same spelling systems as English. Although this difference between the L1 and L2 spelling system was done on purpose for this study, it would be interesting to do the same study with speakers of similar language spelling systems. The final limitation is that this study is more concerned with the psycholinguistic side rather than with the pedagogical aspects of the results.

Based on these results, one of the future psycholinguistic research suggestions is to test the effect of different accents of English on the auditory-presented words comparing simultaneous bilinguals and late L2 learners. Although it could be tempting to think that late bilinguals would be slower than simultaneous L2 bilinguals in terms of recognition times, simultaneous bilinguals have two existing languages in place, and both are almost equally dominant. Therefore, the cross-language co-activation could pose more difficulty for these simultaneous bilinguals; a hypothesis worth testing. In addition to that, a correlational study could be conducted between the learners’ L2 proficiency test scores and their RTs on the recognition of auditory-presented words vs. visually-presented words to see if there is a parallel relation between L2 proficiency and auditory modality of L2 input. Another suggested study is to use the above experiment design, but in the opposite stimuli-target direction. That is, presenting the picture first and then presenting the stimulus word next to test whether a similar pattern would be observed for L1 and L2. Finally, this experiment could be done with words of human voice and be compared to words generated by a text-to-speech software. This would give a clue of whether L2 words with natural human acoustic cues (e.g., pitch, frequency) are easier to recognize by L2 listeners than L2 words generated by a software voice.

REFERENCES

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing?. *Trends in cognitive sciences*, *10*(12), 564-570.

de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition*, *5*(03), 175-197.

Forster, K. I. (1976). Accessing the mental lexicon. *New approaches to language mechanisms*, *30*, 231-256.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior research methods, instruments, & computers*, *35*(1), 116-124.

Gravetter, F. J., & Wallnau, L. B. (2016). *Statistics for the behavioral sciences*. Cengage Learning.

Grosjean, F. (1998). Studying bilinguals: Methodological and semantic issues. *Bilingualism: Language and cognition*, *1*(02), 131-149.

Grosjean, F., & Li, P. (2012). *The psycholinguistics of bilingualism*. John Wiley & Sons.

Hanulová, J., Davidson, D. J., & Indefrey, P. (2011). Where does the delay in L2 picture naming come from? Psycholinguistic and neurocognitive evidence on second language word production. *Language and Cognitive Processes*, *26*(7), 902-934.

Hatzidaki, A., Baus, C., & Costa, A. (2015). The way you say it, the way I feel it: emotional word processing in accented speech. *Frontiers in psychology*, *6*.

Ibrahim, R. (2008). Does visual and auditory word perception have a language-selective input? Evidence from word processing in Semitic languages. *Linguistic Journal*, *3*(2), 82-103.

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production?. *Acta psychologica*, *127*(2), 277-288.

Jain, A., Bansal, R., Kumar, A., & Singh, K. D. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied and Basic Medical Research*, *5*(2), 124.

J Shelton, J., & Kumar, G. P. (2010). Comparison between auditory and visual simple reaction times. *Neuroscience and medicine*, *1*(1), 30.

Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, *9*(2), 119-135.

Kroll, J. F., & De Groot, A. M. (Eds.). (2009). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149.

Linck, J. A., Hoshino, N., & Kroll, J. F. (2008). Cross-language lexical processes and inhibitory control. *The mental lexicon*, *3*(3), 349-374.

Nicol, J. (Ed.) (2001). *One Mind, Two Languages: Bilingual Language Processing.* Blackwell Publishing

Paradis, M. (2004) A Neurolinguistic Theory of Bilingualism. Amsterdam: John Benjamins.

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and semantic representation in beginning and proficient bilinguals. *Journal of verbal learning and verbal behavior*, *23*(1), 23-38.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schulpen, B., Dijkstra, A., Schriefers, H., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 1155- 1178.

Shook, A., & Marian, V. (2013). The bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, *16*(02), 304-324.

Yagi, Y., Coburn, K. L., Estes, K. M., & Arruda, J. E. (1999). Effects of aerobic exercise and gender on visual and auditory P300, reaction time, and accuracy. *European journal of applied physiology and occupational physiology*, *80*(5), 402-408.