2013

Bilingual Lexical Interactions: Inevitable or Malleable?

Rachel Lynn Jobe
Lehigh University

Bilingual Lexical Interactions: Inevitable or Malleable?

by

Rachel Jobe

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Psychology

Lehigh University

July 8, 2013

Thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science in Psychology.

BILINGUAL LEXICAL INTERACTIONS: INEVITABLE OR MALLEABLE?
RACHEL JOBE

_____
Date Approved

Signatures:

_____
Barbara Malt
Thesis Advisor

_____
Catherine Arrington
Committee Member

_____
Jessecae Marsh
Committee Member

_____
Gordon Moskowitz
Department Chairperson

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 1057885 to Barbara Malt and Ping Li. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

List of Tables

List of Figures

Abstract

Many bilingual speakers speak each of their languages in ways that differ from monolingual speakers of those languages, but it is not known if these differences are inevitable or can be influenced by learning conditions. In two experiments, 129 English speakers named objects in English, were trained on the lexical pattern of five Russian words either with or without feedback and either blocked or intermixed training (Experiment 1) or with or without metalinguistic knowledge and with or without English trials (Experiment 2) and then named objects again in English. Additionally, a control group provided English naming data at two different times to compare English consistency with and without second language training. Though there were few differences in the experimental conditions, participants in these conditions were less consistent in their English naming than those in the control condition, suggesting that bilingual speakers' lexicons are inevitably linked.

Many bilingual speakers do not speak either of their two languages in completely monolingual-like ways. In terms of the lexicon (the focus of this paper), a bilingual speaker's lexical choice may differ from the lexical choice of a monolingual speaker of that language. For example, while a monolingual speaker of English may refer to a specific drinking vessel as a *cup*, a bilingual speaker might use the word *glass* to refer to the same object, even if she has *cup* in her English vocabulary. Evidence for these differences comes from Malt and Sloman (2003), who found that second-language learners of English often differed from monolingual English speakers in terms of their lexical choices even after 10 years or more spent in an immersion environment. Differences from monolingual lexical patterns have also been found when bilingual speakers are tested in their first language (Pavlenko & Malt, 2011) and even when bilingual speakers who have learned both languages from birth are tested in either of their two languages (Ameel, Storms, Malt, & Sloman, 2005). In some way, the two lexicons of the bilingual interact so that one lexicon influences the other.

It is not known if these effects are necessarily inevitable or if they might be due to typical bilingual learning conditions that may be able to be mitigated under some circumstances. These possibilities lead to two possible conceptual hypotheses. The first is that some interactions between languages will always occur, in any and all learning environments, due to interconnections between the two lexicons. According to the second hypothesis, there could be mitigation of the interactions under optimal learning conditions. The purpose of these studies is to discriminate between these two possibilities.

There are two contributing factors to the observed interactions: cross-linguistic differences and lexical interactions. I will first discuss these factors and then elaborate on each hypothesis before describing each of the two studies.

Cross-Linguistic Differences

Although some researchers have argued that concrete words are generally truly equivalent across languages (e.g., De Groot, 1993), experimental evidence indicates that this is not the case. Kronenfeld, Armstrong, and Wilmoth (1985) found that speakers of Hebrew, English, and Japanese placed various drinking vessels into linguistic categories in quite different ways. For example, while English speakers used the same name for tall cylindrical objects made out of plastic or paper as for small objects with handles intended for hot drinks, calling both types of items *cup*, Hebrew speakers placed all tall cylindrical objects, regardless of material, into a single category, *cos*, that was generally more aligned with an English speaker's *glass* category. The difference apparently arose because the English category *glass* included the feature "made out of glass" while the Hebrew category *cos* had no such constraint on material. Additionally, Japanese speakers separated the objects into three labeled categories, in contrast to the two categories used by the English and Hebrew speakers.

In a more extensive experiment, Malt, Sloman, Gennari, Shi, and Wang (1999) asked native speakers of English, Spanish, and Chinese to name 60 different common household containers—items generally called *bottle* or *jar* by English speakers or that were similar to bottles or jars. They also asked participants to perform non-linguistic sorting tasks based on the physical or functional characteristics of the objects or their overall similarity to one another. Interestingly, while all participants sorted the objects

very similarly in the non-linguistic sorting tasks, the linguistic labeling task showed marked differences between the speakers of the three languages. For example, there were six objects that were most commonly labeled *bidon* by the native Spanish speakers. Of these objects, three were most commonly called *jug* by native English speakers, one was called *bottle* and two were called *container*. Similarly, the objects labeled *bottle* by native English speakers were split into seven different linguistic categories by the Spanish speakers and three different linguistic categories by the Chinese speakers, most of which also contained objects not labeled *bottle* by native English speakers. These results led Malt et al. to the conclusion that while speakers of different languages view the commonalities among objects similarly, the lexical patterns of different languages are different both from one another and from the pattern obtained in a non-linguistic sorting task.

Lexical Interactions

If, as has been suggested (e.g., MacNamara, 1967), bilingual speakers' two languages are independent of one another, these differences between languages would likely not present much of a problem. Penfield and Roberts (1959) were the first to suggest that this independence between languages occurs through an input-switch mechanism under which one language is "on" at any given time while the other is necessarily "off," preventing any cross-over from one language to another. However, much evidence suggests that bilingual lexicons are, in fact, largely interconnected.

