

# Individual Differences in the Ability to Resolve Translation Ambiguity across Languages

---

Erica B. Michael  
University of Maryland  
*emichae3@umd.edu*  
Natasha Tokowicz  
University of Pittsburgh  
*tokowicz@pitt.edu*  
Tamar Degani  
University of Pittsburgh  
*tdegani@pitt.edu*  
Courtney J. Smith  
University of Pittsburgh  
*cojasmith@gmail.com*

## Abstract

We investigated whether individual differences in working memory (WM) span and the ability to ignore task-irrelevant information were related to the ability to resolve *translation ambiguity*, which occurs when a word has multiple translations. Native English speakers who were intermediate learners of Spanish translated translation-ambiguous (multiple translation) and translation-unambiguous (single translation) words. As in previous studies, translation-ambiguous words were translated less accurately than translation-unambiguous words. Individuals better able to ignore task-irrelevant information translated words more accurately, but only for translation-unambiguous words. The best overall translation performance was for individuals with both higher WM span and better ability to ignore task-irrelevant information; higher WM span was a *disadvantage* for individuals more susceptible to task-irrelevant information. These results suggest that higher WM span and ability to ignore task-irrelevant information contribute to better L2 word learning, but greater word knowledge is problematic if individuals are not able to control the activation generated from multiple translation alternatives.

**Keywords:** bilingual language processing, translation ambiguity, individual differences, working memory, Stroop

## Resumen

En nuestra investigación comprobamos si las diferencias en memoria a corto plazo (MCP) y la capacidad de ignorar información no relevante para la tarea se relacionan con la capacidad de solucionar la ambigüedad en la traducción, que aparece cuando una palabra tiene múltiples traducciones. En el experimento, hablantes nativos de inglés con un nivel intermedio de español traducían palabras ambiguas en la traducción y palabras no ambiguas. Los individuos que fueron más capaces de ignorar la información no relevante traducían las palabras más adecuadamente, pero sólo cuando se trataba de palabras no ambiguas en traducción. El rendimiento mejor en la totalidad fue de los individuos con un alcance mayor de MCP y mejor capacidad de ignorar información no relevante: un alcance mayor de MCP resultó ser una desventaja en los individuos más susceptibles a tener en cuenta información no relevante. Los resultados sugieren que un alcance mayor de MCP y mayor capacidad de ignorar información no relevante contribuye a un aprendizaje mejor de palabras en la L2, pero que un conocimiento mayor de palabras resulta problemático si los individuos no son capaces de controlar la activación generada a partir de múltiples alternativas de traducción.

**Palabras clave:** procesamiento bilingüe de lenguas, ambigüedad en la traducción, diferencias individuales, memoria a corto plazo, Stroop

## 1. Introduction

One of the main barriers to learning a second language (L2) as an adult is that aspects of the new language may differ substantially from aspects of the highly practiced first language (L1). These differences can occur at all levels of language representation, even at the word level between so-called translation equivalents. Previous research has demonstrated that many words have more than a single way of being expressed in another language (e.g., Tokowicz, Kroll, De Groot, & Van Hell, 2002), a phenomenon that is referred to as *translation ambiguity*. Furthermore, translation ambiguity makes the process of translating words slower and less accurate (e.g., Tokowicz & Kroll, 2007), and creates difficulties in the beginning stages of L2 word learning (Degani & Tokowicz, 2010a).

Although differences between the L1 and L2 pose challenges for nearly all learners, some individuals seem to have much less difficulty than others in learning new languages. Part of this skill may be due to the ability to exert inhibitory control, allowing individuals to suppress L1 word forms during the acquisition of L2 vocabulary (Kroll, Michael, & Sankaranarayanan, 1998) and to suppress activation from one

language to speak in the other (e.g., the Inhibitory Control Model, see Green, 1998; Trude & Tokowicz, in press). In the present study, we examined whether individual differences in cognitive skill mitigate the difficulties associated with learning and processing translation-ambiguous words.

Translation ambiguity arises for a variety of reasons. Near-synonymy is one of the most prevalent sources; for example, the Spanish word *sofá* can be translated into English as “sofa” or “couch.” Translation ambiguity also occurs as a result of polysemy (the existence of multiple related senses), as with the word “glass,” which can be translated into Spanish as either *vidrio* (material) or *vaso* (drinking vessel). It also occurs as the result of homonymy (the existence of multiple unrelated meanings); for example, the word “bark” translates into Spanish as *ladrido* (sound a dog makes) and *corteza* (outer layer of a tree). Finally, part-of-speech ambiguity of the type often created in English (e.g., “walk”) typically leads to the presence of translation ambiguity (Prior, MacWhinney, & Kroll, 2007). In contrast, translation-unambiguous words have a single translation equivalent (e.g., *cat-gato*).

