

Investigating the Effects of Transparency and Ambiguity on Idiom Learning

Mehrgol Tiv (met81@pitt.edu)

University of Pittsburgh
4200 Fifth Ave.
Pittsburgh, PA 15260

Evelyn Milburn (eam115@pitt.edu)

University of Pittsburgh
4200 Fifth Ave.
Pittsburgh, PA 15260

Tessa Warren (tessa@pitt.edu)

University of Pittsburgh
4200 Fifth Ave.
Pittsburgh, PA 15260

Abstract

The purpose of this study was to learn how transparency and ambiguity affect idiom learning. To start, 157 French idioms were translated to English and normed for familiarity, transparency, and ambiguity. Experiment 1 was a training study in which 32 of these idioms were taught to 25 native English speakers over two days of training. A cued recall test during a third session showed a reliable effect of transparency, but performance was close to ceiling. In Experiment 2, the amount of training was reduced to one session and a semantic relatedness test was included after the cued recall test. The results of Experiment 2 suggest that high transparency idioms are recalled with greater accuracy in a cued-recall test but low transparency idioms are recalled with greater accuracy in a semantic relatedness test. No significant effect of or interaction with ambiguity was found.

Keywords: Idioms; Language Learning; Figurative Language; Individual Differences

Background

Figurative language, which includes phenomena such as metaphor, metonymy, irony, and idioms, is generally characterized as “going beyond the literal” (Gibbs, 1994). Of these, idioms have been of particular interest to cognitive linguists and psycholinguists. Idioms are traditionally defined by their conventionality; as Nunberg, Sag, and Wasow stated: “their meaning or use can’t be predicted, or at least entirely predicted, on the basis of a knowledge of the independent conventions that determine the use of their constituents when they appear in isolation from one another” (1994). The tension that arises between overall expression meaning and literal word meaning gives rise to several unique characteristics, two of which, transparency and ambiguity, will be considered here.

Most research on idiomatic language has focused on how idioms are mentally represented and processed. There has been less investigation of how idioms are learned, despite the practical ramifications it bears for second language learning applications (e.g. Howarth, 1998; Yorio, 1989).

The current study investigates the way that the properties of transparency and ambiguity affect idiom learning.

Transparency refers to the ease with which the comprehender can make a connection between the idiom’s literal and figurative meanings (Nunberg et al., 1994). The idiom *spill the beans* is highly transparent because the connection between “divulging a secret” and “releasing beans from a container” is very clear. In contrast, the idiom *wet behind the ears* is less transparent because the relationship between the literal and figurative meaning is less obvious. Several factors, including properties of individual words (compositionality) and the capacity to evoke a mental image (imageability), can influence the ease of drawing a connection between literal and figurative meanings and thus contribute to an idiom’s level of transparency (Nunberg et al., 1994).

Ambiguity refers to whether an idiom has both literal and figurative meanings, or only a figurative meaning. The idiom *kick the bucket* is ambiguous because it has both a literal meaning (striking a pail with a foot) and a figurative meaning (dying), but the idiom *under the weather* is unambiguous because it only has a figurative meaning (Titone & Connine, 1994).

We were interested in whether transparency and ambiguity influence how successfully individuals will learn the meanings of idioms. There is evidence that transparency and ambiguity both influence the comprehension of idioms, such that greater transparency tends to facilitate idiom comprehension, whereas greater ambiguity usually impedes comprehension (Foss, 1970), but there is only a small literature on how these factors influence learning. In one of the most relevant studies, Steinhel et al. (2007) investigated the effects of transparency on L2 idiom learning and found minimal effects of transparency but significant effects of imageability. However, their study included very few items

in each condition, so its findings must be interpreted with caution.

Most previous idiom learning studies have focused on learning idioms in a second language (e.g. Steinhel et al., 2007) and on the way that context affects learning (e.g. Zyzik, 2011). The current study is novel because it investigates idiom learning in the participants' native language and uses completely unfamiliar idioms, thereby avoiding any learning advantage from previous exposure. This is critical because there is evidence that comprehenders find more familiar idioms more transparent (Keysar & Bly, 1995). We presented participants with unfamiliar phrases that had no English idiomatic meaning. This process, in which the learner must assign idiomatic meaning to unfamiliar strings of words, mirrored natural idiom acquisition and allowed us to examine how unfamiliar idiomatic meanings are learned.

