Chapter II

A detailed description of the different stages of the experiment is provided in this chapter beginning with a description of the population who participated in the experiment followed by a discussion describing the different stages of the elaboration of the stimuli and concluding with the report of the actual experiment.

2.1. Participants

Forty-six native Spanish speakers from Mexico, enrolled in a private Mexican university, between the ages of 18-25, participated in the experiment. At the time of the study, all the participants were registered in a low-intermediate English as a foreign language (EFL) course. In order to register in this particular course, students must obtain a minimum score of 460 on the TOEFL test.

Although 11 intact English language groups were involved during the pre-testing and the testing phases, participants from five of the eleven intact groups performed the actual experiment and earned extra academic credit for their participation.

2.2. Materials

2.2.1. Stimuli

The final stimuli set consisted of 176 Spanish-English translation pairs. There were 88 true Spanish-English translation pairs and 88 false Spanish-English translation pairs (see Table 1. for examples). The translation pairs were presented in four conditions, including 22 Spanish-English translation pairs per condition. Condition one, the concrete multiple condition (CM), included concrete Spanish-English words which had

multiple translation equivalents in English. Condition two, the abstract multiple condition (AM), included abstract Spanish-English words which had multiple English translations. The multiple translation equivalents in English were either synonyms or homonyms. Condition three, the concrete single condition (CS), included concrete Spanish-English translation pairs. A single translation equivalent for the given Spanish word was identified in English. Condition four, the abstract single condition (AS), included abstract Spanish-English translation pairs. A single dominant equivalent for the given Spanish word was available in English. Appendix A lists the true translation pairs per condition. The false translation set consisted of 44 concrete and 44 abstract Spanish-English translation word pairs. The false translations served as distracters and were not taken into account in the data analysis. The number of translations for these was not evaluated because they were nonsense translation equivalents created for the purpose of the experiment.

Condition	Definition	Spanish	English Translation Equivalent One	English Translation Equivalent Two
Condition One - Concrete Single (CS)	One translation available in the English Language	*labio*	*lip*	*X*
Condition Two - Concrete Multiple (CM)	Two or more translations available in the English Language	*piedra*	*rock*	*stone*
Condition Three - Abstract Single (AS)	One translation available in the English Language	*creencia*	*belief*	*X*
Condition Four - Abstract Multiple (AM)	Two or more translations available in the English Language	*alma*	*soul*	*spirit*

Table 1. A sample list of the stimuli, per condition.

For the translation recognition task, participants needed receptive knowledge of the stimuli in order to determine whether the English words were true translation equivalents of the Spanish words (i.e. *mar – sea*) or were false translation equivalents of the Spanish words (ie. *cocina – lamp*). Receptive knowledge is the ability to recognize a word and its translation equivalent without necessarily being able to produce a written or oral translation equivalent (Nation, 2001). Moreover, since ambiguity effects were under investigation, it was important that participants have receptive knowledge of multiple possible translation equivalents for stimuli in Condition one (CM) and Condition two (AM).

The next section describes the stages involved in the creation of the stimuli, starting with the initial translation production task to the final experiment. These satges are summarized in *Figure 14*.