A growing body of evidence suggests that translation-ambiguous words are more difficult to learn and process than translation-unambiguous words. This finding was first documented by Tokowicz and Kroll (2007), who noted that moderately-proficient English-Spanish and Spanish-English bilinguals were slower and less accurate to produce translations for translation-ambiguous than for translation-unambiguous words. Tokowicz and Kroll noted that this finding could be due to the increased response competition that occurs for translation-ambiguous words, because the speaker must choose a word for production from among multiple valid translation equivalents. Degani and Tokowicz (2010a) later discovered that the disadvantage for translation-ambiguous words occurs even in very early stages of L2 vocabulary learning. In a laboratory training study in which native English speakers were taught a set of Dutch words, participants had more difficulty learning translation-ambiguous than unambiguous words. This finding was evident even in a translation-recognition task that did not require production, indicating that response competition cannot entirely explain the ambiguity effect. As an alternative/additional explanation, Degani and Tokowicz proposed that the associative connections between translations may be weaker for translation-ambiguous than unambiguous words, and that a one-to-many mapping can be more difficult to learn than a one-to-one mapping (see Tokowicz and Degani, in press), for a more detailed review of these and other studies documenting the consequences of translation ambiguity for L2 learning and processing.)

Thus, both moderately-proficient bilinguals and L2 learners have difficulty with translation-ambiguous words. As mentioned above, in the present study we asked whether individual differences in cognitive skill mitigate the translation-ambiguity

disadvantage. If so, what cognitive factors predict successful resolution of translation ambiguity? We focused on two factors potentially important in both L1 and L2 processing: working memory (WM) capacity and the ability to ignore task-irrelevant information.

Differences in the rate of learning and ultimate attainment for adults learning an L2 cannot be explained entirely by motivation, personality, or general intelligence (Harrington & Sawyer, 1992; Miyake, 1998). Research on cognitive factors that contribute to individual differences in L2 processing has focused largely on WM capacity, or the ability to simultaneously store and process information. WM correlates with L2 reading comprehension, L2 grammar acquisition, L2 grammatical rule generalization, and translation latency (Harrington & Sawyer, 1992; Kempe, Brooks, & Kharkhurin, 2010; Michael, Tokowicz, & Kroll, 2003; Miyake, 1998; for a review, see Michael & Gollan, 2005).

WM might be especially critical for learning and processing translation-ambiguous words. Individual differences as a function of WM capacity are often larger for more difficult tasks (e.g., Miyake, Carpenter, & Just, 1994), and WM capacity has also been specifically linked to ambiguity resolution within a language (e.g., Gunter, Wagner, & Friederici, 2003; Miyake, Just, & Carpenter, 1994). Little research has directly addressed the relationship between WM and cross-language ambiguity, but see Degani and Tokowicz (2010b) for a discussion of individual differences in ambiguity processing within and across languages.

Although several studies have demonstrated that suppression of L1 may be important in L2 learning (Kroll et al., 1998; Levy, McVeigh, Marful, & Anderson, 2007; Linck, Kroll, & Sunderman, 2009; Trude & Tokowicz, in press), few studies have directly examined bilingual processing as a function of individual differences in the ability to ignore task-irrelevant information. Even fewer studies have simultaneously examined both WM capacity and the ability to ignore task-irrelevant information to determine the relative contribution of each to L2 processing. Although some views of WM describe it as controlled attention – i.e., the ability to selectively attend to relevant information and inhibit irrelevant information (Conway & Engle, 1994) – there is no consensus about the precise nature of the relationship between the two constructs. Whereas some studies do suggest considerable overlap between measures of WM capacity and the ability to ignore task-irrelevant information (Conway, Cowan & Bunting, 2001; Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003), there is not always a strong correlation between performance on the two types of tasks (Kwong See & Ryan, 1995; Long & Prat, 2002). If indeed the two constructs are separable, it will be important to examine their distinct roles in bilingual processing.

## 2. Overview of the current study

In summary, the goal of the current study was to determine whether the cognitive factors shown to be related to individual differences in overall L2 processing affect L2 learners' ability to cope with translation ambiguity. We examined individual differences in native English speakers who were learning Spanish as an L2. We assessed WM capacity using an Operation-Word Span task (O-Span; Turner & Engle, 1989) and we assessed the ability to ignore task-irrelevant information using a color-word Stroop task (Stroop, 1935). We then examined the relationship between performance on these tasks and the ability to produce translations for translation-ambiguous relative to translation-unambiguous words. Translating ambiguous words was expected to be difficult for the participants in the current study given their limited proficiency in L2. Because difficulty results in relatively low accuracy rates, we emphasize accuracy findings more than reaction time (RT) findings.

Within the available stimulus set there were very few items that were translation-ambiguous from Spanish into English. Therefore, we were not able to examine the effect of ambiguity in Spanish-English (L2-L1) translation, and instead focused on translation from L1 to L2. Moreover, L1-L2 translation is of most theoretical interest because translation ambiguity effects are larger in L1-L2 translation (Tokowicz & Kroll, 2007), which is also the direction of translation most likely to reveal individual differences because it is thought to require more inhibitory control (Green, 1998).

In accordance with the results of Tokowicz and Kroll (2007), we predicted that translation-ambiguous words would be more difficult to translate than translation-unambiguous words, resulting in lower accuracy and slower RTs. For translation-unambiguous words, we predicted that performance would be better for individuals with higher WM capacity, who typically perform better than their lower-span counterparts in various L2 processing tasks (Michael & Gollan, 2005); we also predicted that performance would be better for individuals with better ability to ignore task-irrelevant information, given that translation requires control of activation from two languages (Green, 1998).