We base our predictions on findings from the growing literature on ambiguous word learning, under the assumption that assigning a new conventionalized meaning for a phrase that already has a compositional meaning is like learning a new meaning for a word that already has a meaning. Given evidence that ambiguous words are harder to learn than unambiguous words (Degani & Tokowicz, 2010), we predict that figurative meanings of less ambiguous idioms will be learned better than figurative meanings of more ambiguous idioms. This is because literal meanings of ambiguous idioms are highly accessible and may interfere with learning of figurative meanings. Drawing from work showing that when new meanings are assigned to old words, increased semantic similarity between the meanings leads to increased recall accuracy (Rodd, Berriman, Landau, Lee, Ho, Gaskell, & Davis, 2012), we also predict that figurative meanings of highly transparent idioms will be learned better than figurative meanings of less transparent idioms. This is because participants will be able to more easily draw connections between the idioms' literal and figurative meanings, thereby boosting recall.

Finally, we investigated three individual difference factors that could interact with the learning process and influence performance. The first was performance on the Operation-Span (O-Span) task (Turner & Engle, 1989), a measure of working memory. Because working memory is important for learning, we predict that individuals who score higher on an O-Span task will learn better than individuals that score lower on the O-Span task (Unsworth, Heitz, Schrock, & Engle, 2005). Additionally, we predict that more creative individuals will be able to create more or stronger connections between the literal and figurative meanings of the idioms, making the low transparency idioms seem more transparent to them. Thus, we expect that participants who score higher on the Abbreviated Torrance Test for Adults (ATTA) (Goff, 2002) will learn better and show less of a difference between performance on high and low transparency idioms. Finally, we expected that better performance on an Author Recognition Test (Stanovich & West, 1989), a proxy for language experience and thus

figurative language proficiency, would also be predictive of idiomatic learning.

Norming Idiom Properties

To inform the choice of stimuli for the training study, we conducted multiple norming questionnaires.

Methods

Participants 85 undergraduate students enrolled in Introduction to Psychology at the University of Pittsburgh participated in one of various norming questionnaires for course credit. All participants were over 18 years old, native speakers of American English, and had no prior experience with French.

Materials 305 French idioms were translated word-for-word (literally) and as an idiomatic phrase (nonliterally) using online and print resources (Bonin, Meot, & Bugaiska, 2013; "Idioms," 2006; Lamiroy & Klein, 2010; "Proverbes en français," 2006; "Traduction en contexte," 2006; "Word Reference") For example, the French idiom *avoir un cheveu sur la langue* was translated literally as *to have a hair on your tongue* and nonliterally as *to have a lisp*. French idioms that were extremely similar to English idioms were excluded, leaving 157 translated idioms in the set that was normed.

Procedures We designed online Qualtrics questionnaires to evaluate familiarity, transparency, and ambiguity. Each participant was randomly assigned to one questionnaire. Each questionnaire contained several catch trials to confirm that participants were on task and attentive. All participants completed a brief language history questionnaire.

To verify that the translated idioms were not familiar, 15 participants were shown each idiom and asked to "determine if there is an English phrase that is similar to and means the same thing as the following expression" by answering "Yes" or "No." Any idiom that received more than one "Yes" response was excluded from the set of stimuli.

To measure the level of transparency for each idiom, we presented the literal translation of each idiom (e.g. *to have a hair on your tongue*) with its figurative translation (e.g. *to have a lisp*) and asked participants to rate how likely from 1 (*Very unlikely*) to 5 (*Very likely*) they were to have the same meaning. Each opportunity for rating was followed by a free-response question: "why did you select that rating?" Piloting trials indicated that the questionnaire could take over an hour to complete, so, in order to avoid fatigue, each participant received half of the questionnaire. Fourteen participants completed the first half, and 13 completed the second half.