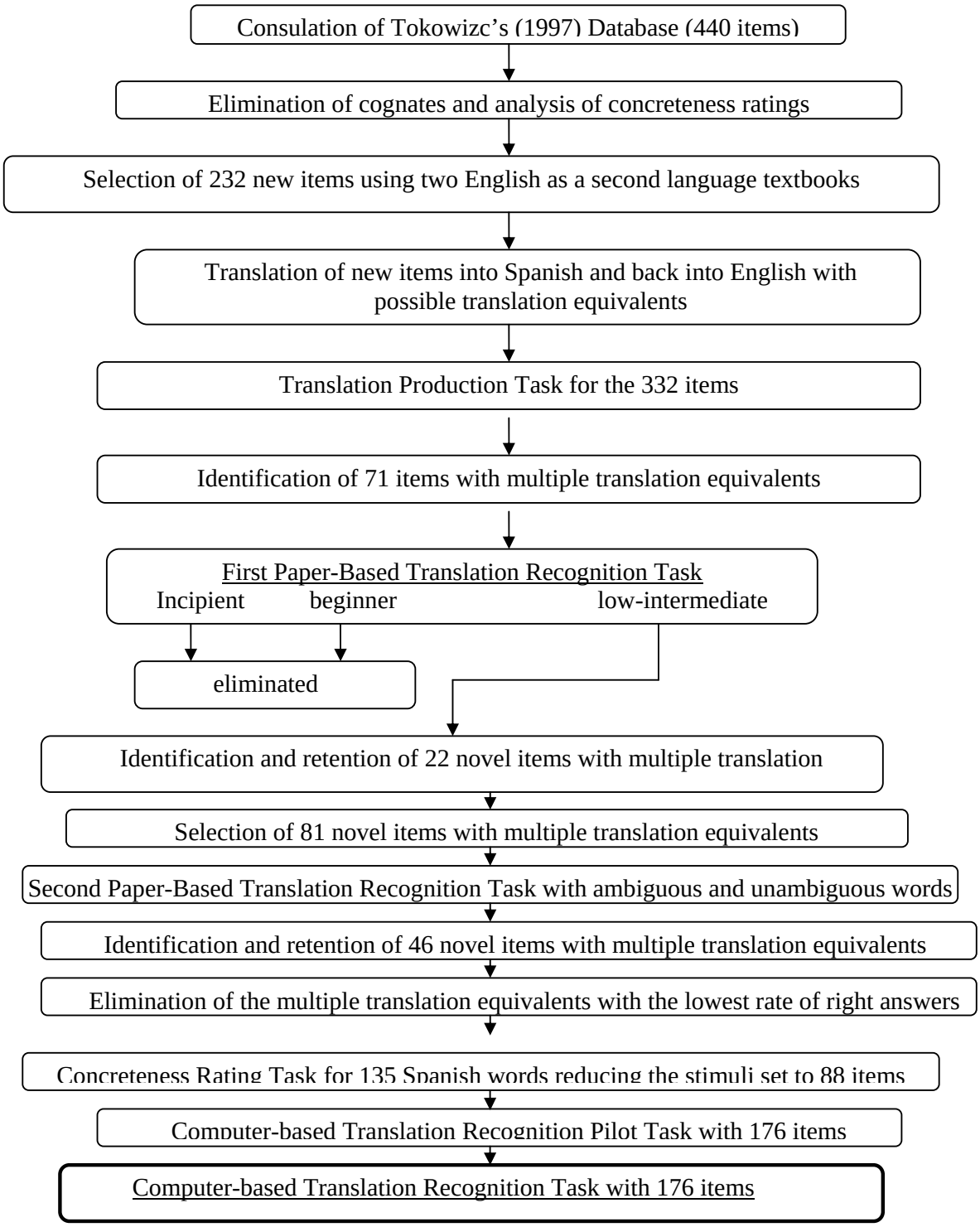


Figure 14. Map of the processes involved in the creation of the stimuli.

In order to create the true translation pairs, a database consisting of 440 items, previously used by Tokowicz (personal communication, November, 2005) in a translation production experiment, was consulted. This database included Spanish-English translation equivalents and number of translation equivalents in each language. Concreteness, context availability, imagery and word length norms were also available. A careful analysis of the items led to the identification of some limitations. A considerable number of words were cognates. Cognates are words that share phonological and/or orthographic form between two languages and that may also be semantically related (Hall, 2002). Given that this study was not interested in typology effects, no cognates were included. The second limitation surfaced after closely analyzing the concreteness ratings. It was not made clear whether the concreteness ratings had been obtained for the English words, the Spanish words or both. Although translation equivalents are believed to be linked to a shared conceptual representation in the bilingual mental lexicon, concreteness ratings can differ. This can primarily be observed with homonyms which are two different concepts that link up to a single word form in one language (Hall, 2005). Let us return to the previously mentioned example *manzana.* In Spanish, the word *manzana* is the translation equivalent of the English word *apple*. Here, both the L1 and the L2 forms share the same syntactic category and link-up to the same semantic features at the conceptual level. Nonetheless, in Spanish, the word *manzana* can also mean a *street block*, a more abstract concept. If concreteness ratings were obtained from a native Spanish speaker for the word form *manzana*, both concepts could have been evoked, affecting the concreteness rating. Finally, cultural experience can affect how word forms are perceived and linked to a conceptual representation. For instance, the concreteness norms from Tokowicz study included the word *market* as abstract. From a North American perspective, this can be

