RESEARCH NOTE

The roles of study-abroad experience and working-memory capacity in the types of errors made during translation*

NATASHA TOKOWICZ
Department of Psychology and the Learning Research and Development Center, University of Pittsburgh
ERICA B. MICHAEL
Carnegie Mellon University
JUDITH F. KROLL
Department of Psychology, The Pennsylvania State University

We examined the effects of study-abroad experience (SAE) and working-memory capacity (WMC) on the types of errors made during single-word translation from the first language to the second language, contrasting non-response with meaning errors (i.e. when individuals translate semantically-related words instead of the target word). SAE and WMC interacted; individuals with more SAE and higher WMC made as many meaning as non-response errors, whereas individuals in the other groups made more non-response than meaning errors. We conclude that SAE encourages the use of approximate translations to communicate, but only higher WMC learners can do so because this strategy requires multiple items to be maintained in memory simultaneously. A speech-production model is adapted to capture our results and demonstrate the effects of differential working memory demands on producing correct translations, meaning errors, and non-response errors.

Two of the most common assumptions about learning a second language as an adult are that some individuals are more gifted than others at second-language learning (e.g. Segalowitz, 1997), and that studying abroad improves one's ability to speak a second language (e.g. Freed, 1995). Despite these very common assumptions, relatively little psycholinguistic research has investigated the SPECIFIC effects of individual differences in cognitive ability or study-abroad experience (SAE) on second language learning and performance; we examined the roles of these two factors in translation production, a common bilingual task that involves seeing a word in one language and translating it aloud into the other language. In particular, we examined the effects of the relative amounts of SAE and working-memory capacity (WMC) on the kinds of errors that bilinguals made during translation from the dominant language (L1) into the non-dominant language (L2). Translation is a task that has been used extensively to investigate lexical access in bilingual production (e.g. De Groot, 1992; De Groot, Dannenburg and Van Hell, 1994; Kroll and Stewart, 1994). Normally, correct response latencies are analyzed. Error rates typically follow the pattern of the reaction time data, but the types of errors are not ordinarily analyzed (cf. Talamas, Kroll and Dufour, 1999). Despite a lack of attention to the types of errors made during bilingual translation, much research has examined speech errors in monolinguals.

Research on monolingual speech has led to the classification of several types of speech errors, based on the level at which the error is thought to have occurred. Specifically, SEMANTIC or MEANING ERRORS, which occur when a word related in meaning to the intended word

* Natasha Tokowicz (NT), Learning Research and Development Center and the Center for the Neural Basis of Cognition, University of Pittsburgh; Erica B. Michael (EBM), Department of Psychology, Carnegie Mellon University; Judith F. Kroll (JFK), Department of Psychology and Program in Linguistics and Applied Language Studies, The Pennsylvania State University.

This research was supported in part by NSF Grants BCS-9905850 and BCS-0111734 and NIMH Grant MH62479 to JFK at The Pennsylvania State University, and from grants to NT from the Leibowitz Fund, The Department of Psychology at The Pennsylvania State University, the grant-in-aid of research program at Sigma Xi, and by a matching fund grant from The Pennsylvania State University Chapter of Sigma Xi. During writing of this manuscript, NT was supported by a National Institutes of Health Individual National Research Service Award (NIH HD42948-01), and EBM was supported by a National Institutes of Health Individual National Research Service Award (NIH HD41307-01). A portion of these data were collected at the Cemanahuac Educational Community. The support and resources of that institution are gratefully acknowledged. We thank Vivian Harvey, Harriet Guerrero-Goff, and Charles Goff for their assistance. We thank Gustavo Perry for his assistance in translating the language history questionnaire into Spanish, and Ana Teresa Pérez-Leroux for translating the task instructions into Spanish. We thank Emily Barth, Ryan Gilligan, Dayne Grove, Adam Issan, José "Tony" Matamoros, Israel Roling, Heather Shrigley, and Jan Toney for their invaluable research assistance. We also thank Robert DeKeyser, Susan Dunlap, Peter Gianaros, Tamar Gollan, Erik Reichle, Ana Schwartz, Simone Sprenger, and three anonymous reviewers for helpful comments on an earlier version of this manuscript. This paper is based on a presentation at the 4th International Symposium on Bilingualism in Tempe, Arizona, in April, 2003.

Address for correspondence
Natasha Tokowicz, 604 Learning Research and Development Center, 3939 O'Hara St., University of Pittsburgh, Pittsburgh, PA 15260, USA
E-mail: *Tokowicz@alumni.umass.edu*

is said (e.g. saying *goat* for *sheep*), are contrasted with PHONOLOGICAL ERRORS, which occur when a word related in sound to the intended word is said (e.g. saying *sheet* for *sheep*). Semantic errors have also been contrasted with OMISSION or NON-RESPONSE ERRORS, which occur when lexical retrieval fails (e.g. Blanken, Dittmann and Wallesch, 2002). Models of language production (e.g. Levelt, 1983; Dell, 1986, 1989; Peterson and Savoy, 1998) have described the general process of converting thoughts into speech (LEXICALIZATION), and are classified into SERIAL and CASCADING PROCESSING MODELS. The main distinction between these two classes of models is that strictly serial models (e.g. Levelt, 1983, 1989) propose that a word is selected to express the intended meaning (LEXICAL SELECTION), and only then is the word's phonological form specified (PHONOLOGICAL ENCODING). In contrast, cascading models (e.g. Dell, 1986; Peterson and Savoy, 1998) allow for overlap between lexical selection and phonological encoding. For the purposes of the present study we remain agnostic regarding this important distinction, but we point out that these models all share the critical idea that during lexical selection in translation, an incorrect, but related, word may be chosen.

The basis for our interest in the relationship between SAE and types of errors was a study in which DeKeyser (1991a, b) compared study-abroad and classroom L2 learners. DeKeyser found that in picture descriptions and interviews conducted in L2, study-abroad learners were more likely than classroom learners to use CIRCUMLOCUTION (i.e. filling in gaps in knowledge by making use of descriptions or definitions that include known words) and RESTRUCTURING (i.e. making adjustments by rephrasing when they realize there is a gap in what they've already started to say). Conversely, classroom learners were more likely than study-abroad learners to use DIRECT and INDIRECT APPEAL (i.e. asking for a translation or indicating that they didn't know the correct word, respectively), and literal translation from L1 to L2, which can lead to errors in L2 (e.g. translating *fireplace* as *lugar del fuego*, which literally means 'place of fire', instead of as *chimenea*). In other words, study-abroad learning may encourage individuals to use an alternative word or phrase with a similar meaning when they don't know the correct word, possibly reflecting a necessity or desire to communicate that is especially salient in the study-abroad context. These findings suggest that the assumption of increased proficiency associated with study abroad may at least partly reflect differences in communication strategies between learners with and without SAE.

Because WMC has been shown to influence performance in both L1 and L2 processing, we were also interested in whether individuals with higher WMC are unique in their ability to take advantage of SAE compared to individuals with lower WMC (e.g. Sunderman and Kroll, 2003). Working memory has been a central part of theories of cognitive processing, because it is used to construct and store intermediate products of language processing (e.g. Just and Carpenter, 1992; Reichle, Carpenter and Just, 2000). For example, within L1, WMC correlates with the ability to retrieve facts from a passage (e.g. Daneman and Carpenter, 1980), resolve ambiguity (e.g. Daneman and Carpenter, 1983), and infer meaning from context (e.g. Daneman and Green, 1986). In bilingual processing, WMC correlates with the acquisition of L2 syntactic cues (e.g. Miyake, 1998) and single-word translation performance (e.g. Kroll, Michael, Tokowicz and Dufour, 2002; see Michael and Gollan, in press, for a review). WMC is also related to the ability to efficiently allocate cognitive resources (e.g. Just and Carpenter, 1992; Engle, 2002; Michael and Gollan, in press). Thus, it is possible that individuals with higher WMC will be better able to take advantage of the study-abroad context because of their greater working-memory capacity *per se* and/or greater resource-allocation ability.