For translation-ambiguous words, we again expected better performance from individuals with better ability to ignore task-irrelevant information, but the prediction for WM capacity was less straightforward. One possibility is that any advantages conferred by higher span would be even greater for translation-ambiguous words than for translation-unambiguous words due to the increased difficulty involved in processing ambiguous words. Alternatively, to the extent that WM span reflects the ability to maintain multiple pieces of information, we might predict that individuals with higher WM span would activate more translation alternatives than individuals

with lower WM span, and this greater activation would lead to greater competition among alternatives, thereby slowing translation and making it less accurate. This prediction is similar to ideas about the role of WM in within-language ambiguity resolution, according to which higher-span individuals activate and maintain both the dominant and the subordinate meanings of ambiguous words (e.g., Miyake et al., 1994, but see Gunter et al., 2003, for an opposing explanation).

Finally, although there is uncertainty in the literature regarding the precise relationship between WM capacity and the ability to ignore task-irrelevant information, the two measures we used likely tap at least slightly different aspects of cognitive ability. We therefore predicted an interaction between the two. Specifically, the ability to ignore task-irrelevant information could mitigate the difficulty associated with maintaining additional alternatives because the individual may be better able to select a particular translation for production, so we would expect that the translation-ambiguity disadvantage would be greatest (perhaps counter-intuitively) for individuals with higher WM span and less ability to ignore task-irrelevant information, and smallest for individuals with lower WM span in general.

### 3. Method

#### 3.1. Participants

Participants were 29 native English speakers (mean age = 19.4 years) who had studied Spanish as an L2 for a minimum of four years. All were recruited from the University of Pittsburgh or Carnegie Mellon University and were paid \$6 for their participation. Data from a total of ten participants were excluded for a variety of reasons: technical errors ( $n = 2$ ), low translation accuracy (less than 33%,  $n = 7$ ), or low operation accuracy in the O-Span task (50%,  $n = 1$ ). The analyses were conducted on the final set of 19 participants.

Participants completed a language history questionnaire (Tokowicz, Michael, & Kroll, 2004) on which they provided self-ratings of their L2 proficiency. Participants had studied Spanish for an average of 6.4 years, beginning at a mean age of 13.1 years. Average self-ratings for Spanish reading ability, writing ability, conversation ability, and speech comprehension ability were 7.0, 6.3, 6.6, and 7.1, respectively (each on a 10-point scale with 10 indicating a high rating).

### 3.2. Procedure

Participants first completed two individual-difference measures: an O-Span task (Turner & Engle, 1989) and a color-word Stroop task (Stroop, 1935). Participants then performed a single-word translation production task in which they saw words on the computer screen and translated them aloud from English to Spanish and Spanish to English in separate blocks. Finally, participants completed the language history questionnaire.

#### 3.2.1. Operation-word span task

In the O-Span task, participants had to simultaneously solve simple mathematical operations and store lists of English words for later recall. Each operation appeared in the center of the computer screen alongside a solution (e.g.,  $4 / 2 + 3 = 5$ ) for 2500 ms, followed by a question-mark probe. At the probe, participants were instructed to indicate whether the solution was correct or incorrect by pressing a Yes or No button on the keyboard with the left or right index finger, respectively. The probe disappeared after the response or after 1250 ms, whichever came first. Following each response or timeout, an English word appeared for 1250 ms. After each set of operations and words, a "RECALL" prompt appeared, at which point participants were instructed to write as many words from that set as they could remember in the order in which they appeared. Operation-word sets were presented in increasing size from two to six, with three sets of each size; in addition, two practice sets were presented prior to the critical sets. Although the sets were presented sequentially, the operations and words within each set were presented in a random order determined by the computer program (E-Prime software, Psychology Software Tools, Pittsburgh, PA). The measure of WM span derived from this task is described in the *Results* section.

#### 3.2.2. Stroop task

In the color-word Stroop task, participants were presented with a series of letter strings displayed in colored ink. Participants were instructed to name aloud the ink color of each stimulus. On *congruent* trials, the ink color matched the color word (e.g., the word "blue" was presented in blue ink). On *incongruent* trials, the ink color did not match the color word (e.g., the word "blue" was presented in red ink). On *neutral* trials, a row of colored X's was presented (e.g., "XXXXX" was presented in red ink). Each trial began with a fixation cross for 1000 ms, followed by the stimulus for 5000 ms. The trial concluded at the initiation of a vocal response or after 10000 ms, whichever came first. The task began with six practice trials for which feedback was

given. There were then 60 critical trials during which no feedback was given. The scoring procedure is described in the *Results* section.

### 3.2.3. Translation task

The stimuli for the translation production task consisted of 478 English words and their Spanish translations. The words varied in their number of available translations from English to Spanish and the reverse, determined by having a group of bilinguals provide the first translation they thought of for each word and then counting all correct translations across participants (norms obtained from Tokowicz, 1997). As in previous research on translation ambiguity, the stimulus set included more translation-unambiguous words than translation-ambiguous words to avoid revealing the purpose of the experiment.