For instance, Schwnanenflugel and Rey (1986) found that cross-language primes, whether presented for 300 or 100 milliseconds, reduced reaction times by the same amount as within-language primes when fluent bilingual participants were asked to

determine whether or not a target letter string was a word or a non-word. More recently, Sahlin, Harding, and Seamon (2005) demonstrated cross-linguistic memory errors in the Deese-Roediger-McDermott false memory task. In the original version of this task, when participants are presented with lists of words related to a single "critical lure" which does not actually appear on the list, they often falsely remember seeing the critical lure. For example, if participants are given a list including the words *bed*, *rest*, *awake*, and *tired*, they often report remembering that the word *sleep* was also on the list (Roediger & McDermott, 1995). In the bilingual version of this task, Sahlin et al. presented Spanish-English bilingual participants with 12 DRM lists of 10 words each. Half of these lists were in Spanish while half were in English. After hearing each of the 12 lists, participants were given a visual recognition test that contained words from the studied lists, the critical lures for the studied lists, and words and critical lures for non-studied lists. For half of the Spanish lists and half of the English lists, the words on the recognition test were translated into the other language. Participants were asked to respond "yes" only if the word presented on the recognition test was in the same language in which it had been heard during the study phase. Although false recognition of a critical lure was more likely to occur if that critical lure was presented in the same language as the studied list to which it was related, participants also falsely recognized words that had been presented in the other language (a word presented in Spanish during the study phase with its English translation presented on the recognition test) and critical lures that had been related to a list presented in the other language (recognizing *montaña* [*mountain*] if they had been presented with a list including the words *hill*, *valley*, *climb*, and *summit*). These findings, along with many other similar findings (e.g., Guttentag, Haith, Goodman, & Hauch,

1984), indicate that rather than maintaining fully separate lexicons, bilingual speakers' word knowledge is linked through a single conceptual system that is language-neutral.

In an attempt to account for the differences between monolingual and bilingual lexicons and for the role of this conceptual system, Malt and Li (2011) proposed a connectionist network in which features map onto words. However, words that are often taught as translation equivalents may not have the same feature mappings across languages. For example, although the Russian word *stakan* is generally translated into English as *glass*, the items that Russian speakers would call *stakan* do not completely align with the items that English speakers would call *glass*. As with the differences between Hebrew and English mentioned earlier, the differences appear to be due to English speakers associating the feature "made out of glass" to the word *glass* while Russian speakers do not have this constraint on *stakan.* When presented with two objects of similar size and shape, but with one object made out of glass and the other made out of plastic, English speakers are likely to call the object made out of glass *glass* and the one made out of plastic *cup* while Russian speakers would call both objects *stakan.* Similarly, while English speakers would apply the word *cup* to both a tall cylindrical object made out of plastic and a small handled container made out of ceramic used for drinking hot tea, Russian speakers would call the tall cylindrical object a *stakan* while the small container would receive the word *chashka*. For a bilingual speaker to speak two different languages in monolingual-like ways, then, one feature (e.g., "made out of plastic") may have to be mapped differentially across the two languages—that is, that feature may have to be mapped to two words that are not generally thought of as being equivalent to one another, such as *cup* and *stakan*.

Because of the existence of this connectionist network, repeated exposure to words in both languages may change the underlying patterns of connections in a bilingual speaker's long-term memory (Malt, 2008). When learning a second language, a native English speaker may assume that the features she associates with the word *glass* in her first language (L1) are identical to the features that native speakers of her second language (L2) associate with that language's translation "equivalent." Eventually, the L2 learner might adjust the features she associates with the L2 word to better match native use of that word. If she does this while still assuming, even implicitly, that the L1 and L2 words are truly equivalent, this may then lead to shifts away from monolingual usage in the L1. Even if the speaker is aware of the differences between the meanings, though, the cross-connections through the conceptual store and through word-to-word direct links are still likely to pull the L1 usage closer to the L2 usage (and vice versa). For example, if the words *glass* and *stakan* are always linked to one another for an English-Russian bilingual, this may prevent complete separation from ever occurring, since any feature associated with *stakan* would also be indirectly linked to *glass* through *stakan*. Over time, through Hebbian learning, the feature that was originally associated only with *glass* may also become linked directly to *stakan*. For a bilingual speaker to use the lexical patterns of both languages in monolingual-like ways, then, the direct link between so-called translation equivalents may need to be weak or non-existent. This lack of direct connections may be particularly difficult to obtain for classroom L2 learners, as they are often taught that words do have direct translations. These cross-connections help explain why L2 speakers' lexical patterns are not identical to native speakers' lexical patterns, even after many years in immersion contexts (e.g., Malt et al., 1999; Pavlenko & Malt,

2011). Real-world bilingual speakers seem to always experience cross-talk between their two languages (e.g., Ameel et al., 2005), preventing, potentially inevitably, either language from becoming truly monolingual-like.

Input Deficits

Alternatively, however, the differences between monolingual and bilingual lexicons might be the result of input deficits—a bilingual speaker may simply not have the available input to learn two languages in completely monolingual-like ways (Malt, 2008). Malt originally named four possibilities relating to input deficits (hearing altered input, the role of feedback, decreased input, and decreased sensitivity to input), which, of course, might occur in any combination. An additional two possibilities are also discussed below. The first of the possibilities originally suggested by Malt is simply that bilingual speakers might hear input that differs from monolingual usage. Cook (2003) suggests that the language of an isolated linguistic community might evolve away from the language spoken in the homeland. Cook gives the example of so-called Pennsylvania Dutch, which has been adapted over time from its original German. If this occurs, L2 users may hear altered L1 from other members of their expatriate community. Alternatively, for children of parents who speak two different languages, the child may grow up hearing atypical usage from one parent when the child and parents are conversing together, as at least one parent will be speaking an L2 in these situations. The child may then adopt these atypical patterns in her own speech, possibly explaining some bidirectional influences for children who grow up in bilingual households.

Another possibility relating to input deficit is that L2 learners often do not receive negative feedback when conversing with native speakers, who may not correct atypical

usage if the meaning can be understood (Malt, 2008). Although children often do not receive negative feedback in these types of situations (Brown, 1973), this lack of feedback might be more disadvantageous for adults who already have established lexical usage for their first language that is initially imported into the L2 (Malt, 2008).