In the present study, we examined the effects of the relative amount of L2 SAE and individual differences in L1 WMC on single-word translation performance from the dominant to the non-dominant language. We focused on two particular types of errors that learners made: NON-RESPONSE ERRORS, which occurred when participants said that they didn't know the answer or said nothing at all, and MEANING ERRORS, which occurred when participants made incorrect responses that were related in meaning to the correct translation. Meaning errors were of interest because they suggest the desire to communicate without knowledge of the correct word (similar to circumlocution). Note that although other types of errors were made by the participants, including FORM-RELATED ERRORS (e.g. when a word that is orthographically or phonologically similar to the correct response was given) and errors that fit into more than one category (e.g. both a meaning and a form error), these kinds of errors were not theorized to differ as a result of SAE and WMC, and therefore were not analyzed in the present study. In addition, although form errors are of particular interest because they are widely studied in speech error research, there were too few in our sample to include in the analyses. There were several trials on which individuals gave more than one response (e.g. gave the correct answer after an incorrect answer); these trials were also excluded from the analyses. In our analyses, we focused on the percentage of the total number of errors that were meaning errors versus non-response errors. In this way, we were able to examine the frequency of these types of errors independent of differences in the total number of errors made.

We hypothesized that individuals with more SAE would know more words than individuals with less SAE due to greater exposure to L2 words, and that this knowledge difference would be reflected in the overall accuracy rate (e.g. Michael, Tokowicz and Kroll, 2003). Similarly, we hypothesized that individuals with higher WMC would know more words than individuals with lower WMC (e.g. Michael et al., 2003) because learning an L2 word requires co-activation of the L2 word form and either the concept or the L1 equivalent (e.g. Kroll, Michael and Sankaranarayanan, 1998); therefore, individuals with greater ability to maintain multiple items in working memory should be better able to form these connections.

We further hypothesized that, ABOVE AND BEYOND THE WORD-KNOWLEDGE EFFECTS DESCRIBED ABOVE, learners with more SAE would be more accustomed to communicating without precise word translations and should therefore be less prone to non-response errors than learners with less SAE. This increase in communicating without precise translations may result in strategic effects on the part of the speaker, or in a shift in the criterion for precision in translation. Whichever the reason for the change, individuals with more SAE, given adequate cognitive resources, should be more likely to make meaning errors than individuals with less SAE. Because great demands are placed on working memory when one attempts to translate a particular word and instead produces a related word, only individuals with more SAE AND higher WMC were predicted to make a high percentage of meaning errors. We made this prediction because meaning errors may require the stimulus to be maintained in working memory while related words are activated. For example, if an individual is asked to translate *dog*, but only knows the translation for *cat*, then *dog*, *cat*, and *gato* (the Spanish translation of *cat*) must be maintained in memory until a response is made. Moreover, meaning errors may tax working memory because the related concept ("cat") must be activated to a greater extent than the concept activated by the stimulus (*dog* in our example), so that *cat* can be selected for translation. In this latter case, making a meaning error requires more efficient allocation of cognitive resources than making a non-response error. Thus, meaning errors may require additional WMC because of a large demand for activation and/or allocation ability, so that individuals with higher WMC should be better able to use techniques like circumlocution to fill in gaps in knowledge, causing these individuals to make a higher percentage of meaning errors than individuals with lower WMC. Thus, we predicted an interaction between SAE and WMC such that individuals with more SAE and higher WMC would be more likely than individuals in the other three groups to make a substantial percentage of meaning errors.

Method

Participants

Participants were 37 individuals who were relatively proficient in English and Spanish (15 native Spanish speakers and 22 native English speakers), and who varied widely in SAE. Participants were paid or given psychology course credit for their participation.

General procedure

The participants were tested at either The Pennsylvania State University or the Cemanahuac Educational Community, Cuernavaca, Morelos, Mexico. To properly compare the participants in the two testing environments, the instructions were given in the primary language of the testing environment. That is, both the verbal and written instructions were given in English to participants tested in the United States and in Spanish to participants tested in Mexico, regardless of the participants' native language.

The participants completed the memory-span task in their native language and then the translation task from L1 to L2. Therefore, native English-speaking participants translated from English to Spanish, and native Spanish-speaking participants translated from Spanish to English. Each participant then completed a language history questionnaire in which she or he reported on her or his L1 and L2 language learning experiences. The language history questionnaire was administered in English in the U.S. and in Spanish in Mexico. The data from the language history questionnaire were later used to group participants as a function of their study-abroad experience. Participants' self-rated L2 proficiency and their translation accuracy were used to control for the *a priori* effects of word knowledge that existed between the groups.

Memory-span task

We used a version of the Turner and Engle (1989) OPERATIONS-WORD TASK to measure individual differences in WMC. Participants solved mathematical expressions while maintaining sets of words in memory. They saw mathematical expressions on the computer screen, displayed one at a time, and then pushed buttons to indicate whether or not each expression appeared with the correct answer. A word was presented after each mathematical expression (see Appendixes A and B for the operation and word stimuli, respectively). For example, a participant may see "$(4 + 2) - 3 = 10$", after which she or he should push the "no" button, and then the word *tiger* would appear. The sets ranged in size from two to six operation–word pairs, and were presented in increasing order of size with three sets of each size. Each set contained an approximately equal number of expressions requiring "yes" and "no" responses.

Two versions of the word materials were created so that the words were always presented in the participant's L1. The words in each set size and in each presentation (first, second, or third) of a given set size were also matched on length in English and Spanish, and on English word

Table 1. *Properties of the stimuli used in the translation task.*

	Concreteness			
	Abstract		Concrete	
Measure	Noncognates	Cognates	Noncognates	Cognates
English length (# of letters)	5.5	6.1	5.2	6.0
Spanish length (# of letters)	6.3	6.4	5.8	6.5
Concreteness rating (English; $N = 13$)	3.8	3.6	6.0	6.0
Word frequency (English)	110.5	110.8	104.4	102.4

Note: Concreteness ratings were performed using a 7-point scale, where 1 represented least concrete and 7 indicated most concrete. English word frequency was taken from Francis and Kučera (1982). In English, the cognates were reliably longer than the noncognates in length, $t(158) = 2.65$, $p < .01$ (all other $ps > .10$).

frequency (Francis and Kučera, 1982). The participants were instructed to attempt to remember all of the words from that set while responding as quickly and accurately as possible to the expressions. Participants were told that they should try to recall the words from each set in the order in which they were presented, but not to be too concerned if they could not remember the correct order.

Participants were given two practice sets (one with four trials and one with six trials). On each trial, participants saw a fixation point (+) for 1000 ms, then an expression for 2500 ms. Participants responded to the expressions using the keys marked either "Y" or "N" during the presentation of a question mark (?), which was displayed for 1250 ms. Participants then saw a word for 1250 ms, and then the fixation point reappeared, which started the next trial. After the last word in the set was displayed, a "RECALL" prompt appeared, at which time participants wrote the words from that set in a booklet. WMC was defined as the set size at which a participant was able to recall all of the words from at least two of three sets. The accuracy of the order of the words was not factored into the calculation of span.

Translation task

The stimuli used in the translation task were 160 English words and their Spanish translations (see Appendix C). All of the words had a clearly dominant translation in English and in Spanish. Half of the words were concrete (i.e. they referred to an imaginable referent, e.g. *cat*), and the other half were abstract (e.g. *beauty*). Approximately one third of the words had translations in the other language that were COGNATES (i.e. similar in meaning and lexical form across languages, e.g. *color* [kul'ər] in English translates to *color* [ko•lór] in Spanish). The items from the four conditions (concrete noncognates, concrete cognates, abstract noncognates, and abstract cognates) were matched on English word frequency (Francis and Kučera, 1982) and on word length in both English and Spanish (see Table 1 for more details about the stimuli). These four kinds of words were included to enhance the generalizability of our findings to words of various types.

The words were presented to each participant in a randomized order that was generated by the computer program at the start of each run. Participants were given practice trials and had an opportunity to ask questions about the task before beginning the critical trials. Each trial began with a fixation cross appearing until the participant pressed a key on the computer keyboard. A word was then presented for 2000 ms or until the participant made a verbal response. Responses were tape recorded for later coding of accuracy. Participants were instructed to respond as quickly and accurately as possible, and to say "no" when they were unable to translate a given word.

Apparatus

The experiment was run on a Macintosh computer using SuperLab software (Cedrus Corp.; Haxby, Parasuraman, Lalonde and Abboud, 1993) to present the stimuli. This program also recorded response times in ms from the onset of the stimulus to the button press (in the WMC task) or to the onset of articulation (in the translation task).

