The effects of transliterations, thematic organization, and working memory on adult L2 vocabulary learning

Alison M. Tseng, Martin C. Doppelt and Natasha Tokowicz
University of Pittsburgh

We examine how L2 vocabulary learning is affected by the information provided to the learner during training, organization of the to-be-learned vocabulary, and working memory capacity of the learner. Native English speakers were taught Arabic vocabulary in seven sessions, during which they heard L1 (English)-L2 (Arabic) translation pairs. Training was manipulated between participants by crossing the presence vs. absence of a transliteration and thematic vs. random organization of vocabulary. Session, working memory capacity, transliteration condition, and organization condition interacted in English-Arabic translation accuracy. Participants with lower working memory capacity performed best in the transliteration-thematic organization condition, whereas participants with higher working memory performed best in the transliteration-random organization and no transliteration-thematic organization conditions. Translation RT and free recall were not related to working memory, and were best in the transliteration conditions. Results suggest that adult L2 vocabulary learning is aided by exploiting well-established spelling-sound connections to improve L2 lexical representations.

Keywords: second language learning, vocabulary learning, transliterations, working memory, thematic organization

1. Introduction

How is adult second language (L2) vocabulary learning affected by the information that is provided to the learner during training, the organization of the to-be-learned vocabulary, and the working memory capacity of the learner? These three factors are examined in the present study on adult native English speakers learning Arabic vocabulary.

https://doi.org/10.1075/jsls.17018.tse
Journal of Second Language Studies 1:1 (2018), pp. 141–165. issn 2542-3835 | e-issn 2542-3843

The information that is provided to the learner during training is thought to influence the likelihood that a high quality lexical representation will be formed. According to the Lexical Quality Hypothesis (LQH; Perfetti & Hart, 2002), a high quality lexical representation is formed when a word's orthography, phonology, and semantics are highly specified and interconnected. In the present study, we examined how the lexical representations that we provide to beginning adult L2 learners during training, as well as the organization of the items being learned, affect beginning adult L2 vocabulary learning. We further examined how individual differences in working memory capacity may mediate these effects. We examined this over the course of four weeks, which is a longer duration of time than is typically examined in laboratory studies on beginning adult L2 vocabulary learning.

A common strategy is to teach L2 learners vocabulary that is organized either semantically or thematically. This may be done for communicative purposes or to try to take advantage of schemas or mental frameworks as an aid to memory (e.g., Brewer & Nakamura, 1984). Thematic organization involves putting words/phrases from a schema together, regardless of their hierarchical/categorical relationship or part of speech. For example, a "beach" schema could include nouns (e.g., "sand"), verbs (e.g., "swim"), adjectives (e.g., "sunny"), and phrases (e.g., "suntan lotion"). This thematic organization contrasts with semantic organization, in which all words have the same hierarchical relationships (e.g., "is a fruit"), are all the same part of speech, and typically have shared features. Whereas there is some evidence that thematic organization of vocabulary can improve learning (Tinkham, 1997; see Elgort, 2011, for discussion), there is evidence that semantic organization can hinder learning and processing (e.g., Finkbeiner & Nicol, 2003; Kroll & Stewart, 1994; see Folse, 2004, for a review).

Tinkham (1997) compared learning vocabulary under thematically- and semantically-organized list conditions. Participants learned lists of English-artificial language word pairs. Within each list, words were semantically related (i.e., "jacket", "shirt"), semantically unrelated (i.e., "beard", "island"), thematically associated (i.e., "beach", "sunny"), or thematically unassociated (i.e., "improve", "triangle"). In the semantically-related and -unrelated conditions, words were the same part of speech, whereas in the thematically-associated and -unassociated conditions, they were not. Learning was improved by thematic, but hindered by semantic, organization, as measured by time-to-criterion.

In a related study by Choi (2003), native English-speaking participants learned a list of English-Korean word pairs that were semantically related, thematically associated, or semantically unrelated and thematically unassociated. Learning was not affected by the organization of items, as measured by time-to-criterion and retention on post-tests. This contrasts with Tinkham's (1997) finding

that learning was improved by thematic, but hindered by semantic, organization (see also Finkbeiner & Nicol, 2003). As a possible explanation for the contrasting findings, Choi noted that in Tinkham word presentation was either spoken or written, but that in Choi word presentation was spoken and written. Therefore, Choi proposed that joint spoken and written word presentation may require more cognitive processing than spoken or written word presentation alone. Increased cognitive processing may not leave participants with sufficient cognitive resources to attend to the organization of items, resulting in a null effect of list organization.

Choi's (2003) explanation is that the increase in cognitive processing from spoken and written word presentation during learning may reduce the usefulness of list organization. This raises the possibility that learners with more cognitive resources, such as higher working memory capacity, may be able to benefit from thematic organization to a greater extent than learners with fewer cognitive resources. This is because learners with higher working memory capacity have additional cognitive resources and, therefore, could have sufficient cognitive resources to attend to the thematic organization of the items. Further, learners with more cognitive resources may benefit more from joint spoken and written word presentation during training than learners with fewer cognitive resources, because their cognitive resources may be sufficient for attending to the presented information. Here, we explore these possibilities by examining whether working memory capacity interacts with our training manipulations. Specifically, we used a version of a thematic organization manipulation as well as the presence/absence of a written form of the Arabic vocabulary using the Latin script during training.

