Alleviating the translation-ambiguity disadvantage: Using a placeholder to signal an upcoming translation

by

Gabriela Terrazas Duarte

BA, University of Texas at El Paso, 2014

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Gabriela Terrazas Duarte

It was defended on

July 27, 2018

and approved by

Natasha Tokowicz, Associate Professor, Psychology Department

Tessa Warren, Associate Professor, Psychology Department

Marc Coutanche, Assistant Professor, Psychology Department

Thesis Advisor: Natasha Tokowicz, Associate Professor, Psychology Department

Alleviating the translation-ambiguity disadvantage: Using a placeholder to signal an upcoming translation

Gabriela Terrazas Duarte, MS

University of Pittsburgh, 2018

Translation-ambiguous words are words with two or more translations across languages. These words are remembered less accurately and at slower rates than words with only one translation (so-called translation-unambiguous words; e.g., Eddington & Tokowicz, 2013; Tokowicz & Kroll, 2007). Previous research has investigated different training methods that could reduce the translation-ambiguity disadvantage. Degani et al. (2014) found that presenting multiple translations in the same session improves retention of translation ambiguous words compared to training translations in different sessions. The current study explores the effects of informing second language learners that a word has multiple translations, and that the second one will be presented later in vocabulary training. We predicted that the use of a placeholder will produce similar accuracy results as training both translations in the same session, however, this is only observed when participants' individual differences are considered.

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Acquisition of vocabulary is a vital part of the process of language acquisition and is particularly important for adult second language (L2) learners. One difficulty L2 learners need to overcome is the acquisition of translation-ambiguous words, which are words with two or more translations across languages. For example, the German word “Kiefer” has two translations in English, “jaw” and “pine”. Such translation-ambiguous words are remembered less accurately and more slowly than words with only one translation (so-called translation-unambiguous words; e.g., Eddington & Tokowicz, 2013; Tokowicz & Kroll, 2007; see Tokowicz, 2014, for a review). Our main goal is to expand the existing research on translation ambiguity by investigating if the translation-ambiguity disadvantage can be alleviated by notifying learners that a word has multiple translations when the first translation is instructed. If, indeed, providing this information to learners can alleviate the translation-ambiguity disadvantage it will inform us about the flexibility of consolidation of new knowledge; particularly it will expand our understanding of the process involved in the consolidation of mental representations for novel L2 vocabulary in adult L2 learners.

Previous research has investigated different training methods to help reduce this translation-ambiguity disadvantage. Degani, Tseng, and Tokowicz (2014) found that L2 learners can benefit from learning early that a word is translation ambiguous, and that a one-to-many mental representation needs to be created for the word, instead of creating a one-to-one mapping representing the word as translation unambiguous. The latter is illustrated by the Revised

Hierarchical Model of Translation Ambiguity (RHM-TA; Eddington & Tokowicz, 2013), which demonstrates how translation-ambiguous and unambiguous words may be represented in memory during the early stages of L2 acquisition (see Figure 1). According to the RHM-TA, when a novel L2 word is introduced, a mental representation is created connecting the L2 word to its L1 translation. If the learner does not know that a word is translation ambiguous, that word will be represented as if it were translation unambiguous (see Figure 2), which will require the learner to later attempt to revise the representation.

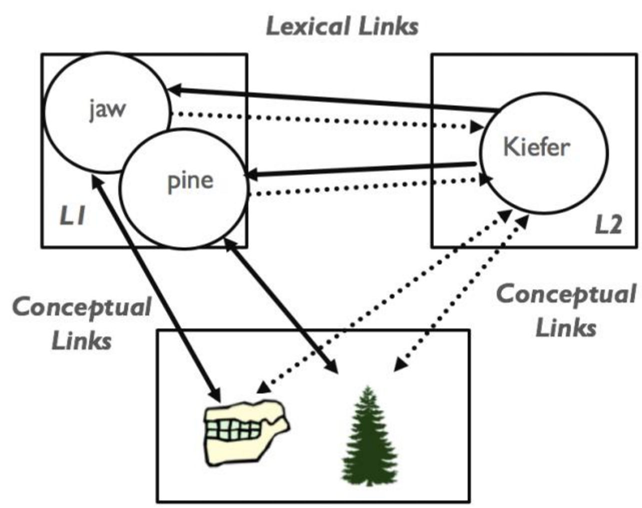

Figure 1. The Revised Hierarchical Model of Translation Ambiguity

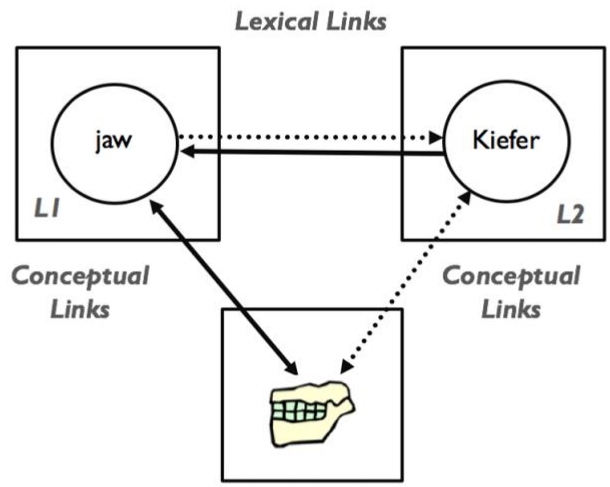

Figure 2. Illustration of a one-to-one representation in terms of the RHM-TA for a word that is actually translation ambiguous

Degani et al. (2014) suggest that presenting participants with both translations consecutively allows them to represent the word as translation ambiguous before solidifying the representation of a translation-ambiguous word as translation unambiguous. On the other hand, not informing participants that a word is translation ambiguous and training the translations in separate sessions strengthens a one-to-one mapping representation favoring that over the appropriate one-to-many mapping. Such a one-to-one mapping for the translation-ambiguous word must then be revised later when a second translation is introduced, which leads to a decrease in accuracy for translation-ambiguous words. Results presented by Degani et al. (2014) also indicate that translations learned first are learned more accurately than translations learned second.

Unfortunately, teaching all translations at once cannot easily be implemented in versions of traditional language instruction in which vocabulary is introduced in sets of themes or situations. Using our previous example, if a language instructor is teaching a class about body parts, only “jaw” would be relevant, and thus “pine” would not typically be introduced. These instructional constraints were a motivation for the current study to investigate new training methods that could be implemented by L2 instructors. Furthermore, the current study has the potential to teach us about the representation of translation-ambiguous words.

Specifically, the proposed study explores the effects of informing L2 learners that a word has multiple translations during their first exposure to the word, and that their second translation will be presented later in a different session during vocabulary training. We predicted that providing this information to participants would allow them to create a "placeholder" representation for the translation to be introduced later (see Figure 3). We hypothesized that using a placeholder would produce similar results as training multiple translations in the same session. It is possible that creating a placeholder representation will prevent learners from forming an inappropriate one-to-one mapping representation for translation-ambiguous words, by holding a place for a representation that cannot be introduced at the moment but will be presented later.

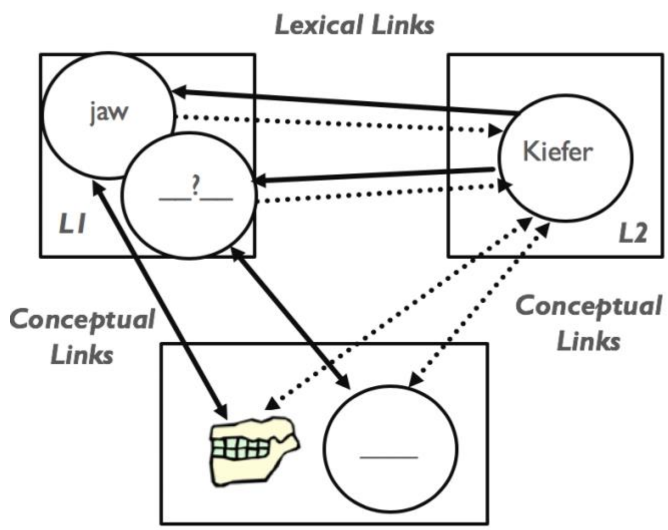

Figure 3. Placeholder in terms of the RHM-TA

Mechanistically, if the placeholder condition leads to higher accuracy than the separate condition, it will indicate that the metalinguistic information provided was used to prevent forming an inappropriate one-to-one mapping for translation-ambiguous words. On the other hand, if the metalinguistic information provided by the placeholder manipulation does not lead to higher accuracy, it will inform us that the information provided was not sufficient to prevent a representation of a translation-ambiguous word as translation-unambiguous. Not finding the placeholder manipulation useful in our current study might also indicate that our manipulation was not strong enough to accomplish our goal.

To investigate the effects of the placeholder manipulation a “fake placeholder” manipulation was included in the current study, which consisted of informing participants that a second translation will be introduced at a later time but not introducing a second translation. We used the fake placeholder condition to investigate if introducing a second translation had an effect on the first translation. The fake placeholder condition was used to compare the effect of this training manipulation to The effects of the fake placeholder condition were compared to performance for the placeholder condition as well as performance for a “single” training manipulation, which consisted of translation-unambiguous words introduced in the second session. We predicted that providing metalinguistic information using the fake placeholder manipulation to aid the acquisition of translation-ambiguous words would produce similar results to training translation-unambiguous words trained under the single condition.