Language history questionnaire

Participants completed a language history questionnaire (see Appendix D) in which they reported on their L2 learning experiences. Each participant rated his or her L1 and L2 reading, writing, conversational, and speech comprehension proficiency on scales from one to ten, and indicated the age at which L2 learning began and the types of exposure he or she had learning the L2. One participant's average self-ratings were higher in her L2 than in her L1, thus we considered L2 to be her dominant

Table 2. *Language history questionnaire data for the revised set of participants.*

	Lower WMC		Higher WMC	
Measure	Less SAE $N = 6$	More SAE $N = 6$	Less SAE $N = 12$	More SAE $N = 8$
Age (years)[a,b,c]	22.88	26.56	21.42	35.38
Age began L2 (years)	15.29	14.13	13.83	14.00
Time studied L2 (years)[b,c]	4.63	9.00	6.25	21.88
L2 study-abroad experience (months)[b]	2.19	65.11	1.07	159.00
L2 proficiency rating[b]	7.17	7.83	7.42	9.22
Set size span[c]	3.83	3.33	5.50	5.75

Note: Reading, writing, conversation, and speech comprehension ability were rated on a 10-point scale, where 1 indicated the lowest level of ability and 10 indicated the highest level of ability.
[a] interaction between SAE and WMC
[b] main effect of SAE
[c] main effect of WMC

language (hereafter referred to as L1). In addition, data from four participants who rated themselves as equal in L1 and L2 but who were living in an L2 environment were recoded as L2-dominant.

Results

Study-abroad experience data

Participants were grouped on the basis of their reported amount of SAE. The 20 individuals with 8 or fewer months of SAE were considered to have "less" SAE, and the 17 individuals with 15 or more months of SAE were considered to have "more" SAE. Although these criteria are arbitrary, they are based on the bimodal distribution of months of reported SAE in our test sample.

Working-memory capacity data

The mean working memory span of participants in this study was 4.4 (range = 0–6). We classified the 17 participants with a span of 0–4 as being "lower WMC" and the 20 participants with a span of 5–6 as being "higher WMC". On average, the participants responded correctly to 76% of the mathematical expressions (range = 30%–98%). Six participants responded below chance on the mathematical portion of this task and all six had lower working-memory capacity, suggesting that their poor performance on the expressions was not due to a strategy of focusing on the words to the detriment of the expressions.

Using the SAE and WMC criteria in combination, there were 8 participants with less SAE and lower WMC, 12 participants with less SAE and higher WMC, 9 participants with more SAE and lower WMC, and 8 participants with more SAE and higher WMC.

Differences between groups

Using an analysis of variance, we compared the four groups of participants on their age, the age at which they began to learn L2, the number of years they had studied L2, their L2 proficiency ratings, the number of months of SAE, and their working memory span. There were reliable group differences on several of these dimensions. Critically, although the same WMC criteria were used for all participants, the group with less SAE and lower WMC had higher working memory span than the group with more SAE and lower WMC, as evidenced by an interaction between SAE and WMC, $F(1, 33) = 8.71$, $p < .01$. Because we are interested in group differences, this confound could be an alternative explanation for our results; therefore, we excluded 5 participants so that there were 6 participants in each of the two lower WMC groups, and these groups were matched on their span [$F(1, 28) = 3.35$, $p > .05$]. As a result, our analyses did not include any participants with a span of 0 or 1. All subsequent analyses that we report are based on the revised set of 32 participants.

The language history questionnaire data are shown in Table 2. There were still some remaining group differences after removing the 5 participants. First, participants in three of the four groups had approximately the same age (between 21 and 23 years of age, on average), but the individuals with more SAE and higher WMC were significantly older (35 years); $F(1, 28) = 6.91$, $p < .05$. This interaction qualifies two main effects on age: individuals with more SAE were older than individuals

Table 3. *Overall accuracy (%) for participants in the four groups.*

	Lower WMC		Higher WMC	
Measure	Less SAE $N = 6$	More SAE $N = 6$	Less SAE $N = 12$	More SAE $N = 8$
Raw accuracy score	62.50	68.75	69.74	81.80
Adjusted accuracy score	69.43	68.67	71.96	71.20

with less SAE, $F(1, 28) = 11.17$, $p < .01$, and individuals with higher WMC were older than individuals with lower WMC; $F(1, 28) = 7.88$, $p < .01$. As expected, participants with more SAE rated their L2 proficiency as higher than participants with less SAE; $F(1, 28) = 11.25$, $p < .01$. Furthermore, participants with more SAE had studied L2 longer than participants with less SAE; $F(1, 28) = 16.05$, $p < .01$. Similarly, participants with higher WMC had studied L2 longer than participants with lower WMC; $F(1, 28) = 6.24$, $p < .05$. The effect of WMC was greater for participants with more SAE, as shown by the marginally-reliable interaction between SAE and WMC on number of years of L2 study; $F(1, 28) = 4.06$, $p = .054$. There were also the expected effects of the grouping factors on working memory span and number of months of SAE; namely, individuals with higher WMC had reliably higher working memory span than individuals with lower WMC; $F(1, 28) = 99.45$, $p < .01$; and individuals with more SAE had reliably more months of SAE than individuals with less SAE; $F(1, 28) = 10.12$, $p < .01$.

Study-abroad activities

The participants who were classified as having more SAE reported spending 72% of their time speaking their second language while in the study-abroad context. Most of the participants (64%) reported reading a newspaper or magazine in L2 three or more times a week; all of the participants reported reading one at least once a week. Most of the participants (71%) reported speaking with native speakers of their L2 at every possible opportunity, and all reported doing so at least occasionally.

General data analysis approach

Because of our focus on SAE and WMC, we decided to covary the effects of the factors that differed between the groups. Although our experimental design lends itself to the analysis of covariance approach to data analysis, there were several reasons that ANCOVA was not appropriate for our data analyses (Cohen and Cohen, 1983, p. 380). Because some of our covariates interacted with our independent variables (and these interactions accounted for additional variance when using the procedure recommended to determine whether ANCOVA is appropriate; pp. 388f.), we followed the data analysis technique suggested by Cohen and Cohen (p. 389). We used a hierarchical linear regression analysis, in which the covariates and their interactions with the independent variables were entered on earlier steps of the analysis so that we could determine whether the independent variables had unique predictive power above and beyond the covariates and their interactions with the independent variables.

We entered the variables that differed between the four groups (age, L2 proficiency ratings,[1] and the number of years of L2 study) on the first step of the analysis so that we could examine the effects of SAE and WMC without confoundings. In the main (type of error) analysis we also included the overall task accuracy as a covariate on the first step to further exclude the proficiency (word knowledge) differences among the groups. In addition, because there are interactions between the covariates and our independent variables, these interaction terms were entered on the second step of the analysis and also served as covariates.

Translation accuracy

We had predicted that individuals with more SAE would know more words than those with less SAE, and that this difference would be observed in overall accuracy rates. In addition, we predicted that individuals with higher WMC would know more words than individuals with lower WMC, and that this difference would be observed in overall accuracy rates. To test these predictions, we compared overall accuracy rates for the four groups of participants. The raw accuracy scores for the four groups (see Table 3) suggest that our predictions were correct; individuals with more SAE were more accurate than those with less SAE, and individuals with higher WMC were more accurate than those with lower WMC. However, these patterns were not confirmed by the statistical

[1] The four skill ratings were highly intercorrelated, and also reliably correlated with the average. Therefore, the average L2 proficiency rating was used in the analyses.

analyses. As described above, we used a hierarchical linear regression analysis in which we covaried the effects of some of the factors that differed between the groups; this method was used so that we could determine what differences in accuracy could be attributed to SAE and WMC without pre-existing differences in L2 knowledge or skill.

We covaried age, number of years of L2 study, and L2 proficiency ratings by including them on the first step of the analysis. In addition, we covaried the effects of the interactions between the covariates and the independent variables by entering them on the second step of the analysis. The independent variables were entered on the third step and the SAE by WMC interaction was entered on the fourth step of the analysis.