understood if one assumes that this word form is linked to the abstract notion of *financial market.* In Mexico, *mercados* are still very present (in a physical sense) in the society and are not considered abstract concepts.

In light of these arguments, the database was highly scrutinized and consequently reduced to 100 items. It was imperative to expand the list since the receptive knowledge of these items and their concreteness rating had yet to be evaluated with the target population. The English as a second language textbooks Northstar (Haugnes and Maher, 1998) and Skyline (Skyline, 2001), presently used at the private Mexican university where the experiment was conducted, were consulted. Two hundred and thirty two new items were identified. These items were translated into Spanish by the researcher (a balanced English-Spanish bilingual with knowledge of the regional and national cultural backgrounds of the participants) and preliminary translation equivalents were confirmed by consulting the Spanish-English Collins Dictionary (2000) and confirmed by a balanced native Spanish-English bilingual.

Since the lexical knowledge of the participants was tested in the translation recognition task, it was important to consider multiple translation equivalents that reflected the participants' English vocabulary knowledge. The initial translation equivalent norms were obtained by conducting a translation production task adapted from Schönpflug's (1997) first translation method, with 23 native Spanish participants from an intact English class with a slightly more advanced level. It was hypothesized that their productive knowledge would resemble the receptive knowledge of the target population. Each participant was given a list of 166 words and was asked to provide the translation equivalent, a task which lasted approximately 10 minutes. Five distinct lists were created (randomized using the random function in Excel), to eliminate any ordering effects. Participants received written and oral instructions in Spanish and were

instructed to write the first English translation equivalent that came to mind for each Spanish word (see Appendix B for the instructions). If a word evoked more than one possible translation, they were asked to provide it.

The results were compiled manually in an Excel worksheet. Each different translation equivalent and occurrence of a given response was recorded. Spelling mistakes were ignored. All words with multiple translations were included in the stimuli list that would later be piloted in a paper-based translation recognition task with the target population. This led to the identification of 71 new items.

The following stages involved a series of paper-based translation recognition tests designed to test the participants' knowledge of the multiple translation words. Participants were given a list of translation pairs and asked to identify true and false translation pairs by writing an *S* (for true) and *N* (for false) (see Appendix C for instructions). Individual participants only saw one of the two possible translation pairs in order to avoid any priming or familiarity effects. For example, List A included the Spanish word *cuadro* and the true English translation equivalent *square* and List B included the true English translation equivalent *painting* (for the word *cuadro*). Moreover, each list included an equal number of false translation pairs that were both concrete words (*piel – tool*) and both abstract words (*bostezo – people*). Ten distinct randomized lists were created to eliminate any ordering effects.

The tests were performed by incipient, beginner and low-intermediate groups in order to identify participants with knowledge of multiple translations. The yes/no answers for each participant were manually recorded into an Excel worksheet. To score the true translation pairs, correct answers received a score of one and wrong answers received a score of zero. For each true translation pair, the correct answer percentage was calculated, and words that had a response rate of 70% and over were kept. This

stage was crucial for a successful timed translation recognition task and confirmed that incipient and beginner learners of English could not perform a test investigating the effects of lexical ambiguity since only one of the possible translation equivalents was known by a high percentage of the participants. After analyzing the responses, the low-intermediate group was chosen for the experiment. The results from 37 participants from the four intact low-intermediate English groups were retained, producing a total of 22 word pairs, namely 11 abstract and 11 concrete Spanish-English pairs, with multiple translation equivalents in English.

