ORTHOGRAPHIC REPRESENTATIONS AFFECT ADULT SECOND LANGUAGE WORD LEARNING

by

Alison Marie Phillips

B.A., Gannon University, 2006

M.A., University of Delaware, 2008

Submitted to the Graduate Faculty of

the College of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

COLLEGE OF ARTS AND SCIENCES

This thesis was presented

by

Alison Marie Phillips

It was defended on

April 21, 2011

and approved by

Michael Walsh Dickey, Assistant Professor, Department of Communication Science and Disorders

Tessa Warren, Assistant Professor, Department of Psychology

Thesis Director: Natasha Tokowicz, Assistant Professor, Department of Psychology

ORTHOGRAPHIC REPRESENTATIONS AFFECT ADULT SECOND LANGUAGE WORD LEARNING

Alison Marie Phillips, M.S.

University of Pittsburgh, 2011

Adults have considerable difficulty learning a second language (L2), and learning to produce L2 words is particularly difficult. In the present study, we examined how the lexical representations available to beginning adult L2 learners during training, and the order in which words are trained, affect their L2 Arabic vocabulary learning. The findings suggest that the presence of orthographic information and thematic groupings during training improves L2 Arabic learning. These findings are discussed in relation to existing models and previous findings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank my academic advisor, Natasha Tokowicz, for her invaluable assistance in conducting this research and in writing this paper. I would also like to thank my mentoring committee, Tessa Warren and Michael Walsh Dickey, for their insightful suggestions. Finally, I would like to thank Martin Doppelt and Kevin Jarbo for their assistance in running the experiment, and Walt Schneider and Chris Schunn for their assistance in designing the experiment.

1.0 INTRODUCTION

Adults have considerable difficulty learning a second language (L2), and one aspect of adult L2 learning that is particularly difficult is learning to produce words. As adults, we have well-established connections between spelling and sound in our first language (L1), and transfer of these connections from the L1 to the L2 may decrease L2 learning difficulty. We also have well-established connections between concepts in our L1, but transfer of these connections from the L1 to the L2 paradoxically may increase L2 learning difficulty. We explore these two issues in the present study in the context of adult L2 word learning.

1.1 QUALITY OF LEXICAL REPRESENTATIONS

In the present study, we examined how the lexical representations available to beginning adult L2 learners during training affect their L2 vocabulary learning. According to the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002), a high quality lexical representation is formed when the spelling (orthography), sound (phonology), and meaning (semantics) of a word are highly specified and interconnected.

In applying the LQH to L1 vocabulary learning, Nelson, Balass, and Perfetti (2005) trained adult native English speakers on rare English words in two critical conditions: (1) orthography and meaning (OM), and (2) phonology and meaning (PM). Participants reached a

100% accuracy criterion significantly faster for words trained in the OM condition than for words trained in the PM condition. The authors concluded that training in the OM condition resulted in the creation of three memory traces (orthography, phonology, and meaning) for the words, because phonology could be derived automatically and reliably from orthography via grapheme-phoneme correspondence rules. However, training in the PM condition resulted in the creation of only two memory traces (phonology and meaning) for the words, because orthography could not be derived automatically and reliably from phonology via grapheme-phoneme correspondence rules. This is presumably because, when reading a word, the phonological representation of the word is readily activated, whereas when hearing a word, the orthographic representation of the word is not necessarily activated. The additional memory trace therefore improved L1 vocabulary learning.

Using event-related potentials (ERPs) in a training paradigm similar to Nelson et al. (2005), Balass, Nelson, and Perfetti (2010) further examined the memory traces involved in L1 vocabulary learning. They also included a third critical condition, in which orthography and phonology (OP) were trained. The P600 is an ERP component that is associated with recognition memory, among other things. Items that are recognized as "familiar" have a larger P600 component than items that are recognized as "unfamiliar." Words trained in the OM condition had a larger P600 component than words trained in the OP and PM conditions, suggesting that words trained in the OM condition were recognized as more "familiar" than words trained in the OP and PM conditions. This further suggests that the memory traces for words in the OM condition were stronger than the memory traces for the words in the OP and PM conditions, likely due to the creation of a phonological memory trace in the OM condition.

In an extension of the LQH to L2 vocabulary learning, Hu (2008) trained native Mandarin Chinese speaking children on the pronunciations of English pseudonames for novel cartoon characters. During training, a spoken presentation of a pseudoname was accompanied by a written presentation of the pseudoname, in which the pseudoname was spelled either with letters (e.g., nof) or with symbols (e.g.,]□). In a subsequent naming test, children more accurately named the cartoon characters whose pseudonames had been spelled with letters than with symbols. The author concluded that, consistent with the LQH, training pseudonames spelled with letters resulted in a high quality lexical representation, because the letters provided meaningful information about the pronunciation of the pseudoname. However, training pseudonames with symbols resulted in a low quality lexical representation, because the symbols provided no information about the pronunciation of the pseudoname. These results therefore provide evidence that orthography improves auditory L2 vocabulary learning.

The present experiment also extended the LQH to L2 vocabulary learning. Across 8 sessions, we trained native English-speaking adults with no prior exposure to Arabic on 96 Arabic words and phrases. During training, participants heard an L1 English word and its L2 Arabic translation. Participants in the Transliteration Conditions saw the L1 English word and its L2 Arabic transliteration, whereas participants in the No Transliteration Conditions saw only the L1 English word. In every session, participants were tested in a free recall test, in which they orally recalled as many English words and their Arabic translations as possible, and in an English-to-Arabic translation test, in which they orally provided the Arabic translation of the visually presented English word.

The LQH would predict that participants who see transliterations would more accurately recall and translate Arabic words and phrases than participants who do not see transliterations.

As a result of having all three lexical components of the words (orthography, phonology, and meaning) available to them during training, participants who see transliterations should form higher quality lexical representations, and should also have three memory traces that they can access during testing. In contrast, as a result of having only two lexical components of the words (phonology and meaning) available to them during training, participants who do not see transliterations should form lower quality lexical representations, and should also have only two memory traces that they can access during testing.

Alternatively, participants who do not see transliterations might more accurately recall and translate Arabic words and phrases than participants who see transliterations. As mentioned earlier, the free recall test and the English-to-Arabic translation test emphasize the oral production of the Arabic words and phrases. Only two lexical components (phonology and meaning) are relevant to oral production. For participants who do not see transliterations, all information that is available during training is relevant during testing. This match between training and testing may be advantageous. In contrast, for participants who see transliterations, some information that is available during training is relevant during testing, and some is irrelevant. This mismatch between training and testing may be disadvantageous.

1.2 SEMANTIC ORGANIZATION

In L2 textbooks and classrooms, L2 vocabulary words are generally organized into semantic categories such as “clothes” and “animals,” so L2 learners generally learn all the “clothes” vocabulary words together and all the “animals” vocabulary words together. However, training vocabulary words in semantic categories has been shown to negatively affect L2 vocabulary

learning. For example, Finkbeiner and Nicol (2003) examined the effect of semantic categorization on learning L2 "alien" vocabulary words. Training the vocabulary words in semantic categories resulted in slower and less accurate translation during testing than training the vocabulary words in a random order.

This finding can be explained by assuming that activating a concept for a word also activates concepts for related words. Concepts for words in the same semantic category will share some, but not all, features. For example, the concepts for "peach" and "nectarine" share some features. The presentation of "peach" will activate concepts related to "peach." These concepts will remain active for a period of time. The subsequent presentation of "nectarine" will activate concepts related to "nectarine." If the concepts activated by "peach" are still active, "nectarine" might not be precisely linked to the features of concepts related to "nectarine," but might instead be linked to the features of concepts related to "peach" in addition to the features of concepts related to "nectarine." This imprecise linking of words to concepts interferes with learning. See Folse (2004) for a review of how training L2 words in semantic categories interferes with L2 learning, and Kroll and Stewart (1994) for an example of category interference effects for proficient L2 speakers in their L2.

The present experiment examined whether training L2 vocabulary words in thematic categories has the same negative effect on L2 vocabulary learning as training L2 vocabulary words in semantic categories. Previous research examining these category interference effects has focused on concrete words. The present experiment examines both concrete and abstract words and phrases (see Appendix A for a list of the stimuli). Therefore, it expands the previous research on category interference effects. Training in Thematic/Semantic Order Conditions was compared to training in No Order Conditions. If previous research found a disadvantage for

training L2 words in semantic categories because participants focused on the relationship between concepts rather than on the mapping between words and concepts, we should find this same disadvantage for training L2 vocabulary words in thematic/semantic categories. Specifically, participants trained with words in a random order should more accurately recall and translate Arabic words than participants trained with words in a thematic/semantic order. Alternatively, if previous research found a disadvantage for training L2 vocabulary words in semantic categories because words in a concrete semantic category share perceptual features, we should not find a disadvantage for training L2 vocabulary words in thematic/semantic categories, because words in an abstract thematic category are less likely to share perceptual features. Specifically, participants trained with words in a thematic/semantic order should recall and translate Arabic words as accurately as participants trained with words in a random order.