Examination of the R^2 statistic for each step showed that the covariates did significantly predict overall accuracy, $R^2 = .286$, $F(3, 28) = 3.736$, $p < .05$. In addition, the covariate by independent variable interactions predicted additional variance in overall accuracy, $\Delta R^2 = .309, F(6, 22) = 2.800, p < .05$. However, counter to our predictions, neither WMC nor SAE predicted additional variance beyond these factors. Similarly there was no interaction between WMC and SAE on overall task accuracy (see Table 3 for adjusted means). Thus, the groups of interest in our study did not differ in their overall accuracy on this task once the covariates and their interactions with the independent variables were included in the analysis.[2]

Types of errors

Our primary hypothesis in the present study was that only individuals with more SAE and higher WMC would have sufficient working-memory capacity and desire to communicate to make a substantial percentage of meaning errors. To test this hypothesis, we examined the percentage of total errors that were non-response and meaning errors made by each participant group in each word-type condition. We performed a hierarchical linear regression analysis in which we entered the covariates (age, number of years of L2 study, L2 proficiency ratings, and overall task accuracy) on the first step of the analysis. On the second step of the analysis we entered the interactions between the covariates and the independent variables. These procedures allow us to measure the effects of SAE and WMC INDEPENDENT OF L2 PROFICIENCY AND OVERALL TASK ACCURACY (I.E. WORD KNOWLEDGE) by examining the change in the proportion of variance (ΔR^2) accounted for by each step in the analysis as well as the associated semi-partial correlation coefficients (*sr*s). For purposes of brevity, the ßs and *sr*s for each factor are listed in Table 4.

Examination of the R^2 statistic for the first step showed that the covariates did significantly predict error types as expected, $R^2 = .045$, $F(4, 251) = 2.98$, $p < .05$. The interactions between the covariates and the independent variables also accounted for a significant proportion of variance, $\Delta R^2 = .388$, $F(20, 231) = 7.93$, $p < .01$. The third step, which included the independent variables using dummy coding (concreteness, cognate status, type of error, SAE, and WMC), also accounted for a significant proportion of the remaining variance, $\Delta R^2 = .046$, $F(5, 226) = 3.98$, $p < .01$. The fourth step of the analysis, which included the two-way interactions between the independent variables, was also significant, $\Delta R^2 = .060$, $F(10, 216) = 2.81$, $p < .01$. All of the preceding main effects and interactions are qualified by the three-way interactions in the significant fifth step, $\Delta R^2 = .044, F(10, 206) = 2.16$, $p < .05$. In this step, the only factors that were significant were concreteness by cognate status by type of error, $t = -2.34, p < .05$, and type of error by SAE by WMC, $t = 3.30$, $p < .01$. Because these three-way interactions qualify the previously described main effects and interactions, we will explore them in more detail.

To demonstrate the nature of the effects, for each interaction, estimated error percentages were calculated using the regression equation from the fifth step, following the procedure suggested by Aiken and West (1991). Because all of the critical factors were categorical, means were estimated for each group separately by entering the appropriate dummy code for these factors. For example, for the type of error by SAE by WMC interaction, 0 indicated non-response errors, less SAE, and lower WMC, and 1 indicated meaning errors, more SAE, and higher WMC. The average value was entered for all other variables. This procedure provides us with descriptive information about the errors for each group, adjusted for the covariates.

The estimated error percentages for each type of error are plotted in Figure 1. The overall pattern suggests that all of the groups except the more SAE higher WMC group made a higher percentage of non-response than meaning errors, whereas the more SAE higher WMC group made as many meaning as non-response errors. To test the reliability of this pattern, we performed follow-up tests by running hierarchical regression analyses separately for the four groups to determine whether the percentage of the two error types differed for each group. We entered the same covariates on the first step as we had in the overall

[2] However, the groups did differ in their mean reaction time. The mean reaction times for correct trials (excluding voice key errors) for the four groups were as follows: less SAE higher WMC = 1099 ms, less SAE lower WMC = 1127 ms, more SAE higher WMC = 1017 ms, and more SAE lower WMC = 1123 ms. After covarying age, non-dominant proficiency ratings, years of L2 study, and task accuracy, and their interactions with SAE and WMC, a hierarchical linear regression analysis showed that there was no effect of SAE ($p > .10$). However, individuals with higher WMC translated more quickly than individuals with lower WMC, $p < .05$. There was no interaction between the two factors on reaction time ($p > .10$).

Table 4. *Overview of the hierarchical regression analysis for variables predicting types of errors.*

	Step 1		Step 2		Step 3		Step 4		Step 5	
Variable	ß	*sr*	ß	*sr*	ß	*sr*	ß	*sr*	ß	*sr*
Age										
Non-dominant proficiency (Prof)			0.430*	0.112			0.413*	0.097	0.442*	0.103
Years studied L2 (Years)										
Task accuracy (Acc)	– 0.166*	– 0.138	– 0.84**	– 0.192	– 0.584*	– 0.097	– 0.655*	– 0.100	– 0.618*	– 0.094
Acc × Concreteness (Conc)	–	–								
Acc × Cognate status (Cog)	–	–	0.771*	0.122	0.883*	0.115	0.827*	0.099	0.893*	0.105
Acc × Type of error (Type)	–	–			1.058**	0.138				
Acc × SAE	–	–					1.398*	0.095	1.398*	0.095
Acc × WMC	–	–	1.674**	0.133						
Years × Conc	–	–								
Years × Cog	–	–								
Years × Type 2	–	–	0.699**	0.277	0.316*	0.095				
Years × SAE	–	–								
Years × WMC	–	–								
Prof × Conc	–	–								
Prof × Cog	–	–								
Prof × Type	–	–	– 0.651*	– 0.105						
Prof × SAE	–	–	– 1.317*	– 0.107						
Prof × WMC	–	–			– 1.314*	– 0.1	– 1.524*	– 0.112	– 1.524*	– 0.112
Age × Conc	–	–								
Age × Cog	–	–								
Age × Type	–	–	– 0.658**	– 0.141					– 0.647*	– 0.115
Age × SAE	–	–								
Age × WMC	–	–			– 1.654*	– 0.115			– 1.359*	– 0.089
Conc	–	–	–	–						
Cog	–	–	–	–						
Type	–	–	–	–	– 1.655**	– 0.176	– 1.274**			
SAE	–	–	–	–						
WMC	–	–	–	–						
Conc × Cog	–	–	–	–	–	–	0.196*	0.113	0.446**	0.119
Conc × Type	–	–	–	–	–	–				
Conc × SAE	–	–	–	–	–	–				
Conc × WMC	–	–	–	–	–	–				
Cog × Type	–	–	–	–	–	–				
Cog × SAE	–	–	–	–	–	–				
Cog × WMC	–	–	–	–	–	–				
Type × SAE	–	–	–	–	–	–	0.302**	0.123		
Type × WMC	–	–	–	–	–	–			– 0.351*	– 0.093
SAE × WMC	–	–	–	–	–	–			– 0.474*	– 0.098
Conc × Cog × Type	–	–	–	–	–	–	–	–	– 0.278*	– 0.105
Conc × Cog × SAE	–	–	–	–	–	–	–	–		
Conc × Cog × WMC	–	–	–	–	–	–	–	–		
Conc × Type × SAE	–	–	–	–	–	–	–	–		
Conc × Type × WMC	–	–	–	–	–	–	–	–		
Conc × SAE × WMC	–	–	–	–	–	–	–	–		
Cog × Type × SAE	–	–	–	–	–	–	–	–		
Cog × Type × WMC	–	–	–	–	–	–	–	–		
Cog × SAE × WMC	–	–	–	–	–	–	–	–		
Type × SAE × WMC	–	–	–	–	–	–	–	–	0.476**	0.148

Note: $*p < .05$, $**p < .01$. Empty cells represent non-significant ($p > .05$) values.