As mentioned above, the information that is provided to the learner during training can influence the likelihood that a high quality lexical representation will be formed. Therefore, we manipulated the information that we provided to beginning adult L2 learners during training. Specifically, we provided the spoken forms of L2 Arabic vocabulary, with or without written forms using the Latin script instead of the Arabic script. These written forms are hereafter referred to as *transliterations*. Transliterating refers to substituting one script for another (e.g., Clauson, 2002). In this case, transliterations may improve the lexical representations of L2 Arabic vocabulary in several ways, which may subsequently facilitate L2 Arabic vocabulary learning. First, transliterations provide an orthographic representation that is not present when only a spoken form of the L2 Arabic vocabulary is provided. Second, transliterations may clarify the spoken forms of the L2 Arabic vocabulary, because some of the sounds in the L2 Arabic vocabulary words will be unfamiliar to the learners. Third, transliterations may help the learners better remember the spoken forms of the L2 Arabic vocabulary. All of these may then improve L2 Arabic vocabulary learning. However, if Choi's (2003) explanation is correct, the presentation of both visual and auditory information during training

may negate or reduce the effectiveness of our thematic organization manipulation entirely, or only for the subset of learners who have fewer cognitive resources (i.e., lower working memory capacity).

The idea that manipulating the information that is provided to the learners during training will affect word learning is consistent with the LQH, according to which a high quality lexical representation is formed when a word's orthography, phonology, and semantics are highly specified and interconnected. In related research, Nelson, Balass, and Perfetti (2005) taught adult native English speakers rare English words, presenting their orthography and meaning (OM) or phonology and meaning (PM). Participants reached a 100% accuracy criterion significantly earlier for words trained in the OM than PM condition. Nelson et al. concluded that OM training resulted in the creation of orthographic, phonological, and semantic representations, because phonology could be derived automatically and reliably from orthography via grapheme-phoneme correspondence rules. In contrast, PM training resulted in the creation of only phonological and semantic representations, because orthography could not be derived automatically and reliably from phonology. Thus, the availability of an additional representation improved L1 vocabulary learning.

A related study was conducted by Hu (2008), who trained native Mandarin Chinese-speaking children on the pronunciations of English pseudonames for novel cartoon characters. During training, a spoken and written form of a pseudoname was provided; the pseudoname was either written using letters from the English alphabet (e.g., nof) or using symbols (e.g.,]✚□). In a subsequent naming test, children more accurately named the cartoon characters whose pseudonames had been written using letters than using symbols. Hu concluded that, consistent with the LQH, training pseudonames written using letters resulted in a higher quality lexical representation than training pseudonames written using symbols. This is because the letters provided meaningful information about the pseudoname's pronunciation, whereas the symbols did not. These results therefore provide evidence that orthography improves auditory L2 vocabulary learning. However, because children and adults may differ in how they process an L2, it is as yet unclear whether these results would generalize to an adult learner population.

1.1 The present study

In the present study, we examined how L2 vocabulary learning is affected by the information that is provided to the learner during training, the organization of the to-be-learned vocabulary, and the working memory capacity of the learner. To do so, we trained adult native English speakers with no prior exposure to Arabic on English-Arabic translations. Participants attended eight sessions; training took

place in the first seven sessions and testing took place in all eight sessions. During training, participants heard an L1 English word/phrase and its L2 Arabic translation. Training was manipulated between participants in four conditions formed by crossing the presence vs. absence of a transliteration and the thematic vs. random organization of items. In every session, participants were tested in a free recall test, in which they orally recalled as many English words/phrases and their Arabic translations as possible (in any order), and in an English-Arabic translation test, in which they attempted to orally provide the Arabic translation of the visually presented English words/phrases, which were presented in a new random order on each occasion.

1.2 Research questions and hypotheses

We asked three research questions. First, we asked whether, when compared to random organization, thematic organization would provide an advantage (as in Tinkham, 1997), a disadvantage, or a null effect (as in Choi, 2003), in learning L2 Arabic vocabulary. Second, we asked how the presence/absence of transliterations would affect learning L2 Arabic vocabulary. Third, we asked if individual differences in cognitive abilities (specifically working memory capacity) would interact with our training manipulations.

Because some previous research has shown that training words in themes positively affects learning, we expected that thematic organization would lead to better retention, at least for some of our participants. Beyond any inherent benefit to this type of training, participants trained with items in themes could use the theme as a recall cue during free recall testing, which could result in better performance.

Based on the LQH, we predicted that participants who saw transliterations during training would recall more items and translate more accurately because of having formed higher quality lexical representations. As mentioned earlier, this could be the case because participants who see transliterations would have had all three lexical components of items (orthography, phonology, and semantics) provided to them during training, and should also have three lexical representations they can access during testing. In addition to, or alternatively, the presence of transliterations may have helped to clarify the spoken forms of the L2 vocabulary that the participants heard during training. The transliterations may have also improved memory for the spoken forms of the L2 Arabic vocabulary. Alternatively, because the free recall and English-Arabic translation tests emphasize oral production, the tests may favor more focused training with only the two relevant lexical representations (phonology and semantics; e.g., Transfer Appropriate Processing; Morris, Bransford, & Franks, 1977).

However, if the presence of both a written and spoken form requires more cognitive processing than the presence of one form alone (Choi, 2003), then participants who see transliterations during training may have fewer (and perhaps insufficient) cognitive resources to attend to the organization of items, compared to participants who do not see transliterations during training. Therefore, we may expect an effect of thematic organization only in the absence of transliterations. Further, we might expect participants with higher working memory capacity to demonstrate larger effects of our thematic organization manipulation, because they may have more (and perhaps sufficient) cognitive resources to attend to the thematic organization of items when both written and spoken forms of the items are present, compared to participants with lower working memory capacity. Previous research has also suggested that monolinguals with higher phonological working memory have an advantage in learning phonologically unfamiliar L2 words relative to their peers with lower phonological working memory (Kaushanskaya, 2012). Although we did not specifically measure phonological working memory in the present study, we would expect an overall benefit for individuals with higher working memory based on this study and other related research (see review in Linck, Osthus, Koeth, & Bunting, 2014).

2. Method

2.1 Participants

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

Data from four participants were not complete: two due to scheduling conflicts, one due to experimenter error, and one due to a recorder error. Analyses were therefore conducted on a final set of 32 participants (10 male; mean age = 20.1 years) divided evenly among the four training conditions. All participants gave informed consent and the study was carried out with approval from the university's Institutional Review Board.