To restate our hypotheses, the LQH would predict that participants who see transliterations will more accurately recall and translate than participants who do not see transliterations. This is because transliterations will allow learners to form higher quality lexical representations, and will provide a third memory trace that learners can access during testing. Alternatively, if the similarity between training and testing (e.g., Transfer Appropriate Processing; Morris, Bransford, & Franks, 1977) is important in L2 vocabulary learning, participants who do not see transliterations might more accurately recall and translate than participants who see transliterations.

Further, if the category interference disadvantage in L2 vocabulary learning is due to less attention to mappings between words and concepts than to relationships between concepts, we would predict that participants who are trained with words in a random order will be more accurate than participants who are trained with words in a thematic/semantic order. However, if

the category interference disadvantage is due to overlap across exemplars at the perceptual feature level, we will expect this disadvantage to be reduced or eliminated.

The present study extends existing research by examining L2 rather than L1 vocabulary learning in adults, and by using an extended training protocol. Further, the present study used thematic groupings that, unlike semantic categories, include more than concrete, pictureable objects.

2.0 EXPERIMENT 1A

2.1 METHOD

2.1.1 Participants

Participants were 36 native speakers of American English recruited from the University of Pittsburgh community. All participants were right-handed and had normal or corrected-to-normal vision, and normal hearing. Participants were additionally screened to have no previous exposure to any Semitic language, including Arabic, Hebrew, and Turkish. Participants were paid $10 per hour for up to 16 hours, and received a $50 bonus upon completion of the last session.

Of the 36 participants, data from four were not complete: two due to scheduling conflicts, one due to experimenter error, and one due to a technical (recorder) error. Analyses were therefore conducted on a final set of 32 participants divided evenly among the four training conditions (10 male; mean age 20.1 years).

2.1.2 Design

We used an 8 Session (1 through 8) x 2 Transliteration Condition (Transliterations vs. No Transliterations) x 2 Order Condition (Thematic/Semantic Order vs. Random Order) mixed

design, with Session as a within-subjects factor and Transliteration Condition and Order Condition as between-subjects factors.

2.1.3 Stimuli

The stimuli were 96 English words and phrases and their Iraqi Arabic translations, selected from the Iraqi Basic Language Survival Kit materials from the United States Defense Language Institute (DLI). These materials include a booklet that provides a list of translations and their transliterations in thematic/semantic groupings and sound files that provide pronunciations in English and Iraqi Arabic. These 96 items were selected by five research assistants to be important vocabulary for survival in a foreign country. Characteristics of these 96 items, which included 61 words and 35 phrases, are summarized in Table 1. Standard deviations are in parentheses. English length was calculated using number of letters, and does not include spaces or punctuation marks. Arabic length was calculated from the number of letters in the transliteration, and does not include spaces or punctuation marks, except for punctuation marks that represent a sound in Arabic. Frequency (Kučera & Francis, 1967) and familiarity, concreteness, and imageability ratings were obtained using the MRC Psycholinguistic Database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm). Frequency information was available for 54 words, familiarity and imageability ratings were available for 45 words, and concreteness ratings were available for 42 words. Familiarity, concreteness, and imageability ratings are on a scale ranging from 100 (less) to 700 (more).

Table 1: Stimulus Characteristics

	Words	Phrases
English length (number of letters)	5.74 (2.26)	12.34 (6.90)
Arabic transliteration length (number of letters)	6.10 (2.20)	11.06 (6.20)
English frequency	393.00 (674.34)	
English familiarity	558.53 (59.93)	
English concreteness	444.45 (133.55)	
English imageability	464.76 (122.07)	

2.1.4 Procedure

Participants were trained in one of the four training conditions over eight sessions (two sessions per week for four weeks). On every session, participants completed a free recall test, an English-to-Arabic translation production test, and an individual difference test. In the free recall test, participants typed the English words that they recalled, and then pronounced the English words they had listed and their Arabic translations. In the English-to-Arabic translation test, participants saw English words and pronounced the Arabic translations of the English words that they saw. The individual difference tests were variations of the Stroop test (Stroop, 1935), the operation span (o-span) test (Turner & Engle, 1989), the Waters reading span test (Waters & Caplan, 1996), and the Flankers test (Eriksen & Eriksen, 1974). [1]

In Session 1, participants were trained before testing. In Sessions 2 through 7, participants were tested before training. In Session 8, participants were tested only. An individual difference test separated testing and training on all sessions, and the free recall test always

preceded the English-to-Arabic translation test. At the end of Session 1, participants completed a language history questionnaire (from Tokowicz, Michael, & Kroll, 2004), which asked participants about their prior language learning experiences (see Table 2 for a summary of the responses, including average age at which participants began learning their L2 and average ratings of reading, writing, conversation, and comprehension ability in L1 and L2) Standard deviations are in parentheses. Reading, writing, conversation, and comprehension ability ratings are on a scale ranging from 0 (low) to 10 (high). [2]

Table 2: Participant responses on Language History Questionnaire

	Experiment 1A	Experiment 1B
Age began learning L2	11.0 (5.3)	12.3 (6.5)
L1 Reading	9.7 (0.7)	9.8 (0.5)
L2 Reading	4.3 (2.3)	3.3 (1.7)
L1 Writing	9.3 (0.9)	9.9 (0.4)
L2 Writing	3.0 (1.9)	2.3 (1.4)
L1 Conversation	9.8 (0.4)	9.8 (0.5)
L2 Conversation	3.5 (2.5)	3.3 (2.7)
L1 Comprehension	9.8 (0.5)	9.8 (0.5)
L2 Comprehension	4.2 (2.7)	3.3 (2.4)

The sessions were arranged such that there was one day between sessions within a week and four days between sessions between weeks. However, several violations to this arrangement

occurred due to severe weather conditions; thus, the average number of days between Sessions 5 and 6 was 0.97, between Sessions 6 and 7 was 4.19, and between Sessions 7 and 8 was 0.91.

2.1.5 Training

On training trials, a fixation cross was presented at the center of the screen until the participant pressed a button on the button box. In all conditions, the participant then heard an English word (pronounced once) and its Arabic translation (pronounced twice, the second time slower than the first) and saw the English word centered on the top half of the screen. In the Transliteration Conditions, the participant also saw the Arabic transliteration centered on the bottom half of the screen. Participants were instructed to repeat the word pair aloud twice and to press a button on the button box after repeating the pair, at which point the next fixation cross appeared. In the Thematic/Semantic Order Conditions, the words were presented in the same thematic/semantic groupings as in the Iraqi Basic Language Survival Guide, and in the Random Order Conditions, the words were presented in a new randomly generated order determined by E-Prime (Psychology Software Tools, Pittsburgh, PA) on each presentation. There were three training runs per training session. The 96 words and phrases were presented once per training run.

2.1.6 Free Recall

In the free recall test, participants typed the English words that they recalled in a spreadsheet, and then pronounced the English words and their Arabic translations. Vocal responses were recorded digitally and were later coded for accuracy. A pair was considered "correct" if the

participant recalled the English word and gave a reasonable pronunciation of its Arabic translation (i.e., it was clear which translation the participant intended).

2.1.7 English-to-Arabic Translation Production

In the English-to-Arabic translation test, a fixation cross was presented at the center of the screen until the participant pressed a button on the button box. An English word then appeared centered on the top half of the screen until the onset of a vocal response, at which point the fixation cross reappeared. E-prime recorded the response times in ms from the onset of the stimulus to the onset of articulation. Vocal responses were recorded digitally and were later coded for accuracy. As in the free recall test, response was considered "correct" if the participant gave a reasonable pronunciation of its Arabic translation. The 96 words and phrases were presented in a new random order on each session.

2.2 RESULTS

The data from Session 8 for one participant in the Transliterations/Random Order Condition were lost due to experimenter error. Therefore, the missing free recall accuracy data and the missing English-to-Arabic translation accuracy and reaction time (RT) data were replaced with the corresponding means for that condition.

Only RTs from correct trials were included in the RT analysis for the English-to-Arabic translation test. Voice key errors and RTs faster than 300 ms or slower than 6000 ms were removed from the RT analysis as outliers, resulting in the exclusion of 1.09 % of the data.

Because testing followed training in Session 1, and training followed testing in Sessions 2 through 8, it was not appropriate to include the data from Session 1 in the same analyses as the data from Sessions 2 through 8. [3] A mixed ANOVA was performed on the data from Sessions 2 through 8 with Session (2 through 8) as a within-participant variable in the analysis by participants (reported as F_1) and in the analysis by items (reported as F_2); Transliteration Condition (Transliterations vs. No Transliterations) and Order Condition (Thematic/Semantic Order vs. Random Order) were entered as between-participant variables in the analysis by participants (reported as F_1) and as between-items variables in the analysis by items (reported as F_2). For the mixed ANOVA, significant interactions were probed with *t*-tests, using the Bonferroni Correction for multiple comparisons (alpha level of .05 divided by the number of comparisons). When the assumption of sphericity was violated, we applied the Greenhouse and Geisser (1959) non-sphericity correction, and report uncorrected degrees of freedom, the corrected *p*-value, and the corrected mean square error values.