The English words ranged in their ratings for concreteness (Paivio, Yuille, & Madigan, 1968), context availability (e.g., Schwanenflugel & Shoben, 1983), and familiarity (e.g., Gernsbacher, 1984), with each set of ratings collected from a separate set of native English-speaking monolinguals ( $N_s = 16, 15, 15$ , respectively). The words were divided into two lists matched on these three dimensions as well as on English and Spanish word length (number of letters), English word frequency (Francis & Kucera, 1982), and number of translations in both directions, all  $F_s < 1$ . Each participant translated one list in each direction of translation; assignment of list to language block was counterbalanced across participants.

The original set of 478 items included noncognates (words that are not similar in form to their translation equivalents, e.g., *hecho*-fact), cognates (words that are similar in form to their translation equivalents, e.g., *foto*-photo), and false cognates (words that look like a word in the other language but have a different meaning, e.g., *fábrica*, which looks like the English word “fabric” but means factory). This balance was included to maintain a naturalistic sampling of words, but as in previous research on translation ambiguity our primary interest was in noncognates. For translation-ambiguous words that share form closely with a word in the other language (e.g., cognates), it is often the case that only one of the possible translations has high form overlap (e.g., the Spanish word *sofá* translates into English as “couch” or “sofa”); we therefore excluded the 207 cognates and false cognates from analysis.

We also eliminated the 87 noncognate items for which no participant provided a correct response. For the purposes of analysis, we removed an additional 17 words so that the remaining 135 translation-unambiguous and 32 translation-ambiguous words matched each other on English word frequency (untransformed and logged;

Francis & Kucera, 1982), English length (number of letters), Spanish length (number of letters), and concreteness, context availability, and familiarity ratings, all  $F_s < 1.4$ , all  $p_s > .23$  (see Table 1).

**Table 1.** Mean characteristics of the critical stimuli for the translation task (SDs). The 135 translation-unambiguous words had one translation from English to Spanish and the reverse. The 32 translation-ambiguous words had more than one translation from English to Spanish.

	Translation-Unambiguous	Translation-Ambiguous
English Word Frequency (occurrences per million)	124.81 (17.33)	105.50 (35.33)
English Length (# letters)	5.14 (.13)	4.94 (.27)
Spanish Length (# letters)	5.72 (.15)	5.81 (.30)
Concreteness (1-7 rating*)	4.91 (.11)	4.98 (.23)
Context Availability (1-7 rating*)	5.26 (.06)	5.37 (.12)
Familiarity (1-7 rating*)	4.68 (.06)	4.83 (.11)

\*Note: For all ratings, 7 indicates a high rating along the given dimension.

In the analyses presented here, we accepted only the expected response as correct, because accepting alternate valid translations would violate our matching procedures. The expected responses were based on the stimuli used in previous research (Tokowicz, 1997; Tokowicz & Kroll, 2007). The consequences of this scoring procedure were minimal, as will be described in the *Results* section.

Participants were presented with a series of individual words at the center of the computer screen and instructed to translate each word aloud into the target language as quickly and accurately as possible. On each trial participants saw a fixation cross and then pressed a key on the keyboard to initiate the trial. The stimulus appeared for 4000 ms or until the participant initiated a vocal response. Stimuli in each block were presented in a random order determined by E-Prime (Psychology Software Tools, Pittsburgh, PA). Participants were instructed to guess or to say “I don’t know” if they could not retrieve the translation. RTs were recorded by E-Prime, and responses were

taped for later accuracy coding. Each participant completed one block of trials from L1 (English) to L2 (Spanish) and another block from L2 to L1, with order of language blocks counterbalanced across participants.

## 4. Results

### 4.1. Individual-difference tasks

WM span was calculated by counting the number of words correctly recalled *only in sets that were entirely correctly recalled*. The range of possible span scores was 0-60. Scoring did not consider order, and recall was computed regardless of the accuracy of the associated operation. However, we did examine overall accuracy on the operation component of the task because low operation accuracy may indicate that a participant ignored the operations to focus more attention on memorizing the words. As mentioned above, data from one participant were removed from analysis due to an operation accuracy score of 50%. For the remaining 19 participants, operation accuracy ranged from 88.3% to 100% ( $M = 96.67\%$ ;  $SD = 3.42\%$ ) and span scores ranged from 14 to 60 words ( $M = 40.68$ ,  $SD = 15.33$ ).

In the Stroop task, correct trials with RTs under 300 ms or over 3000 ms were treated as voice key errors and discarded from analysis. Of the remaining data, RTs more than 2.5 standard deviations above or below the mean were discarded as outliers. These procedures resulted in removal of 2.6% and 2.3% of the data, respectively. Average RTs for congruent, neutral, and incongruent trials were 653 ms ( $SD = 86.2$ ), 656 ms ( $SD = 86.4$ ), and 755 ms ( $SD = 126.1$ ), respectively. The measure of ability to ignore task-irrelevant information derived from this task was based on the trimmed RT for correct critical trials, and was computed as each participant's inhibition score (incongruent RT minus the average of congruent and neutral RTs) divided by his/her overall average RT. Higher scores indicate a higher degree of interference experienced on incongruent trials; therefore, individuals who experienced less interference are said to be better at ignoring task-irrelevant information. Interference scores ranged from 0.03 to 0.27 ( $M = 0.15$ ,  $SD = 0.07$ ). There was no correlation between performance on the O-Span Task and performance on the Stroop task,  $r(19) = .026$ ,  $p > .9$ .