2.2 Design

We used an 8 Session (1–8) x 2 Transliteration Condition (transliterations vs. no transliterations) x 2 Organization Condition (thematic vs. random) mixed design.

Session was a within-participants factor and Transliteration Condition and Organization Condition were between-participants factors.

2.3 Stimuli

The stimuli were 51 English words and 34 English phrases (for a total of 85 items)[1] and their Iraqi Arabic translations, selected from the Iraqi Basic Language Survival Kit materials from the United States Defense Language Institute (DLI; see Appendix). These materials include a booklet with a list of translations and their transliterations, organized thematically, and sound files that provide pronunciations in English and Arabic (these materials are available for download at <http://fieldsupport.dliflc.edu/products/iraqi/ir_bc_LSK/default.html>). We chose to use these materials because they are freely available and because they are used by actual language learners. The specific items were selected by five research assistants to be important for survival in a foreign country, and were drawn from eight broad themes. Characteristics of these items are summarized in Table 1.

Table 1. Stimulus characteristics

Measure	Words ($n = 51$)	Phrases ($n = 34$)
English length (number of letters)	5.96 (2.24)	15.76 (8.72)
Transliteration length (number of letters)	7.43 (2.58)	13.29 (7.42)
English frequency	3.27 (1.15)	4.98 (.63)

Note. Mean (Standard Deviation). English frequency is the $SUBTLEX_{US}$ Lg10 word frequency from Brysbaert and New (2009). Mean word frequency was calculated by averaging the frequencies of the 51 words. Mean phrase frequency was calculated by averaging the frequency of each of the words in each of the 34 phrases, and then by averaging those averages.

2.4 Procedure

Participants were assigned to one of the four training conditions and attended two sessions per week for four weeks. Sessions 1–7 lasted approximately two hours each. Session 8 lasted approximately one hour. There was one day between ses-

1. The original set of items included 61 English words and 35 English phrases (for a total of 96 items), but three of the organizations included items that could be considered semantically related as well as thematically associated, therefore, these were removed. Additional analyses were performed to determine whether this removal had an impact on the outcome of the study. The pattern of significance and direction of the findings was nearly identical. The only change was that in the free recall analysis, the interaction between transliteration condition and session that had been fully significant with all items included became marginally significant at the $p = .06$ level with the removal of these items.

sions within a week and four days between sessions between weeks. Several minor violations to this occurred due to severe weather conditions; the average number of days between Sessions 5 and 6 was .97, between Sessions 6 and 7 was 4.19, and between Sessions 7 and 8 was .91.

In Session 1, participants were tested after training to promote long-term retention (e.g., Roediger & Karpicke, 2006). To examine the retention of the vocabulary between sessions, in Sessions 2–7, participants were tested before training, and in Session 8, participants were tested only (i.e., were not also trained).

See Table 2 for a summary of the procedure. In every session, participants first completed free recall (so that the test with the least information provided preceded the test with the most information provided). This was followed by an English-Arabic translation production test and an individual difference test, which was a Stroop task variant (Sessions 1 and 5; Stroop, 1935), the operation-word span task (Sessions 2 and 6; Turner & Engle, 1989), the Waters reading span task (Sessions 3 and 7; Waters & Caplan,1996), or the Flankers task (Sessions 4 and 8; Eriksen & Eriksen, 1974). The primary purpose of these individual difference tests was to use these data to examine whether our training manipulations differentially influence learners of different cognitive abilities. Secondarily, they served as a buffer between testing and training. The Waters reading span task yielded the largest number of significant findings in our analyses,[2] therefore the other tasks will not be mentioned further. At the end of Session 1, participants completed a language history questionnaire (Tokowicz, Michael, & Kroll, 2004), which asked participants about their prior language-learning experiences (see Table 3).

2.4.1 *Training*

On training trials, a fixation (+) was presented at the center of the screen until a button press. In all conditions, participants then heard an English word/phrase (pronounced once) and its Arabic translation (pronounced twice, the second time slower than the first), and saw the English word/phrase centered on the top half of the screen. In the transliteration conditions, participants also saw the transliteration centered on the bottom half of the screen. All participants were instructed to repeat the pair aloud twice and to press a button after repeating the pair, at which

2. To determine which individual difference task to explore, the three dependent measures were correlated with the individual difference measures on the first session on which these tasks were administered. We therefore had seven sessions of data per task x three dependent measures x four individual difference tasks to examine. The Waters reading span task correlated with the data significantly on 14 occasions, the Stroop task eight, the operation-word span task seven, and the Flankers task five. We therefore used the Waters task in our primary analyses.

Table 2. Summary of the procedure

Session 1	Sessions 2–7	Session 8
Training	Free recall test	Free recall test
Stroop Individual difference task	English-Arabic translation production test	English-Arabic translation production test
Free recall test	Individual difference task (varied)	Flankers Individual difference task
English-Arabic translation production	Training	
Language history questionnaire		

Table 3. Participant responses on the language history questionnaire

Age began learning L2	11.0 (5.3)
L1 Reading	9.7 (0.7)
L2 Reading	4.3 (2.3)
L1 Writing	9.3 (0.9)
L2 Writing	3.0 (1.9)
L1 Conversation	9.8 (0.4)
L2 Conversation	3.5 (2.5)
L1 Comprehension	9.8 (0.5)
L2 Comprehension	4.2 (2.7)

Note. Standard deviations are in parentheses. Reading, writing, conversation, and comprehension ability ratings are on a scale ranging from 0 (worst) to 10 (best). L2 varied across participants, but was not Arabic, Hebrew, or Turkish.

point the next fixation appeared. In the thematic organization conditions, items were presented in the same order as in the Iraqi Basic Language Survival Guide, and in the random organization conditions, items 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 session. All items were presented once per training run.