Reported means are from the analysis by participants unless otherwise indicated. Some effects were significant only in the analysis by items. We only report effects that are significant by both participants and items (or fully significant in one analysis and marginally significant at $p < .10$ in the other). A list of all effects that are fully or marginally significant in at least one analysis is provided in Table 3 for reference.

Table 3: List of Effects

Test/Measure	Effect	F1 Significance	F2 Significance
Free Recall	Session	< 0.05	< 0.05
	Transliteration Condition	< 0.05	< 0.05
	Order Condition	n.s.	< 0.05
	Session x Transliteration Condition	< 0.05	< 0.05
	Session x Order Condition	< 0.05	< 0.05
	Transliteration Condition x Order Condition	n.s.	< 0.05
	Session x Transliteration Condition x Order Condition	n.s.	< 0.05
Translation (Accuracy)	Session	< 0.05	< 0.05
	Transliteration Condition	< 0.05	< 0.05
	Order Condition	n.s.	< 0.05
	Session x Transliteration Condition	< 0.05	< 0.05
	Session x Order Condition	Marginal	< 0.05
	Transliteration Condition x Order Condition	n.s.	< 0.05
Translation (RT)	Session	< 0.05	< 0.05
	Transliteration Condition	n.s.	< 0.05
	Order Condition	Marginal	n.s.
	Transliteration Condition x Order Condition	n.s.	Marginal
Translation No Free Recall (Accuracy)	Session	< 0.05	< 0.05
	Transliteration Condition	< 0.05	< 0.05
	Order Condition	n.s.	< 0.05
	Session x Transliteration Condition	< 0.05	< 0.05
	Session x Order Condition	n.s.	< 0.05

Free recall accuracy is shown in Figure 1. Accuracy increased across sessions, F_1 (6, 168) = 148.34, *MSE* = 0.02, $p < 0.01$; F_2 (6, 570) = 389.46, *MSE* = 0.05, $p < 0.01$. Participants who saw transliterations recalled significantly more words than participants who did not see transliterations (32.3 vs. 20.2 %), F_1 (1, 28) = 6.81, *MSE* = 0.12, $p < 0.05$; F_2 (1, 95) = 68.69, *MSE* = 0.14, $p < 0.01$. Session and Transliteration Condition interacted, with accuracy increasing more across sessions for participants who saw transliterations than for participants who saw no transliterations, F_1 (6, 168) = 8.58, *MSE* = 0.02, $p < 0.01$; F_2 (6, 570) = 37.83 *MSE* = 0.02, $p < 0.01$. This interaction was probed with *t*-tests, using 0.007 as the corrected alpha level for 7 comparisons. Participants who saw transliterations correctly recalled significantly more words in Session 8, t (30) = 3.42, $p < 0.007$. Session and Order Condition interacted significantly in the analysis by items and marginally in the analysis by participants, F_1 (6, 128) = 3.06, *MSE* = 0.02, $p = 0.07$; F_2 (6, 570) = 18.66, *MSE* = 0.01, $p < 0.01$. Paired comparisons for this interaction did not yield any significant effects. However, the source of this interaction was participants who were trained in a thematic/semantic order correctly recalling increasingly more words across sessions than participants who were trained in a random order.

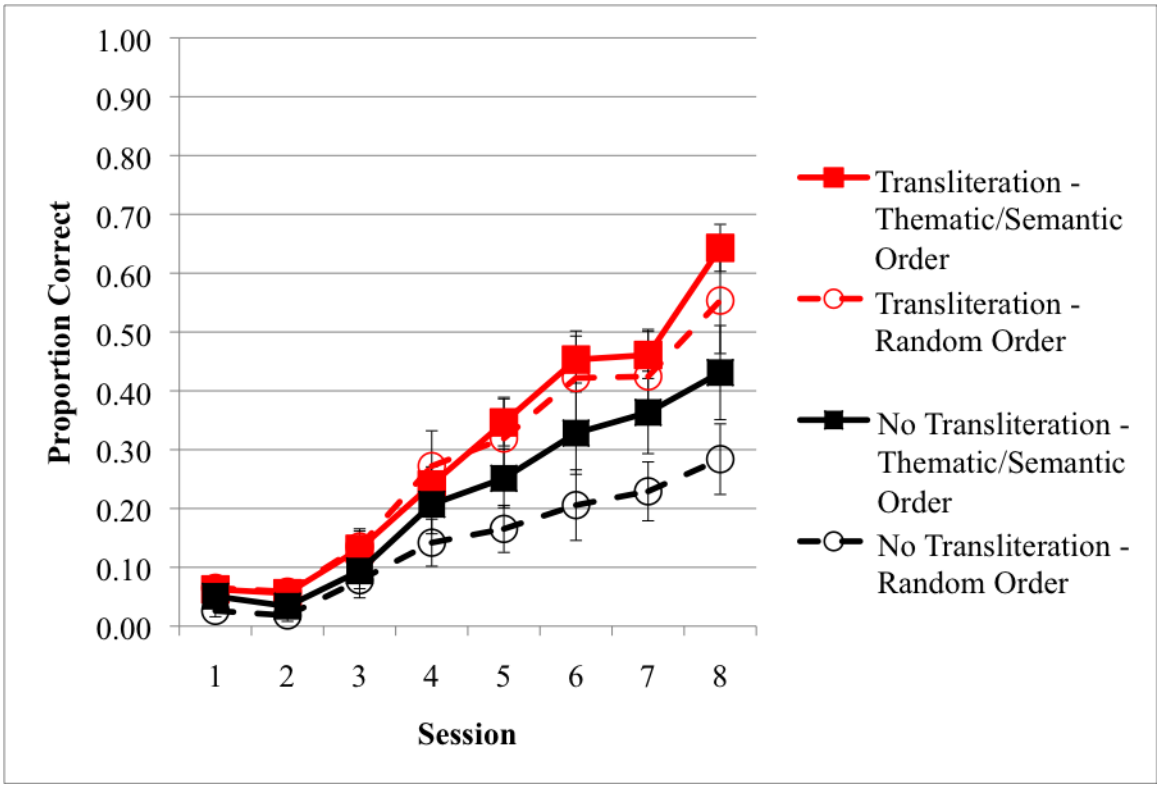

Figure 1: Free recall test (accuracy)

English-to-Arabic translation accuracy is shown in Figure 2. Accuracy increased across sessions, F_1 (6, 168) = 210.13, *MSE* = 0.02, $p < 0.01$; F_2 (6, 570) = 512.16, *MSE* = 0.06, $p < 0.01$. Participants who saw transliterations correctly translated more words than participants who did not see transliterations (43.1 vs. 25.0 %), F_1 (1, 28) = 12.37, *MSE* = 0.15, $p < 0.01$; F_2 (1, 95) = 126.70, *MSE* = 0.18, $p < 0.01$. Session and Transliteration Condition interacted, with accuracy increasing more across sessions for participants who saw transliterations than for participants who did not see transliterations, F_1 (6, 168) = 16.83, *MSE* = 0.02, $p < 0.01$; F_2 (6, 570) = 67.42, *MSE* = 0.03, $p < 0.01$. This interaction was again probed with *t*-tests, using 0.007 as the corrected alpha level for 7 comparisons. Participants who saw transliterations correctly translated significantly more words in Sessions 5 through 8, all *t*s (30) > 2.96, all *p*s < 0.007. Session and

Order Condition marginally interacted, such that participants who were trained with words in a thematic/semantic order recalled more words in the later sessions than participants who were trained with words in a random order, F_1 (6, 128) = 2.71, *MSE* = 0.02, p = 0.08; F_2 (6, 570) = 14.08, *MSE* = 0.02, $p < 0.01$. Paired comparisons for this interaction did not yield any significant effects.