### 4.2. Translation task

Repeated measures analyses of variance (ANOVAs) were used to examine the effects of translation ambiguity, WM span, and Stroop interference; separate

ANOVAs were conducted with accuracy and RT as dependent variables. Data were analyzed with participants ( $F_1$ ) and items ( $F_2$ ) as random factors. The  $F_1$  analyses treated translation ambiguity as a within-participants factor, and the individual-difference measures as covariates of interest; these measures were treated as covariates so their effects could be examined without needing to dichotomize the factors (see, e.g., Tokowicz & Warren, 2008). The  $F_2$  analyses treated translation ambiguity as a between-items factor and the individual-difference measures as within-items factors. Because it was not possible to use the individual-difference measures as covariates in the  $F_2$  analyses, we created groups based on median splits (see Table 2). All reported means are from the  $F_1$  analyses except where otherwise noted. Please note that although the  $F_1$  and  $F_2$  for each effect are reported together,  $F_1$ s reflect continuous factors and  $F_2$ s reflect categorical factors.

**Table 2.** Participants grouped by WM span and amount of Stroop interference. Those participants with less Stroop interference are considered better at ignoring task-irrelevant information.

	Amount of Stroop Interference	
	Less (.03-.16)	More (.17-.27)
Higher WM Span (48-60 words)	n = 6	n = 4
Lower WM Span (14-37 words)	n = 3	n = 6

#### 4.2.1. Accuracy

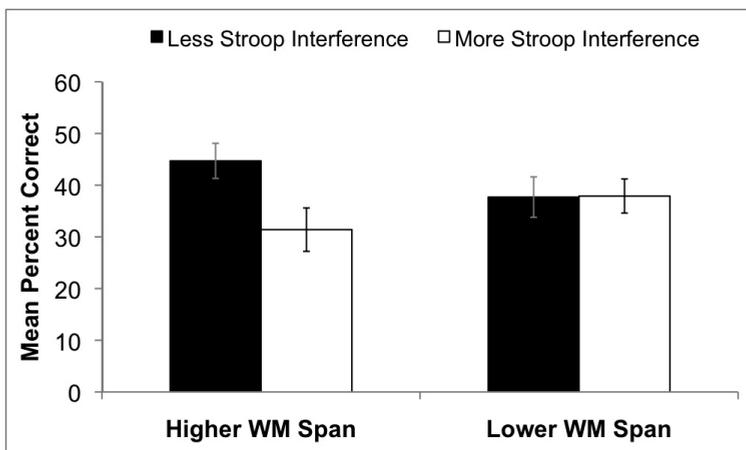
As mentioned above, we only accepted the “expected” translation as correct to avoid violating our matching of items across conditions and lists. All other answers were considered to be incorrect. This procedure resulted in the non-acceptance of 1.8 % of the data (trials in which valid translation equivalents were given that would have been considered correct under a more liberal scoring procedure).

Overall, translation-ambiguous words were translated less accurately than translation-unambiguous words (28 vs. 53 %),  $F_1(1,16) = 11.45$ ,  $MSE = .009$ ,  $p < .01$ ;  $F_2(1, 165) = 6.36$ ,  $MSE = .303$ ,  $p < .01$ . This main effect was qualified by an interaction between translation ambiguity and Stroop score in the participant analysis,  $F_1(1, 16) = 5.52$ ,  $MSE = .009$ ,  $p < .05$ ;  $F_2(1, 165) = 1.27$ ,  $MSE = .096$ ,  $p = .26$ . We explored this interaction by computing the ambiguity effect for each participant (mean translation-unambiguous accuracy minus mean translation-ambiguous accuracy) and correlating that effect with Stroop score. Stroop score and translation ambiguity effect were

significantly correlated,  $r(19) = -.50, p < .05$ , demonstrating that individuals who experienced higher Stroop interference were less affected by ambiguity. Examining the correlations with Stroop separately for the translation-unambiguous ( $r = -.34$ ) and translation-ambiguous ( $r = .13$ ) words suggests that the translation-unambiguous words drive this effect. In other words, the moderate negative correlation shows that for translation-unambiguous words, individuals who were less susceptible to Stroop interference had higher accuracy; in contrast, the small positive correlation shows that for translation-ambiguous words, individuals who were less susceptible to Stroop interference no longer had an advantage (and if anything, had slightly lower accuracy than individuals who experienced more Stroop interference).

According to the item analysis only, individuals with less Stroop interference had higher accuracy than those with more Stroop interference (by items: 41 vs. 35 %),  $F_1 < 1$ ;  $F_2(1, 165) = 4.64, MSE = .096, p < .05$ . There was also an interaction between WM span and Stroop score in the analysis by items,  $F_2(1, 165) = 6.32, MSE = .075, p < .05$  (see Figure 1). Exploration of the means and 95% confidence intervals to interpret this interaction indicated that individuals with lower WM span had similar accuracy regardless of their Stroop interference (both means = 38% and within each other's confidence intervals). By contrast, individuals with higher WM span and less Stroop interference had higher accuracy than individuals with higher WM span and more Stroop interference (45 vs. 31 %, which are not within each other's confidence intervals).