These findings called for a second pre-testing phase with low-intermediate participants to expand the stimuli set. In order to generate a new stimuli set with multiple translation equivalents, 81 English words were taken from the British National Corpus (Leech, Rayson, and Wilson, 2001). Because high-frequency words are often learned in the earlier stages of vocabulary development, it was hypothesized that the low-intermediate participants would already have knowledge of translation equivalents and possibly have learned some second translation equivalents. Words from the entire corpus with a minimum frequency of 75 per million words could be included. In order to obtain English translation equivalents, the researcher first provided a Spanish equivalent for the English words and translated these back into English while providing multiple translations (when possible).

The translation production task stage was eliminated at this point because the multiple translation word pairs obtained during the production task were often not recognized in the translation recognition task. The knowledge discrepancy of ambiguous English words between low-intermediate and advanced learners of English was noticeable. The second, paper-based translation recognition task was performed by 38 participants from two intact low-intermediate classes. This list further integrated

possible single translation equivalents that had been established during the translation production phase. Here, 6 distinct randomized lists were created, using the random function in Excel, to eliminate any ordering effects.

The responses were tabulated using the same method as the first translation recognition task. Again, Spanish words with single or multiple translation equivalents, that had a mean correct response rate between 70% and 100%, were accepted. Forty-six new items with multiple translations were identified, creating a word bank of 68 items with multiple translations.

For the multiple translation pairs, only one of the possible translation equivalents was included in the final stimuli set. Therefore, the translation equivalent with the lowest correct answer rate was eliminated. If results from the experiment demonstrate that ambiguous words are recognized slower than unambiguous words, keeping the dominant translation equivalent would provide stronger evidence for ambiguity effects. If the translation pairs with the lowest correct answer rate were kept and an ambiguity effect would surface, it would remain unclear whether this could be attributed to the weaker connections between the lexical and the conceptual level. Figure 15 presents the distribution of the percentage of correct answers for the translation pairs with multiple translation equivalents that were included in the study. Sixty-four percent of the abstract translation pairs and 61 percent of the concrete translation pairs were known to all the participants.

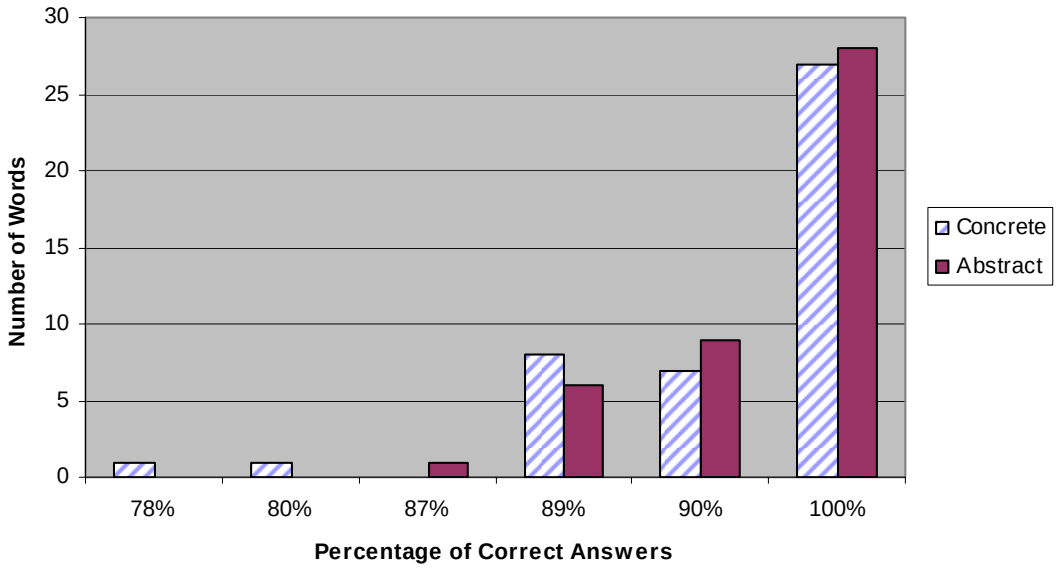


Figure 15. Distribution of the percentage of correct answers for paper-based translation recognition task for the 88 translation pairs separated by concreteness.