To index ambiguity, we evaluated the salience of the literal interpretation of the idiom by asking participants to "decide which [interpretation] you think is the more likely interpretation by assigning each option with a percentage (0-100). A higher percentage indicates the more likely

interpretation.” For example, for the idiom *to have a hair on your tongue* participants would have to assign a percentage to each of the two options: (1) *to have a hair resting on your tongue* and (2) *to have a lisp*. Increasing the likelihood of the literal meaning increases the chance that it will compete with the figurative meaning. Therefore, we used the rated likelihood of the literal meaning (option 1) as our index of the ambiguity of the idiom. Participants were additionally asked to “indicate how likely (from *very unlikely* to *very likely*) you would be to encounter these expressions when reading a novel or talking with someone.” This questionnaire was also time-consuming so it was similarly split in half. Fifteen participants completed each half.

Results

Average ambiguity and transparency scores for each idiom were calculated. We selected 32 idioms that varied across the full range of transparency and ambiguity scores, yet also satisfied a 2x2 design crossing transparency and ambiguity. Table 1 gives examples from each condition.

Table 1: Example Idioms Used in Training Study

Idiom	Translation	Transparency	Ambiguity
To have a hair on your tongue	To have a lisp	High	High
To fall like apples	To faint	High	Low
To do your hair	To worry yourself a lot	Low	High
To not miss the air	To be bold	Low	Low

Experiment 1: Learning of Unfamiliar Idioms

Methods

Participants 26 undergraduate students from the same population as the norming participated for course credit. None had previously participated in the norming.

Materials The 32 items described above.

Procedures The training study and Operation Span test were presented electronically using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA); the ATTA, Author Recognition test, and language history questionnaire were administered on paper.

Participants signed up for three separate sessions about two to three days apart. For each session, presentation of all items was randomized. During the first session (approximately 30-45 minutes), participants viewed an idiom and typed what they thought it meant. After they typed their answer, the correct meaning of the idiom appeared, and participants were asked to indicate whether

their answer was correct or not. Then participants were asked to generate a novel sentence that included the idiom. This procedure was repeated for each idiom.

The second session (approximately 20-30 minutes) consisted of three parts. First, each idiom and its definition were presented on the screen for 12 seconds. Then, participants saw each idiom again, and rated their confidence in their memory of the idiom’s definition. Then, the idiom and its definition reappeared on the screen for another twelve seconds. The metacognitive confidence rating was included to help ensure that participants were paying attention and to aid learning (Kang, 2007). During the second part, participants verbalized each idiom and its meaning to add a physical and auditory memory trace. Participants concluded the session by generating a novel sentence that included the idiom. This time, the participants saw the idiom and had to type the sentence without a reminder of the definition.

In the third session, participants were tested via a cued-recall test. Participants were instructed to “type the learned definition of the expression to the best of your ability.” Emphasis was put on recalling the correct meaning, instead of verbatim recall. On each screen, the participant saw an idiom and typed their answer below. Next, participants completed the Operation Span task, the Abbreviated Torrance Test for Adults, and Author Recognition Test. Finally all participants completed a brief language history questionnaire.

In the Operation Span task, participants viewed variably sized sets of mathematical computations followed by single words. The task was to judge whether the answer to the math problem was correct or incorrect and then remember the subsequent word. At the end of each set, the participants were asked to recall as many words as they could remember from that set. Set size ranged from 3-6 words.

In the Abbreviated Torrance Test, participants had three minutes to complete each of three activities. They had to answer thought questions and create images or pictures with incomplete figures. Participants only worked on one activity at a time.

In the Author Recognition test, participants were given a piece of paper that contained eighty names. They were instructed to “read the names and put a check mark next to the names of the individuals you know to be writers.”

Results

Responses from the cued-recall test were coded for accuracy. A verbatim or slightly reworded response of the learned definition was counted as correct. For example, for the idiom *to jump from the rooster to the donkey*, which means “to switch from one subject to another”, the response “to switch sides” was marked as incorrect. These criteria are strict enough to measure learning over guessing, yet flexible enough to measure competence over memorization. Table 2 shows means and standard deviations for accuracy by condition. It is important to note that even with this relatively strict coding, there was limited variability in the data. Participants performed with perfect accuracy on

approximately half the items, and approximately half of participants made no errors across all items.