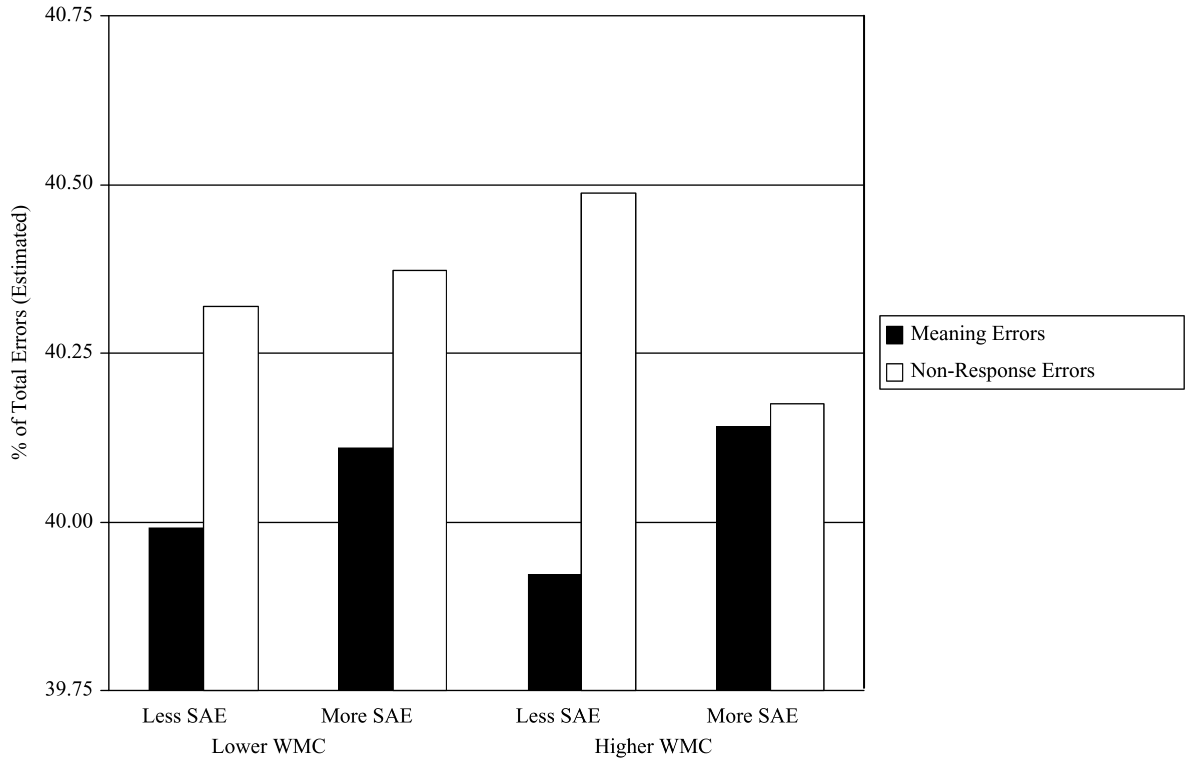

Figure 1. Graph depicting the estimated percent of total errors made of each type as a function of the WMC (lower or higher) and the amount of SAE (less or more). Because the graph does not depict other types of errors, the sum of the errors for each participant group may not equal 100%. The data were estimated from the regression equation, and then 40% was added to make the mean more similar to the raw data.

analysis described above. We then entered the independent variables on the second step, the interaction terms on the third step, and the three-way interaction on the fourth step. We found that the second step, which is the step that indicates whether there was a difference between meaning and non-response errors for that group of participants, was significant for every group except the more SAE higher WMC group [for less SAE lower WMC, step 2 $\Delta R^2 = .329$, $F(3, 40) = 8.38$, $p < .01$; type $t = -4.96$, $p < .01$; for more SAE lower WMC, step 2 $\Delta R^2 = .298$, $F(3, 40) = 6.07, p < .01$; type $t = -3.06, p < .01$; for less SAE higher WMC, step 2 $\Delta R^2 = .622$, $F(3, 88) = 51.41$, $p < .01$; type $t - 12.32$, $p < .01$; for more SAE higher WMC, step 2 $\Delta R^2 = .076$, $F(3, 56) = 1.66, p > .10$; type $t = -.691$, $p > .10$]. These tests confirm the pattern of data we observed – only the more SAE higher WMC group made the same percentage of meaning as non-response errors; individuals in the other groups made more non-response than meaning errors.

It is worth noting that the difference between non-response and meaning errors was greater for the individuals with less SAE and higher WMC than for the other groups. We did not make specific predictions regarding the magnitude of the difference for each of the four groups. However, it is not surprising that the difference is greater for this group than for either of the groups with more SAE, because, as we have posited, individuals with less SAE have less desire to communicate and are therefore less likely to make meaning errors. What is especially interesting is that the difference is greater for this group than for the less SAE lower WMC group. It is possible that having higher WMC makes one more aware of not knowing the correct answer, and without higher SAE, makes one more willing to make a non-response error.

We estimated the means for the concreteness by cognate status by type of error interaction from the regression equation. The overall pattern shows that there were fewer non-response errors for concrete cognate translations than for any other word type, most likely because the participants were more aware of the translations for this type of word (e.g. De Groot et al., 1994). For the abstract cognates there were 40.35% non-response errors and 39.97% meaning errors, for the concrete cognates there were 40.16% non-response errors and 40.03% meaning errors, for the abstract noncognates there were 40.37% non-response errors and 40.04% meaning errors, and for the concrete noncognates there were 40.45% non-response errors and 40.10% meaning errors (note that 40% was added to the adjusted means).

Discussion

Our experiment demonstrates that only individuals with relatively high working-memory capacity and more study-abroad experience make as many meaning errors as non-response errors when translating into their non-dominant language. This finding is consistent with the notion that study-abroad learners are accustomed to being in situations in which they need to communicate; however, considerable working-memory capacity is required to make responses that are related in meaning to the expected translation. The three-way interaction between type of error, amount of SAE, and WMC is consistent with our predictions that: (1) SAE leads individuals to attempt to make meaning-related responses rather than not responding, but (2) only individuals with relatively high WMC have the cognitive resources required to make meaning errors. Thus, examining the kinds of errors individuals make is more informative than only analyzing overall accuracy data.

A remaining question is whether individuals are aware that they are making errors when they make meaning errors, or whether, for example, they actually believe that *gato* means "dog". There are several reasons why the latter is unlikely. The group with the highest proficiency and overall task accuracy made the most meaning errors; therefore, because this group knows more words, they should be less likely to mis-map word forms to semantic neighbors (e.g. to think that *gato* means "dog") than the individuals in the other groups. Further, because there is no *a priori* reason to predict that the incidence of mis-mapping of word forms to semantic neighbors would be highest for this group, in the discussion that follows we assume that, although mis-mapping may occasionally be the cause of meaning errors, the rate of mis-mapping does not differ across the four groups. Rather, we will assume that the participants usually know that they are not giving the correct translation of the target word. In future research, confidence ratings or post-task vocabulary tests may be used to pursue this issue further.

In addition, although in our study we were interested in group differences in the types of errors once pre-existing differences were partialled out, it is important to consider that there may be interesting differences as a result of some of the word knowledge factors we removed. For example, Blanken et al. (2002) showed that certain types of words (e.g. those with high semantic competitiveness) were more likely to elicit meaning errors than non-response errors. It is therefore possible that individuals with richer semantic networks (e.g. higher WMC, more SAE) would be more likely to make meaning than non-response errors. This issue deserves attention in future research. Furthermore, although we included several types of translation stimuli, we did not observe any word group that was more likely to elicit meaning than non-response errors. It is difficult to say how our word groups would differ in their semantic competitiveness so it is not possible to examine this hypothesis with our data.

In the section that follows, we attempt to explain our findings by adapting an existing model of bilingual language processing. In particular, we explore the circumstances that might allow for meaning errors to be made by some individuals but not by others.

A number of models of bilingual lexical processing propose that multiple candidates are activated in parallel, and therefore might lead to the prediction that errors could be made among the activated set. However, only language production models, in which processing is initiated by an event that produces conceptual activation and corresponding competition among semantic alternatives, can capture the result we have reported. For example, models such as the BIA (Dijkstra, Van Heuven and Grainger, 1998) or BIA+ (Dijkstra and Van Heuven, 2002) characterize the process whereby a visually-presented word is recognized. In the case of word recognition, the alternative candidates are related by virtue of similarity across their lexical form. Because words that are lexical neighbors are rarely semantically related (but see Dell and O'Seaghdha, 1991, for a discussion of mixed errors in speech production), it is unlikely that the BIA or BIA+ models can account for our results. Other models that specifically address aspects of translation production (e.g. De Groot et al., 1994; Kroll and Stewart, 1994) do not make specific predictions about how alternative candidates might be generated. The Distributed Feature Model (De Groot et al., 1994) assumes that abstract noncognate translations are less likely to share semantic features than other words. The model therefore predicts that translation will be slower and less accurate for this category of words and prior studies have indeed verified that finding. However, our data also show that this category produced the fewest meaning errors and therefore this model is unlikely to account for the differential effects of SAE and WMC. We turn instead to a model of bilingual language production that we have adapted to explain our data.