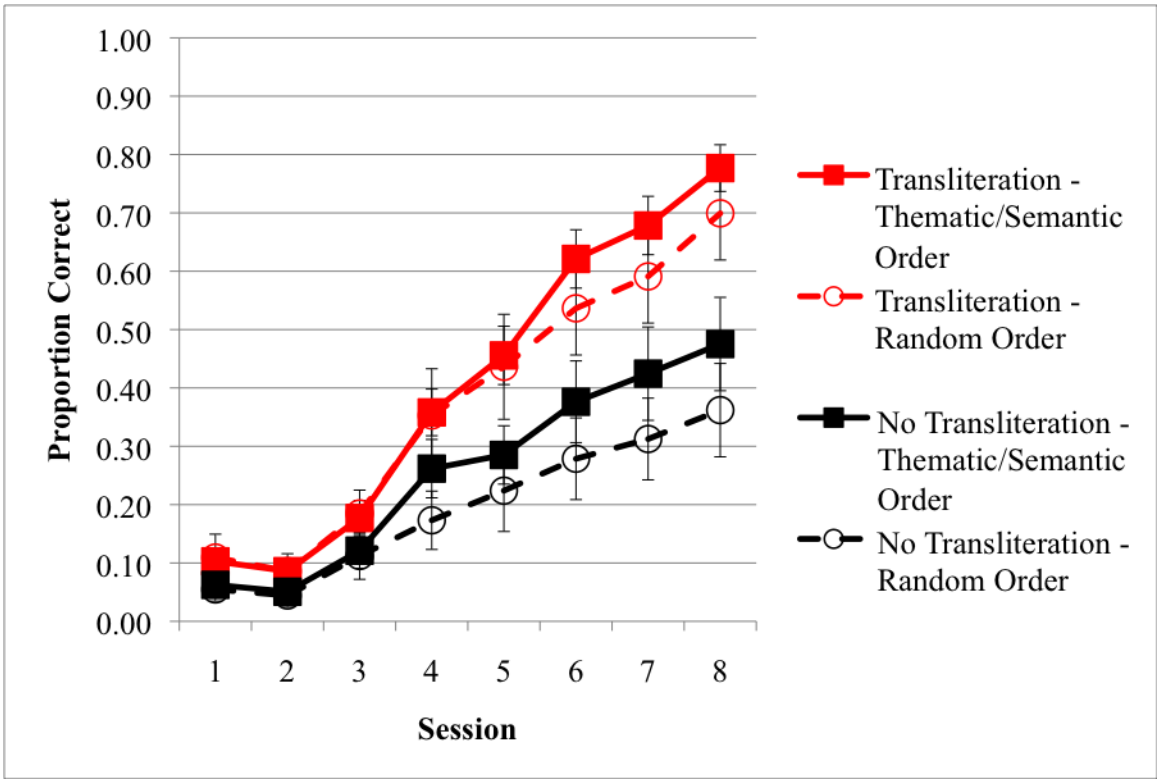

Figure 2: English-to-Arabic translation test (accuracy)

English-to-Arabic translation RTs are shown in Figure 3. RT decreased across sessions F_1 (6, 168) = 13.96, *MSE* = 179259, $p < 0.01$; F_2 (6, 42) = 11.36, *MSE* = 312464, $p < 0.01$. No other effects were significant.

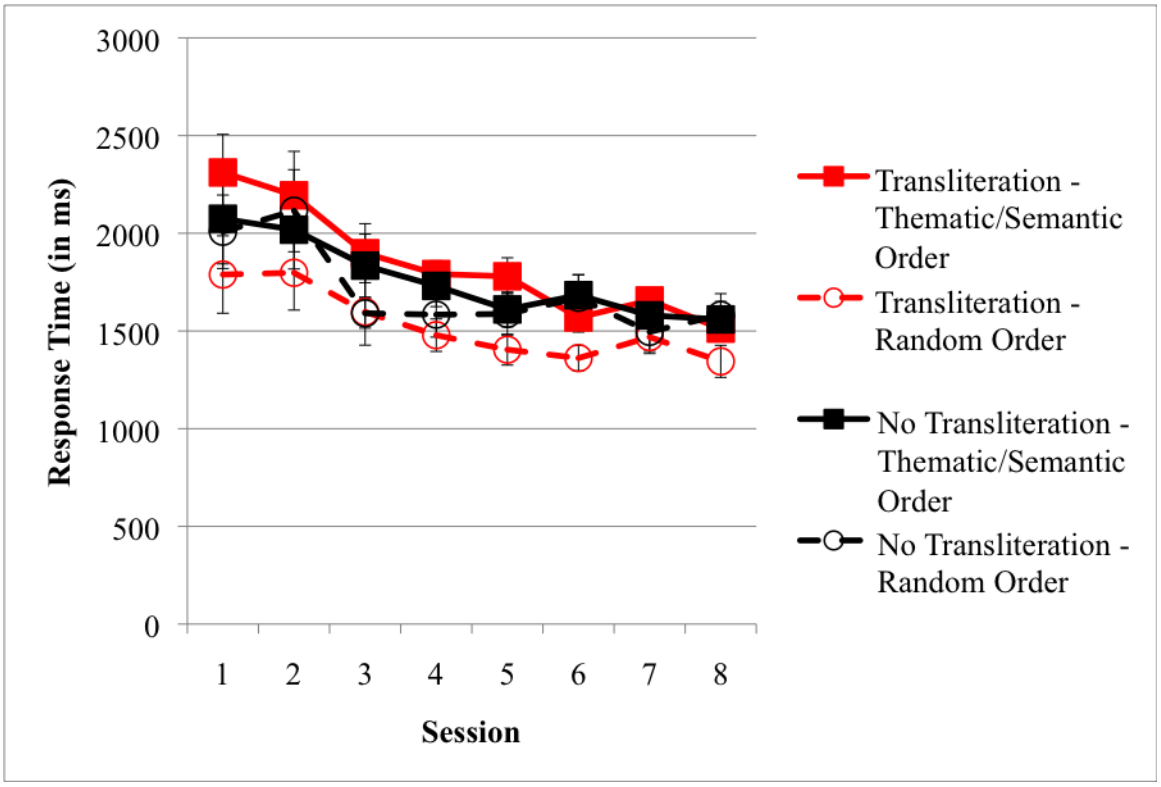

Figure 3: English-to-Arabic translation test (response time)

The interactions between Session and Order Condition observed for free recall and translation accuracy suggest that participants who were trained with words in a thematic/semantic order had an advantage in learning over participants trained with words in a random order. However, the free recall test allows participants to recall the words in any order. If thematic/semantic categories can serve as memory cue during free recall, participants who were trained with words in a thematic/semantic order may have been more aware of the relationship between the items, and may have used this awareness to their advantage. This advantage would transfer to the English-to-Arabic translation test, because words that participants correctly recalled were probably also correctly translated. Therefore, we examined whether the advantage for words that were trained in a thematic/semantic order would hold for words that were not correctly recalled. To do so, for each participant, we removed the words that he or she had

correctly recalled from his or her English-to-Arabic translation data. We then re-analyzed the data from the English-to-Arabic translation test with the free recall data removed.

Figure 4 shows English-to-Arabic translation accuracy from this re-analysis. We examined only effects of or interactions with Order Condition. Because this analysis is underpowered by participants, we report effects that are significant only by items. Words trained in a thematic/order were translated more accurately than words trained in a random order (17.3 vs. 13.7 % by items), but only in the analysis by items, F_1 (1, 27) = 0.10, MSE = 0.08, p = 0.75; F_2 (1, 78) = 10.16, MSE = 0.07, $p < 0.01$. Session and Order Condition interacted in the analysis by items, F_1 (6, 162) = 0.35, MSE = 0.01, p = 0.75; F_2 (6, 468) = 3.08, MSE = 0.03, $p < 0.05$. This interaction was probed with *t*-tests, using 0.007 as the corrected alpha level for 7 comparisons. In Sessions 4 and 6, words trained in a thematic/semantic order were correctly recalled more than words trained in a random order, all *t*s (95) < 3.157, all *p*s < 0.007. Because of missing cells, it was not possible to examine these effects in the reaction time data.