Additionally, speakers of two languages receive less input for each language overall and have only half as many opportunities as monolingual speakers to appropriately learn word-object mappings (Gollan, Montoya, Fennema-Notestine, & Morris, 2005). Finally, second-language learners may pay less attention to the positive input that they do receive. Communicative intent can generally be understood even without a monolingual-like understanding of a word, so bilingual speakers may not focus on the ways that monolingual speakers linguistically categorize objects or compare how this categorization differs from their L1 categorization.

In addition to these four issues, two more considerations may affect a bilingual speaker's ability to maintain two completely separate patterns. Related to the possibilities of decreased input and decreased sensitivity to that input, bilingual speakers may be unable to explicitly compare and contrast the categories as they are used by native speakers, due to the long delay between hearing exemplars named by native speakers—or at least may be unable to do so sufficiently to overcome the initial L1 assumptions. For example, for an L2 English learner to learn to discriminate between the words *cup, glass,* and *mug* in the same way that a native English speaker would, she has to pick up the information from a limited amount of input that is spread out over months or years, as she is likely to hear each word used only once or twice in a given day, and perhaps not at all on some days. Additionally, even when she does hear one of these words, there may not

be an exemplar of the item present. This gap between presentations of exemplars is likely to make it quite difficult for an L2 learner to pick up on the commonalities of each term and the differences that are important for distinguishing between, for example, *cup* and *glass*, especially when both terms are applied to tall, cylindrical objects, as the learner may not be able to extract the meaningful difference between the two (namely, material in this case) when the objects are named by native speakers over a time span potentially as long as a week or more.

The other consideration regards frequent switching between languages. Although some bilingual speakers may become completely immersed in their L2 environment and only rarely revert to using their L1, many bilingual speakers immigrate to the L2 environment with their families and may use their L1 at home while using L2 at work, school, or within the larger community. This continual switching back and forth may lead to changes in both languages (Wolff & Ventura, 2009). Because of the cross-talk between the two languages of a bilingual speaker, if second-language learners encounter evidence that an L2 word is used differently than its L1 "equivalent," this may result in memory changes for the meanings of both words (Ameel, Malt, Storms, & van Assche, 2009). In effect, an exemplar of a category in a speaker's L2 may also be encoded as an exemplar of that category's rough translation in L1, even if monolingual speakers of that L1 would not consider that exemplar to be a member of that category. Depending on the strength of the direct connections between the two words, feature updating may be unable to effectively update the word in only one language, resulting in the features of both words being continually updated. Continual switching between the two languages may

additionally work to strengthen the direct connection between the two words, as the language not currently in use will still be available in short-term memory.

Strong vs. Weak Interconnection Hypotheses

The real-world data lead to two possible conceptual hypotheses: a strong interconnection hypothesis and a weak interconnection hypothesis. Under the strong interconnection hypothesis, bilingual speakers' lexical patterns will inevitably differ from the monolingual patterns of either language due to unavoidable cross-connections between the two lexicons. That is, it may be impossible for bilingual speakers to avoid any links between related words in their two languages (i.e., it may be impossible for a Russian-English bilingual to completely eliminate the link between *stakan* and *glass*, even if she knows that the two words are not exactly equivalent to one another), leading to Hebbian learning and changes in the speaker's usage of the words. Under the weak interconnection hypothesis, the differences that have been observed in real-world bilingual speakers are due to learning conditions as well as the structure of the lexical network. The learning conditions may affect speakers' lexical patterns because some learning conditions allow for those interconnections to affect one another in terms of the lexical choices of bilingual speakers, while other types of learning conditions may prevent these effects from occurring. For example, one way to potentially increase an L2 learner's sensitivity to the positive input she receives is by providing her with the metalinguistic knowledge that many of the words in her two languages are not exact translations of one another. With this knowledge, the L2 learner might be able to pay more attention to the ways that native speakers linguistically categorize objects in order to keep her two lexical patterns separate. In terms of the network, this metalinguistic

knowledge might help her maintain weaker links between the translation "equivalents," allowing the network's adjustment mechanism to update the words in only one language. If the weak interconnection hypothesis is true, bilingual speakers can obtain monolingual-like lexicons in each of their languages if they learn each language under optimal conditions.

Pilot Study

In a first attempt to determine which of these hypotheses is more accurate, a pilot study (Jobe, 2012) was conducted in which participants were trained on the Russian lexical pattern for the words *chashka, stakan,* and *kruzhka*—words that are generally translated as *cup, glass,* and *mug* respectively, but which differ in application from the English usage (Pavlenko & Malt, 2011). Factors related to input deficits were controlled for as much as possible to allow optimal L2 learning. The possibility of altered input was eliminated, as participants learned words based solely on naming patterns from monolingual speakers of Russian. The experiment was designed to diminish the effect of reduced positive input by only teaching participants three words, with each observation paired with an object that would receive that name and by requiring them to meet a given learning criterion before moving on in the experiment, minimizing the possibility that failure to adjust the features relevant to the connectionist network was due to a lack of input. Similarly, reduced sensitivity to this positive input should also have been diminished, as participants were asked to pay attention to the exemplars presented and had to do so in order to meet the criterion to move on to the next stage of the experiment. The role of feedback was directly manipulated by providing corrective feedback during the learning stage to only half the participants, while the effect of reduced sensitivity was

manipulated by providing half of the participants with explicit metalinguistic knowledge that the Russian and English lexical patterns differed while providing the other half with no such knowledge, resulting in a 2 x 2 design.