**Figure 1.** Mean translation accuracy scores as a function of WM span and the amount of Stroop interference (means taken from the  $F_2$  analysis).



#### 4.2.2. Reaction time

Correct trials with RTs under 300 ms or over 3000 ms were treated as voice key errors and discarded from further analysis. Of the remaining data, all trials with RTs more than 2.5 standard deviations above or below the mean were discarded as outliers. These two procedures resulted in removal of 3.5% and 2.3% of the data, respectively.

In the  $F_1$  analysis, the translation-ambiguous RT for one participant was replaced with the mean because the participant had 0% accuracy for that condition. The  $F_2$  analysis had many missing cells due to a large number of words that did not have valid RTs in at least one of the four Stroop x WM span conditions; this analysis was run with only 4 translation-ambiguous words and 57 translation-unambiguous words, and therefore should be interpreted with caution.

Individuals with less Stroop interference translated more quickly than individuals with more Stroop interference (1242 vs. 1441 ms),  $F_1 < 1$ ;  $F_2(1, 59) = 5.95$ ,  $MSE = 99372.95$ ,  $p < .05$ . Although this effect was not significant in the  $F_1$  analysis, amount of Stroop interference moderately correlated with translation-unambiguous RT,  $r(19) = .36$ , but not with ambiguous RT,  $r(19) = .09$ . No other main effects or interactions were statistically significant.

### 5. Discussion

As we predicted, and as first demonstrated by Tokowicz and Kroll (2007), translation-ambiguous words were more difficult to translate than their unambiguous counterparts. In the current study, even when translation-ambiguous and unambiguous words were matched on important word-specific characteristics, accuracy for translation-ambiguous words was 25% lower than accuracy for translation-unambiguous words. Although previous research has also demonstrated that translation-ambiguous words are translated more *slowly* than translation-unambiguous words, the RT difference was not significant in the present study, most likely because of low overall accuracy and the resulting empty cells in the analysis. Notably, the accuracy of participants in the study by Tokowicz and Kroll was considerably higher than in the current study. Importantly, the present study exemplifies that the disadvantage associated with translation-ambiguity is not limited to proficient bilinguals (e.g., Tokowicz & Kroll, 2007) or to learners at the very early stages of learning (Degani & Tokowicz, 2010a), but holds even for learners at the intermediate stages of learning an L2.

Most interestingly, the current results show that overall performance on the

translation task was related to individual differences in performance on the Stroop task. Individuals who were better able to control Stroop-type interference translated more accurately, and tentative evidence from the RT analysis suggests that they also translated more quickly. These findings can be explained by two different (but not necessarily incompatible) accounts: a processing account and a learning account.

A processing account would suggest that the ability to deal with interference is generally advantageous in bilingual *processing* tasks such as translation because inhibitory control is necessary to manage activation from two languages (Green, 1998). In the current experiment, individuals who were less able to ignore task-irrelevant information may have had more difficulty inhibiting L1 to retrieve and produce responses in L2.

A learning account is consistent with recent claims that inhibitory control is critical for *acquiring* an L2. Levy et al. (2007) demonstrated that L1 phonology is inhibited when learners name objects in their relatively weak L2. Similarly, Linck et al. (2009) found that learners immersed in their L2 showed evidence of inhibiting L1 phonology and outperformed their classroom-only counterparts on L2 comprehension and production. Furthermore, Michael and MacWhinney (2003) found that individuals who experienced less Stroop interference more successfully learned foreign language vocabulary words in a training study.

Although participants in the current study were not at the very beginning stages of L2 learning, they also were not highly proficient. To successfully translate, they not only needed to access L2 words and produce them under time pressure, but they first needed to know the L2 translations. If the ability to ignore task-irrelevant information helps learners acquire L2 vocabulary, then participants who showed less Stroop interference may have been more successful language learners in general and thus able to translate more accurately.

Importantly, this accuracy advantage held only for translation-unambiguous words. For translation-ambiguous words, accuracy was only weakly correlated with Stroop performance; individuals with less Stroop interference actually experienced a slight (though non-significant) decrement in accuracy when translating ambiguous words relative to individuals with more Stroop interference. This finding is not well explained by the processing account, which would predict that, if anything, the advantages associated with successful management of interference should be greater for ambiguous than unambiguous words because translating ambiguous words is more difficult and requires more control.

The learning account may provide a better explanation of these findings. If

individuals with less Stroop interference are also better at L2 vocabulary learning, they would be expected to know more words than their counterparts who have more Stroop interference. A larger L2 vocabulary would lead to higher accuracy in general, as seen in the main effect of Stroop performance on accuracy and the particularly large effect for unambiguous words. However, being better vocabulary learners means that the individuals with less Stroop interference are also more likely to know more of the alternate translations for the ambiguous words. A so-called translation-ambiguous word is only ambiguous if the individual knows more than one translation equivalent. For individuals with more Stroop interference, some of the ambiguous words may have been functionally unambiguous. If the participants with less Stroop interference had larger L2 vocabularies, they might have been more likely to suffer from response competition as well as from weaker associative connections associated with multiple translations. This link between increased word knowledge and the translation-ambiguity disadvantage has also been proposed to explain why more-proficient bilinguals experience a larger disadvantage (Tokowicz, 2005; see Tokowicz & Degani, in press).