In addition to our vocabulary training manipulation, we included some individual difference tasks to assess participants’ cognitive skills so that we could examine whether our manipulations would similarly affect individuals of varying cognitive skills. These tasks were also used as fillers between vocabulary training and testing. In addition, such individual difference tasks served as distractors between the short-term retention test and training, which was particularly important in Session 2.

We used the *Operation Span* (O-Span) task (Tuner & Engle, 1989), which measures working memory skills, as well as the Stroop task (Stroop, 1935) which measures the ability to ignore task-irrelevant information. Michael, Tokowicz, Degani, and Smith (2011) found that L2 learners with higher WM span and a lower Stroop effect had higher translation accuracy than individuals with higher WM span and a higher Stroop effect. This is, better performance in O-Span, or higher WM span, without the ability to ignore task irrelevant information, or higher Stroop interference, was associated with intermediate L2 learners' ability to translate translation-ambiguous and unambiguous words. Michael et al. (2011) also found that individuals with lower WM spans showed similar performance regardless of their ability to ignore task-irrelevant information. Therefore, we used this working memory task, O-Span, along with the Stroop task, to assess participants' ability to learn L2 vocabulary. Based on the aforementioned results we expected to see higher accuracy rates overall for individuals with higher WM span and lower Stroop effect, compared to individuals with lower WM and a higher Stroop effect.

In addition to the O-Span and Stroop tasks, we also used the AX-CPT task (Rosvold, Mirsky, Sarason, Bransome, Edwin, & Beck, 1956) to assess participants' ability to ignore task-irrelevant information. The AX-CPT task is designed to measure the use of proactive and reactive control mechanisms associated with the ability to update and maintain relevant information (proactive), as well as the ability to disregard information irrelevant to the task at hand (reactive). We predict that early L2 learners who rely more on proactive control, that is, learners with the ability to maintain and update information to achieve the goal at hand (Braver, Gray, & Burgess, 2007) will be better skilled to integrate the second translation into their mental representation of translation-ambiguous words in memory. This is, we expect that participants who rely the most on maintaining and updating information, will be able to use the metalinguistic information provided by the placeholder condition, by avoiding the consolidation of a one-to-one mapping of translation-ambiguous words leaving some flexibility for a new translation to be integrated into the representation.

Lastly, we included the Peabody Picture Vocabulary Test IV (PPVT-IV) which is used to measure individuals' English vocabulary knowledge (Dunn, Dunn, & Pearson Assessments, 2007). Participants' performance in this task correlates with their ability to produce the correct English translation of the corresponding L2 word (Koch, 2015). Therefore, we expect that participants with a larger English vocabulary will perform better than participants with a smaller English vocabulary. It consists of the presentation of 4 pictures, one of which corresponds to the meaning of a word read aloud by a researcher; participants are asked to point to the picture that correctly describes the meaning of the target word.

2.0 Method

2.1 **Participants**

Our sample consisted of 41 individuals who identified as English monolinguals, and who had no previous exposure to German, or to Dutch because it is a highly related language. Participants were recruited from the University of Pittsburgh and received credit toward an Introduction to Psychology requirement. Data from 16 participants were not included in the final analyses due to zero percent accuracy in one or more training manipulations ($N = 5$) computer errors ($N = 4$), missed sessions ($N = 3$), or researcher errors ($N = 4$). Therefore, analyses were performed on a final set of 25 participants (14 female; $M = 18.8$ years).

2.2 **Materials**

Stimuli consisted of 48 German words with two English translations per word, and 16 German words with only one English translation. Stimulus characteristics were matched on their psycholinguistic properties (word-length, frequency, concreteness, Translation Semantic Variability[1]). See Table 1 for summarized characteristics of stimuli.

---

[1] Translation Semantic Variability is a measure of semantic similarity between the different English translations of translation-ambiguous words, ranging from 1-7, ratings closer to 7 indicate pairs are highly related (Bracken et al., in press).

Table 1. Stimulus characteristics

Measure	Translation 1	Translation 2
Word Length (# letters)	6.30 (1.93)	6.29 (2.04)
Concreteness	244.08 (233.0)	256.02 (218.16)
Word Frequency	30.15 (37.75)	25.91 (37.29)
TSV (German words)	4.95 (1.29)	N/A

Note. Mean (Standard Deviation)

2.3 **Procedure**

Our study consisted of three sessions, completed over the course of a week, separated by one day between sessions (i.e., Monday, Wednesday, and Friday or Tuesday, Thursday, and Saturday). Participants were trained on the German-English word pairs during the first two sessions; at the end of each vocabulary training session participants completed a learning reinforcement test to enhance learning (e.g., Karpicke & Roediger, 2008; Pyc & Rawson, 2010). Sessions 2 and 3 started with a short-term retention test to assess learning of the previously-trained translations.

During the vocabulary training phase, vocabulary was trained under four different conditions: separate (training each translation in different sessions; Degani et al., 2014), placeholder, fake placeholder, and single. See Table 2 for examples of the conditions. For both the placeholder and fake placeholder training methods, a line was added indicating that a second translation was going to be presented later, whereas for the separate vocabulary training method there was no indication of another translation coming, and for the single condition we presented

German words with only one translation (i.e., translation unambiguous), during the second training session. In addition, there were no presentation differences between the placeholder and fake placeholder conditions during Session 1, however in Session 2, the placeholder condition introduced a second English translation, whereas for the fake placeholder condition, no second translation was introduced. On the other hand, for Session 2, we introduced the single condition to serve as a baseline.

Table 2. Examples of training manipulations

Condition	Session 1	Session 2
Separate	German = English 1	German = English 2
Placeholder	German = English 1 German = _______	German = English 2
Fake Placeholder	German = English 1 German = _______	N/A
Single	N/A	German = English 1

Translation-ambiguous word pairs were divided into six lists consisting of eight German-English word pairs each. Each translation-ambiguous word pair was counterbalanced across training methods such that some participants saw the first translation on Session 1 whereas other participants saw the second translation on Session 1, and vice versa. For an overview of the task order by session see Table 3.

Table 3. Task order by session

Session 1	Session 2	Session 3
Vocabulary Training 1	Short-Term Retention Test 1	Short-Term Retention Test 2
Distractor Task	Stroop	AX-CPT
Learning Reinforcement Test	PPVT	LHQ
O-Span	Vocabulary Training 2	
	Distractor Task	
	Learning Reinforcement Test	

2.3.1 Vocabulary training 1

Session 1 started with vocabulary training 1, which consisted of the presentation of 48 German words and their English translations. German-English word pairs (e.g., Kiefer = pine) were presented to the participant in a new random order during each cycle, and vocabulary training concluded after completion of three cycles of 48 trials. Each German-English word pair was presented for 8 seconds at the center of the computer screen, followed by a fixation cross, and participants then cued the next trial by pressing the "5" key[2].

[2] The procedure used followed Lotto and De Groot (1998) and Degani et al. (2014).

2.3.2 Vocabulary training 2

At the end of Session 2, participants were presented with vocabulary training 2, which consisted of training participants on the second English translations in the separate and placeholder conditions (e.g., Kiefer = jaw), as well as the introduction of the single condition German-English pairs. It is important to mention that the only differences between this second training and the first one, was that this second vocabulary training did not train the second English translation for “fake placeholder” trials, and this time “placeholder” trials looked like “separate” trials. These trials were all intermixed in random order.

2.3.3 Distractor task

After completion of vocabulary training 1 and 2 a brief distractor task was administered (Karpicke & Roediger, 2008). This task consisted of instructing participants to indicate whether an equation presented on the computer screen was correct or incorrect. According to the testing effect literature, learning is enhanced by adding a delay between studying and testing (Bahrick, 1979), therefore in the current study we used the distractor task as a buffer between vocabulary training and a learning reinforcement test to enhance participants’ performance (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013).

2.3.4 Learning reinforcement test

After completion of the distractor phase, participants completed the learning reinforcement test, which instructed participants to retrieve and type the correct English translation of the German word presented on the center of the screen. Each German word was presented one at a time and remained on the screen until the participant typed its corresponding translation, or on the other hand, if the participant was not able to recall the translation they were able to press "ENTER" to see the next German word.

2.3.5 Short-term retention test 1

Session 2 started with a short-term retention test, which we used to assess learning of previously trained translation-ambiguous word pairs. The short-term retention test consisted of presentation of each German word previously studied in random order, one at a time, and remained in the center of the screen until participants verbally provided a response. Participants were instructed to verbally produce the English translation or to respond "I don't know" if they didn't remember. Each German word remained on the screen until a response was registered.

2.3.6 Short-term retention test 2

Session 3 started with the second, and final short-term retention test. Just as in the short-term retention test 1 administered in Session 2, participants were asked to verbally produce the English translation for the German word presented on the screen. At this point in the study, participants had learned two translations for most of the German words, and each German word should have activated both translations learned. Therefore, participants were informed that if a German word had two translations it would be presented twice consecutively, to give them a chance to provide one English translation each time the German word was presented. This test was used to assess the effects of our different training methods and overall learning.

2.3.7 Individual Difference Measures

Operation-word span task (O-Span). After completion of the first learning reinforcement test we administered the operation-word span task, which is designed to assess working memory. The task requires participants to simultaneously solve a simple mathematical equation and to store an English word for later recall. Participants were instructed to indicate, by pressing a “YES” or “NO” button on the response box, if a mathematical equation (e.g., 9 / 3 + 1 = 8) was correct or incorrect, each equation was shown in the center of the computer screen for 2500 ms. Following each mathematical equation an English word was presented at the center of the computer screen for 1250 ms. Mathematical equations and English words were presented in sets varying in size from 2 to 6, with three sets of each size. At the end of each set a “RECALL” prompt appeared indicating participants to type all the English words they remembered from that set, in the order in which they were presented. We calculated Working Memory (WM) span by using participants’ set size score, which was measured as the set size at which participants recalled all the words in at least two of the three sets (e.g., Tokowicz et al., 2004).