Unexpectedly, participants from all four conditions largely showed ceiling effects: the English speaking participants all learned the Russian pattern quite quickly and were then able to generalize it accurately to novel objects. Participants were also asked to name the objects in English both before and after their Russian training. Although real-world evidence suggests that learning a second language can often have effects on the first (e.g., Pavlenko & Malt, 2011), the participants in this study were relatively consistent with their English choices both before and after learning the Russian pattern. Further, this consistency was not affected by the experimental manipulations. These two findings together suggest that it may be possible for bilingual speakers to maintain two separate lexical patterns, given appropriate learning conditions, providing evidence for the weak interconnection hypothesis.

However, limitations of the pilot study make those findings tentative. The experiment was designed to make learning relatively easy in order to allow participants to complete the experiment within an hour, as they were only receiving credit for that amount of time, which led to ceiling effects in many aspects of the experiment. This ease is most clearly evidenced by how many trials participants needed to meet the 90% criterion to move on to the testing set of objects. On average, participants only needed 2.5 trials to move on, suggesting that participants learned the pattern quite quickly—much more quickly than real-world bilingual speakers, who often fail to match monolingual norms even after many years in immersion environments (Malt & Sloman, 2003). Instead

of truly learning the lexical pattern, however, participants may have memorized the names used for the exemplars presented during the training runs and then simply chosen a name during the testing run based on the object's similarity to the objects that had been seen during the training runs. If this is the case, the participants in the pilot were not engaged in typical word learning processes, potentially negating any real-world applications of the study. There were obviously differences between the learning in this experiment and real-world L2 learning, as the participants in this experiment were only required to learn three words rather than the many thousands that a true second-language learner would need to master. Each of these words was learned by viewing multiple exemplars of items that would be called by that name by native Russian speakers. This method of learning, even for participants who were not given explicit metalinguistic knowledge, is quite likely to have facilitated some awareness that the Russian lexical pattern differed from the English lexical pattern. Because real-world L2 learners will generally not see many different exemplars from one or a small number of categories presented together in a short amount of time, this knowledge may be more difficult for them to pick up on than it was for the participants in this experiment. Additionally, participants here did not have to simultaneously remember other L2 words that they would need to carry on a full conversation in the second language, as real-world L2 learners would have to do. Although the goal of the pilot study was to provide an optimal learning task, the ease of the task suggests that participants may not have been engaged in the same word-learning processes as real-world L2 learners, preventing the results obtained in the pilot study from providing much relevant information outside of a laboratory.

Finally, the focus of the pilot study was largely on how well participants were able to learn the Russian pattern, rather than how well they could maintain the English pattern given acquisition of the Russian. Because of this, there was no control condition in which participants named objects in English, completed a non-linguistic task, and then named objects in English again to determine how consistently monolingual speakers name objects when they have not been influenced by the lexical patterns of another language. The absence of a control condition made it impossible to evaluate whether mastery of the Russian had an impact on English naming.

Current Studies

The goal of the studies here is to minimize some of the limitations of the pilot study to better determine which conceptual hypothesis—the strong interconnection hypothesis or the weak interconnection hypothesis—is more accurate. If the participants in any condition of these two experiments are able to accurately learn the Russian lexical pattern and simultaneously maintain their English lexical pattern, this would suggest that the strong interconnection hypothesis may be too strong. If the participants in none of the conditions are able to learn the Russian pattern while maintaining the English pattern, this would suggest that the results of the pilot study were obtained largely due to the low external validity of that study. In other words, if participants' lexical patterns show evidence of lexical interactions in learning conditions that are only slightly more difficult than the learning conditions of the pilot study, then the differences that have been observed between monolingual and bilingual speakers may be inevitable, suggesting that the cross-connections are an unavoidable side effect of speaking two languages.

In order to address this goal, two experiments were conducted. Although the overarching goal of both experiments was the same, each experiment was intended to address language learning and the interconnection hypotheses for different populations of L2 learners. In Experiment 1, variables that were consistent with classroom L2 learning were manipulated while in Experiment 2, variables that were consistent with immersion L2 learning were manipulated. The two experiments were therefore independent; the outcomes of Experiment 1 did not influence the goals or methodology of Experiment 2.

In both experiments, participants were exposed to five new Russian names, rather than three as in the pilot study. This change was intended to make the learning slightly more challenging, as participants were exposed to a greater number of words that they needed to learn. Though this is still not as many words as a real-world second language learner would be exposed to, learning a greater number of words increased the difficulty—and therefore the realism—of the task as much as possible within the limits of a one-hour study. This increase in quantity was intended to decrease the likelihood of participants simply memorizing which exemplars belong to which category, requiring them to abstract the similarities within categories and differences among categories. Additionally, in Experiment 2, participants were not provided with the general dictionary translations of the Russian words, preventing them from directly building from their English knowledge. As in the pilot study, the input deficit issues of being exposed to altered input and having decreased input were eliminated or minimized in both experiments due to the experimental design: Participants learned the Russian terms based solely on the names used by native monolingual speakers of Russian and were required to meet a certain criterion of correct answers before moving on to the testing phase of the

experiment. Each experiment also manipulated two aspects of the learning conditions, one from the pilot study and one additional one. The input deficit issues of lack of negative feedback and delay between exemplar presentations were addressed by the specific experimental manipulations of Experiment 1 while the issues of decreased sensitivity to input and language switching were addressed in Experiment 2. Specifically, in Experiment 1, half of the participants received feedback on their lexical choices during the training portion of the experiment while half did not. Additionally, half of the participants viewed the Russian terms via a blocked presentation schedule while half viewed the terms via an intermixed presentation schedule. This resulted in a 2 x 2 design that tested the role of feedback and the role of a time delay in between presentations of exemplars. In Experiment 2, half of the participants were given explicit metalinguistic knowledge that the Russian and English lexical patterns differ; half were given no such knowledge. Furthermore, half of the participants completed English-language trials in between the Russian trials while half completed only the Russian trials. Again, this resulted in a 2 x 2 design that tested the roles of sensitivity to the input and of switching between languages.