2.4.2 *Free recall test*

During free recall, participants typed the English words/phrases that they recalled, and then pronounced the English words/phrases that they had typed and their Arabic translations. Vocal responses were digitally recorded and later coded for accuracy. Throughout this study, responses were considered "correct" if participants recalled the English word/phrase and gave a reasonable pronunciation of its

Arabic translation (i.e., it was clear to the coder that the participant intended the pronunciation of the correct Arabic translation and not a different trained item).

2.4.3 *English-Arabic Translation test*

During English-Arabic translation, participants saw English words/phrases and pronounced the Arabic translations of the English words/phrases that they saw. A fixation was presented at the center of the screen until a button press. An English word/phrase then appeared centered on the top half of the screen until the onset of a vocal response, at which point the fixation reappeared. Response times were recorded in ms from the onset of the stimulus to the onset of articulation. Vocal responses were digitally recorded and later coded for accuracy. The items were presented in a new randomly-generated order determined by E-Prime on each session.

2.4.4 *Waters reading span task*

In the Waters reading span task (Waters & Caplan, 1996), participants read 80 sentences. The sentences were presented in sets of two to six sentences, with four sets per set size. A fixation was presented at the center of the screen for 1000 ms, when it was replaced by a sentence. Participants were to respond by pressing "yes" with the right hand if the sentence was semantically plausible or "no" with the left hand if the sentence was semantically implausible; the sentences was presented for 5000 ms or until a response was made.

At the end of a set, participants typed as many of the final words of the sentences in the set as possible and pressed the escape key when finished. E-Prime recorded responses and reaction time. Alternative forms of a word (e.g., wrong tense) were considered incorrect. Set size span was calculated as the maximum set size for which the participant correctly recalled all words in at least two of the four sets for that set size (order was irrelevant). The span was "0" for participants who did not correctly recall all words in at least two sets of set size two.

3. Results

Session 8 data for one participant were lost due to an error. Because testing followed training in Session 1, and preceded training in Sessions 2–8, it was not appropriate to include the data from all sessions in the same analyses. Analyses were therefore performed on the data from Sessions 2–8. Only reaction times from correct trials were included in the reaction time analysis. Based on the distribution, reaction times slower than 6000 ms were removed from analysis. Voice key errors were defined as trials on which the voice key was triggered prior to 300

ms or by a sound other than the vocal response; these were also removed from the analysis. These procedures resulted in the exclusion of 3.54% of the data.

The data were analyzed using linear mixed effects models. These models include fixed effects for the variables of interest (transliteration condition, organization condition, working memory span) and random effects for the variables that were sampled from a larger population (participants and items). All models were run using the lmer and glmer commands of the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2013) in RStudio Software (RStudio Team, 2016). The categorical variables were dummy coded as 0 and 1, and the linear variables were z-scored to place them on a similar scale to help the models converge. Binomial models were fit using the bound optimization by quadratic approximation algorithm (BOBYQA; Powell, 2009). The full models are given in the notes to the tables of the fixed effects.

We provide the raw data for the three dependent measures, but note that the model includes working memory span, which is not represented in the raw data figures; therefore, these figures should be interpreted cautiously. The mean reading span on the Waters task was 3.50 (SD = 1.27), with a range from 0 to 6. In the case of an interaction with working memory, we graphically represent estimated means taken from the regression equation using a procedure suggested by Aiken and West (1991).

3.1 Free recall

The fixed effects for the free recall analysis are shown in Table 4 along with the final model, and the raw data are shown in Figure 1. Free recall accuracy increased across sessions, $\beta = .90$, $z = 10.18$, $p < .001$. Transliterations led to better recall, $\beta = 1.14$, $z = 2.71$, $p < .01$. No other effects were significant, although the interaction between session and transliteration condition was marginally significant, $\beta = .21$, $z = 1.85$, $p = .065$. If it had reached traditional levels of statistical significance, this interaction would have demonstrated that the benefit of transliterations increased over sessions.

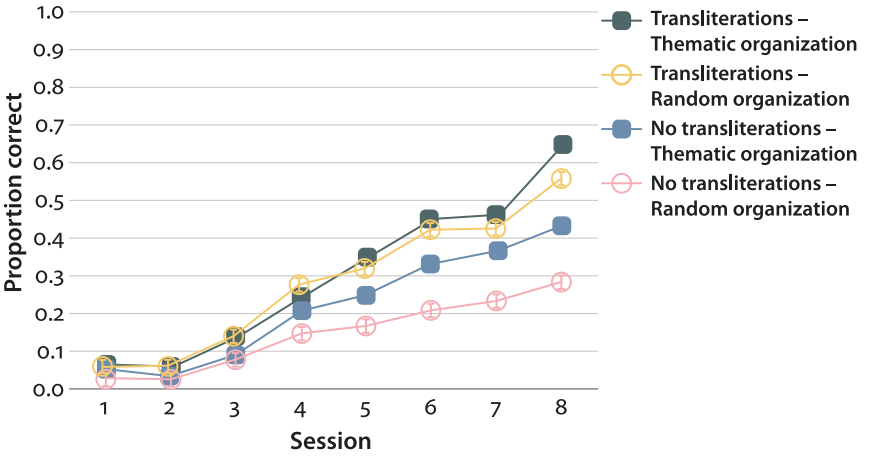


Figure 1. Raw free recall test accuracy data by condition and session
Error bars represent the standard error of the mean (because it is quite small the bars are difficult to see in some cases).