Accuracy was analyzed in R (R Development Core Team, 2015; ver 3.2.3) with the lme4 package. We used linear mixed effect logit models with participants and items as crossed random factors (Baayen, 2008; Bates, Maechler, Bolker, & Walker, 2015). Fixed factors (e.g. ambiguity, transparency) were treated as continuous, and were centered in all analyses. When maximal models failed to converge, the random slopes that captured the least variance were dropped until the model converged (Barr, Levy, Scheepers, & Tily, 2013).

Analyses showed a significant effect of transparency ($\beta=0.59$, $SE=0.28$, $p=0.04$), such that more transparent idioms were learned better. There was no effect of ambiguity ($\beta=0.01$, $SE=0.01$, $p=0.63$) and no interaction between ambiguity and transparency ($\beta=0.01$, $SE=0.01$, $p=0.32$).

Individual difference measures were evaluated according to standardized procedures. For the Operation Span Task, set size was used instead of total words recalled because it is more representative of an individual's maximal working memory potential (Unsworth et al., 2005). The range of scores across subjects for each test was as follows: Operation Span Task Set Size (3-6), Abbreviated Torrance Test for Adults Creativity Index (26-90), and Author Recognition Test hits (4-26). Before inclusion in mixed models, all measures were centered and the ATTA Creativity Index was scaled down by a factor of 10 to make its scale more similar to the others.

Table 2: Cued Recall Accuracy in Experiment 1

Transparency	Ambiguity	Mean	SD
High	High	0.94	0.24
High	Low	0.93	0.26
Low	High	0.85	0.36
Low	Low	0.85	0.36

We found a significant positive interaction between transparency and O-Span ($\beta=0.42$, $SE=0.18$, $p=0.02$). However, the limited variability in the data made us hesitant to draw any firm conclusions from the current results. Therefore we decided to run a follow-up study with less intensive training.

Experiment 2: Less Training of Idioms

Methods

Participants 17 undergraduate students from the same population participated for course credit. None had participated in the norming or Experiment 1.

Materials The same 32 idioms as Experiment 1.

Procedures Experiment 2 consisted of two sessions: a learning session and, two days later, a testing session. Because Experiment 1 hinted at an interaction between

transparency and the O-Span test, this task was also administered in Experiment 2. We did not administer the other individual differences tasks for lack of time.

Experiment 2 only included a subset of Experiment's training procedures. In Session 1, participants viewed an idiom and typed what they thought it meant. The correct meaning of the idiom then appeared on the next screen, and participants indicated whether their answer was correct or not. They did this for all idioms. Next, participants went through the same flashcard activity as Experiment 1, however the duration of each flashcard was reduced from twelve seconds to six seconds. Finally, participants generated a novel sentence that included the idiom. This procedure was also repeated for each idiom.

In Session 2, participants were tested via the same cued-recall test as Experiment 1. After the participant typed the definitions to all idioms, they completed a semantic relatedness test. In this test, participants saw one of the idioms they had learned paired with one of two possible words: one related to the idiom's learned figurative meaning and one related to the idiom's literal meaning. Two versions of the semantic relatedness test were created for counterbalancing purposes; this ensured that across the entire experiment, idioms were paired with both related and unrelated words, but individual participants saw each idiom only once, and had an equal number of related and unrelated trials. Participants indicated whether the idiom and the presented word were or were not related in meaning by pressing keys on the keyboard. Finally, participants completed the O-Span test and language history questionnaire.

Results

All analyses were done in R with the same procedures as Experiment 1.

Cued Recall Accuracy Recall accuracy was coded according to the same criteria as Experiment 1; Table 3 shows means and standard deviations for accuracy by condition. Note that accuracy was lower, and there was more variability in the data than in Experiment 1. Our model included random slopes by subject for transparency and ambiguity, and replicating Experiment 1, it showed a significant positive effect of transparency ($\beta=0.61$, $SE=0.26$, $p=0.02$), but no effect of ambiguity ($\beta<0.01$, $SE=0.01$, $p=0.95$) and no interaction ($\beta<0.01$, $SE=0.01$, $p=0.79$).