In addition to these experimental groups, a control group of participants provided English naming data at two different times in order to compare consistency in naming for participants who were not exposed to a lexical pattern that differed from that of their own native language. Because the first and last stages of both experiments were identical, a single control group which only completed these first and last stages, with a non-linguistic filler task in between, served as the control for both experiments. From the data obtained from this control group, it was possible to determine if the brief Russian lexical

training presented in the two experiments changed speakers' English lexical patterns more so than would be expected without the Russian lexical training. Specific experimental predictions for each experiment are discussed below.

Experiment 1

Many L2 learners learn their second language in a classroom environment. The first experiment explored variables that may be easy to manipulate in classroom learning. The first variable of interest for this experiment was the role of feedback on learning. Evidence from the categorization literature suggests that feedback is often beneficial for categorization of artificial stimuli (e.g., Ashby, Maddox, & Bohil, 2002), so if the weak interconnection hypothesis is correct, participants who received feedback were predicted to perform better, in terms of their accuracy in Russian naming and consistency in English naming, than those who did not. This prediction was based on the idea that, by providing participants with feedback, they may have been better able to adjust the weights of their connectionist networks, resulting in higher levels of performance, particularly for learning the Russian lexical pattern. Since the feedback in this experiment was only provided in one language, if the two patterns were kept separate, then the feedback should have only adjusted the feature weights for the Russian pattern. If, though, the strong interconnection hypothesis is more accurate and the two patterns were linked directly through the translation "equivalents" provided, the feedback could have simultaneously adjusted the feature weights of both the English and the Russian patterns. By comparing participants' English naming consistency to the English naming consistency of the control group, it was also possible to determine which of these two possibilities was more accurate. Even if the group that received feedback was more

consistent in their English naming than the group that did not receive feedback, there may still have been room for both groups to be less consistent in their English naming than the control group. If this were the case, then this would suggest that the two lexical patterns may be inevitably linked to one another through the conceptual store.

The other variable of interest in this experiment was the role of blocked vs. intermixed presentation of the Russian categories during the learning phase. This manipulation tested whether a long time period between presentations of exemplars is inadequate for L2 learners to abstract the commonalities of one term or learn the contrasts between different terms. The weak interconnection hypothesis predicts that the way in which the input is presented can affect learners' ability to master the L2 lexical pattern while maintaining the L1 lexical pattern. Because the stimuli that will be used in this experiment are familiar to participants, unlike the artificial stimuli often created for categorization tasks, participants may be able to see the differences among the stimuli relatively easily, likely resulting in an advantage for a blocked presentation schedule, as participants will be better able to see within-category similarity (Carvalho & Goldstone, 2011). However, an alternative argument could be made that because participants will have to learn categories that are not completely distinct from one another and that have some similarities to their previously-known categories but are still different, an intermixed presentation may be more beneficial, as it will allow participants to better compare how the Russian categories relate to each other (Carvalho & Goldstone, 2011) If the strong interconnection hypothesis is more accurate, there will always be an L1-L2 interaction no matter how the input is presented. If either blocked or intermixed

presentation increases Russian learning, there should be a corresponding decrease in English consistency for the same conditions.

The weak interconnection hypothesis predicts that these manipulations will lead to main effects for both feedback and presentation order. Specifically, participants who received feedback were predicted to perform significantly better than those who did not. Similarly, participants who learned the Russian pattern via a blocked presentation schedule were predicted to perform better than those who learned the pattern via an intermixed schedule (although it is possible that the advantage would be in the other direction, depending on whether it is more important for participants to see the similarities within categories or the differences between categories). Although no interaction effects were expected here under either hypothesis, the two variables were combined into a single experiment in order to make the design more efficient and use fewer participants while maintaining a relatively high level of statistical power. If the strong interconnection hypothesis is correct, any differences found between the experimental groups would be overshadowed by the differences in the English consistency scores of the control group as compared to the experimental groups. The English naming consistency among the experimental groups was therefore compared to the English naming consistency of the control group. If the lexical interactions that real-world bilingual speakers have experienced are inevitable, the control group's English naming consistency (measured as a percentage of objects to which individual participants give the same name at times 1 and 2 of the English naming) would be higher than that of all the experimental groups.

Method

Participants

Participants were 73 native monolingual English speakers from the participant pool at Lehigh University. In the control condition and the no-feedback, blocked presentation order condition, there were 14 participants each; the remaining three conditions had 15 participants each.

Materials

A Language History Questionnaire was used to obtain demographic information. Similar questionnaires have been used in related studies (e.g., Malt et al., 1999). This questionnaire consisted of questions regarding the participant's age and sex, knowledge of languages other than English, and relative use of English vs. another language in daily life. See Appendix A for a copy of this questionnaire.

A set of 60 photographs of drinking containers created for a related study (Pavlenko & Malt, 2011) was used to teach participants Russian names for these drinking vessels. These containers included objects made from a variety of materials (Styrofoam, glass, plastic, etc.) and designed to hold a variety of beverages (tea, coffee, alcohol, cold beverages, etc.). These photographs were taken in front of a neutral background with a ruler used to provide sizing information and will be referred to as the training set. Participants were trained using a subset of 45 pictures, including all of the pictures that were most commonly called *chashka* (n=11), *kruzhka* (n=9), *fuzher* (n=4), and *riumka* (n=6) by monolingual speakers of Russian and a portion (n=15) of items that were most commonly called *stakan*. Although the original set of 60 items had a total of 23 items that were most commonly called *stakan*, 8 items were eliminated from the training set so that participants would be unable to use frequency information in their judgments. Typicality

information from monolingual speakers of Russian (Pavlenko & Malt) was used to determine which *stakan* items were eliminated from the testing set: From the 16 items in the middle of this typicality range, 8 items were eliminated, so that participants in the experiment were exposed to the full range of typicality.