Table 4. Fixed effects from the model for free recall

Fixed effect	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	−2.37	.34	−6.97	.00	***
TransliterationCondition	1.14	.42	2.71	.01	**
OrganizationCondition	0.41	.43	0.95	.34	
ZSession	0.90	.09	10.18	<2.00E-16	***
ZWaters	0.55	.45	1.21	.22	
TransliterationCondition:OrganizationConditon	−0.27	.59	−0.46	.64	
TransliterationCondition:ZSession	0.21	.12	1.85	.06	†
OrganizationCondition:ZSession	0.10	.12	0.80	.43	
TransliterationCondition:ZWaters	−0.06	.49	−0.12	.90	
OrganizationCondition:ZWaters	0.36	.69	0.52	.60	
ZSession:ZWaters	0.10	.13	0.79	.43	
TransliterationCondition: OrganizationCondition:ZSESSION	0.05	.16	0.33	.74	
TransliterationCondition: OrganizationCondition:ZWaters	−0.92	.79	−1.17	.24	
TransliterationCondition:ZSession:ZWaters	−0.06	.14	−0.44	.66	
OrganizationCondition:ZSession:ZWaters	0.19	.19	0.99	.32	
TransliterationCondition: OrganizationCondition:ZSession:ZWaters	−0.27	.22	−1.25	.21	

Note: glmer(Recall~1+TransliterationCondition*OrganizationCondition*ZSession*ZWaters+ (1+ZSession|Participant)+ (1|EngWord), data=freerecall, family=binomial, glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5000)))
** $p < .01$.*** $p < .001$.† $p < .10$.

3.2 Translation accuracy

The fixed effects for the translation accuracy analysis are shown in Table 5 along with the final model, and the raw data are shown in Figure 2. In English-Arabic translation, accuracy increased across sessions, $\beta = 1.00$, $z = 11.27$, $p < .001$. Transliterations led to higher accuracy, $\beta = 1.38$, $z = 3.08$, $p < .01$. These main effects were qualified by a two-way interaction between session and transliteration condition, $\beta = .42$, $z = 3.52$, $p < .01$, which was itself qualified by a higher-order four-way interaction between session, transliteration condition, organization condition, and working memory capacity, $\beta = -.57$, $z = -2.61$, $p < .01$.

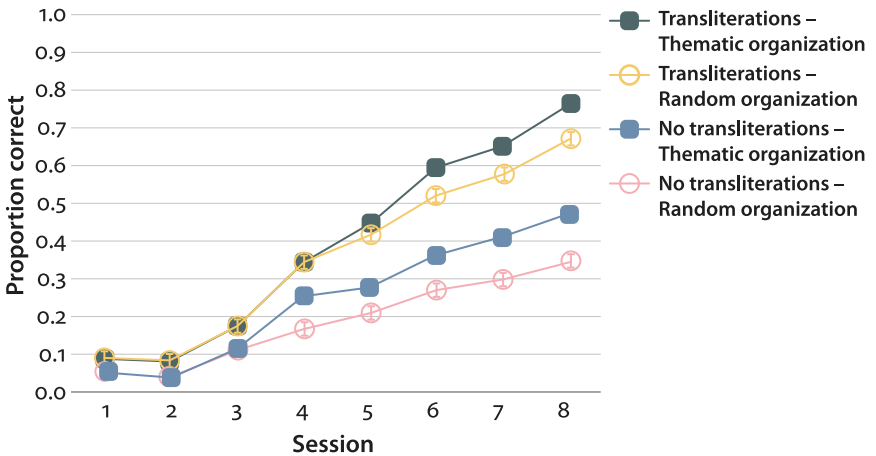


Figure 2. Raw English-Arabic translation test accuracy data by condition and session Error bars represent the standard error of the mean (because it is quite small the bars are difficult to see in some cases).

Table 5. Fixed effects from the model for English-Arabic translation accuracy

Fixed effect	β	Std. Error	z value	Pr(>|z|)	
(Intercept)	−2.02	.37	−5.49	.00	***
TransliterationCondition	1.38	.45	3.08	.00	**
OrganizationCondition	0.23	.46	0.50	.62	
ZSession	1.00	.09	11.27	<2.00E-16	***
ZWaters	0.68	.48	1.41	.16	
TransliterationCondition:OrganizationConditon	−0.02	.63	−0.02	.98	
TransliterationCondition:ZSession	0.42	.12	3.52	.00	***
OrganizationCondition:ZSession	0.04	.12	0.35	.73	
TransliterationCondition:ZWaters	−0.10	.52	−0.18	.85	
OrganizationCondition:ZWaters	0.42	.74	0.57	.57	
ZSession:ZWaters	0.13	.13	1.01	.31	

Table 5. *(continued)*

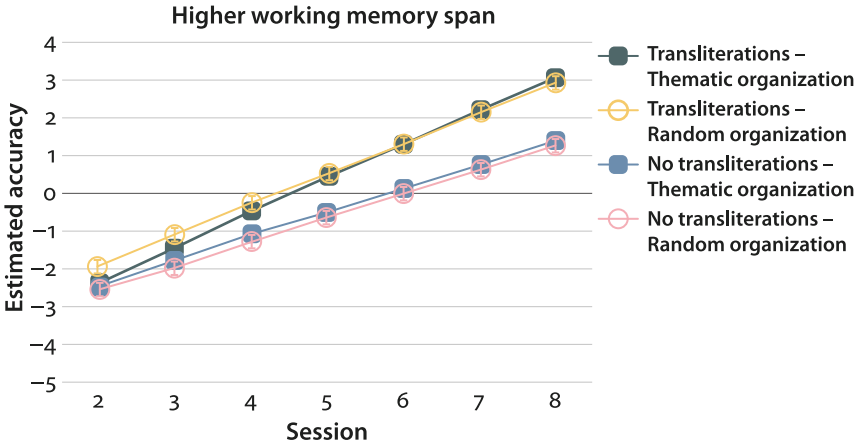
Fixed effect	β	Std. Error	z value	Pr(>|z|)	
TransliterationCondition:OrganizationCondition: ZSESSION	0.20	.16	1.22	.22	
TransliterationCondition:OrganizationCondition: ZWaters	−1.04	.84	−1.24	.21	
TransliterationCondition:ZSession:ZWaters	−0.02	.14	−0.18	.86	
OrganizationCondition:ZSession:ZWaters	0.25	.19	1.31	.19	
TransliterationCondition:OrganizationCondition: ZSession:ZWaters	−0.57	.22	−2.61	.01	**

Note: glmer(Accuracy~1+TransliterationCondition*OrganizationCondition*ZSession*ZWaters+ (1+ZSession|Participant)+(1|EngWord), data=translation, family=binomial, glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 5000)))
** $p < .01$.*** $p < .001$.