Table 3: Cued Recall Accuracy in Experiment 2

Transparency	Ambiguity	Mean	SD
High	High	0.84	0.37
High	Low	0.88	0.33
Low	High	0.68	0.47
Low	Low	0.65	0.48

Semantic Relatedness Accuracy Means and standard deviations for accuracy by condition on the semantic relatedness test are shown in Table 4. A model including random slopes by subject for transparency and ambiguity showed a significant negative effect of transparency ($\beta = -0.52$, $SE = 0.22$, $p = 0.02$), but no effect of ambiguity ($\beta = -0.01$, $SE = 0.01$, $p = 0.25$) and no interaction ($\beta < 0.01$, $SE = 0.01$, $p = 0.64$).

Table 4: Semantic Relatedness Accuracy in Experiment 2

Transparency	Ambiguity	Mean	SD
High	High	0.79	0.41
High	Low	0.79	0.41
Low	High	0.90	0.30
Low	Low	0.92	0.27

Transparency x Test Type Interaction We conducted a repeated-measures ANOVA to investigate the interaction of transparency and test type. It indicated a significant main effect of test mode ($F(1, 16) = 18.54$; $p < 0.05$) and a significant interaction between test mode and transparency ($F(1, 16) = 41.78$; $p < 0.05$). The main effect of transparency approached significance ($F(1, 16) = 3.54$; $p = 0.08$). Pairwise comparisons indicated that recall was more accurate for high transparency idioms than low transparency idioms. However, this was reversed for semantic relatedness judgment accuracy: low transparency idioms were more accurate than high transparency idioms.

Operation Span Test We found no significant effect of O-Span ($\beta = 0.80$, $SE = 0.60$, $p = 0.18$) and no interaction of O-Span with transparency ($\beta = -0.04$, $SE = 0.12$, $p = 0.70$).

Discussion

Transparency

The current results suggest that transparency plays an important role in idiom learning. In both Experiments 1 and 2, higher transparency idioms were learned better than lower transparency idioms when learning was indexed by performance on a cued-recall test. One plausible explanation for this is that transparency facilitates a useful memorization strategy: finding connections and links between items.

Transparency may aid learning at both the word and phrase levels. For cases in which compositionality or decomposability contributes to higher transparency, spreading activation from a single word in the phrase could aid in recall. For example, in the experimental item *there is an eel under the rock*, the word *eel* could prompt recall of the adjective *fishy*, thus bringing the learner to the meaning *something fishy is going on*. At the phrasal level, the relationship may rely less on lexical associations, and instead be more illustrative or story-like. Participants may elaborate on the pairing between the literal and idiomatic meaning through mental imagery or storytelling.

However, our findings suggest that transparency's effect on learning may depend on the test. On the semantic relatedness task in Experiment 2, accuracy was higher for less transparent idioms. This result is surprising, and could potentially be caused by some inadvertently uncontrolled property of the semantic relatedness test itself. But if we take it seriously, a potential mechanism for explaining this data pattern could be borrowed from the Construction-Integration model (Kintsch, 1988), or other accounts according to which comprehenders activate wide networks of inferred relationships when trying to relate less-related concepts (e.g. Mason & Just, 2004).

We speculate that a similar process could come into play during our training sessions, such that participants might generate a larger set of weak possible relations when trying to relate low transparency idioms to their meanings than when trying to relate higher transparency idioms to their meanings. Although activating this wider network of possible relations will not necessarily improve performance on cued-recall tests, it could lead to improved performance on a semantic relatedness test. This is because having activated a wider network of related concepts during training may make it more likely that the concepts corresponding to the related words at test had already received some activation.

Ambiguity

The current experiments found no effect of ambiguity on recall of idiomatic meaning. It is possible that the way ambiguity was operationalized and the way the current experiments were structured could have contributed to this lack of effect. Unlike in previous word-learning studies in which ambiguity was manipulated by teaching either one or more meanings (e.g. Degani & Tokowicz, 2010), in the current study, ambiguity served as a measure for the potential of the literal meaning of an expression to compete with a newly taught figurative interpretation; this kind of ambiguity has also been referred to as *literal plausibility* (Titone, Columbus, Whitford, Mercier, & Libben, 2015). Critically, although this kind of ambiguity would be expected to have a strong effect during comprehension when multiple meanings compete for activation, it may have a less important role in the context of a training study. Participants in the current study were focused on learning idiomatic meanings for all of the items, and at test they knew that only the idiomatic meanings were relevant. This strong experiment-internal emphasis on the idiomatic meaning could have overwhelmed any potential effects of competition from a literal meaning.