Stroop task. After completion of the short-term retention test 1 in Session 2, participants completed the Stroop task, which is designed to measure participants’ ability to ignore task-irrelevant information. Participants were asked to name the color of the ink of each letter string presented in the center of the screen. Three different types of trials were presented to the participants: (1) congruent trials, in which the ink color matched the color word (e.g., the word “PURPLE” in purple ink); (2) incongruent trials, in which the ink was different from the color word (e.g., the word “YELLOW” in purple ink); and, neutral trials in which there was no match or mismatch possible between letter strings and ink color (e.g., the letter string “XXX” in purple ink). Participants had to overcome the interference produced by the incongruent trials, performance is facilitated by the congruent trials, and a baseline was obtained from performance on the neutral trials. Participants' ability to ignore task-irrelevant information was calculated by subtracting the average of congruent and neutral RTs from incongruent RTs and dividing this score by each participant's overall average RT. This inhibition score was calculated using only RTs from correct trials. Thus, lower Stroop scores reflect individuals' better ability to ignore task-irrelevant information.

Continuous performance test (AX-CPT). After completion of the short-term retention test 2 in Session 3, participants were asked to complete a measure of non-verbal cognitive control, the AX-CPT task (Ophir, Nass, & Wagner, 2009). The AX-CPT paradigm consisted of a cue-probe presentation in which participants were required to provide a probe-response for each trial based on a particular combination of a cue (first letter) and a probe (last letter). Each of those trials consisted of 5 letters, presented one at a time, starting with a cue, followed by 3 distractor letters and ending with a probe; each letter was presented for up to 300 ms (see Figure 4 for an illustration of the paradigm).

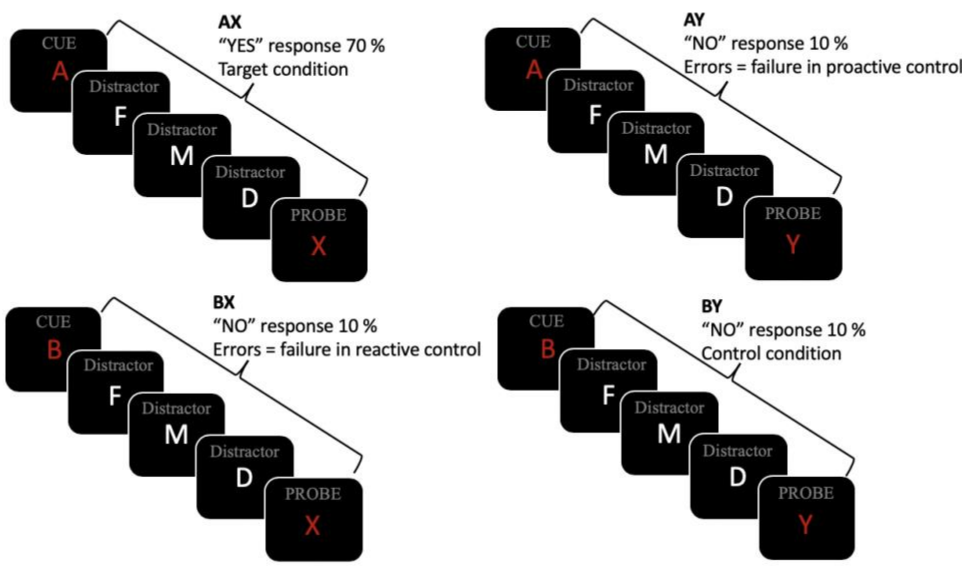

Figure 4. Type of trials for the AX-CPT task

Participants were required to provide a “YES” response only when the cue-probe trial was “A-X”. Participants were asked to press the “NO” key for all other stimuli. This is, if the cue-probe trial was “B-X”, participants were required to provide a “NO” response; the same was true for an “A-Y” cue-probe trial. Dependent measures derived from this task were response time and error proportion, therefore participants were asked to respond as quickly and as accurately as possible.

Through the AX-CPT task we can measure two types of cognitive control: proactive and reactive. Proactive control refers to the participants’ ability to monitor and maintain active goal-relevant information, failure in the activation of proactive control produces more errors in the “A-Y” cue-probe trials. On the other hand, reactive control refers to the ability to suppress and prevent interfering responses; failure to activate reactive control produces more errors for the “B-X” cue-probe trials. In general, successful completion of the AX-CPT task requires an adequate adjustment of both measures of cognitive control. The AX-CPT task was analyzed by computing the Behavior Shift Index (BSI[3]) for errors. BSI was calculated as (AY-BX)/(AY+BX), resulting in scores ranging from -1 to 1, where more positive values reflected higher proactive control engagement, also thought of as goal maintenance.

Peabody Picture Vocabulary Test IV (PPVT-IV). After completion of the AX-CPT task, participants were asked to complete the Peabody Picture Vocabulary Test IV (PPVT-IV). This test was used to assess participants’ English vocabulary. For this test, the researcher read a list of English vocabulary words, one at a time. Participants were then instructed to select the picture, out of four different choices, that best described the meaning of the English word the researcher had

[3] Behavior Shift Index, also known as Proactive Behavioral Index (PBI) is a composite measure of individuals’ control style in the AX-CPT task (Braver, Paxton, Locke, & Barch, 2009).

read for them. PPVT-IV standard score was computed as participants' raw score (ceiling score – number of errors) by age.

3.0 Results

Only reaction times from correct trials were included in the reaction time analyses; responses triggered prior to 100 ms were removed as voice key failures, and those longer than 10,000 ms were excluded from the analyses based on the reaction time distribution. Mean accuracy by training manipulations and test is shown in Table 4.

Table 4. Means and SDs for translation number by training manipulation and tests

Test	Training Manipulation	Translation #1	Translation #2	Mean
Test 1	Fake Placeholder	.47 (.50)	N/A	.47 (.50)
	Placeholder	.45 (.50)		.45 (.50)
	Separate	.45 (.50)		.45 (.50)
	Single		N/A	
Test 2	Fake Placeholder		N/A	
	Placeholder	.54 (.50)	.59 (.49)	.57 (.50)
	Separate	.52 (.50)	.62 (.49)	.57 (.50)
	Single	.50 (.50)	N/A	.50 (.50)

3.1 Linear mixed effects models

We used RStudio Software (RStudio Team, 2016), as well as the lmer and glmer commands of the lme4 package (Bates, Maechler, & Bolker, 2012) and the lmerTest package (Kuznetsova, Brokhoff, & Christensen, 2013), to perform four different linear mixed effects analyses. To investigate if our participants in fact benefited from the placeholder training manipulation by improving their accuracy rates on items trained under this manipulation, we compared participants' performance for each training manipulation. Specifically, our first set of models investigated if our training manipulations (placeholder vs. separate) influenced participants' accuracy (for first vs second translation learned), and if an effect was found, whether such effect was qualified by our participants' performance in our individual difference tasks. The second set of models explored the same question but in terms of decreased reaction time rather than increased accuracy, although generally effects for beginning learners are more likely to be observed in accuracy (e.g., Degani et al., 2014). The third set of models investigated if our participants benefited from the placeholder training manipulation specifically as a function of test (first vs. second), and if an effect was found, whether such effect was qualified by our participants' performance in our individual difference tasks.

We also used linear mixed effects models to investigate if there was a difference between translation-ambiguous words, trained using our fake placeholder manipulation, and translation-unambiguous words, trained under the single translation manipulation; and if our training manipulations similarly affected individuals of varying cognitive skills. Finally, the last set of models look into the effects of our placeholder training manipulations (placeholder vs fake placeholder) and in order to make sure these training manipulations showed similar effect and such effects similarly affected individuals of varying cognitive skills. Because each of our individual difference measures tap into a different cognitive skill, results for all of the models will be reported below. Interactions were visually represented using the estimated means taken from each of our regression equations. Extreme observed values for each individual difference measure were used as minimum and maximum values, and the observed mean was used as the mid value in the figures presented below, based on the procedure suggested by Aiken and West (1991).

3.1.1 Models for Short-term retention test 2

O-Span. Test 2 accuracy was significantly higher for the second translation than the first translation learned, $\beta = .42$, $z = 7.87$, $p < .001$. A main effect of WM span was found, $\beta = .44$, $z = 2.13$, $p < .05$, such that higher span was associated with higher accuracy. Training manipulation interacted with translation number, $\beta = -.24$, $z = -9.24$, $p < .001$, such that the second translation trained was remembered more accurately than the first translation trained, however this effect was greater for the separate training condition (see Figure 5).

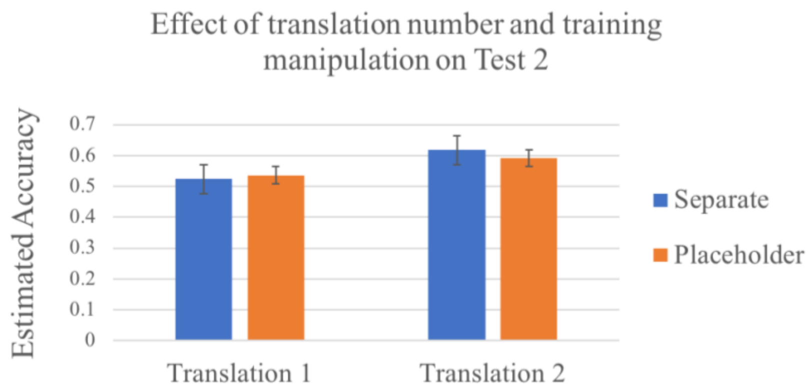

Figure 5. Effect of translation number and training manipulation on Test 2

Also, translation number interacted with working memory, $\beta = -.57$, $z = -43.37$, $p < .001$; our results suggest that the effect of translation number mainly affected individuals with lower WM span (see Figure 6). No other effects were significant in this model.