One of the most intriguing aspects of the current results is that WM span did not correlate significantly with performance on either the Stroop task or the translation task. The current finding of no correlation between WM span and Stroop interference is somewhat surprising in light of numerous studies suggesting that inhibition is an important component of WM (see Michael & Gollan, 2005, for a review). For example, in a set of five experiments using measures very similar to the current tasks, Kane and Engle (2003) demonstrated that individual differences in WM span predicted performance on a Stroop task. Kwong See and Ryan (1995) also observed overlap between the two constructs. However, not all researchers have found a relationship between WM span and Stroop performance. Michael and MacWhinney (2003) found no correlation between the two measures, and Long and Prat (2002) found that the relationship depended on the proportion of incongruent trials in the Stroop task. This lack of consensus in the literature suggests that although the constructs of WM capacity and the ability to ignore task-irrelevant information are probably closely related, they may be different enough to make distinct contributions to L2 acquisition and processing, as in the current study.

We predicted that both Stroop performance and WM span would be associated with better performance on the translation task, but as previously noted the prediction held true only for Stroop performance. However, WM span still played an important role in the overall pattern of results in the form of an interaction between WM span and performance on the Stroop task. As seen in Figure 1, individuals with lower WM span performed similarly regardless of their susceptibility to Stroop interference. In contrast, the performance of individuals with higher WM

span was related to their ability to ignore task-irrelevant information. In particular, individuals with higher WM span and less Stroop interference performed better than the individuals with lower WM span. However, higher WM span was actually a *disadvantage* for individuals who had more Stroop interference. This pattern of results is consistent with both learning and processing accounts of the mechanism by which individual differences affect translation performance. Having higher span may have allowed individuals to both know and activate a large number of related words for each stimulus, but that activation is only useful if the individual is able to control it. With too much activation from incorrect responses, participants with higher span but difficulty ignoring task-irrelevant information may have been unable to successfully activate the target word above the threshold required to make a response (e.g., Tokowicz et al., 2004).

It is interesting to note that the interaction between WM span and Stroop performance did not vary as a function of translation ambiguity, although processing of translation-ambiguous words should have entailed activation of competing alternative translations. However, it is possible that individuals with higher WM span activated a number of related words even for translation-unambiguous stimuli (e.g., Tokowicz et al., 2004).

## 6. Conclusion

The translation accuracy results in the current study confirm previous findings that translation-ambiguous words are more difficult to translate than unambiguous words. The individual-difference results suggest that both WM span and the ability to ignore task-irrelevant information may have been related to translation accuracy not only because of the processing demands of the task, but also because individual differences in cognitive skill may affect vocabulary learning. The more L2 words an individual knows, the larger the potential effect of translation ambiguity. The different patterns of results for the O-Span and Stroop tasks – and the lack of correlation between the two – call for a better understanding of the nature of these tasks and the roles of WM capacity and ability to ignore task-irrelevant information in acquiring and using an L2.

## Acknowledgements

This research was supported in part by a National Research Service Award (NIMH HD41307-02) to Erica B. Michael. TD and NT were supported by NSF-BCS 0745372 to Natasha Tokowicz during the writing of this manuscript. We also thank

Lisa Auslander and members of the PLUM Lab at the University of Pittsburgh for their assistance in collecting and coding the data, and we thank all of the participants for their time and effort.

## References

- Conway, A. R. A., N. Cowan, and M. F. Bunting. 2001. "The cocktail party phenomenon revisited: The importance of working memory capacity." *Psychonomic Bulletin & Review* 8 (2): 331-335.
- Conway, A. R. A., and R. W. Engle. 1994. "Working memory and retrieval: A resource-dependent inhibition model." *Journal of Experimental Psychology: General* 123 (4): 354-373.
- Degani, T., and N. Tokowicz. 2010a. "Ambiguous words are harder to learn." *Bilingualism: Language and Cognition* 13 (3): 299-314.
- \_\_\_\_\_. 2010. "Semantic ambiguity within and across languages: An integrative review." *Quarterly Journal of Experimental Psychology* 63 (7): 1266-1303.
- E-Prime [Computer software]. Pittsburgh, PA: Psychology Software Tools.
- Francis, W. N., and H. Kucera. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin Co.
- Gernsbacher, M. A. 1984. "Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy." *Journal of Experimental Psychology: General* 113 (2): 256-281.
- Green, D. W. 1998. "Mental control of the bilingual lexico-semantic system." *Bilingualism: Language and Cognition* 1 (2): 67-81.
- Gunter, T. C., S. Wagner, and A. D. Friederici. 2003. "Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories." *Journal of Cognitive Neuroscience* 15 (5): 643-657.
- Harrington, M., and M. Sawyer. 1992. "L2 working memory capacity and L2 reading skill." *Studies in Second Language Acquisition* 14 (1): 25-38.
- Kane, M. J., M. K. Bleckley, A. R. A. Conway, and R. W. Engle. 2001. "A controlled-attention view of working-memory capacity." *Journal of Experimental Psychology: General* 130 (2): 169-183.
- Kane, M. J., and R.W. Engle. 2003. "Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference." *Journal of Experimental Psychology: General* 132 (1): 47-70.
- Kempe, V., P. J. Brooks, and A. Kharkhurin. 2010. "Cognitive predictors of

generalization of Russian grammatical gender categories.” *Language Learning* 60 (1): 127-153.