Individual Differences

The current results are inconclusive regarding the influence of individual difference measures on idiom learning. Although there was some indication that working memory might be related to the influence of transparency on idiom learning in Experiment 1, this did not hold in Experiment 2, which had more variability in accuracy and a more even

distribution of O-Span scores. Therefore we are hesitant to draw any strong conclusions. Future research with larger samples will be important for addressing the roles that individual differences play in idiom learning.

Acknowledgments

We would like to thank the University of Pittsburgh Honors College for providing the funding for this project and Dr. Scott Fraundorf for his help with analyses in R.

References

- . Word Reference.
(2006). Idioms. The Free Dictionary, Fairlex.
(2006). "Proverbs en francais." from <http://parolesfr.kiev.ua/idiomsfr.php>.
(2006). Traduction en Contexte. Reverso.
Baayen, R. H. (2008). "Analyzing linguistic data: A practical introduction to statistics using R." Cambridge University Press.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.
Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1-48.
Bonin, P., et al. (2013). "Norms and comprehension times for 305 French idiomatic expressions." *Behav Res Methods* 45(4): 1259-1271.
Degani, T. and N. Tokowicz (2010). "Ambiguous words are harder to learn." *Bilingualism: Language and Cognition* 13(03): 299-314.
Foss, D. J. (1970). "Some effects of ambiguity upon sentence comprehension." *Journal of Verbal Learning and Verbal Behavior* 9(6): 699-706.
Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
Goff, K. (2002). "The abbreviated Torrance test for adults (ATTA)." IL Scholastic Testing Service.
Howarth, P. (1998). "Phraseology and second language proficiency." *Applied Linguistics* 19(1): 24-44.
Kang, S. H., McDermott, K.B., Roediger III, H.L. (2007). "Test format and corrective feedback modify the effect of testing on long-term retention." *European Journal of Cognitive Psychology* 19(4-5): 528-558.
Keysar, B., Bly, B. (1995). "Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans?" *Journal of Memory and Language* 34(1): 89-109.
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
Lamiroy, B., Klein, J.R. (2010). "Les expressions verbales figées de la francophonie: Belgique, France, Québec, et Suisse." OPHRYS.
Mason, R. A., & Just, M. A. (2004). How the brain processes causal inferences in text: a theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychological Science*, 15(1), 1-7.
Nunberg, G., Sag, I.A., Wasow, T. (1994). "Idioms." *Language* 70(3): 491-538.
Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M.G., Davis, M.H. (2012). "Learning new meanings for old words: effects of semantic relatedness." *Memory and Cognition* 40(7): 1095-1108.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime computer software and manual. *Pittsburgh, PA: Psychology Software Tools Inc.*
Stanovich, K. E., West, R.F. (1989). "Exposure to print and orthographic processing." *Reading Research Quarterly*: 402-433.
Steinel, M. P., Hulstijn, J.H., Steinel, W. (2007). "Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency." *Studies in Second Language Acquisition* 29(3): 449-484.
Titone, D. A., Connine, C.M. (1994). "Descriptive Norms for 171 Idiomatic Expressions: Familiarity, Compositionality, Predictability, and Literality." *Metaphor and Symbolic Activity* 9(4): 247-270.
Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. *Bilingual figurative language processing*, 171-207.
Turner, M. L., Engle, R.W. (1989). "Is working memory capacity task dependent?" *Journal of Memory and Language* 28(2): 127-154.
Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior research methods*, 37(3), 498-505.
Yorio, C. (1989). Idiomaticity as an Indicator of second language proficiency. *Bilingualism Across the Lifespan*. L. K. O. K. Hyltenstam. Cambridge, Cambridge University Press: 55-72.
Zyzik, E. (2011). "Second language idiom learning: The effects of lexical knowledge and pedagogical sequencing." *Language Teaching Research* 15(4): 413-433.