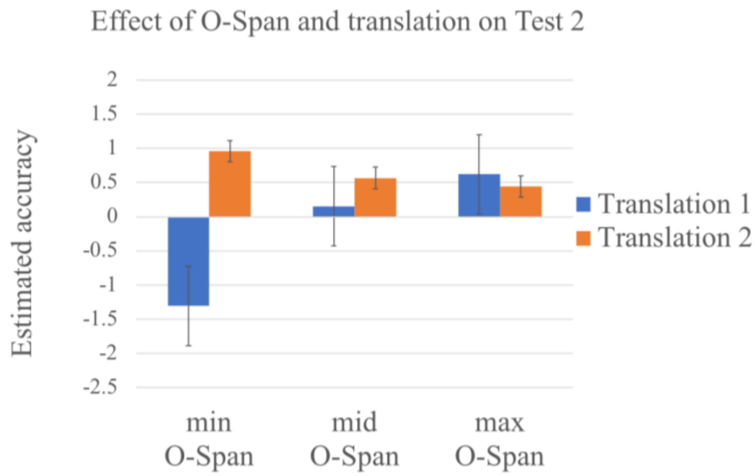

Figure 6. Effect of training manipulation and O-Span on translation number. Mean and extreme values observed were used to plot the minimum, middle, and maximum scores respectively for O-Span estimates (Aiken & West, 1991)

AX-CPT. Just as the results reported in the previous model, this model showed that Test 2 accuracy was significantly higher overall for the second translation compared to the first translation learned, $\beta = .39$, $z = 6.90$, $p < .001$. Translation number interacted with BSI, $\beta = .11$, $z = 4.55$, $p < .001$; here, we observed a benefit for the second translation trained for the participants with the highest BSI values. By contrast, participants with the lowest BSI values showed relatively higher and similar accuracy for both translations. Training manipulation and translation number also interacted, $\beta = -.16$, $z = -4.11$, $p < .001$; here we observed a bigger difference between the first and second translation for the separate condition than the placeholder condition, with the second translation being more accurate. A marginally significant two-way interaction between training manipulation and BSI was found, $\beta = 37$, $z = 1.96$, $p = .05$; had this interaction reached conventional levels of significance it would have shown that individuals with lower BSI values were most affected by the separate training manipulation. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and BSI,

β = -.11, z = -2.21, p < .03; our results suggest that the placeholder training manipulation benefits accuracy for the first translation for individuals with bigger BSI values, this benefit was also observed for participants with average BSI. The effect of translation number was similar for both training conditions for individuals with lower BSI, however the separate condition produced higher accuracy than the placeholder condition for these individuals (see Figure 8).

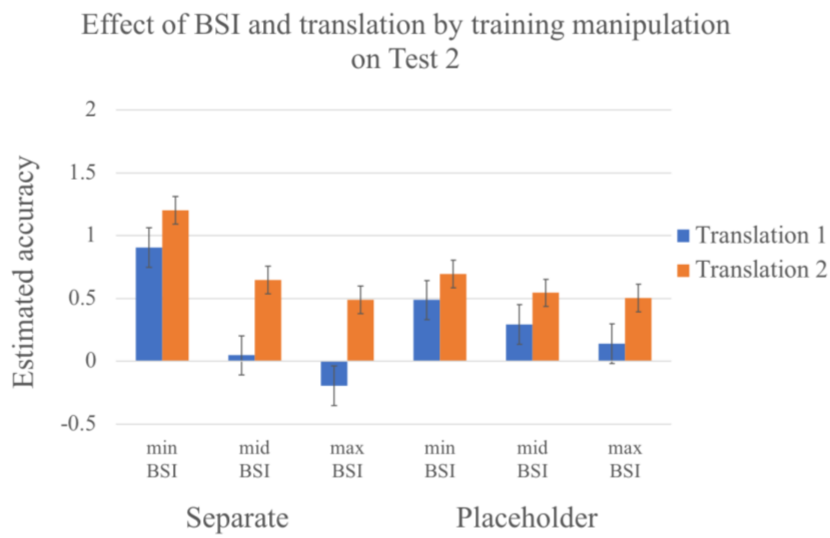

Figure 7. Effect of BSI and translation by training manipulation on Test 2. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for BSI estimates (Aiken & West, 1991)

Stroop. Just as the results reported in the previous models, this model showed that Test 2 accuracy was significantly higher for the second translation than the first translation learned, β = .44, z = 8.22, p < .001. Training manipulation interacted with translation number, β = -.22, z = -8.45, p < .001, such that there was a bigger difference in performance between the first and second translation for the separate condition than the placeholder condition, with the second translation being more accurate. Also, translation number interacted with Stroop effect, β = .08, z = 6.47, p < .001; here we observed a benefit for the second translation trained for the participants with the largest Stroop effect, who had quite low accuracy for the first translation. By contrast, participants with the smallest Stroop effect showed relatively higher and similar accuracy for both translations. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and Stroop effect, $\beta = .08$, $z = 3.19$, $p < .001$. Our results suggest that the placeholder condition benefits accuracy for the first translation for individuals with a smaller or average Stroop effect. The effect of translation number was similar for both training conditions for individuals with a larger Stroop effect (see Figure 10). No other effects were significant in this model.

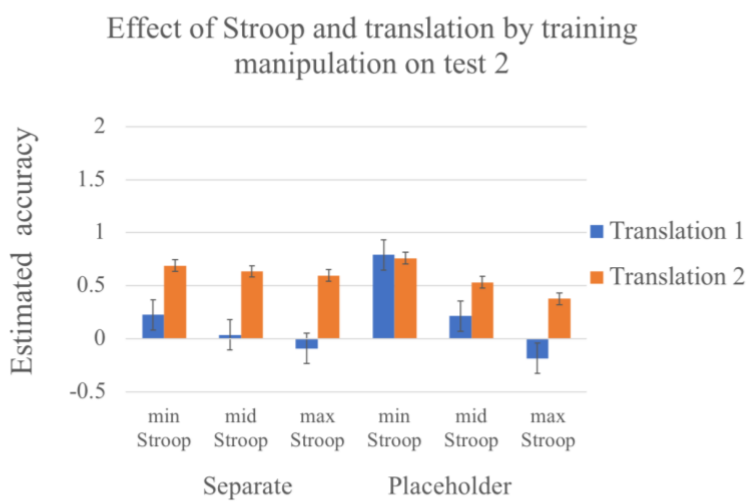

Figure 8. Effect of Stroop and translation by training manipulation on Test 2. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for Stroop estimates (Aiken & West, 1991)

PPVT. Just as the results reported in the previous models, this model showed that Test 2 accuracy was significantly higher for the second translation compared to the first translation learned, $\beta = .39$, $z = 6.82$, $p < .001$. Training manipulation and translation number interacted, $\beta = -.29$, $z = -10.4$, $p < .001$, such that there was a bigger difference in performance between the first and second translation for the separate condition than the placeholder condition, with the second translation being more accurate. In addition, translation number interacted with PPVT score, β = .19, z = 12.53, p < .001, such that there was a benefit for the second translation trained for the participants with the largest PPVT score, who had quite low accuracy for the first translation. By contrast, participants with the smallest PPVT scores showed relatively lower and similar accuracy for both translations. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and PPVT score, β = -.13, z = -4.49, p < .001. Here, we observed that the placeholder condition benefits accuracy for the first translation to an increasing degree the higher the PPVT score. second translation accuracy was higher overall, increased with PPVT, and was especially high for the separate condition. (see Figure 12). No other effects were significant in this model.

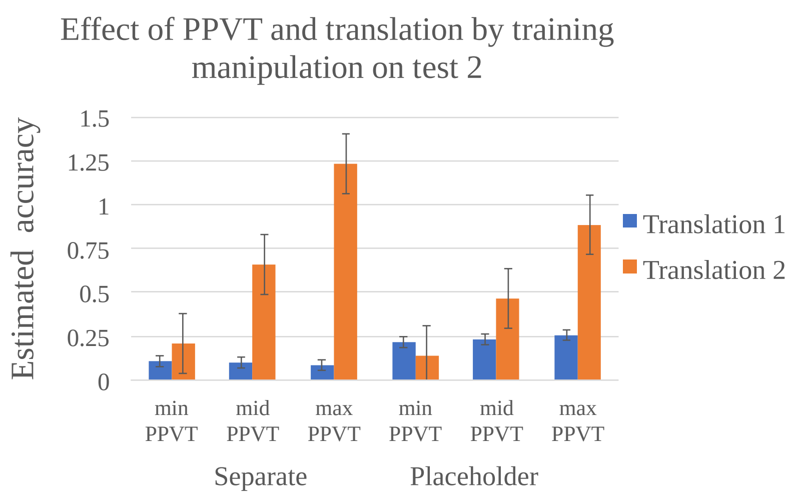

Figure 9. Effect of PPVT and translation by training manipulation on Test 2. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for PPVT estimates (Aiken & West, 1991)

3.1.2 Models for test 2 Reaction Times

O-Span. Test 2 reaction times were significantly slower for the second translation than the first translation, β =1790, t = 17.36, p < .001. Translation number interacted with WM span, β = -243, t = -12.76, p < .001. Here we observed a bigger difference in performance between the first and second translation for the placeholder training manipulation compared to both translations for the separate training manipulation, first translation being faster across training manipulations. Training manipulation and WM span interacted, β = 60, t = 3.83, p < .001, such that the separate condition benefited individuals with higher WM span compared to the placeholder training manipulation. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and WM span, β = -126, t = -6.25, p < .001; here, we observed that the separate condition benefits reaction time for the second translation for individuals with higher and average working memory, whereas participants with lower working memory performed similarly across conditions. A visualization of these results can be found in Figure 13.