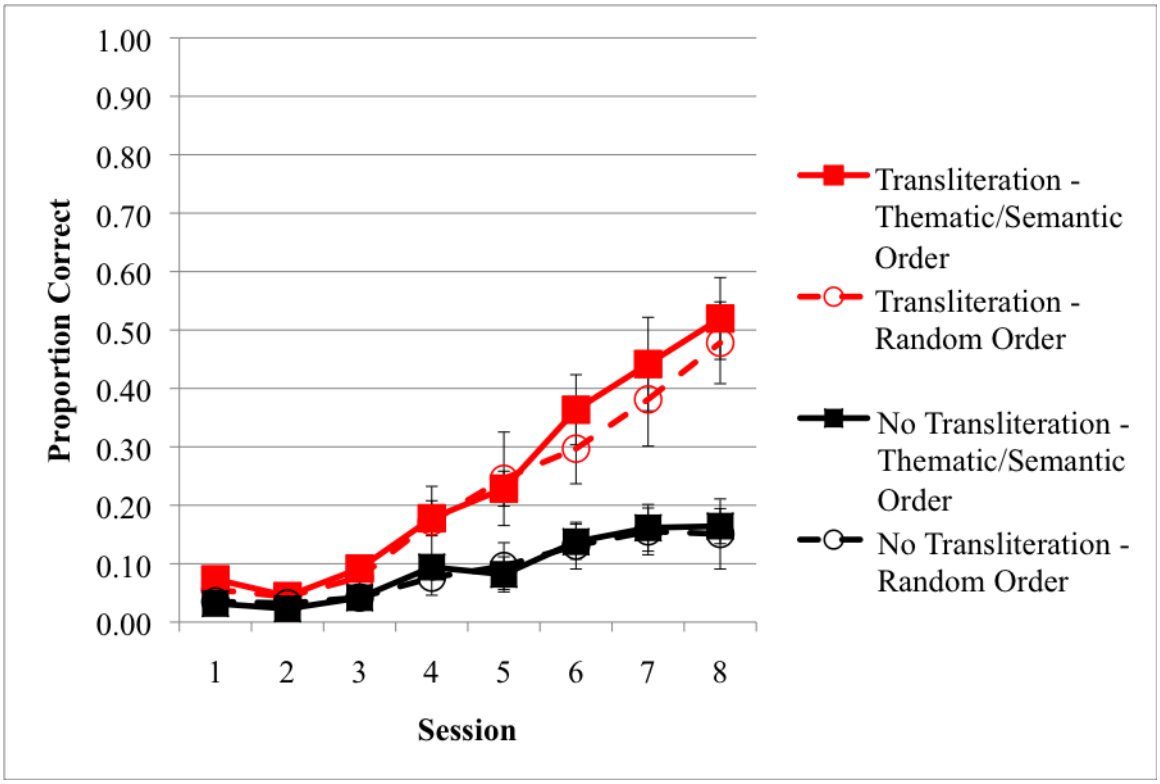

Figure 4: English-to-Arabic translation test (accuracy) with correctly recalled words removed

2.3 DISCUSSION

The LQH predicted that participants who saw transliterations would be more accurate on the free recall test and the English-to-Arabic translation test than participants who did not see transliterations. Participants who saw transliterations should have formed higher quality lexical representations as a result of having all three lexical components of the words (orthography, phonology, and meaning) available to them during training, and as a result of having two memory traces (orthography and phonology) available to them during testing. The results of this experiment are consistent with these predictions. Specifically, participants who saw transliterations recalled and translated more accurately than participants who did not see

transliterations. This transliteration advantage was observed in all sessions, including Session 1 (based on separate analyses; see Footnote 3), and increased across sessions, suggesting that transliterations provided an immediate boost to L2 Arabic learning, and continued to boost L2 Arabic learning throughout training.

Previous research has shown that training L2 vocabulary words in semantic categories has a negative effect on L2 vocabulary learning (Finkbeiner & Nicol, 2003; Folse, 2004; Kroll & Stewart, 1994). Therefore, if the category interference disadvantage in L2 vocabulary learning were due to lesser attention to word-meaning mappings than to meaning-meaning mappings, we predicted that participants who were trained with words in a random order would have been more accurate than participants who were trained with words in a thematic/semantic order. However, if the category interference disadvantage were due to overlap across exemplars at the perceptual feature level, we would expect this disadvantage to have been reduced or eliminated. The data demonstrate that participants who were trained with words in a thematic/semantic order recalled and translated more accurately than participants who were trained with words in a random order. This effect went beyond the free recall task and remained significant by items in the analysis of translation with correctly recalled words removed. These findings suggest that training L2 vocabulary in thematic/semantic categories does not have the same effect on L2 vocabulary learning as training L2 vocabulary in semantic categories, perhaps because thematic/semantic categories have less perceptual feature overlap than semantic categories.

The finding that participants who saw transliterations recalled and translated more accurately than participants who saw no transliterations suggests that participants who saw transliterations formed higher quality lexical representations as a result of having all three lexical components of the words (orthography, phonology, and meaning) available to them during

training, whereas participants who did not see transliterations formed lower quality lexical representations as a result of having only two lexical components of the words (phonology and meaning) available to them during training. This finding further suggests that participants who saw transliterations may have formed three memory traces (orthography, phonology, and meaning) that they accessed during testing, whereas participants who did not see transliterations may have formed only two memory traces (phonology and meaning).

Even though the primary task was oral production, and not written production, participants who were trained with transliterations recalled and translated more accurately than participants who were trained with no transliterations. These results suggest that beginning L2 Arabic learners use all lexical components available to them during training and testing, regardless of their relevance to the primary task. Other studies have also demonstrated that learners use all lexical components available to them during learning. For example, a study by Ricketts, Bishop, and Nation (2009) demonstrated that, if orthography was present during vocabulary learning, children used it to improve their vocabulary learning, even though they were never instructed to do so. In this study, child native English speakers were trained to associate the phonology of non-words with pictures of an object. The orthography of the non-words was provided for half of the pictures, but the children were never instructed to attend to it. Reaction times on a non-word/picture-naming posttest were shorter for non-words for which the orthography had been provided than for non-words for which the orthography had not been provided, suggesting that the children had used orthography when learning the vocabulary, which improved their vocabulary learning.

Even though transliterations were beneficial for L2 Arabic vocabulary learning, they may also have drawn cognitive resources away from associating phonology with meaning and from

learning the Arabic vocabulary. In the current study, participants who saw transliterations could have attempted to learn the unfamiliar correspondences between the spellings and the pronunciations of the transliterations, which could have drawn cognitive resources away from the primary task of oral production. The correspondences between the spelling and the pronunciations of the transliterations were unfamiliar because several Arabic phonemes do not exist in English, and the letter combinations that represent these phonemes in the transliterations provided by the DLI are not always permissible in English. As an example, in the Arabic transliteration "muDhamid" (meaning "medic"), the phoneme [dˤ] does not exist in English, and the letter combination <Dh> that represents this phoneme is not permissible in English. Further, the letters that represent phonemes in Arabic may represent different phonemes in English. Using the Arabic transliterations "Tabeeb" (meaning "doctor") and "ilteehaab" (meaning "infection") as examples, <T> represents the phoneme [tˤ] in Arabic, whereas <t> represents the phoneme [t] in both Arabic and English. Although there is a phonemic distinction between <T> and <t> in Arabic, there is not in English.

The results of two studies suggest that unfamiliar orthography and phonology affect L2 vocabulary learning. In a study by Kaushanskaya and Marian (2008), when the phonology of the L2 matched the phonology of the L1, participants more accurately mapped the phonology of the L2 onto the orthography of the L2. Further, a study by Bird and Williams (2002) found that orthography improved vocabulary learning when the phonology of the words was unfamiliar, but not when the phonology of the words was familiar. The authors suggest that, when the phonology of the words was familiar, the orthography of the words provided no more information than the phonology provided. However, when the phonology of the words was unfamiliar, the orthography of the words provided more information than the phonology of the

words provided. Transliterations improved L2 Arabic vocabulary learning in our experiment, suggesting that the transliterations provided more information than the pronunciations provided.

We conducted Experiment 1B to test the hypothesis that the unfamiliar grapheme-phoneme correspondences of the transliterations drew cognitive resources away from the primary task of oral production. In this experiment, we trained native English speakers on the same 96 spoken Arabic words and phrases used in Experiment 1A, with accompanying transliterations generated by the participants (e.g. "muvamed," meaning "medic"). Because the participants generated the transliterations on which they were trained, the grapheme-phoneme correspondences between the spellings and the pronunciations of the transliterations should be more familiar than those used in Experiment 1A. If the unfamiliar grapheme-phoneme correspondences of the transliterations drew cognitive resources away from the primary task of oral production in Experiment 1A, performance should be better when participants are trained on the transliterations that they generated than on the transliterations that the DLI provided. We also hypothesized that hearing an Arabic word and then generating a transliteration that matched what was heard would result in a higher quality lexical representation than hearing an Arabic word and then seeing a transliteration that may or may not have matched what was heard, because orthography and phonology would be more interconnected for generated transliterations than for provided transliterations. However, the phonological representation of the Arabic word may change as participants become more familiar with Arabic phonology, which may result in a corresponding change in the orthographic representation of the Arabic word. We also examine this possibility by allowing participants to indicate how they would have spelled the words at the end of the final experimental session. We directly compare the data collected in Experiment 1B to the data collected in Experiment 1A.

3.0 EXPERIMENT 1B

3.1 METHOD

3.1.1 Participants

Participants were 10 native speakers of American English from the same population as the participants in Experiment 1A. They were paid $10 per hour for up to 16 hours, and received a $50 bonus upon completion of the last session.

Data from two participants were excluded due to technical (recorder and computer) errors. Thus, the analyses were conducted on a final set of 8 participants (3 male; mean age 21.8 years).

3.1.2 Design

We used an 8 Session (1 through 8) within subjects design. Transliterations were always generated, and Order was always random. We chose to train words in a random order so that we could examine the advantage for training words with generated transliterations when no other experimental manipulations could assist learning.

3.1.3 Stimuli

The stimuli were the same 96 English words and phrases and their Iraqi Arabic translations used in Experiment 1A.