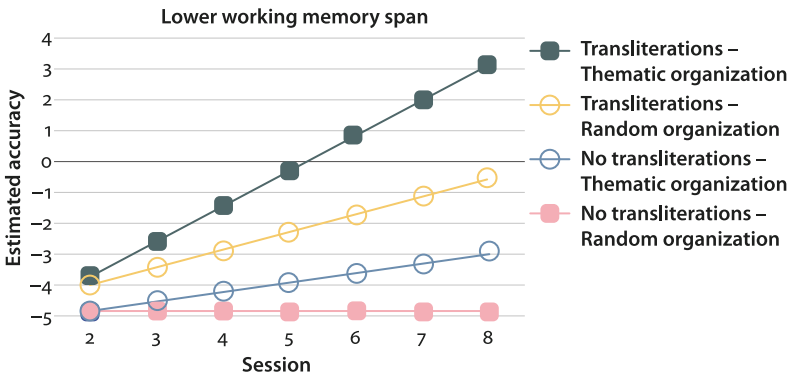
To visualize this four-way interaction, estimates from the regression equation at each session have been plotted in Figure 3 (Aiken & West, 1991). Participants with higher working memory capacity generally have higher translation accuracy and are less influenced by the training manipulations. However, there is one condition in which participants with lower working memory capacity outperform their higher working memory capacity peers – the transliterations-thematic organization condition. Lastly, participants with lower working memory capacity benefited more overall from the presence of transliterations. We return to this finding in the general discussion.

3.3 Translation reaction times

The fixed effects for the translation reaction time analysis are shown in Table 6 along with the final model, and the raw data are shown in Figure 4. Translation reaction times decreased across sessions, $\beta = -116.27$, $t = -2.59$, $p < .05$. Transliterations led to faster translation times, $\beta = -253.82$, $t = -2.16$, $p < .05$. No other effects were significant, although the interaction between transliteration condition and organization condition was marginally significant, $\beta = 299.74$, $t = 1.83$, $p = .08$. If it had reached traditional levels of statistical significance, this interaction would have demonstrated that there was a speedup for the transliteration-random condition relative to the other three conditions. The three-way interaction between session, organization condition, and working memory capacity was also marginally significant, $\beta = -172.25$, $t = -1.79$, $p = .09$. This seems to be due to participants in the lower and higher working memory capacity-random organization conditions speeding up across sessions. This speedup is exaggerated in the higher working memory capacity-thematic organization condition (i.e., there is a steeper



a.



b.

Figure 3. Estimated accuracy from the regression equation for the English-Arabic translation task as a function of session and condition.
Higher working memory span is plotted in panel a and lower working memory span is plotted in panel b. Extreme observed values were used as the higher and lower scores for working memory span estimates.

speedup such that these participants start at a slower speed and end at approximately the same place as individuals in the lower and higher working memory capacity-random organization conditions). This overall effect is reversed in the lower working memory capacity-thematic organization condition (see Figure 5).

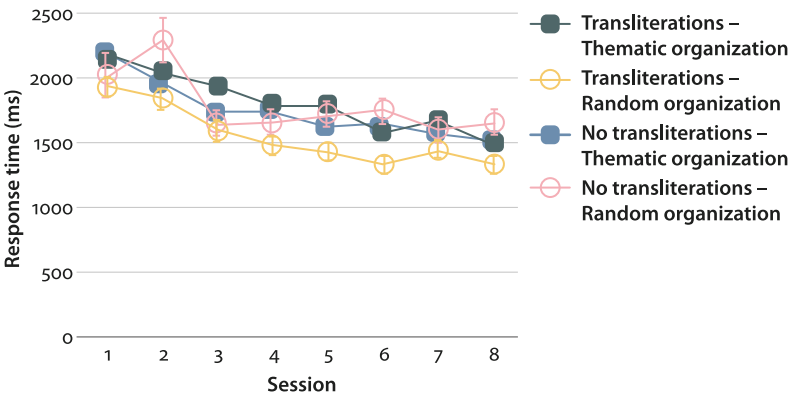


Figure 4. Raw English-Arabic translation test reaction time data by condition and session. Error bars represent the standard error of the mean (because it is quite small the bars are difficult to see in some cases).

Table 6. Fixed effects from the model for English-Arabic translation reaction time

Fixed effect	β	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	1956.49	98.94	37.22	19.77	<2e-16	***
TransliterationCondition	−253.82	117.43	23.20	−2.16	.04	*
OrganizationCondition	−43.55	121.99	24.15	−0.36	.72	
ZSession	−116.29	44.93	23.30	−2.59	.02	*
ZWaters	−64.97	129.14	26.16	−0.50	.62	
TransliterationCondition: OrganizationConditon	299.74	163.67	22.62	1.83	.08	†
TransliterationCondition:ZSession	−64.85	56.80	20.18	−1.14	.27	
OrganizationCondition:ZSession	−45.09	61.33	24.50	−0.74	.47	
TransliterationCondition:ZWaters	−2.53	138.55	25.61	−0.02	.99	
OrganizationCondition:ZWaters	131.84	194.42	23.59	0.68	.50	
ZSession:ZWaters	58.53	67.64	31.82	0.87	.39	
TransliterationCondition: OrganizationCondition:ZSESSION	−15.36	79.04	20.58	−0.19	.85	
TransliterationCondition: OrganizationCondition:ZWaters	−87.86	218.99	23.11	−0.40	.69	
TransliterationCondition:ZSession:ZWaters	−45.01	71.70	29.66	−0.63	.53	
OrganizationCondition:ZSession:ZWaters	−172.25	96.06	24.07	−1.79	.09	†
TransliterationCondition:OrderCondition: ZSession:ZWaters	183.50	107.45	22.98	1.71	.10	

Note: lmer(RT~1+TransliterationCondition*OrganizationCondition*ZSession*ZWaters+(1+ZSession|Participant)+(1|EngWord), data=translation)

* $p < .05$.*** $p < .001$.† $p < .10$.