2.2.2. Concreteness Norms

In order to create a reliable stimuli set, concreteness ratings were obtained for the Spanish words from the true translation pairs. Twelve other randomly selected Spanish native speakers provided concreteness ratings for 135 Spanish words, using a 1 to 7 point Likert scale, 7 being the most concrete and 1 the most abstract (Paivio, Yuille and Madigan, 1968). Participants were given oral and written instructions in Spanish to ensure that they had the same interpretation of the terms *concrete* and *abstract* (described in Appendix D). For this experiment, a concrete term was defined as a word for a tangible concept that could be experienced by our senses. An abstract concept is one that cannot be perceived by one of our senses. Participants were reminded to use the entire scale providing the first rating that came to mind. They were given three

sample words, to get them to think about the subtle differences between concrete and abstract words. Again, five distinct randomized lists were created, using the random function in Excel, to eliminate any ordering effects. Concreteness ratings were manually recorded in an Excel worksheet. For each item, the sum was calculated and divided by the number of participants. The results from one participant were eliminated because the instructions were misinterpreted and the grading scale was inverted. Although the majority of the abstract words obtained a rating of four or below, one item, whose rating was slightly above the 4.0 cut-off, namely 4.2, was included in the stimuli set. Similarly, a word whose rating fell slightly below 5 points was included in the concrete category, with a rating of 4.83 (see Appendix E for a complete listing). Figure 16 presents the distribution of the concreteness ratings and the number of words included under each range.

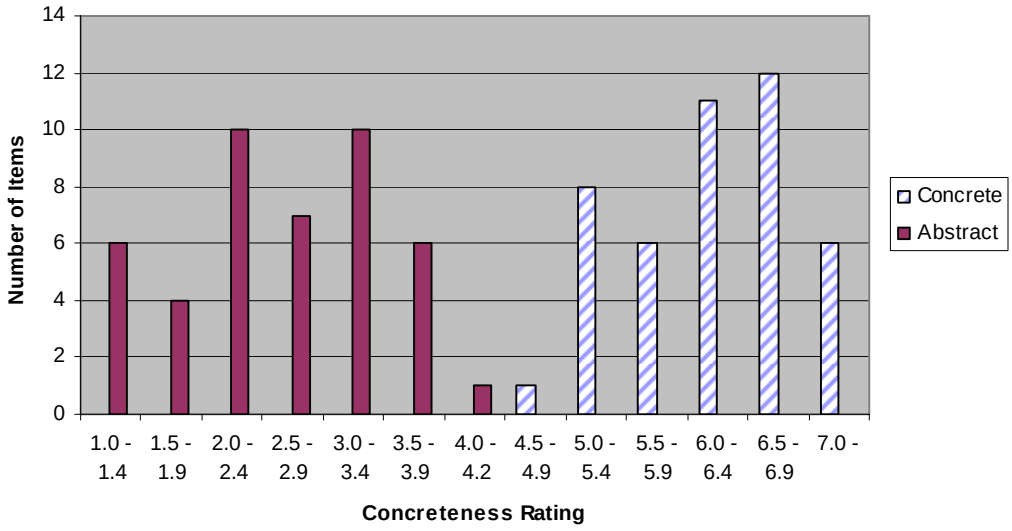


Figure 16. Comparison of concreteness rating distribution for abstract and concrete words.

The distribution shows how concepts are seldom perceived as purely concrete or purely abstract, since words are distributed across the scale of 1-7. Moreover, the concrete word distribution tends to be denser in the upper scale (between 6 and 7), and the abstract words are more equally distributed across the scale. Again, this can be explained by the greater perceptual salience for concrete words.

Another variable that was controlled was the number of letters for both the Spanish word and the target translation equivalent. The majority of the words had between 4 and 6 letters. Figure 17 illustrates the distribution of letters across the words for the two languages. A higher number of longer Spanish words was expected since Spanish is a highly inflectional language. In order to reduce any possible effects related to word length, the number of letters was balanced across false translation pairs which produced a mirror effect of this distribution.