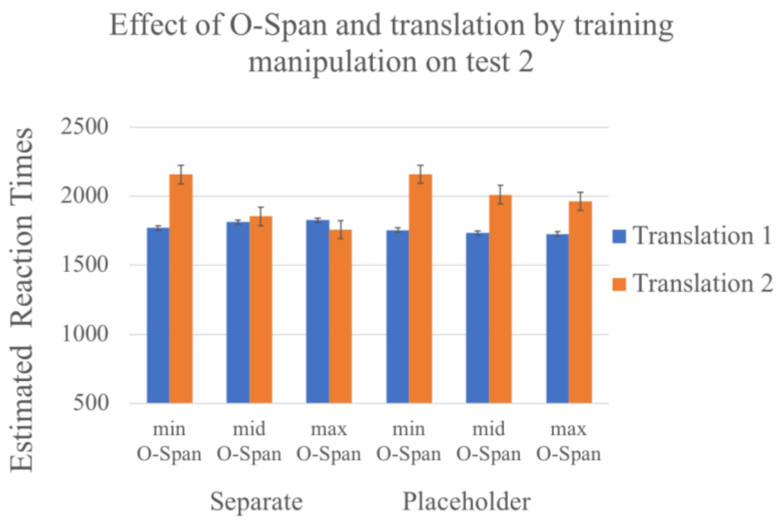

Figure 10. Effect of training manipulation and O-Span on translation number. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for O-Span estimates (Aiken & West, 1991)

AX-CPT. Test 2 reaction times were significantly slower for the second translation than first translation, β =400, t = 20.01, p < .001. Training manipulation interacted with translation number, such that second translations trained in the separate condition were recalled more slowly than first translations, β = -327, t = -11.99, p < .001. Also, BSI interacted with translation number, β = -296, t = -11.33, p < .001, such that a lower BSI value was associated with faster reaction times for the first translation learned. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and BSI, β = 117, t = 3.24, p < .001, such that the placeholder condition benefits reaction times for the first translation for individuals with lower BSI values (and to a small extent with average BSI values), whereas second translations were always translated more slowly in the placeholder condition (see Figure 14).

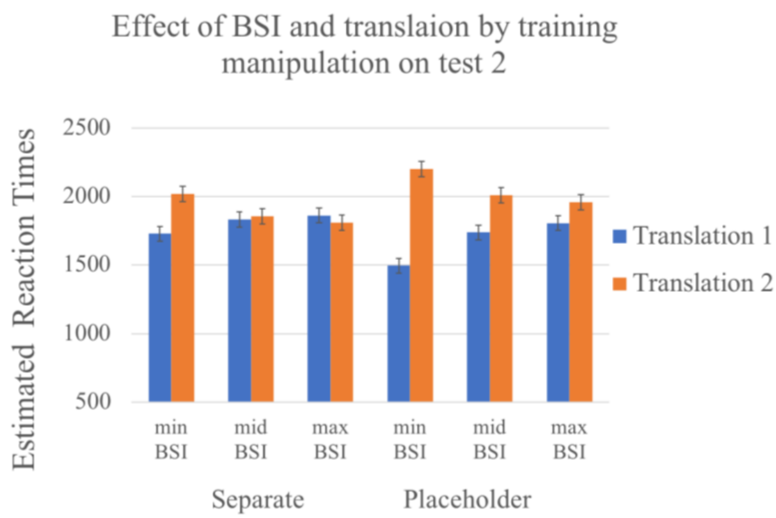

Figure 11. Effect of training manipulation and BSI on translation number. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for BSI estimates (Aiken & West, 1991)

Stroop. Test 2 reaction times were significantly slower for the second translation than the first translation, β =238, t = 17.70, p < .001. A main effect of training manipulation was found, β =33, t = 2.37, p < .05, such that responses in the separate condition were faster than responses in the placeholder condition. Training manipulation interacted with translation number, such that the second translations trained in the placeholder condition were translated faster than the first translations trained, β = -254, t = -13.48, p < .001, however there was not an effect of translation number for the separate condition. Second, the Stroop effect interacted with translation number, β = 260, t = 19.40, p < .001; here, we observed a bigger difference in performance between the first and second translation for the placeholder condition compared to both translations for the separate condition, the first translation being faster across conditions. Another two-way interaction was found between training manipulation and Stroop effect, β = 45, t = 3.39, p < .001, such that the placeholder condition mainly affected individuals with a bigger Stroop effect, whereas the effect was not as prominent for the separate condition. Finally, these two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and Stroop effect, $\beta = -207$, $t = -11.25$, $p < .001$, such that individuals with lower Stroop interference scores benefit from the placeholder condition for the second translation. By contrast, individuals demonstrating higher levels of Stroop interference seem to benefit from the separate condition for the second translation (see Figure 15).

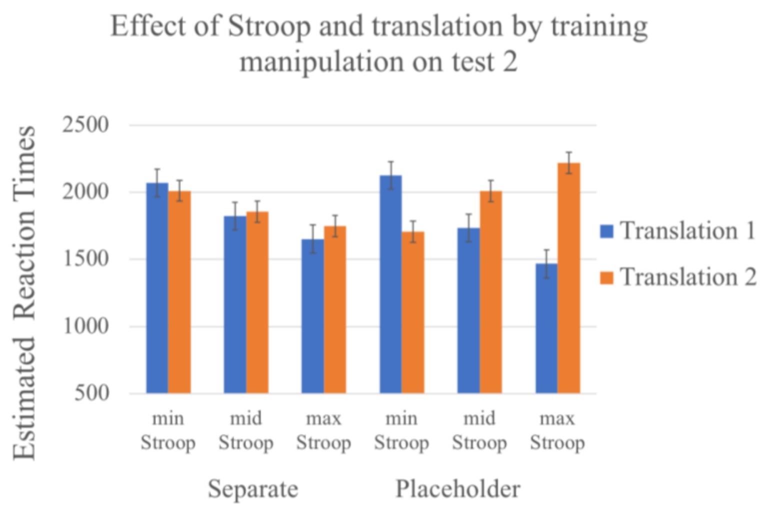

Figure 12. Effect of training manipulation and Stroop on translation number. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for Stroop estimates (Aiken & West, 1991)

PPVT. Test 2 reaction times were significantly slower for the second translation than the first translation, $\beta = 294$, $t = 22.02$, $p < .001$. Training manipulation interacted with translation number, such that second translations trained in the separate condition were translated slower than first translations, $\beta = -361$, $t = -18.50$, $p < .001$. PPVT score interacted with translation number, $\beta = 60$, $t = 4.35$, $p < .001$, such that a bigger PPVT score was associated with slower reaction times for the second translation learned. In addition, training manipulation and PPVT score interacted,

$\beta = 214$, $t = 14.41$, $p < .001$. These two-way interactions were qualified by a three-way interaction between training manipulation, translation number, and PPVT score $\beta = -223$, $t = -11.35$, $p < .001$, such that the placeholder condition benefited individuals with higher PPVT for the first translation. Furthermore, in the placeholder condition, the second translation was generally translated more slowly than the first, and this effect got larger the higher the PPVT. In the separate condition, the first translations trained were not affected by PPVT, whereas for second translations, the higher the PPVT, the faster the reaction time (see Figure 16).

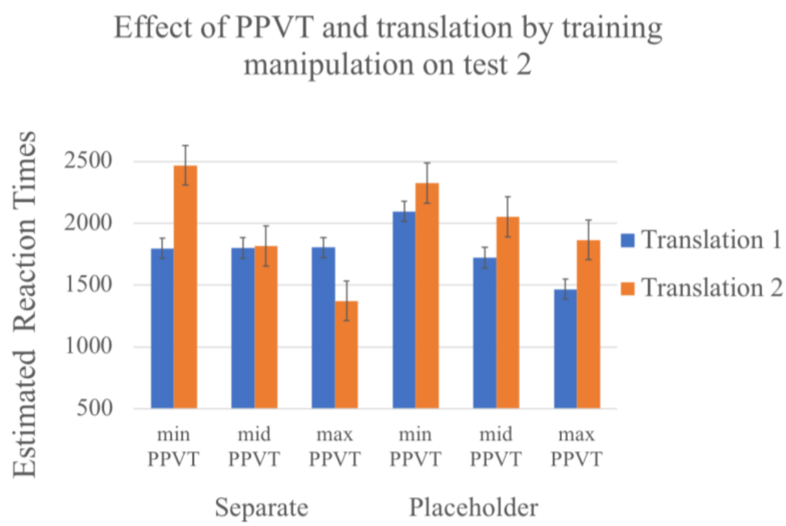

Figure 13. Effect of PPVT and translation by training manipulation on Test 2. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for PPVT estimates (Aiken & West, 1991)

3.1.3 First translation trained

In the following analyses, we examined performance on test 1 and test 2 only for the first English translation trained and the effects of each of our individual difference measures. The main goal of these analyses was to understand the effect of our training manipulations on the first translation trained over the course of the experiment, and to investigate if performance was affected by individuals' varying cognitive skills. In this analysis we only examined accuracy because in the first test only one translation was provided, whereas in the second test, two translations were provided on consecutive trials; this difference between the tests makes reaction time comparisons difficult.