3.1.4 Procedure

Participants were trained over eight sessions (two sessions per week for four weeks). The procedure was the same as in Experiment 1, with the following exceptions. In Session 1, participants first completed a spelling test, in which they typed how they would spell the Arabic words. These participant-generated spellings were used in all subsequent training. Participants then completed the language history questionnaire (from Tokowicz et al., 2004), and were then trained. The spelling test was considered one cycle of training, as the 96 words and phrases were presented one time, so during the training portion of Session 1, the 96 words and phrases were presented only twice. In coding the free recall data, a pair was considered correct if the participant recalled the English word and gave a reasonable pronunciation of the Arabic translation, and in coding the translation data, a response was considered correct if the participant gave a reasonable pronunciation of the Arabic translation. Pronunciations based both on the transliteration provided by the DLI and generated by the participant were accepted as correct.

The sessions were again arranged such that there was one day between sessions within a week and were four days between sessions between weeks. There were no violations to this arrangement.

3.2 RESULTS

We report the data in the same manner as we reported the data for Experiment 1A. To examine whether participants who saw transliterations that they generated were more accurate in a free recall test and an English-to-Arabic translation test than participants who saw transliterations that the DLI provided, we then compared the data from this experiment to the data from the Transliteration/Random Order Condition from Experiment 1A. We performed a 7 Session (2 through 8) x 2 Transliteration Condition (Generated Transliterations vs. Provided Transliterations) mixed ANOVA. In both the analysis by participants (reported as F_1) and the analysis by items (reported as F_2), Session was a within-participants variable and Transliteration Condition was a between-participants variable. For the RT analysis, 3.42 % of the data were excluded as outliers.

Free Recall accuracy is shown in Figure 5. Accuracy increased across sessions, F_1 (6, 84) = 69.96, MSE = 0.02, $p < 0.01$; F_2 (6, 570) = 275.83, MSE = 0.04, $p < 0.01$. No other effects were significant in both analyses. However, words trained with the transliterations that the DLI provided were correctly recalled significantly more often than words trained with the transliterations that the participants generated (31.2 vs. 24.9 %), according to the analysis by items, F_1 (1, 14) = 0.82, MSE = 0.14, $p = 0.38$; F_2 (1, 95) = 29.76, MSE = 0.05, $p < 0.01$. This effect increased across sessions as indicated by a Session x Transliteration Condition interaction in the analysis by items, F_1 (6, 84) = 0.50, MSE = 0.02, $p = 0.57$; F_2 (6, 570) = 2.95, MSE = 0.02, $p < 0.05$. This interaction was probed with *t*-tests, using 0.007 as the corrected alpha level for 7 comparisons. Words trained with the transliterations that the DLI provided were correctly recalled significantly more in Sessions 3 through 8, all *t*s (95) > 2.81, all *p*s < 0.007.

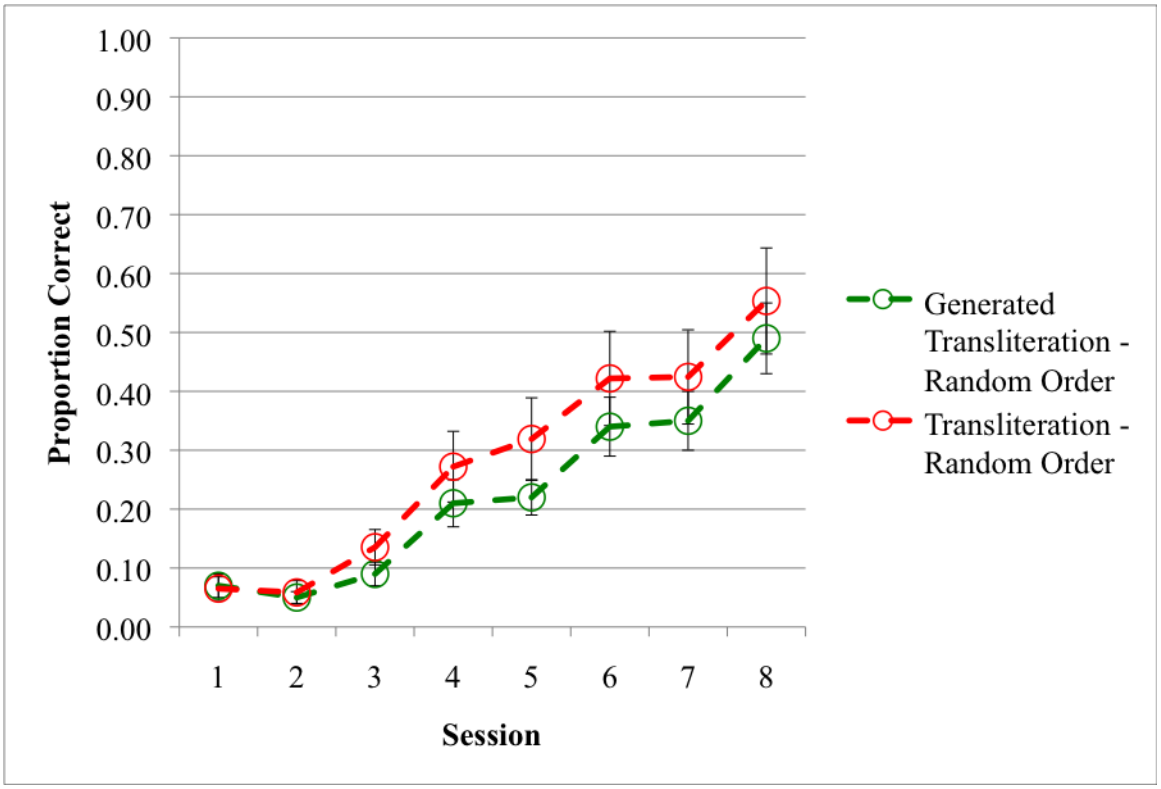

Figure 5: Free recall test (accuracy)

English-to-Arabic translation accuracy data is shown in Figure 6. Accuracy increased across sessions, F_1 (6, 84) = 97.87, *MSE* = 0.02, $p < 0.01$; F_2 (6, 570) = 408.39, *MSE* = 0.04, $p < 0.01$. No other effects were significant according to both analyses, however, words trained with the transliterations provided by the DLI were translated more accurately than words trained with the transliterations that the participants generated (41.2 vs. 34.7 %), according to the analysis by items, F_1 (1, 14) = 0.61, *MSE* = 0.20, $p = 0.45$; F_2 (1, 95) = 24.20, *MSE* = 0.06, $p < 0.01$. As for the free recall task, this effect became larger across sessions, as indicated by a Session x Transliteration Condition interaction in the analysis by items, F_1 (6, 84) = 0.35, *MSE* = 0.02, $p = 0.68$; F_2 (6, 570) = 2.37, *MSE* = 0.02, $p < 0.05$. This interaction was probed with *t*-tests, using 0.007 as the corrected alpha level for 7 comparisons. Words trained with the transliterations that

the DLI provided were correctly translated significantly more in Sessions 3 through 8, all *t*s (95) > 2.76, all *p*s ≤ 0.007.

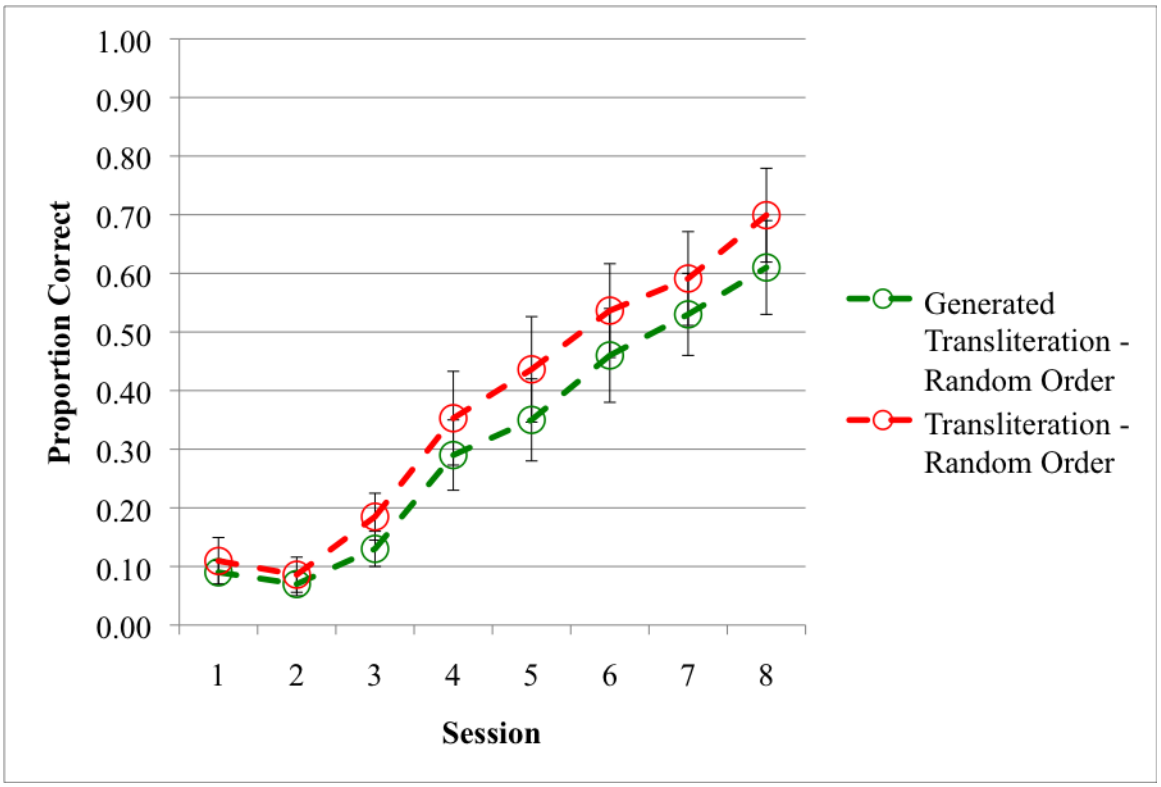

Figure 6: English-to-Arabic translation test (accuracy)

RTs decreased across sessions, F_1 (6, 84) = 6.15, *MSE* = 131454, $p < 0.01$; F_2 (6, 78) = 7.63, *MSE* = 697561, $p < 0.01$. No other results were significant.