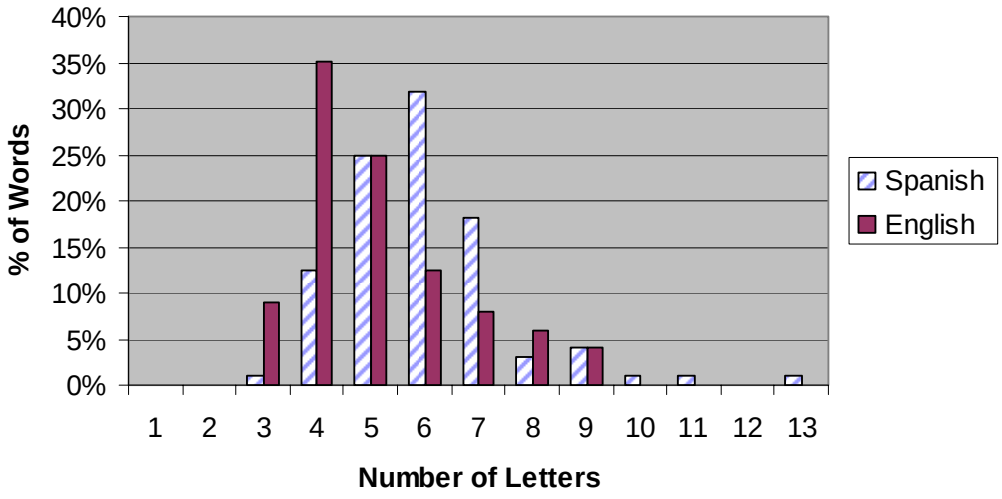


Figure 17. Number of letters for the true translation pairs.

Originally, only concrete and abstract nouns would make up the stimuli set, yet because a word and its translation equivalent do not always share the same syntactic category, the final stimulus set included both words that were either nouns, verbs or adjectives.

2.2.3. Pre-Test: Piloting the Software

A pilot session with the final stimuli set was conducted in the psycholinguistic laboratory at the Universidad de las Américas, Puebla, on 9 Dell PCs, using DMDX software. The clarity of instructions, the length of the task and the speed of appearance and disappearance of individual items on the screen was evaluated. Nine undergraduate students, registered in a psycholinguistics course, performed two versions of the task. For group A, after 88 items, a prompt appeared on the screen informing them that they had reached the half-way mark. They could choose to break and resume with the experiment at their convenience. For group B, such a prompt was not presented. Feedback, obtained from a questionnaire (see Appendix F and G), showed that continuous running was preferred by participants from group A and group B confirmed that such a break was no necessary. Given that a few participants would have preferred a longer practice trial, extra practice translation pairs were incorporated. The final modifications were made to the software instructions and the final stimuli list was embedded in the DMDX software specifications.

2.3. Procedure

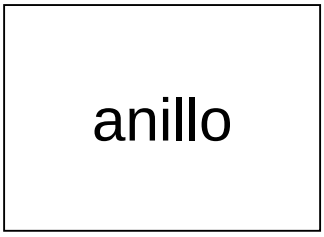
2.3.1. DMDX Software

The experiment was run on 4 Dell PCs in the psycholinguistics laboratory at the Universidad de las Américas, Puebla. The stimuli were presented on computer screens

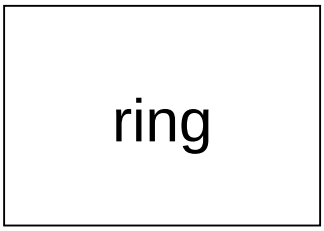
using DMDX, which permitted the recording of reaction times per participant and per item across the four conditions. The successful recording of the reaction times is guaranteed by first running the TimeDX software (see Appendix H for a complete description of the TimeDX).