Of the 35 objects in the training set which received a name of *stakan, chashka*, or *kruzhka*, only 15 received the name which would have been expected based on the dictionary translation of the most common English name—that is, there were only 15 objects that were called *stakan* in Russian and *glass* in English, *chashka* in Russian and *cup* in English, or *kruzhka* in Russian and *mug* in English. The other two object names that participants were asked to learn do not have good general English translations, so this type of analysis is not possible for the 10 objects that received one of those two names. Because more than half of the objects differed from this expectation, this data set should therefore provide enough information for participants to learn the Russian pattern for these categories.

Because of the relatively small number of objects in each category from Pavlenko and Malt (2011), there were not enough members of each category to provide a sufficient amount of training and a number of objects from which generalization could be measured, so an additional testing set was also created. To provide this testing set, additional photographs of drinking vessels likely to be called *chashka, stakan, kruzhka, riumka*, or *fuzher* were also taken in front of a neutral background along with a ruler used to provide size information. The objects in the testing set, as in the training set, varied in material, including objects made of glass, Styrofoam, plastic, and paper, and in intended use (i.e., intended to hold coffee or tea, water or other cold beverages, or as units of

measure). The objects were found in the experimenters' homes or purchased in nearby stores. These objects were thought to be likely to receive the names *chashka, stakan, kruzhka, riumka,* or *fuzher* based on the factors listed by Pavlenko and Malt and similarity to the objects that made up the training set. The Russian names for each object were confirmed by collecting data from monolingual Russian speakers (n=11) through the use of Mechanical Turk and asking native speakers of Russian on campus to forward a survey link to monolingual friends and family in Russia. The data from these monolingual speakers indicate that of the 88 objects originally included in the testing set, 71 received one of the five names used in the experiments more often than any other name (that is, the modal name for 71 objects was one of these five names): *chashka* (11 items), *stakan* (36 items), *kruzhka* (19 items), *fuzher* (1 item), or *riumka* (4 items), with an additional five objects receiving multiple names equally often (1 *chashka*/*stakan*, 1 *stakan*/*stopka*, 1 *riumka*/*vaza*/*kremanka*, and 2 *riumka*/*stopka*). The remaining 12 objects most commonly received a name that was not included in the training set. From these 76 objects, a final testing set was selected. The primary criterion to be included in the testing set was that there had to consensus on the name from a minimum of seven of the eleven (64%) monolingual Russian speakers. This criterion left a total of 61 possible objects. From these 61 objects, the final set was selected to include a wide range of typicality for each name and to ensure that a number of the objects received a name that would not be expected based on the translation equivalent of the most common English name. The final testing set of 35 objects consisted of 10 *chashka*, 13 *stakan*, 10 *kruzhka*, 1 *fuzher*, and 1 *riumka*. Of the 33 objects that received the names *chashka, stakan,* or *kruzhka*, 21 received the name which would have been expected based on the dictionary translation of

the most common English name. While this proportion may seem high, much of it is driven by the high number of objects which received the name *kruzhka* in Russian and its translation equivalent, *mug*, in English. When considering the objects called *stakan* and *chashka*, only 11 out of 23 objects received the name that would be predicted if *chashka* were treated as a direct translation of *cup* and *stakan* were treated as a direct translation of *glass*, providing ample opportunity to test for knowledge of the Russian lexical pattern. See Figure 1 for sample pictures.

Procedure

Experimental groups. Participants first filled out the Language History Questionnaire. After completing this, they were directed to a computer where the experiment proper began. The experiment was conducted using E-Prime version 2.0. The first screen was an instruction screen telling the participants that the experiment would be composed of several parts.

Baseline testing. In this part of the experiment, participants viewed both the training and testing photographs and were asked to name each of the objects in English in order to provide a baseline measure of a participant's already-established English lexical pattern. This section began with an instruction screen about the first part of the experiment. Participants were told that they would see pictures of common objects and be asked to label these objects as either *cup, mug,* or *glass* by pressing a given key. In previous experiments (Pavlenko & Malt, 2011), the dominant name for every object tested has been *cup, mug,* or *glass*. Participants were therefore limited to a forced choice of these three words, which they selected by pressing a key labeled either "c," "m," or "g." This forced-choice procedure kept the English portions of the experiment parallel to

the Russian portions, in which participants were also limited to a forced choice of the names *chashka, stakan, kruzhka, riumka,* or *fuzher*.

Training phase. Participants then saw another instruction screen in which they were asked to imagine that they would soon be leaving to study abroad in Russia and therefore needed to begin learning the Russian language (Malt, 2008). Participants were told that they would be seeing some of the photographs they had previously seen, this time labeled with the Russian name for the object. Similar to classroom instruction situations, all of the participants were given the typical dictionary translations for these objects, i.e., *stakan* is translated as *glass*, *chashka* as *cup*, and *kruzhka* as *mug*. They were also told that they would see objects labeled with the names *riumka* and *fuzher* but that these names do not have good English translations. Participants were not explicitly told that the translations are not exact matches.

All of the participants then saw the objects from the training set, presented one at a time for three seconds each and labeled with its appropriate Russian name. For half of the participants, the exemplars of each lexical category were presented in a blocked fashion (i.e., participants saw all the examples of *stakan*, followed by all the examples of *kruzhka*, etc.). Because there are 5 Russian names that participants were exposed to, there were 5 blocks of categories within the experiment. Five versions of the experiment were created so that the order of the blocks was not the same for all participants. However, each individual participant saw the blocks in the same order each time he or she completed a training run, although the order of the exemplars within a block varied randomly. For the other half of participants, the exemplars of each lexical category were presented in an intermixed fashion, with the exemplars presented in a random order.