In this model, we make the assumption that translation from the dominant to the non-dominant language is a meaning-mediated task. This assumption is consistent with most models of bilingual language processing (e.g. the Concept Mediation Model of Potter, So, Von Eckhardt and Feldman (1984); the Distributed Feature Model of De Groot (1992); the Revised Hierarchical Model of Kroll and Stewart (1994)). To demonstrate the differential working memory and resource allocation demands on correct translations, meaning errors, and non-responses, we have adapted a model of bilingual production initially proposed by Poulisse and Bongaerts (1994) and later developed by Hermans (2000). Because there are no

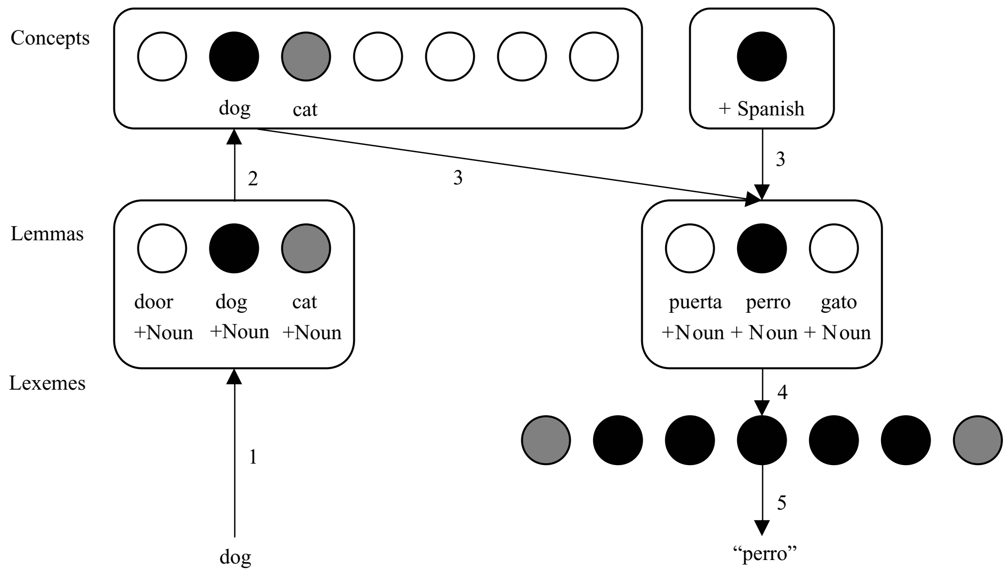

Figure 2. Schematic diagram of the activation/allocation model. The figure illustrates the flow of activation during accurate translation; activation flows in the direction of the arrows. Numbers next to the arrows indicate the time step of activation flow.

existing models that predict the types of errors made during translation, we will propose two variations of the model that can explain our results in alternative ways. The first variant (the ACTIVATION variant) proposes that individuals with higher WMC have more activation available to cycle through the translation process until a meaning error is made instead of a non-response error; the second variant (the ALLOCATION variant) proposes that only individuals with higher WMC are able to allocate activation to the related concept (e.g. "cat") to make it more active than the concept denoted by the stimulus (e.g. Just and Carpenter, 1992; Engle, 2002; Michael and Gollan, in press). These variants are not mutually exclusive; some combination of added activation and allocation ability may be needed to make meaning errors. Although the present study does not attempt to discriminate between these variants, in future research it may be possible to do so by developing separate measures of memory capacity and allocation ability.

Figure 2 shows accurate translation. Figure 3 shows two kinds of errors: the top panel shows a meaning error and the bottom panel shows a non-response error. In each case, we assume that an English word is presented and the task is to translate that word into Spanish. Like other models of spoken word production (e.g. Levelt, 1989; Levelt, Roelofs and Meyer, 1999), distinct levels are assumed to represent conceptual information, abstract lexical representations (lemmas) that contain syntactic and conceptual information, and word forms (lexemes) that contain orthographic and phonological information. The darkness of each node indicates its level of activation; darker nodes are more highly activated than lighter nodes. The models also include a language cue to indicate the language that must be selected for production.

According to the model, a target word will activate its associated lexical features and its lemma. The lemma is assumed to be language specific, at least in part because the connections to language-specific syntactic constraints must be considered within any complete model of production. Note that when translation proceeds accurately, semantic associates of the stimulus word are activated (in this example, "cat" is the associate of the stimulus *dog*). However, because the concept for dog is more active than the concept for its associates, and because the translation for *dog* is known, the individual successfully produces *perro* (which means "dog" in Spanish). In both panels of Figure 3, note that there is no lemma-level representation for the word *perro* because it is either not known or unavailable. The top and bottom panels show two possible outcomes for this situation, both of which apply to individuals with a strong desire to communicate (i.e. high SAE individuals). The specific models for low SAE individuals are not shown because it is assumed that these individuals have a less strong desire to communicate, and thus would be most likely to make a non-response error when faced with the inability to produce the correct translation. Again, a strong desire to communicate could be considered a threshold or change in the criterion for precision in translation; this setting is determined by such factors as the desire or necessity of communication and the past experience of the learner (see Dell, Lawler, Harris and Gordon, 2004, and Laine, Tikkala and Martti, 1998, for threshold-type speech production models that predict meaning and non-response errors).

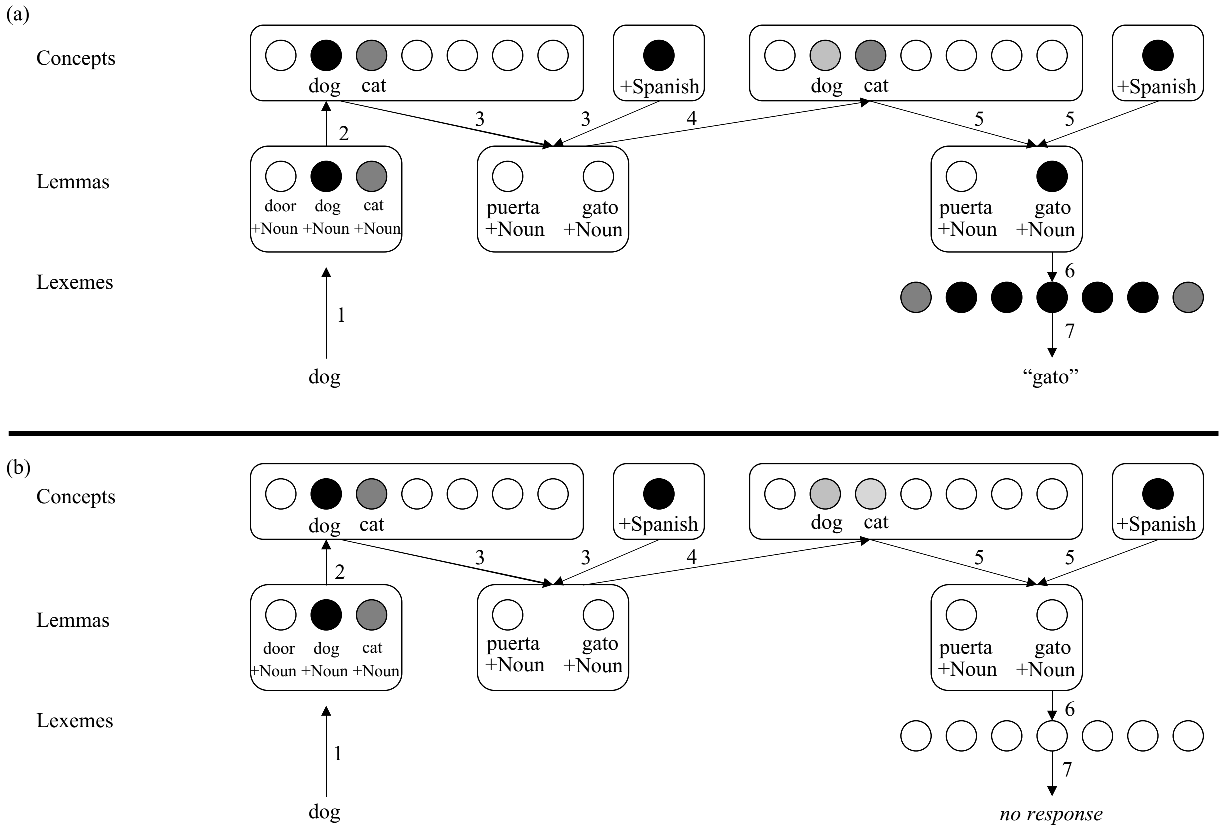

Figure 3. Schematic diagram of the activation/allocation model. The figure illustrates the flow of activation during (a) a translation trial on which a meaning error is made (top panel: more SAE higher WMC), and (b) a translation trial on which no response is produced (bottom panel: more SAE lower WMC). Activation flows in the direction of the arrows. Numbers next to the arrows indicate the time step of activation flow. The left and right sides of the figure represent the same system at two different times.