Figure 5. Estimated English-Arabic reaction time data by condition and session. Error bars represent the standard error of the mean (because it is quite small the bars are difficult to see in some cases).

4. General discussion

To summarize our findings, performance improved across sessions (i.e., free recall and translation accuracy increased and translation reaction times decreased) in every analysis, as would be expected with repeated training and testing. The benefit of transliterations was also consistent across analyses. There was a tendency for this benefit to increase across sessions; this interactive effect was significant in the translation accuracy analysis and was marginally significant in the free recall analysis, although it was not significant in the translation reaction time analysis. The only statistically significant effect of organization condition was observed in a four-way interaction in English-Arabic translation accuracy, which contrasts with our predictions of an overall benefit of thematic organization based on Tinkham's (1997) findings.

One question is why our findings varied across measures. A general characterization of the overall pattern of results is that beyond the effects of session and transliteration, English-Arabic translation accuracy was the most sensitive measure in that it was the only one that had additional effects that reached conventional levels of statistical significance. This is not unusual for studies of beginning L2 learning like the current one (Degani & Tokowicz, 2010; Degani, Tseng, & Tokowicz, 2014). Therefore, it is not surprising that this task was the only one to yield significant interactions. Because the interactions in the other measures did not reach conventional levels of statistical significance we do not interpret them to err on the side of caution, but we provide information about them for the interested reader.

One of our most interesting effects was a four-way interaction in the English-Arabic translation accuracy analysis between session, transliteration condition, organization condition, and working memory capacity. In this case, we found

a benefit for the transliteration-thematic organization condition for participants with lower working memory capacity such that in this condition, lower working memory span participants performed better than their higher working memory capacity peers by the end of training. Specifically, the benefit for the transliteration-thematic organization condition was greater and increased more across sessions for participants with lower working memory capacity. This benefit does not occur in any other condition, and as can be seen in Figure 3, these individuals started out at a lower level of performance than their peers with higher working memory capacity on Session 2. This is important to note for two reasons. First, it shows that the transliterations-thematic organization condition provides these individuals with an important advantage that the other conditions do not (i.e., that the combination of training manipulations is required). Second, it demonstrates that this advantage is not observable until the participants have had quite a bit of training; the participants with lower working memory capacity begin to outperform their peers with higher working memory capacity at Session 6. In addition to being interesting in its own right, it demonstrates the value of examining L2 learning over a longer periods of time than are typically examined, although we acknowledge that such studies are not always feasible.

Why might the transliterations-thematic organization condition be particularly helpful for participants with lower working memory capacity? This is interesting to consider in light of Choi's (2003) suggestion that individuals with fewer cognitive resources may not have sufficient cognitive resources to attend to the organization manipulation if presented with auditory and visual representations simultaneously. We believe that the repeated presentation of the multiple representations aided the participants in forming strong lexical representations as noted above. In addition, by presenting the items in the same thematic organization, the participants were able to better form semantic relationships among the items. These semantic relationships may have been slow to form because beginning learners are thought to focus more on form-level connections (e.g., Kroll & Stewart, 1994).

On the other hand, the participants with higher working memory capacity performed better overall, and were less sensitive to our training manipulations in general. But, they did show better performance in two particular conditions than the others. In particular, they performed best in the two conditions that had only one training manipulation – the transliterations-random organization condition and the no transliterations-thematic organization condition. Perhaps the combination of the two manipulations was somewhat overwhelming for participants with higher working memory capacity precisely *because* they are able to keep more information in mind.

In the context of the present study, this could be because there was no break given between themes. It may be easier for these participants in particular if a break is given so that there is no potential for interference between themes. In other words, participants with higher working memory capacity may still be holding information from a preceding theme in memory when a new theme begins. Of note, the performance of these individuals is not extremely low in the transliterations-thematic organization condition, it is just lower than in the two conditions that had only thematic organization *or* only transliterations, and lower than the performance of the individuals with lower working memory capacity in the same condition.

How would the LQH explain these results? According to the LQH, it is critical to have a strong orthographic, phonological, and semantic representation for a word, and for these three representations to be strongly interconnected (Perfetti & Hart, 2002). Thus, we extended this hypothesis to suggest that training that included both the spoken form of the Arabic vocabulary words as well as the transliterations would lead to better performance on our retention tests. This was borne out by the data. One of the most consistent findings was that transliterations improved retention in that they improved free recall and translation accuracy, and sped up translation reaction times.

As described in the introduction, a number of cognitive mechanisms may be responsible for this improvement. In the framework of the LQH, the provision of the transliterations can be said to have led to the formation of higher quality lexical representations. This may be because all three lexical components of the L2 Arabic vocabulary (orthography, phonology, and meaning) were provided during training for participants in the transliterations conditions, but only two lexical components of the L2 Arabic vocabulary (phonology and meaning) were provided during training for participants in the no transliteration conditions. Transliterations may also have clarified the spoken forms that participants heard during training. They may have also helped participants to remember the spoken words, forming stronger memory traces.

An implication of these results is that all three lexical components should be provided during L2 vocabulary learning, to form high quality lexical representations of L2 vocabulary, even when the primary focus is on spoken, rather than written, production of L2 vocabulary. In the beginning stages of instruction of a language with a script that is unfamiliar to the learner, transliterations may be used to scaffold instruction. However, it is important to note that although transliterations can help to clarify and strengthen the phonological representation, they can also be detrimental to learning in the sense that the learner may map a sound to the L1 phonological category more strongly than would occur in the absence of the transliteration (e.g., Bassetti, 2008; Park, 2015). Because our

interest was in early vocabulary learning and retention and not specifically in the development of highly accurate phonological distinctions, we did not examine this issue in the present study. However, future investigations may wish to focus on this in more detail, particularly because of the availability of alternate transliterations in some languages (e.g., *Pinyin* and *Zhuyin* for Mandarin Chinese; see Hayes-Harb & Cheng, 2016).