After seeing each photograph in the training set once, participants saw another instruction screen. Participants were told that they would now see the photographs they had just seen and needed to label them with the appropriate Russian name by selecting a given key on the keyboard. All participants saw objects in a random order during this stage of the experiment. Additionally, half of the participants in the blocked condition and half in the intermixed condition were given feedback after each decision, told either "Correct! The correct answer was chashka," or "Incorrect. The correct answer was chashka." When the participant labeled the object correctly, the feedback was in blue; when the object was labeled incorrectly, the feedback was in red. The other half of participants did not receive any feedback regarding their answers. After labeling each of the objects, each participant was told the overall percentage correct for that run. The training runs of the experiment—seeing photographs with Russian labels followed by labeling the same photographs with Russian names—were repeated until the participant reached 90% or better accuracy on a single session of his or her labeling.

Test phase. Once this criterion was met, the participant saw a congratulatory screen and moved on to the test phase of the experiment. In this test phase, participants were again asked to label pictures as *chashka, stakan, kruzhka, riumka,* or *fuzher*. The pictures in the test phase included all of the test objects which the participants had seen in the first portion of the experiment (English naming) but which had not been presented with Russian names, as well as a portion of the training set. The entirety of the training set was not used in the test phase due to time constraints. The pictures from the training set that were included were half of the objects from the original set that were called either *chashka, stakan, kruzhka, riumka,* or *fuzher*: those which had a file name that included an

odd number (each picture was identified with a particular number in otherwise identical file names). Participants were asked to name these old pictures as well as novel ones because it seemed most similar to real-world bilingual learners, who will sometimes name objects which a native speaker has already named and sometimes name objects for which they have not heard a native speaker's lexical choice. Participants were asked to indicate whether they thought the object in the picture would be likely to be called *stakan, chashka, kruzhka, riumka,* or *fuzher* by Russian speakers based on their knowledge of the training items. Participants indicated their choice by selecting a key on the keyboard labeled *s, c, k, r,* or *f.* Although participants were told that they could take as long as they needed to decide on their answer, reaction times were recorded. No participants received feedback on this portion of the experiment.

Posttest. To allow for comparison of participants' English lexical patterns before and after training, participants then saw each item from both the training and testing sets again and were asked to choose from *cup*, *mug*, or *glass* as the name they would use in English to label that object. See Appendix B for the instructions used throughout this experiment.

Feature listing. After naming each object in English again, participants were given a questionnaire with each Russian word listed and asked to list what features they had associated with each of the Russian categories at test. Participants were allowed to write as much or as little as they desired. The purpose of this task was to determine what features participants were using to distinguish between the different Russian categories and how explicitly they were differentiating the Russian lexical categories from the

English lexical categories. After completing this task, participants were debriefed and dismissed.

Control group. In addition to the experimental groups, a control group of participants labeled objects in English at two different times but were not exposed to any language-learning tasks. This control group allowed for a comparison of consistency for participants who had and had not been trained on a new lexical pattern to determine if the training did lead to decreased consistency apart from the differences between the experimental groups. The early stages of the experiment were identical for the experimental and control groups. The control group participants filled out the Language History Questionnaire and then labeled objects in English as *cup, mug,* or *glass*. After naming each of the objects from both the training and the sets, participants completed a filler activity. This activity was designed to take approximately the same amount of time as the training and testing runs of the experimental groups and consisted of participants completing a same-different task with pictures from the training and testing sets for three minutes, solving 50 simple math problems, and then completing the same-different task again for another three minutes. This procedure was intended to provide the control group with approximately the same amount of exposure to the objects as participants in the experimental conditions, but in a non-linguistic task. After completing the filler activity, participants were again asked to label the objects as *cup, mug,* or *glass*. They were then debriefed and dismissed. This group served as a control for both Experiment 1 and Experiment 2, as the first and last stages of the experiments were identical.

Results

A 2 (feedback vs. no feedback) by 2 (blocked vs. intermixed presentation) fully between-subjects ANOVA was used for all analyses among the experimental groups. Both accuracy in Russian naming (based on naming choices obtained from native Russian speakers) and consistency in English naming were analyzed.

As discussed above, the weak interconnection hypothesis predicts that participants who received feedback would be more accurate in Russian naming and more consistent in English naming than those who did not. The weak interconnection hypothesis also predicts that those who learned the Russian terms via a blocked category presentation would be more accurate in their Russian choices and more consistent in their English choices than those who learned the lexical pattern via an intermixed presentation. Furthermore, the participants who received feedback and who were in the blocked conditions were expected to require fewer training trials and make fewer errors in their initial trial of Russian learning than those who did not receive feedback and those who were in the intermixed conditions. No interactions were predicted for this experiment. Under the strong interconnection hypothesis, differences between experimental groups in either of these measures would still be predicted, but if mastery of the Russian terms is acquired, then the experimental groups would all be expected to be less consistent in their English naming choices than the control group because Russian learning will exert an influence on participants' English lexical patterns.

Measures of Russian Accuracy and Learning

Accuracy in Russian naming in the test phase was measured by giving each participant a score of 1 for every object that they named with the same name as monolingual speakers of Russian and a score of 0 for any other name chosen. As a

reminder, participants in the blocked conditions and those who received feedback were expected to be more accurate than those in the intermixed conditions and who did not receive feedback. Because participants had differing amounts of exposure to the training and testing objects, accuracy in Russian naming was analyzed separately for the training objects and for the novel testing objects, with the sum of scores for each participant divided by 21 (the total number of training objects) or 35 (the total number of testing objects), respectively, and then the resulting number multiplied by 100 to convert it into a percentage and create a single accuracy score for each participant. See Table 1 for the mean accuracy scores and Table 2 for *F*- and *p*-values and measures of effect size. For the training objects, there was a non-significant trend in the expected direction for the participants in the blocked conditions to be more accurate (M = 94.90%, SD = 5.01) than participants in the intermixed conditions (M = 92.47%, SD = 5.18). The size of this effect was medium, suggesting that experiment may have lacked sufficient statistical power to detect the effect, given the small sample size. There was no significant effect of feedback nor a significant interaction. The measure of primary importance, however, is participants' accuracy scores for the novel objects, as this is a measure of how well participants were able to accurately generalize to new exemplars based on their previous Russian exposure. For these novel objects, there were no significant differences of feedback or of presentation order, nor was there a significant interaction (see Table 2 for statistical values). The predicted results for the Russian accuracy scores were therefore not supported by the results of the analyses.