The top panel of Figure 3 represents an individual with sufficient WMC to make a meaning error, either because she or he still has activation of related nodes after one translation attempt (activation variant), or because s/he has the ability to activate the concept for cat sufficiently so that it surpasses the level of activation of dog (allocation variant). As shown in the top panel of Figure 3 (moving in the direction of the numbered arrows), the speaker determines that the translation for dog is not available because no nodes are active above criterion. The speaker then allows for the translation of another, similar, concept. At this point, the concepts for dog and cat are still activated, and cat is chosen for lexicalization because it is the most highly activated node. Thus, the speaker has allowed the threshold for selection to be lowered; when precision is required, the threshold will be set high, when precision is not required, the threshold will be lowered.

The bottom panel represents an individual with lower WMC. In this case, after determining that the translation for dog is not available, the speaker attempts to activate a related concept by lowering the threshold, but because no nodes are sufficiently active to lead to lexicalization, the speaker indicates that the translation is not known.

Conclusion

Our data are consistent with the notion that qualitative differences between groups, as a function of SAE and WMC, are important predictors of the types of errors made during translation. We have demonstrated that individuals with higher WMC have the available working memory (or can allocate resources appropriately) to take advantage of a particular communicative strategy that may lead to increased communicative success in the study-abroad context. Thus, it is important to consider qualitative differences between groups in addition to quantitative differences (e.g. in overall accuracy). We have further shown that error-type data are valuable because the qualitative analysis we presented led to interesting results with respect to individual differences in translation, which were not revealed by the reaction time and accuracy data alone. Therefore, it is essential that models of language processing predict the qualitative differences in types of errors that we have reported.

We have adapted a model of language production to account for our results. In future research, we hope to examine the types of errors made by L2 learners in more naturalistic settings, such as the interview context used by DeKeyser (1991a, b). In addition, it will be interesting to examine the CONSEQUENCES of different communicative styles or strategies for comprehension by the interlocutor. It is possible that individuals who make more meaning than non-response errors are also better understood by the individuals to whom they speak. In future research it will be critical to examine these relations.

References

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

Blanken, G., Dittmann, J. & Wallesch, C.-W. (2002). Parallel or serial activation of word forms in speech production? Neurolinguistic evidence from an aphasic patient. *Neuroscience Letters*, 325, 72–74.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ: Lawrence Erlbaum.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450–466.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 561–584.

Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, 25, 1–18.

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1001–1018.

De Groot, A. M. B., Dannenburg, L. & Van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 33, 600–629.

DeKeyser, R. M. (1991a). Foreign language development during a semester abroad. In B. Freed (ed.), *Foreign language acquisition research and the classroom*, pp. 104–119. Lexington: D. C. Heath.

DeKeyser, R. (1991b). The semester overseas: What difference does it make? *ADFL Bulletin*, 22, 42–48.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.

Dell, G. S., Lawler, E. N., Harris, H. D. & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology*, 21, 125–145.

Dell, G. S. & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review*, 98, 604–614.

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

Dijkstra, T., Van Heuven, W. J. B. & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, 38, 177–196.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23.

Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin Co.

Freed, B. F. (1995) (ed.). *Second language acquisition in a study abroad context*. Philadelphia: John Benjamins.

Haxby, J. V., Parasuraman, R., Lalonde, F. & Abboud, H. (1993). SuperLab: General-purpose Macintosh software for human experimental psychology and psychological testing. *Behavior Research Methods, Instruments, & Computers*. 25, 400–405.

Hermans, D. (2000). *Word production in a foreign language*. Ph.D. dissertation, University of Nijmegen.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.

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

Kroll, J. F., Michael, E., Tokowicz, N. & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18, 137–171.

Kroll, J. F. & Stewart, E. (1994). Concept interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.

Laine, M., Tikkala, A. & Martti, J. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics*, 11, 275–294.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*. 22, 1–75.

Michael, E. B. & Gollan, T. (in press). Being and becoming bilingual: Individual differences and consequences for language production. Chapter to appear in J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

Michael, E. B., Tokowicz, N. & Kroll, J. F. (2003). 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 & L. E. Bourne, Jr. (eds.), *Foreign language learning: Psycholinguistic studies on training and retention*, pp. 339–364. Mahwah, NJ: Lawrence Erlbaum Associates.

Peterson, R. R. & Savoy, P. (1998). Lexical selection and phonological encoding during language production: Evidence for cascaded processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 539–557.

Potter, M. C., So, K.-F., Von Eckhardt, B. & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning & Verbal Behavior*, 23, 23–38.

Poulisse, N. & Bongaerts, T. (1994). First language use in second language production. *Applied Linguistics*, 15, 36–57.

Reichle, E. D., Carpenter, P. A. & Just, M. A. (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology*, 40, 261–295.

Segalowitz, N. (1997). Individual differences in second language acquisition. In A. M. B. De Groot & J. F. Kroll (eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*, pp. 85–112. Hillsdale, NJ: Lawrence Erlbaum Publishers.

Sunderman, G. & Kroll, J. F. (2003). Speaking words in a second language: The role of proficiency and individual differences in negotiating cross-language competition. Paper presented at the Fourth International Symposium on Bilingualism, Tempe, AZ.

Talamas, A., Kroll, J. F. & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, 2, 45–58.

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

Received June 17, 2003
Revisions received January 12, 2004; March 21, 2004
Accepted March 24, 2004

Appendix A. ***Operation stimuli for working memory capacity task.***

"No" stimuli	"No" stimuli	"Yes" stimuli	"Yes" stimuli
$(10/1) - 2 = 3$	$(20/5) + 5 = 5$	$(10/1) - 1 = 9$	$(20/4) - 3 = 2$
$(10/5) + 3 = 9$	$(3 * 2) + 1 = 3$	$(10/5) + 1 = 3$	$(3 * 1) + 1 = 4$
$(12/2) - 5 = 6$	$(3 * 3) - 6 = 8$	$(12/2) - 4 = 2$	$(3 * 2) + 3 = 9$
$(12/6) + 3 = 1$	$(4 * 1) + 2 = 2$	$(12/4) + 4 = 7$	$(4 * 1) + 1 = 5$
$(14/7) + 1 = 9$	$(4 * 2) - 2 = 2$	$(14 * 1) - 8 = 6$	$(4 * 1) + 3 = 7$
$(15 * 1) - 7 = 2$	$(5 * 2) - 5 = 9$	$(14/2) - 1 = 6$	$(5 * 1) + 3 = 8$
$(15/3) - 1 = 9$	$(5 * 3) - 9 = 2$	$(14/7) + 2 = 4$	$(5 * 2) - 7 = 3$
$(15/5) + 2 = 9$	$(6 * 1) + 2 = 3$	$(15/3) - 4 = 1$	$(6 * 1) + 1 = 7$
$(16/8) + 4 = 1$	$(6 * 2) - 8 = 9$	$(16 * 1) - 9 = 7$	$(6 * 2) - 3 = 9$
$(18 * 1) - 9 = 4$	$(7 * 2) - 9 = 9$	$(16/2) - 5 = 3$	$(7 * 2) - 6 = 8$
$(18/2) - 6 = 8$	$(7/7) + 5 = 2$	$(16 * 1) - 8 = 8$	$(7/7) + 1 = 2$
$(18/9) + 5 = 3$	$(8 * 2) - 9 = 3$	$(18/3) - 4 = 2$	$(8 * 1) + 1 = 9$
$(2 * 3) + 2 = 4$	$(8/8) + 6 = 3$	$(2 * 2) + 5 = 9$	$(8/2) - 2 = 2$
$(2 * 4) - 2 = 2$	$(9/3) - 1 = 6$	$(2 * 4) + 1 = 9$	$(9/3) + 2 = 5$
$(20/4) + 3 = 3$	$(9 * 2) - 9 = 5$	$(20/2) - 9 = 1$	$(9/3) - 2 = 1$