The words appeared in Arial Black font, size 24 on a white background as illustrated in Figure 18. Data presentation in a translation recognition task can follow various timing sequences. Under an SOA of 0 milliseconds, words in both languages are presented at the same time. Under an SOA condition greater than 0 milliseconds, words are presented at two different times. Based on previous work (de Groot, 1992a), it was decided that the Spanish word would appear on the screen first for 472 milliseconds and then disappear to be immediately followed by the English translation for 708 milliseconds. The next item followed after 2810 milliseconds[1]. This timing sequence was selected to provide participants with enough time to read and recognize the word in their L1, but not have enough time to think of a translation. Each sequence lasted 4000 msec. If no answer was provided, the next item would automatically appear on the screen. An SOA 0 condition was not chosen because it would be impossible to discriminate between the time needed to read and process the word in the L1 and the time needed by the participant to recognize the translation in the L2.

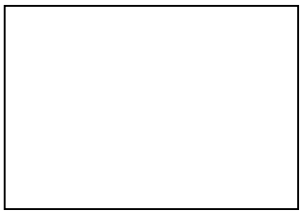
[1] These numbers are not round numbers since DMDX requires that the SOA be measured in a unit labeled ticks, where one tick equals 11.80 milliseconds.



Screen 1 appearing for 472 milliseconds and disappearing



Screen 2 appearing for 708 milliseconds and disappearing



Screen 3 appearing for 2810 milliseconds and disappearing

Figure 18. The sequence of three screens that each participant saw for the 176 experimental items.

Each session began with a greeting written in Spanish. In order to begin the experiment, participants had to press the wheel of the mouse. This was specified at the bottom of the screen. The next slide explained that they were participating in an activity designed to help the Language Department gain insight into students' knowledge of English vocabulary. It specified that they would first see a word written in Spanish, followed by a word written in English. Their task was to decide, as quickly as possible,

whether the word pairs were true or false translation equivalents. Next, they were told to keep their finger on the wheel of the mouse at all times, between answers. They were told that when a true translation pair appeared on the screen, they had to click on the right button of the mouse, marked by a green sticker. When a false translation appeared on the screen, they had to click on left button of the mouse, marked by a red sticker. At this point, they were informed that they would begin with a practice trial and were reminded that they had to provide an answer as quickly as possible. After the practice trial, a message appeared on the screen informing them that the researcher would enter the room shortly to provide any clarifications. Next, participants were informed that they could begin the experiment by pressing the wheel of the mouse. Once they had viewed the 176 items, a final message appeared on the screen asking them to remain silent until all the participants had concluded the task and the researcher entered the room.

Each student was received in the control room in the psycholinguistic laboratory at the Universidad de las Américas, Puebla. They were invited to go into the experiment room and choose one of the four computers. They were also asked to turn off their cellular phones. They were then orally informed that they would be taking part in an experiment conducted by the Language Department in which they were asked to determine, as quickly as possible, whether word pairs were true or false translation pairs by pressing the right button of the mouse for a positive response and the left button of the mouse for a negative response. At this point, they were informed that they would receive written instructions, followed by a practice trial, in order to familiarize themselves with the task and the hardware. From the control room, the researcher could identify when all participants had completed the practice trial. She then entered the experimental room in order to clarify any questions or concerns they might have. After

clarifying any doubts, they were asked to press the wheel of the mouse for the next written instructions and were reminded to keep quiet throughout the entire experiment. Once all participants completed the task, the researcher entered and informed them that they had concluded the experiment.

In the last stage of the experiment, participants were asked to fill out a language questionnaire adapted from Tokowicz (2000), which was designed to measure individual language learning experiences (see Appendix I). Although participants from intact English classes were invited to participate in the study to ensure homogeneity across proficiency level, the results showed that participants had very distinct experiences learning English. The mean number of years learning English was 10 years. Previous studies considered their participants to be advanced or fluent after such a high number years of exposure which could lead one to believe that their proficiency level was underestimated. Yet, it is important to note that more than 50% of the participants had some experience living in an English speaking country. Nonetheless, only 60% of these participants obtained formal instruction classes. Moreover, these participants are not required to use English for educational purposes. Because the TOEFL test requirement for this class is 460, it was concluded that the participants did not have a high proficiency level at the time of the study (see Appendix J for a discussion of the results). The claim that they have a low-intermediate level of English is sustained.

This concluded the experimental task. Participants were then thanked for their interest and for their voluntary participation. The next chapter presents the results obtained from the experiment.