O-Span. In this analysis, we examined performance on test 1 and test 2 only for the first English translation trained. We found a main effect of test such that accuracy increased significantly for translation one from test 1 to test 2, $\beta = .47$, $z = 20.51$, $p < .001$. Test interacted with working memory span, $\beta = .13$, $z = 5.94$, $p < .001$, such that higher working memory span was associated with higher accuracy in the second test than in the first, and the highest accuracy overall (see Figure 17). No other effects were significant in this model.

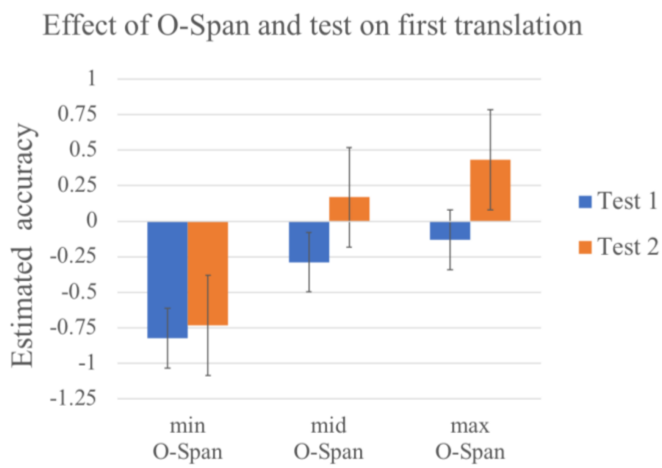

Figure 14. Effect of test and O-Span and test on first translation trained. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for O-Span estimates (Aiken & West, 1991)

AX-CPT. Again, accuracy increased significantly for translation one from test 1 to test 2, $\beta = .40$, $z = 11.99$, $p < .001$. Our results showed an interaction between training manipulation and BSI, $\beta = -.22$, $z = -2.10$, $p < .05$, such that individuals with lower BSI values had the highest accuracy, which was higher in the separate condition than the placeholder condition (see Figure 18).

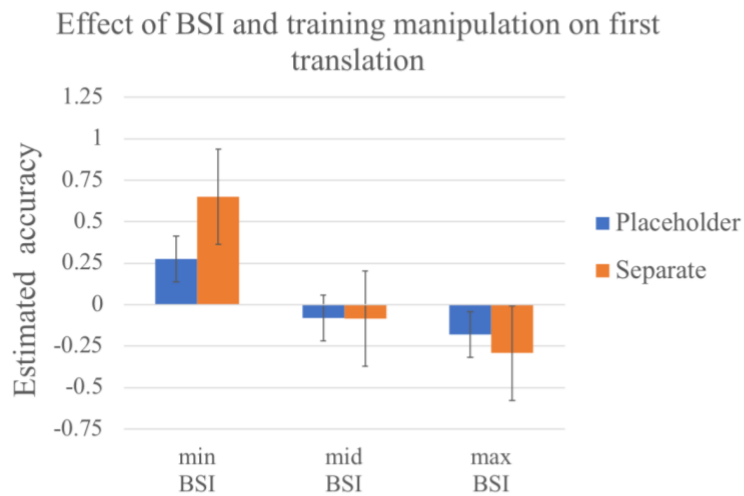

Figure 15. Effect of BSI and training manipulation on first translation. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for BSI estimates (Aiken & West, 1991)

Test interacted with BSI, $\beta = .11$, $z = 2.56$, $p < .05$, such that higher BSI was associated with higher accuracy on the second test, whereas lower BSI was associated with lower accuracy in general, but especially on the first test; the difference between tests increased slightly with BSI (see Figure 19). No other effects were significant in this model.

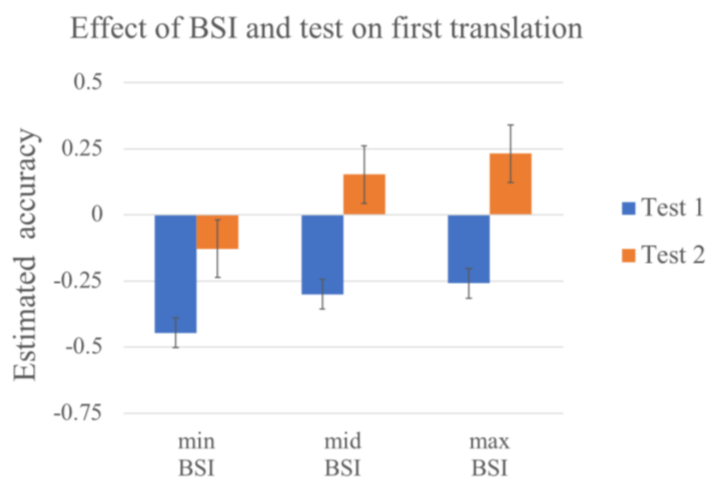

Figure 16. Effect of test and BSI and test on first translation trained. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for BSI estimates (Aiken & West, 1991)

Stroop. Our analysis showed higher accuracy for test 2 than test 1, $\beta = .46$, $z = 20.70$, $p < .001$. Test and Stroop effect interacted, $\beta = .14$, $z = 6.09$, $p < .001$, such that a bigger Stroop effect was associated with lower accuracy for Test 1, whereas a smaller Stroop effect was associated with a smaller influence of test and generally higher accuracy (see Figure 20). No other effects were significant in this model.

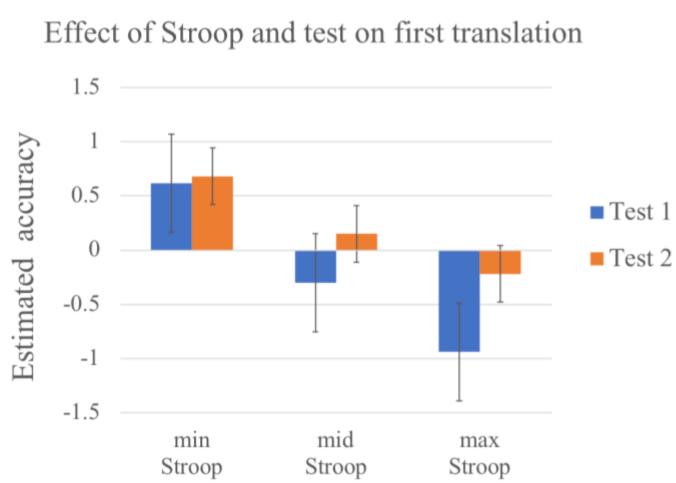

Figure 17. Effect of Stroop and test on first translation. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for Stroop estimates (Aiken & West, 1991)

PPVT. Our analysis showed higher accuracy for test 2 than test 1, $\beta = .55$, $z = 23.24$, $p < .001$. Test interacted with PPVT scores, $\beta = .08$, $z = 3.05$, $p < .001$, such that bigger PPVT scores were associated with higher accuracy for Test 2 (see Figure 21) . No other effects were significant in this model.

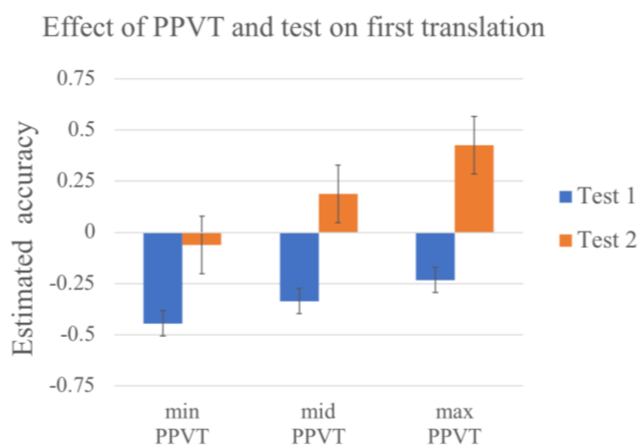

Figure 18. Effect of PPVT and test on first translation. Mean and extreme values observed were used to plot the minimum, middle and maximum scores respectively for PPVT estimates (Aiken & West, 1991)

3.1.4 Performance on Single and Fake Placeholder training manipulation: accuracy

No statistically significant effects were found in any of the models comparing accuracy for the "fake placeholder" and "single" conditions (see Table 17).

3.1.5 Performance on Placeholder and Fake Placeholder training manipulation: accuracy

No significant effects were found when comparing performance for the placeholder and fake placeholder conditions in any of the models (see Table 18).

4.0 Discussion

The goal of our study was to advance our understanding of the underlying mechanisms involved in the acquisition and representation of translation-ambiguous words by using a novel training manipulation for translation-ambiguous words. Specifically, the main goal of the current study was to investigate whether letting participants know a word has multiple translations during their first exposure can alleviate the translation-ambiguity disadvantage, and thus, inform us about what type of information is useful during vocabulary acquisition to properly represent translation-ambiguous words. Our results, however, provide a more complex picture of how the training manipulations used in our study can influence the acquisition of translation-ambiguous words. Generally speaking, in several instances our placeholder condition was beneficial for individuals of relatively higher cognitive skill (e.g., with higher BSI, lower Stroop interference, higher PPVT), particularly for accuracy of the first translation trained.