3.3 DISCUSSION

If the unfamiliar grapheme-phoneme correspondences of the transliterations drew cognitive resources away from the primary task of oral production in Experiment 1A, performance should

have been better when participants were trained on the transliterations that they generated than on the transliterations that they were provided. When participants were trained on the transliterations that they generated, the transliterations that they saw should have matched what they heard, because the grapheme-phoneme correspondences were familiar. Therefore, they should have formed a high quality lexical representation. However, when participants were trained on the transliterations that they were provided, the transliterations that they saw may or may not have matched what they heard, because the grapheme-phoneme correspondences may have been unfamiliar. Therefore, they may have formed a low quality lexical representation.

Participants who were trained with transliterations that they generated recalled and translated less accurately than participants who were trained with transliterations that were provided. Therefore, the results of this experiment do not support the hypothesis that performance should be better when participants are trained with the transliterations that they generated than with the transliterations that they were provided, suggesting that the unfamiliar grapheme-phoneme correspondences of the transliterations did not draw cognitive resources away from the primary task of oral production in Experiment 1A.

There are several possible explanations for why participants who were trained with the transliterations that they generated were less accurate than participants who were trained with the transliterations that they were provided. By having participants generate transliterations, we may have encouraged participants to explicitly attend to orthography over phonology. Because the primary task was oral, not written, production, explicitly attending to orthography over phonology may have negatively affected performance during testing. We are unable to test this possibility with our existing data, but would expect that these participants who generated

transliterations would be better able to spell the transliterations than participants who were provided transliterations.

A second possible explanation for why participants who were trained with the transliterations that they generated were less accurate than participants who were trained with the transliterations that they were provided is that, as participants became more familiar with Arabic phonology, their phonological representations changed. To test this hypothesis, we compared the transliterations that were generated before the experiment to the transliterations that were generated after the experiment. On average, 55% of the generated transliterations changed in spelling, and 54% of those also changed in pronunciation. This demonstrates that the transliterations that participants generated at the start of the experiment did not match what the participants were hearing by the end of the experiment. Therefore, during testing, participants who were trained with the transliterations that they generated may have had access to three memory traces (orthography, phonology, and meaning) for approximately half of the words, and two memory traces (phonology and meaning) for approximately half of the words. Also consistent with this explanation is the finding that individuals who learned with provided transliterations are at an even higher advantage than individuals who learned with generated transliterations in later sessions than they were in early sessions.

Note that although generated transliterations are not better than provided transliterations, a comparison of means suggests that they are still more accurate than the participants in Experiment 1A who were trained without any transliterations (see Figures 7 and 8). This may be because having three representations, even if one is imperfect, is better than having only two representations.

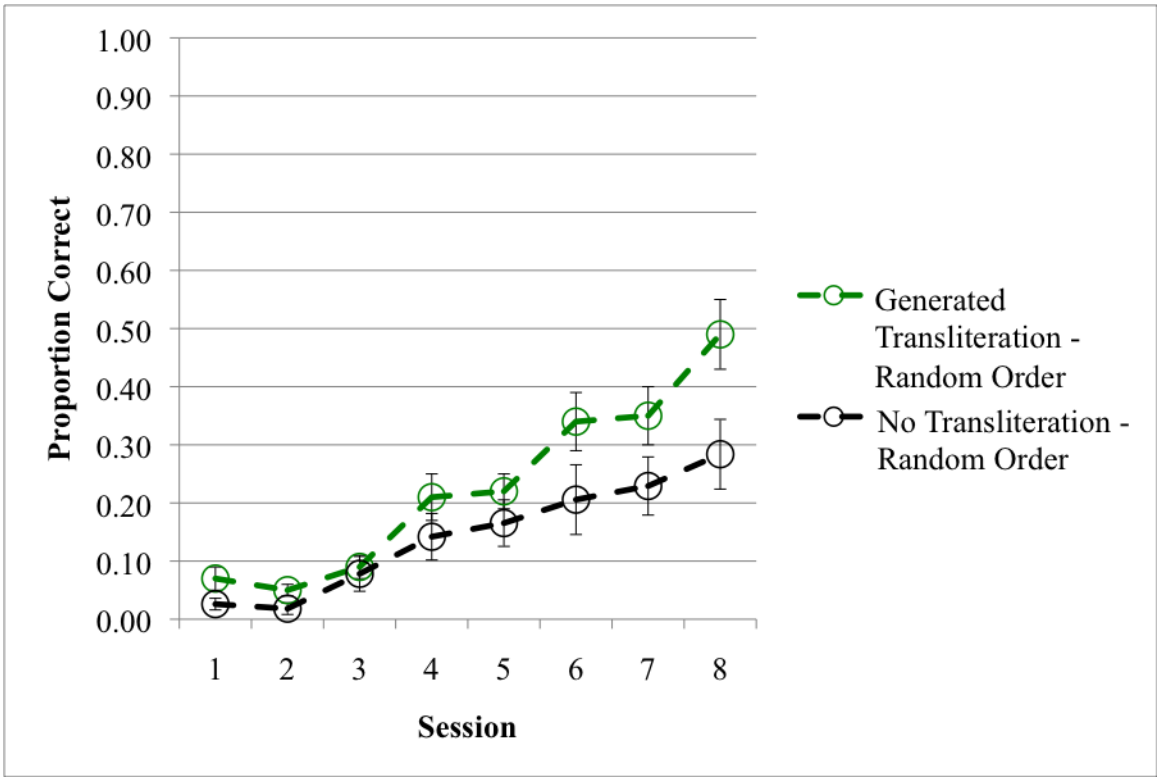

Figure 7: Free recall test (accuracy)

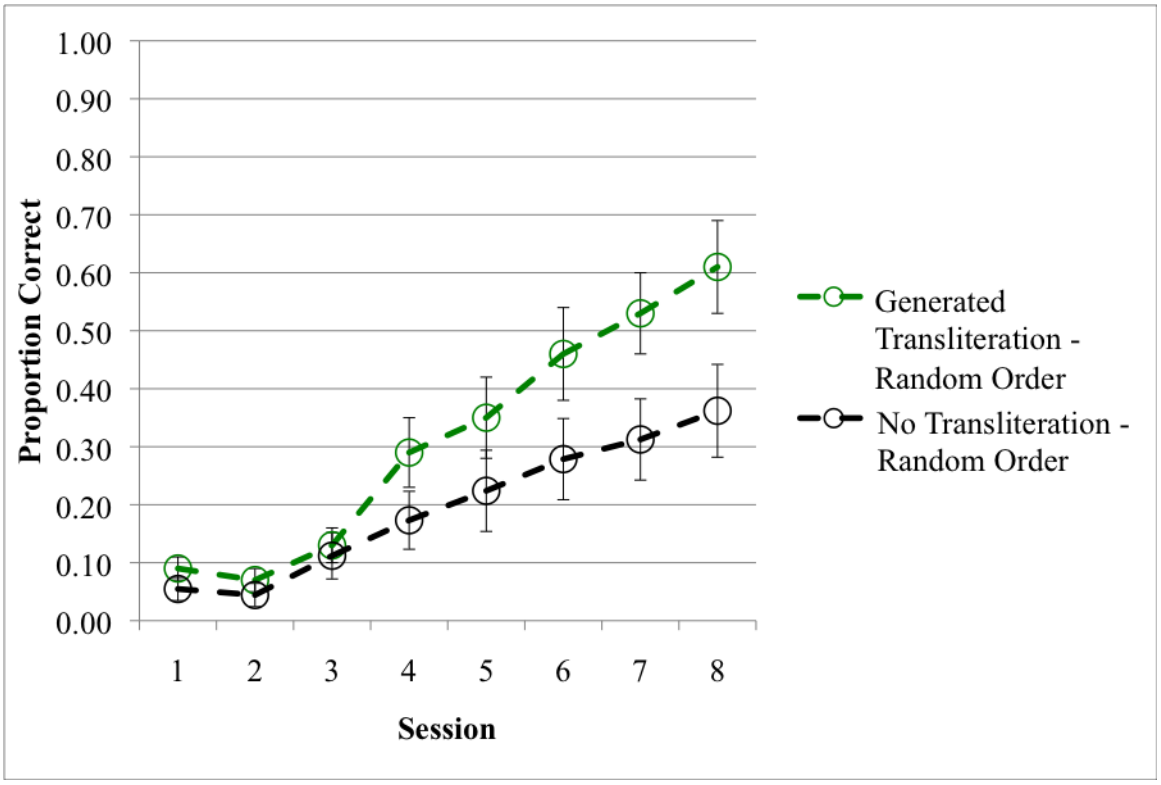

Figure 8: English-to-Arabic translation test (accuracy)

4.0 GENERAL DISCUSSION

One of the most robust effects in this study was the main effect of session. In every condition, accuracy increased across sessions, and reaction times decreased across sessions. Furthermore, participants in the most accurate condition, the Transliteration-Thematic/Semantic Order Condition, correctly recalled 64.3 % of the words and correctly translated 77.7 % of the words in Session 8, which is impressive, given that the participants had been presented with each word only 21 times.