5. Conclusion

The results of this study generally suggest that providing transliterations may be useful in the very beginning stages of adult L2 vocabulary learning when the goal of instruction may be to provide the learner with a base set of vocabulary in a language with an unfamiliar script. Our results further suggest that individual differences in cognitive skill such as working memory may relate to the effectiveness of our instructional manipulations and that these are important to take into account in our investigations. Notably, our data proved most interesting in a later session of our study, demonstrating that studies of learning over longer periods of time are important to advancing knowledge, albeit often infeasible.

Acknowledgements

We thank Emily Braun, Kevin Jarbo, Adrienne McGrail and members of the PLUM Lab for research assistance. We are grateful to Michael W. Dickey, Charles Perfetti, Tessa Warren, and two anonymous reviewers for feedback on this project and the manuscript. And, we are especially thankful to Scott Fraundorf for assistance with the statistical analyses. NT was funded by NIH R01 HD075800; Division of Behavioral and Cognitive Sciences; Eunice Kennedy Shriver National Institute of Child Health and Human Development during the writing of this manuscript.

References

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

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 191–206). Clevedon: Multilingual Matters.

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. *Center for the Study of Reading Technical Report*, no. 325.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Choi, T. H. (2003, July). *The effects of semantic and thematic grouping on learning efficiency and retention of the L2 vocabulary*. Tokyo, Japan: Pan-Pacific Association of Applied Linguistics.

Clauson, G. (2002). *Studies in Turkic and Mongolic linguistics*. New York, NY: Routledge.

Defense Language Institute Foreign Language Center. (2005). *Iraqi basic language survival guide*. Presidio of Monterey, CA.

Degani, T., & Tokowicz, N. (2010). Ambiguous words are harder to learn. *Bilingualism: Language and Cognition*, 13, 299–314. https://doi.org/10.1017/S1366728909990411

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

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61, 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

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. https://doi.org/10.3758/BF03203267

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24, 369–383. https://doi.org/10.1017/S0142716403000195

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press. https://doi.org/10.3998/mpub.23925

Hayes-Harb, R., & Cheng, H. -W. (2016). The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Frontiers in Psychology*, 7 (785). https://doi.org/10.3389/fpsyg.2016.00785

Hu, C. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing*, 21, 823–841. https://doi.org/10.1007/s11145-007-9094-6

Kaushanskaya, M. (2012). Cognitive mechanisms of word learning in bilingual and monolingual adults: The role of phonological memory. *Bilingualism: Language and Cognition*, 15(3), 470–489. https://doi.org/10.1017/S1366728911000472

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. https://doi.org/10.1006/jmla.1994.1008

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0–3. 2013. Available from: <http://CRAN.R–project.org/package=lmerTest>

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-013-0565-2

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. https://doi.org/10.1016/S0022-5371(77)80016-9

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. https://doi.org/10.1075/wll.8.2.04nel

Park, M. S. (2015). Orthographic input and the acquisition of second language phonology: A review. *Korean Journal of Linguistics*, 40(3), 389–401.

Perfetti, C. A. & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 67–86). Amsterdam: John Benjamins. https://doi.org/10.1075/swll.11.14per

Powell, M. J. D. (June 2009). The BOBYQA algorithm for bound constrained optimization without derivatives Department of Applied Mathematics and Theoretical Physics, Cambridge University. DAMTP 2009/NA06. Retrieved 2014-02-14.

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

Roediger, III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA. <http://www.rstudio.com/>.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. https://doi.org/10.1037/h0054651

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, 13, 138–163. https://doi.org/10.1191/026765897672376469

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. https://doi.org/10.1017/S1366728904001634

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154 https://doi.org/10.1016/0749-596X(89)90040-5

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*, 49A, 51–79. https://doi.org/10.1080/713755607

Appendix

Thematic organizations	English word / phrase
Commands, warnings, and instructions	Be quiet.
	Don't shoot.
	Follow me.
	Give me.
	Help me
	Keep away.
	Put your weapon down.
	Stay here.
	Stop or I will shoot!
	Stop!
	Unload
	We must search you.
Helpful words, phrases, and questions	Can someone assist us?
	Danger
	Do you have___ ?
	Excuse me / I'm sorry.
	He / She is
	Here
	How?
	I am
	I do not want.
	I want.
	No
	Please.
	Thank you.
	There
	They are
	We are
	What?
	When?
	Where?
	Who?
	Why?
	Yes
	You are
	Good bye.
	Do you have any identification papers?
	Do you understand?
	I don't understand.
Emergency terms	Distress signal
	Emergency!
	Evacuate the area!
	Help! (help me)
	We need a doctor!

(continued)

Thematic organizations	English word / phrase
Medical / general / body parts	Antibiotics
	Bandage
	Burn
	Clean
	Dead
	Doctor
	Fever
	Hospital
	I am a doctor.
	I am not a doctor.
	I will take you to the hospital.
	Infection
	Injured
	Medic
	Medicine
	Poison
	Sick
	Wound
	Foot
	Head
	Leg
Lodging	Is there a telephone available?
	We need ___ gallons of potable water.
Customs (Port of Entry)	False
	Owner
	Passport
	Permission
	Prohibited
	Property
	Visa
Relatives	Family
	Man
	Relatives
General Military	Ammunition
	Commander
	Gun
	Mine
	Minefield
	Mortar
	Refugee
	Shelter

Address for correspondence

Alison M. Tseng
University of Maryland
Center for Advanced Study of Language
7005 52nd Avenue
College Park, MD 20742
USA

atseng1@umd.edu

Co-author information

Martin C. Doppelt
University of Chicago
Department of Anthropology
mdoppelt@uchicago.edu

Natasha Tokowicz
University of Pittsburgh
Learning Research & Development Center
tokowicz@pitt.edu