An interesting finding was that participants with lower levels of Stroop interference showed better recall of the first translation trained in the placeholder condition, compared to translations trained in the separate condition. This finding supported our prediction, which was based on previous findings (e.g., Michael et al., 2011) that have found that participants with lower levels of Stroop interference are better able to translate words across languages. Particularly, our results suggest that individuals with a better ability to ignore task-irrelevant information might also be able to benefit from the metalinguistic information provided by the placeholder condition. Our study also supports our prediction that individuals with higher English vocabulary (Koch, 2015), as measured by the PPVT-IV task, showed higher levels of accuracy for the second translation compared to the first translation—particularly for items trained in the separate condition.

More specifically, our results show that individuals with lower working memory span acquired the second translation trained better than the first translation. This effect is consistent with previous findings, which have found that higher working memory span results in greater competition of related items leading to lower accuracy overall when translating words across languages or retrieval of items related to the target word within the same language (Miyake et al., 1994). However, the opposite effect was also observed, such that, individuals with higher working memory span recalled the first translation slightly better than the second translation. It is also important to point out that higher working memory span was associated with better accuracy when assessing the overall the acquisition of the first translation. Unfortunately, a difference in accuracy for our training manipulations was not associated with working memory. However, we do see a difference in reaction times, where individuals with average working memory retrieve second translations faster when those translations are trained in the separate training manipulation.

One of our predictions when designing our study was that individuals who relied the most on maintaining and updating information, as shown by higher BSI values, would be able to use the metalinguistic information provided by the placeholder condition, by avoiding the consolidation of a one-to-one mapping of translation-ambiguous words and using their skills to update information by incorporating a second translation when introduced. However, this prediction was only supported for the first translation learned but not for the second one, when comparing individuals with higher BSI values performance for the first translation in both training manipulations. This pattern is also reflected in the significant difference between first translation and second translation in the separate training manipulation, where participants' accuracy on first translation trained seemed to be more impacted by their engagement of proactive control. For the separate training condition, accuracy was much better for participants with lower BSI values. Thus, overall, individuals who relied more on goal-maintenance than response suppression showed lower accuracy for the first translation learned compared to the second translation. A bigger training effect was observed for individuals with a greater ability to suppress information irrelevant to the task at hand who benefited more from the separate condition than the placeholder.

Overall, our results show that participants' performance improved for test 2 compared to test 1, more surprisingly the first translation trained was recalled significantly less accurately than the second translation trained, which is inconsistent with previous studies (Degani & Tokowicz, 2010; Degani et al. 2014) that have found that the first translation learned was recalled more accurately than the second translation. This might be explained by the pair-associated paradigm, which requires participants to learn two lists, A-B and A-C, in which the first list, A-B, is learned first and later required to be replaced by a second list, A-C. It might have been the case that even without instructing participants to replace the first list, German-first English translation, for a second list, German-second English translation, participants nonetheless favored the second list, leading to lower recall of the German-first English translation list. Particularly, this effect was found for items trained under the separate training manipulation, which might be an explanation of why when L2 learners attempt to acquire translation-ambiguous words they encounter problems creating a one-to-many representation.

Unfortunately, this means that our results are limited, and do not provide an easy suggestion to L2 instructors to implement when introducing translation-ambiguous words in a classroom environment. However, these results should not discourage us to keep thinking about ways in which we can improve the acquisition of translation-ambiguous words. In retrospect, we would had liked to have investigated the long-term effects of our training manipulations by bringing participants back after a couple of weeks to see if the effects of our vocabulary training stayed the same. Another manipulation we believe could had been useful to add to the current study is to increase the amount of vocabulary training.

In conclusion as mentioned before, our placeholder manipulation was only beneficial for participants with a particular set of cognitive skills, which might reflect their ability to use the type of metalinguistic information provided by the training manipulation. Our results suggest that participants with bigger L1 vocabulary benefit from the placeholder manipulation when acquiring the first. This is, a greater L1 vocabulary might aid participants to access English translations more accurately, which in combination with participants ability to ignore task-irrelevant information, as well as their ability to update and maintain goal-relevant information allows participants to use the metalinguistic information provided by the placeholder condition and temporarily integrate it in their mapping of the translation-ambiguous word.

Appendix A

Table 5. Fixed effects for the model for Test 2: accuracy with O-Span as a fixed effect

Fixed effect	β	SE	z value	*p*	
Intercept (Baseline accuracy)	.15	.22	.70	.48	
Training Manipulation	.14	.11	1.24	.22	
Translation No.	.42	.05	7.87	3.49e-15	***
Working Memory Span	.44	.21	2.13	.03	*
Training Manipulation:Translation No.	-.24	.03	-9.24	<2e-16	***
Training Manipulation:Working Memory Span	-.09	.11	-.81	.42	
TranslationNo:Working Memory Span	-.57	.01	-43.37	<2e-16	***
Training Manipulation:TranslationNo:OSpan	-.02	.03	.86	.39	

Note 1: glmer (Accuracy ~ 1 + TrainingManipulation * TranslationNo. * Ospan + (1 + TrainingManipulation |Subject) + (1 + TranslationNo.|Item), data = Test2, family = binomial) p < .10 * p < .05 ** p < .01 *** p < .001

Table 6. Fixed effects for the model for Test 2: accuracy with BSI as a fixed effect

Fixed effect	*β*	SE	z value	*p*	
Intercept (Baseline accuracy)	.34	.27	1.24	.21	
Training Manipulation	-.10	.15	-.65	.52	
Translation No.	.39	.06	6.90	5.07e-12	***
BSI	-.33	.33	-1.02	.31	
Training Manipulation:Translation No.	-.16	.04	-4.11	4.03e-05	***
Training Manipulation:BSI	.37	.19	1.96	.05	.
TranslationNo:BSI	.11	.02	4.55	5.38e-06	***
Training Manipulation:TranslationNo:BSI	-.11	.05	-2.21	.03	*

Note 2. glmer (Accuracy ~ 1 + TrainingManipulation * TranslationNo. * BSI + (1 + TrainingManipulation | Subject) + (1 + TranslationNo.|Item), data = Test2, family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

Table 7. Fixed effects for the model for Test 2: accuracy with Stroop as a fixed effect

Fixed effect	***β***	**SE**	**z value**	***p***	
Intercept (Baseline accuracy)	.14	.22	.64	.52	
Training Manipulation	.11	.11	1.05	.29	
Translation No.	.44	.05	8.22	<2e-16	***
Stroop Effect	-.14	.22	-.64	.52	
Training Manipulation:Translation No.	-.22	.03	-8.45	<2e-16	***
Training Manipulation:Stroop Effect	-.14	.11	-1.29	.20	
TranslationNo:Stroop Effect	.08	.01	6.47	9.65e-11	***
Training Manipulation:TranslationNo:Stroop Effect	.08	.03	3.19	.00	**

Note 3. : glmer (Accuracy ~ 1 + TrainingManipulation*TranslationNo.*Stroop + (1 + Training Manipulation |Subject) + (1 + TranslationNo.|Item), data = Test2, family = binomial) p < .10 * p < .05 ** p < .01 *** p < .001

Table 8. Fixed effects for the model for Test 2: accuracy with PPVT as a fixed effect

Fixed effect	β	SE	z value	*p*	
Intercept (Baseline accuracy)	.18	.23	.75	.45	
Training Manipulation	.10	.11	.85	.40	
Translation No.	.39	.06	6.82	8.95e-12	***
PPVT	.00	.22	.02	.99	
Training Manipulation:Translation No.	-.29	.03	-10.4	<2e-16	***
Training Manipulation:PPVT	.03	.11	.27	.78	
TranslationNo:PPVT	.19	.02	12.53	<2e-16	***
Training Manipulation:TranslationNo:PPVT	-.13	.03	-4.49	7.14e-06	***

Note 4. : glmer (Accuracy ~ 1 + TrainingManipulation*TranslationNo.*PPVT + (1 + Training Manipulation| Subject) + (1 + TranslationNo.|Item), data = Test2, family = binomial) p < .10 * p < .05 ** p < .01 *** p < .001

Table 9. Fixed effects from the model for Test 2 reaction times with O-Span as a fixed effect

Fixed effect	*β*	SE	df	t value	*p*	
Intercept (Baseline RTs)	1790	99.65	27	17.97	<2e-16	***
TranslationNo.	235	13.51	73030	17.36	<2e-16	***
Training Manipulation	17	14.17	73099	1.23	.22	
Working Memory Span	-17	96.51	23	-.17	.87	
TranslationNo:Working Memory Span	-243	19.08	73021	-12.76	<2e-16	***
TranslationNo:TrainingManipulation	-3	13.66	73033	-.23	.82	
TraniningManipulation:Working Memory Span	60	15.69	73076	3.83	.00	***
TrainingManip:TranslNo:WorkingMemorySpan	-126	20.12	73032	6.25	4.19e-10	***

Note 5. lmer (RT ~ 1 + TrainingManipulation * Translation * OSpan + (1 + TrainingManipulation | Subject) + (1 + TranslationNo | Item), data = Test2RT) . p < .10 * p < .05 ** p < .01 *** p < .001

Table 10. Fixed effects from the model for Test:2 reaction times with BSI as a fixed effect