In Experiment 1A, participants who saw transliterations recalled and translated more accurately than participants who saw no transliterations, and, further, participants who were trained with words in a thematic/semantic order recalled and translated more accurately than participants who were trained with words in a random order. The finding that participants who saw transliterations recalled and translated more accurately than participants who saw no transliterations suggests that participants who saw transliterations formed higher quality lexical representations of the Arabic words than participants who did not see transliterations. Participants who saw transliterations formed higher quality lexical representations of the Arabic words because all three lexical components of the Arabic words (orthography, phonology, and meaning) were available during training. Participants who did not see transliterations formed lower quality lexical representations of the Arabic words because only two lexical components of the Arabic words (phonology and meaning) were available during training. Therefore, adults

can use well-established connections between spelling and sound to form high quality lexical representations of L2 vocabulary words that are useful for L2 vocabulary learning.

Furthermore, participants who were trained with words in a thematic/semantic order recalled and translated more accurately than participants who were trained with words in a random order. This finding suggests that training L2 vocabulary in thematic/semantic categories does not have the same effect on L2 vocabulary learning as training L2 vocabulary in semantic categories, perhaps because the perceptual features of members of thematic/semantic categories overlap less than the perceptual features of members of semantic categories.

The present study had a few limitations. First, it was not possible to cleanly examine performance on the English-to-Arabic translation test without any influence from the free recall test. However, we addressed this issue in a limited way in Experiment 1A by removing the words that had been correctly recalled in the free recall test from the English-to-Arabic translation data and then calculating a new proportion of words that had been correctly translated. Also, we did not have a native speaker of Iraqi Arabic assess the accuracy of the pronunciations of the L2 words, which prevented a fine-grained examination of the preciseness of the pronunciations of the L2 words and of the changes in oral production over time. Future research could address these issues in more detail.

In summary, the present study demonstrated learning advantages for training L2 vocabulary words with transliterations and in thematic groupings. This study therefore extends existing research by examining L2 vocabulary learning in adults over an extended protocol, and with thematic groupings that involved more than concrete, pictureable objects.

5.0 FOOTNOTES

[1] The data from these tests are outside the scope of the present study and will not be reported here.

[2] At the end of Session 8, participants completed a spelling test, in which they typed how they would have spelled the Arabic words, based on their pronunciations.

[3] The data for Session 1 were analyzed separately to examine whether effects emerged early in training. Although the effects were not fully significant in the all analyses, the advantage for transliteration conditions emerged in Session 1 for free recall, F_1 (1, 28) = 3.20, MSE = 0.002, p = 0.09; F_2 (1, 95) = 4.10, MSE = 0.02, $p < 0.05$, and for translation accuracy, F_1 (1, 28) = 4.20, MSE = 0.004, p = 0.05; F_2 (1, 95) = 13.97, MSE = 0.02, $p < 0.01$.

APPENDIX A

STIMULI

Category	English Word or Phrase	Arabic Translation
Commands, Warnings,	Be quiet.	iskut
and Instructions	Don’t shoot.	latermee
	Follow me.	ilHagnee
	Give me.	inTeenee
	Help me	sa'ednee
	Keep away.	ibta'ed
	Put your weapon down.	Thib slaaHak
	Stay here.	ibqa hna
	Stop or I will shoot!	ogaf tara armee!
	Stop!	ogaf!
	Unload	faregh
	We must search you.	laazim infatshak
Helpful Words, Phrases,	Can someone assist us?	yegdar aHad yeesa'edna?
and Questions	Danger	KhaTar
	Do you have___ ?	'idkum ___ ?

	Excuse me / I’m sorry.	il 'afu / anee mit-asif
	He / She is	huwa / heeya
	Here	hna
	How?	shlon?
	I am	anee
	I do not want.	maa areed
	I want.	areed
	No	laa
	Please.	rajaa-an
	Thank you.	shukran
	There	hnaak
	They are	huma
	We are	iHna
	What?	shinoo?
	When?	shwakit?
	Where?	wayn?
	Who?	minoo?
	Why?	laysh?
	Yes	na'am
	You are	inta
Greetings / Introductions	Good bye.	ma'a as salama
Interrogation	Do you have any identification papers?	'indak haweeya?
	Do you understand?	datiftehim?
	I don’t understand.	mad daftehim
Days of the Week / Time	Day	nahaar

	Today	il yom
	Tomorrow	bachir
	Yesterday	il baarHa
Directions	Down	jawa
	Left	yesra
Emergency Terms	Distress signal	ishaaret najda
	Emergency!	Tawaari!
	Evacuate the area!	iKhloo il manTaqa!
	Help! (help me)	liHgoolee!
	We need a doctor!	niHtaaj Tabeeb!
Food and Sanitation	Drink	ishrab
	Food	akel
	Water	maay
	Where is the latrine?	wayn il maraafiq iS SeHeeya?
Fuel and Maintenance	Gasoline	banzeen
Medical / General	Antibiotics	duwa maal ilteehaab
	Bandage	lafaaf
	Burn	Harig
	Clean	naDheef
	Dead	mayit
	Doctor	Tabeeb
	Fever	Haraara
	Hospital	mustashfa
	I am a doctor.	anee Tabeeb
	I am not a doctor.	anee moo Tabeeb

	I will take you to the hospital.	raH aaKhThek lil mustashfa
	Infection	ilteehaab
	Injured	majrooH
	Medic	muDhamid
	Medicine	duwa
	Poison	sam
	Sick	mareeDh
	Wound	jareH
Medical / Body Parts	Foot	rejil
	Head	raas
	Leg	rejil
Lodging	Is there a telephone available?	aku taleefon?
	We need ___ gallons of potable water.	iHna miHtaajeen __ galanaat maay shurib
Customs (Port of Entry)	False	moo SaHeeH
	Owner	SaaHib
	Passport	jawaaz
	Permission	muwaafaqa
	Prohibited	memnoo'
	Property	amlaak
	Visa	veeza
Relatives	Family	ahel
	Man	rijaal
	Relatives	garaayib
General Military	Ammunition	'etaad
	Commander	qaa-id

Gun	musedas
Mine	lughum
Minefield	Haqel alghaam
Mortar	haawen
Refugee	laaji
Shelter	malja

BIBLIOGRAPHY

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology, 35,* 126-140.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23,* 509-533.

Defense Language Institute Foreign Language Center. (2005). *Iraqi Basic Language Survival Guide.* Presidio of Monterey, CA.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16,* 143-149.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24,* 369-383.

Folse, K. S. (2004). *Vocabulary Myths: Applying Second Language Research to Classroom Teaching.* Ann Arbor, MI: University of Michigan Press.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrics, 24,* 95-112.

Hu, C. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing, 21,* 823-841.

Kaushanskaya, M., & Marian, V. (2008). Mapping phonological information from auditory to written modality during foreign vocabulary learning. *Annals of the New York Academy of Sciences, 1145,* 56-70.

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

Kučera, & Francis, W.N. (1967). *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

MRC Psycholinguistic Database. Available from http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

Nelson, J. R., Balass, M., & Perfetti, C. A. (2005). Differences between written and spoken input in learning new words. *Written Language & Literacy, 8,* 25-44.

Perfetti, C. A. & Hart, L. (2001). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), Precursors of Functional Literacy (Vol. 11, pp. 67-86). Amsterdam: John Benjamins.

Psychological Software Tools, Inc. (2000). E-Prime [Computer software}. Pittsburgh, PA.

Ricketts, J., Bishop, D. V. M., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *The Quarterly Journal of Experimental Psychology, 62,* 1948-1966.

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

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

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

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology-Human Experimental Psychology, 49,* 51-79.