Kroll, J. F., E. Michael, and A. Sankaranarayanan. 1998. “A model of bilingual representation and its implications for second language acquisition.” In A. F. Healy and L. E. Bourne, Jr. (eds) *Foreign language learning: Psycholinguistic experiments on training and retention*, Mahwah, NJ: Lawrence Erlbaum Associates, 365-395.

Kwong See, T., and E. B. Ryan. 1995. “Cognitive mediation of adult age differences in language performance.” *Psychology and Aging* 10 (3): 458-468.

Levy, B. J., N. D. McVeigh, A. Marful, and M. C. Anderson. 2007. “Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition.” *Psychological Science* 18 (1): 29-34.

Linck, J. A., J. F. Kroll, and G. Sunderman. 2009. “Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning.” *Psychological Science* 20 (12): 1507-1515.

Long, D. L., and C. S. Prat. 2002. “Working memory and Stroop interference: An individual differences investigation.” *Memory & Cognition* 30 (2): 294-301.

Michael, E. B., and T. Gollan. 2005. “Being and becoming bilingual: Individual differences and consequences for language production.” In J. F. Kroll and A. M. B. de Groot (eds) *Handbook of Bilingualism: Psycholinguistic Approaches*, New York: Oxford University Press, 389-407.

Michael, E. B., and B. MacWhinney. 2003, November. “Span, Stroop, and second language acquisition.” Poster presented at the 44<sup>th</sup> Annual Meeting of the Psychonomic Society, Vancouver, British Columbia, Canada.

Michael, E. B., N. Tokowicz, and J. F. Kroll. 2003, April. “Modulating access to L2 words: The role of individual differences and language immersion experience.” Paper presented at the Fourth International Symposium on Bilingualism, Tempe, AZ.

Miyake, A. 1998. “Individual differences in second language proficiency: The role of working memory.” In A. F. Healy and L. E. Bourne, Jr. (eds) *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*, Mahwah, NJ: Erlbaum, 339-364.

Miyake, A., P. A. Carpenter, and M. A. Just. 1994. “A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients.” *Cognitive Neuropsychology* 11 (6): 671-717.

Miyake, A., M. A. Just, and P. A. Carpenter. 1994. “Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral context.” *Journal of Memory and Language* 33 (2): 175-202.

Paivio, A., J. C. Yuille, and A. Madigan. 1968. "Concreteness, imagery, and meaningfulness values for 925 nouns." *Journal of Experimental Psychology Monograph Supplement* 76 (1, Pt. 2): 1-25.

Prior, A., B. MacWhinney, and J. F. Kroll. 2007. "Translation norms for English and Spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity." *Behavior Research Methods* 39 (4): 1029-1038.

Schwanenflugel, P. J., and E. J. Shoben. 1983. "Differential context effects in the comprehension of abstract and concrete verbal materials." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9 (1): 82-102.

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

Tokowicz, N. 1997. *Reevaluating concreteness effects in bilingual translation*. Unpublished MS thesis. Department of Psychology, The Pennsylvania State University.

Tokowicz, N., and T. Degani. In press. "Translation ambiguity: Consequences for learning and processing." In B. VanPatten and J. Jegerski (eds) *Research on Second Language Processing and Parsing*. Amsterdam: John Benjamins.

Tokowicz, N. and J. F. Kroll. 2007. "Number of meanings and concreteness: Consequences of ambiguity within and across languages." *Language and Cognitive Processes* 22 (5): 727-779.

Tokowicz, N., J. F. Kroll, A. M. B. De Groot, and J. G. Van Hell. 2002. "Number-of-translation norms for Dutch-English translation pairs: A new tool for examining language production." *Behavior Research Methods, Instruments & Computers* 34 (3): 435-451.

Tokowicz, N., E. B. Michael, and J. F. Kroll. 2004. "The roles of study-abroad experience and working-memory capacity in the types of errors made during translation." *Bilingualism: Language and Cognition* 7 (3): 255-272.

Tokowicz, N., and T. Warren. 2008. "Quantification and statistics." In L. Wei and M. Moyer (eds) *The Blackwell Guide to Research Methods in Bilingualism and Multilingualism*, Malden, MA: Blackwell Publishers, 214-231.

Trude, A. M., and N. Tokowicz. In press. "Transfer from Spanish and English to Portuguese pronunciation: The roles of inhibition and working memory." *Language Learning*.

Turner, M. L. and R. W. Engle. 1989. "Is working memory capacity task dependent?" *Journal of Memory and Language*, 28 (2): 127-154.