Fixed effect	β	SE	df	t value	*p*	
Intercept (Baseline RTs)	1711	144	25	11.89	8.64e-12	***
TranslationNo.	400	20	73030	20.01	<2e-16	***
Training Manipulation	36	20	73100	1.76	.08	.
BSI	143	178	23	.80	.43	***
TranslationNo:BSI	-327	27	73020	-11.99	<2e-16	***
TranslationNo:TrainingManipulation	-296	26	73030	-11.33	<2e-16	***
TraniningManipulation:BSI	776	27	73070	.00	1.00	
TrainingManip:TranslNo:BSI	117	36	73020	3.24	.00	**

Note 6. lmer (RT ~ 1 + TrainingManipulation * Translation * BSI + (1 + TrainingManipulation | Subject) + (1 + TranslationNo | Item), data = Test2RT) p < .10 * p < .05 ** p < .01 *** p < .001

Table 11. Fixed effects from the model for Test:2 reaction times with Stroop as a fixed effect

Fixed effect	*β*	SE	df	t value	*p*	
Intercept (Baseline RTs)	1786	100	27	17.88	<2e-16	***
TranslationNo.	238	13	73034	17.70	<2e-16	***
Training Manipulation	33	14	73098	2.37	.02	*
Stroop	-145	97	23	-1.49	.15	
TranslationNo:Stroop	-254	19	73021	-13.48	<2e-16	***
TranslationNo:TrainingManipulation	260	13	73025	19.40	<2e-16	***
TraniningManipulation:Stroop	45	13	73065	3.39	.00	***
TrainingManip:TranslNo:Stroop	-207	18	73017	-11.25	<2e-16	***

Note 7. lmer (RT ~ 1 + TrainingManipulation * Translation * Stroop + (1 + TrainingManipulation | Subject) + (1 + TranslationNo | Item), data = Test2RT) p < .10 * p < .05 ** p < .01 *** p < .001

Table 12. Fixed effects from the model for Test:2 reaction times with PPVT as a fixed effect

Fixed effect	*β*	SE	df	t value	*p*	
Intercept (Baseline RTs)	1765	101	24	17.48	2.47e-15	***
TranslationNo.	294	14	67446	22.02	<2e-16	***
Training Manipulation	62	14	67508	4.30	1.69e-05	***
PPVT	-199	98	21	-2.02	.06	.
TranslationNo:PPVT	-361	20	67433	-18.50	<2e-16	***
TranslationNo:TrainingManipulation	60	14	67432	4.35	1.37e-16	***
TraniningManipulation:PPVT	214	15	67477	14.41	<2e-16	***
TrainingManip:TranslNo:PPVT	-223	20	67427	-11.35	<2e-16	***

Note 8. lmer (RT ~ 1 + TrainingManipulation * Translation * PPVT + (1 + TrainingManipulation | Subject) + (1 + TranslationNo | Item), data = Test2RT) p < .10 * p < .05 ** p < .01 *** p < .001

Table 13. Fixed effects for the model for First translation: accuracy with O-Span as a fixed effect

Fixed effect	β	**SE.**	**z value**	***p***	
Intercept (Baseline accuracy)	-.32	.27	-1.20	.23	
TrainingManipulation	-.00	.06	-.02	.99	
Test	.47	.02	20.51	<2e-16	***
Working Memory Span	.24	.25	.95	.34	
TrainingManipulation:Test	-.02	.03	-.70	.50	
TrainingManipulation:WorkingMemorySpan	.09	.06	1.51	.13	
Test:WorkingMemorySpan	.13	.02	5.94	2.82e-09	***
TrainingManipulation:Test:WorkingMemorySpan	-.04	.03	-1.29	.20	

Note 9. glmer(Accuracy ~ 1 + TrainingManipulation * Test * OSpan + (1 + Condition | Subject) + (1 | Item), data = FirstTranslation , family = binomial) . p < .10 * p < .05 ** p < .01 *** p < .001

Table 14. Fixed effects for First translation model: accuracy with BSI as fixed effect

Fixed effect	***β***	**SE.**	**z value**	***p***	
Intercept (Baseline accuracy)	-.15	.37	-.41	.68	
TrainingManipulation	.13	.08	1.51	.13	
Test	.40	.03	11.99	<2e-16	***
BSI	-.29	.45	-.64	.52	
TrainingManipulation:Test	.00	.05	.18	.86	
TrainingManipulation:BSI	-.22	.11	-2.10	.04	*
Test:BSI	.11	.04	2.56	.01	*
TrainingManipulation:Test:BSI	-.04	.06	-.73	.46	

Note 10. glmer(Accuracy ~ 1 + TrainingManipulation * Test * BSI + (1 + Condition | Subject) + (1 | Item), data = FirstTranslation , family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

Table 15. Fixed effects for First translation model: accuracy with Stroop as fixed effect

Fixed effect	***β***	**SE.**	**z value**	***p***	
Intercept (Baseline accuracy)	-.32	.26	-1.25	.21	
TrainingManipulation	.00	.06	0.04	.97	
Test	.46	.02	20.70	<2e-16	***
Stroop	-.25	.25	-.99	.32	
TrainingManipulation:Test	-.02	.03	-.65	.51	
TrainingManipulation:Stroop	.05	.06	.82	.41	
Test:Stroop	.14	.02	6.09	1.17e-09	***
TrainingManipulation:Test:Stroop	.02	.03	.54	.59	

Note 11. glmer(Accuracy ~ 1 + TrainingManipulation * Test * Stroop + (1 + Condition | Subject) + (1 | Item), data = FirstTranslation , family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

Table 16. Fixed effects for First translation model: accuracy with PPVT as fixed effect

Fixed effect	***β***	**SE.**	**z value**	***p***	
Intercept (Baseline accuracy)	-.34	.27	-1.29	.20	
TrainingManipulation	.03	.07	0.47	.64	
Test	.55	.02	23.24	<2e-16	***
PPVT	-.00	.26	-.02	.99	
TrainingManipulation:Test	-.05	.03	-1.61	.11	
TrainingManipulation:PPVT	-.03	.07	-.50	.62	
Test:PPVT	.08	.02	3.05	.00	***
TrainingManipulation:Test:PPVT	-.01	.04	-.41	.68	

Note 12. glmer(Accuracy ~ 1 + TrainingManipulation * Test * PPVT + (1 + Condition | Subject) + (1 | Item), data = FirstTranslation , family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

Table 17. Fixed effects from model for Single and Fake Placeholder training manipulations

Fixed effect	***β***	**SE**	**z value**	***p***
Intercept (Baseline accuracy)	.04	.36	.10	.92
TrainingManipulation	-2.25e-06	.05	.00	1.00
Individual Difference Measure	-.12	.43	-.28	.78
TrainingManipulation:ID Measure	1.23e-06	.06	.00	1.00

Note 13. glmer (Accuracy ~1 + TrainingManipulation * BSI + (1+TrainingManipulation | Subject) +(1 | Item), data = One , family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

Table 18. Fixed effects from the model for Placeholder and Fake Placeholder training manipulations

Fixed effect	***β***	**SE**	**z value**	***p***
Intercept (Baseline accuracy)	-.37	.66	-.56	.58
TrainingManipulation	1.03	.82	1.25	.21
Individual Difference Measure	-.01	.64	-.02	.99
TrainingManipulation:ID Measure	.54	.99	-.55	.58

Note 14. glmer (Accuracy ~ 1 + TrainingManipulation * BSI + (1 + TrainingManipulation | Subject) + (1 | Item) , data = PlaceholderVs.FakePlaceholder , family = binomial). p < .10 * p < .05 ** p < .01 *** p < .001

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage Publications.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296-308. doi:http://dx.doi.org/10.1037/0096-3445.108.3.296

Bracken, J., Degani, T., Eddington, C. M., & Tokowicz, N. (2013). *Meaning similarity affects the learning of translation-ambiguous words.* In press

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351-7356. doi:http://doi.org/10.1073/pnas.0808187106

Braver, T. S. (2012) The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences, 16*, 106-113.

Degani, T., Tseng, A., & Tokowicz, N. (2014). Together or apart: Learning of translation-ambiguous words. *Bilingualism: Language and Cognition, 17*(4), 749-765. doi:10.1017/S1366728913000837

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Eddington, C. M. & Tokowicz, N. (2013). Examining English-German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition.*

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319(5865), 966-968.*

Koch, G. E. (2015). Translation-ambiguity and individual differences in L2 vocabulary learning Between mono-and bilinguals. (Unpublished Undergraduate Thesis). University of Pittsburgh, Pittsburgh, PA.

Lotto, L., & De Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48,* 31–69.

Michael, E. B., Tokowicz, N., Degan, T., & Smith, C. J. (2011). Individual Differences in the Ability to Resolve Translation Ambiguity across Languages. *Vigo International Journal of Applied Linguistics, 8.*

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*, 15583-15587.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330* (6002), 335-335.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, J., Edwin, D., & Beck, L. H. (1956). A continuous performance test of brain damage. Journal of Consulting Psychology, 20, 343-350.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology 18* (6): 643-662.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*, 727-779

Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working memory capacity in the types of errors made during translation. *Bilingualism:*

Language and Cognition, *7*, 255-272.

Tokowicz, N. (2014). Translation ambiguity affects language processing, learning, and representation. In R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, A. Tuninetti & D. Walter (Eds.), *Selected proceedings of the 2012 Second Language Research Forum: Building bridges between disciplines* (pp. 170-180). Somerville, MA: Cascadilla Press.

Turner, M. L., & Engle, R. W. (1989) Is working memory capacity task dependent? *Journal of Memory & Language*, *28*, 127-154.