Appendix B. ***Word stimuli for working memory capacity task.***

English word	Spanish word	English word	Spanish word	English word	Spanish word
author	autor	group	grupo	pipe	pipa
band	banda	guide	guía	plan	plan
bank	banco	hand	mano	poem	poema
beach	playa	hill	colina	radio	radio
black	negro	hotel	hotel	rain	lluvia
boat	barco	island	isla	rifle	rifle
bottle	botella	king	rey	rock	roca
bridge	puente	knife	cuchillo	rule	regla
chain	cadena	lake	lago	sign	signo
clock	reloj	leaf	hoja	site	sitio
cloud	nube	line	línea	skirt	falda
coast	costa	lion	léon	street	calle
dinner	cena	month	mes	team	equipo
dust	polvo	moon	luna	tooth	diente
face	cara	mouth	boca	train	tren
finger	dedo	nail	clavo	tree	árbol
floor	suelo	oven	horno	valley	valle
flower	flor	paper	papel	wall	pared
foot	pie	pear	pera	wine	vino
girl	chica	piano	piano	world	mundo

Appendix C. ***Translation stimuli.***

English Word	Spanish Word	English Word	Spanish Word	English Word	Spanish Word
ability	habilidad	gold	oro	pirate	pirata
action	acción	grandmother	abuela	pound	libra
advantage	ventaja	guitar	guitarra	pride	orgullo
age	edad	hat	sombrero	promise	promesa
ambulance	ambulancia	head	cabeza	rage	rabia
army	ejército	health	salud	rice	arroz
arrival	llegada	hell	infierno	rose	rosa
art	arte	honey	miel	sadness	tristeza
attention	atención	honor	honor	scissors	tijeras
beauty	belleza	hope	esperanza	shame	vergüenza
beer	cerveza	horses	caballos	sheep	oveja
belief	creencia	hospital	hospital	shirt	camisa
blanket	manta	hour	hora	shoe	zapato
body	cuerpo	house	casa	silence	silencio
bone	hueso	hunger	hambre	sin	pecado
boot	bota	idea	idea	size	tamaño
brick	ladrillo	influence	influencia	snow	nieve
bullet	bala	inheritance	herencia	soap	jabón
butter	mantequilla	innocence	inocencia	song	canción
captain	capitán	joke	chiste	soul	alma
cat	gato	kitchen	cocina	south	sur
century	siglo	leg	pierna	station	estación
chairs	sillas	library	biblioteca	stench	hedor
chaos	caos	lie	mentira	struggle	lucha
circle	círculo	loan	préstamo	student	estudiante
color	color	love	amor	sugar	azúcar
comedy	comedia	loyalty	lealtad	summer	verano
confession	confesión	luck	suerte	sweat	sudor
corn	maíz	machine	máquina	table	mesa
corner	esquina	mail	correo	talent	talento
cotton	algodón	map	mapa	tea	té
cow	vaca	market	mercado	thought	pensamiento
crisis	crisis	material	material	threat	amenaza
cruelty	crueldad	meat	carne	throat	garganta
culture	cultura	method	método	tiger	tigre
danger	peligro	mind	mente	tobacco	tabaco
death	muerte	miracle	milagro	triangle	triángulo
deceit	engaño	mirror	espejo	truck	camión
door	puerta	movie	película	truth	verdad
dream	sueño	music	música	umbrella	paraguas
ease	facilidad	needle	aguja	victim	víctima
effort	esfuerzo	newspaper	periódico	village	aldea
elephant	elefante	noise	ruido	violin	violín
fame	fama	nonsense	tonterías	war	guerra
favor	favor	oath	juramento	warmth	calor
fever	fiebre	office	oficina	water	agua
fire	fuego	pain	dolor	weakness	debilidad
fist	puño	panic	pánico	week	semana
fruit	fruta	park	parque	woman	mujer
furniture	muebles	passion	pasión	word	palabra
future	futuro	peace	paz	yesterday	ayer
garden	jardín	pearl	perla	youth	juventud
gift	regalo	pencil	lapiz		
glory	gloria	person	persona		

Appendix D. ***Language history questionnaire.***

Sex: M / F **Age (in years) ______** **Native country ______**

Years spent in the U.S. ______ **Years spent in U.S. schools ______**

This questionnaire is designed to give us a better understanding of your experience learning a second language. We ask that you be as accurate and thorough as possible when answering the following questions and thank you for your participation in this study.

What is your first language (i.e., **language first spoken**)? If more than one, please briefly describe the situations in which each language was used.

____

____

Which language do you consider your <u>second</u> language (English or Spanish)?

____

If you have ever lived in or visited a country where languages other than your native language are spoken, please indicate below the name of the country (countries), the duration of your stay in number of months, and which languages you used while you were in the country (please indicate if you were spoken to in a language other than your first language, even if you never actually spoke that language).

Country visited	# Months there	Language(s) used

How many years have you studied your second language? Please indicate the setting(s) in which you have had experience with the language (i.e., classroom, with friends, foreign country...)

Number of years:

____

Setting(s):

____

____

List below, from most fluent to least fluent, <u>all</u> of the languages you know. Also specify the age in years at which you began to learn the language and the context in which you learned it. For example, "English, birth, home". Include <u>all</u> languages to which you have been exposed, although you may never have had formal training in them and may not be able to read, speak or write them.

Please remember to list your native language.

Language	Age in yrs.	Learning Situation

<u>**For the next eight questions, please circle the number of your response: **</u>

Please rate your **English** reading proficiency on a ten–point scale.
(1 = not literate, 10 = very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

Please rate your **Spanish** reading proficiency on a ten–point scale.
(1 = not literate, 10 = very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

Please rate your **English** writing proficiency on a ten–point scale.
(1 = not literate, 10 = very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

Please rate your **Spanish** writing proficiency on a ten–point scale.
(1 = not literate, 10 = very literate)

1 2 3 4 5 6 7 8 9 10
not literate very literate

Please rate your **English** conversational fluency on a ten–point scale.
(1 = not fluent, 10 = very fluent)

1 2 3 4 5 6 7 8 9 10
not fluent very fluent

Please rate your **Spanish** conversational fluency on a ten–point scale.
(1 = not fluent, 10 = very fluent)

1 2 3 4 5 6 7 8 9 10
not fluent very fluent

Please rate your **English** speech comprehension ability on a ten–point scale.
(1 = unable to understand conversation, 10 = perfectly able to understand conversation)

1 2 3 4 5 6 7 8 9 10
no comprehension perfect comprehension

Please rate your **Spanish** speech comprehension ability on a ten–point scale.
(1 = unable to understand conversation, 10 = perfectly able to understand conversation)

1 2 3 4 5 6 7 8 9 10
no comprehension perfect comprehension

Have you ever been immersed in your second language culture?
(please circle one) yes no

If yes, please answer the questions below.

** While you were in the immersion environment: ** (if you are still in your second language immersion environment, please respond based upon your experience so far.

While you were in the immersion environment,
What percentage of the time did you spend speaking your <u>second</u> language? ____%
What percentage of the time did you spend speaking your <u>first</u> language? ____%

How much time did you spend watching TV per day (please circle)?
less than 1 hour 1–2 hours 3–4 hours more than 4 hours

How frequently did you read a newspaper or magazine in your second language (please circle)?
never once a week three or more times a week

How much time did you spend speaking with native speakers of your second language?
hardly ever occasionally often at every opportunity