However, due to the criterion participants were required to meet to continue on the test of novel Russian objects, more sensitive measures of how the variables affected

participants' scores are measures of learning difficulty. Accordingly, the number of trials it took participants to reach the 90% criterion to advance to the testing stage of the experiment and the number of errors participants made on their first round of Russian naming were analyzed. Again, participants in the blocked conditions and those who received feedback were expected to need fewer trials to criterion and to make fewer mistakes on their first training run than participants in the intermixed conditions and those who did not receive feedback. (See Table 3 for the statistical values for the following analyses.) For number of trials to reach criterion, there was a significant effect of presentation order. Specifically, contrary to the prediction originally thought to be more likely (that the blocked presentation order would be more beneficial), participants in the intermixed conditions needed fewer trials to meet this criterion ($M = 3.10$, $SD = 0.79$) than participants in the blocked conditions ($M = 3.93$, $SD = 1.36$). There was no effect of feedback nor a significant interaction. The number of mistakes that participants made in their first round of Russian naming was also analyzed. Again, there was a significant effect of presentation order. Again contrary to the prediction thought to be more likely (that the blocked presentation order would be more beneficial than the intermixed presentation order), participants in the intermixed conditions made fewer errors ($M = 16.68$, $SD = 7.05$) than those in the blocked conditions ($M = 22.14$, $SD = 7.37$). There was no significant effect of feedback nor a significant interaction. For these measures of Russian learning, the predictions for feedback were not supported at all, while the predictions for presentation order were in the wrong direction.

Measures of English Consistency

English consistency was measured by comparing the name each participant had chosen for a particular object at Time 1 to the name that he or she had chosen at Time 2. For each object that received the same name at both times, a score of 1 was given to that object. Any object that received a different name at times 1 vs. 2 was given a score of 0. The scores for each object were summed and divided by 80 (the total number of objects) to obtain a single consistency score for each participant. Each score was then multiplied by 100 to convert it into a percentage. This method of obtaining consistency scores compared each participant's responses to his or her own responses rather than to the most common response across participants or any normed data. See Table 4 for mean English consistency scores. Initially, the English consistency scores of the experimental and control groups were compared to one another with a one-way ANOVA on the scores of the four experimental groups and the control group (see Table 5 for statistical values). The strong interconnection hypothesis predicts that the control group would have significantly higher consistency scores than any of the experimental groups both for the data as a whole and for any subsets of the data that were analyzed. There was a significant effect of condition with a planned comparison indicating that the control group, as predicted by the strong interconnection hypothesis, was significantly more consistent (M = 87.14%, SD = 6.38) than the experimental groups (M = 79.22%, SD = 7.50). This analysis was also conducted for only the training items, as these items were the ones that participants in the experimental conditions had thought of in terms of the Russian lexical pattern most often, with the same pattern of results: a significant effect of condition with a planned comparison indicating that the control group was significantly more consistent (M = 87.14%, SD = 6.39) than the experimental groups (M = 76.50%,

SD = 8.16. An analysis of the consistency scores for only the testing items did not indicate any significant effects between groups.

Additionally, the consistency scores for the discrepant items—those whose most common Russian name was not predicted by the direct translation of its most common English name (e.g., an item called *stakan* in Russian but *cup* in English)—and for the compatible items—those whose most common Russian name was predicted by the direct translation of its most common English name (e.g., an item called *stakan* in Russian and *glass* in English)—were analyzed separately. For the discrepant items, there was a significant difference between groups (see Table 5 for statistical values), with a planned comparison indicating that the control group was significantly more consistent (M = 85.71%, SD = 7.26) than the experimental groups (M = 78.81%, SD = 9.11). For the compatible items, there was a significant difference between groups, with a planned comparison indicating that the control group was significantly more consistent (M = 88.89%, SD = 7.55) than the experimental groups (M = 79.28%, SD = 9.76). Because differences in consistency scores for the compatible items were not expected, as the Russian training should presumably not have changed participants' answers for the objects which did receive compatible names in both languages, consistency scores were further broken down into separate scores for the discrepant and compatible items for the training and testing objects separately. A one-way ANOVA for the discrepant training items was significant, with a planned comparison indicating that the control group (M = 81.90%, SD = 9.58) was significantly more consistent than the experimental groups (M = 76.05%, SD = 10.01). For the compatible training items, there were again significant differences between groups, with a planned comparison indicating that the control group

(M = 89.52%, SD = 6.25) was significantly more consistent than the experimental groups (M = 76.38%, SD = 12.46). Among the discrepant testing items, there were no significant differences between groups. Among the compatible testing items, there was a significant difference of group. However, a planned comparison showed only a non-significant trend for the control group to be more consistent (M = 87.41%, SD = 12.19) than the experimental groups (M = 81.36%, SD = 11.71). Generally speaking, these results were as predicted, with the control group often being more consistent than the experimental groups.

An alternate explanation for the differences between groups could be that the participants in the control group completed the experiment faster than participants in the experimental groups and so had less time to forget their original responses. Although the filler task of the control condition was designed to take approximately the same amount of time as the experimental conditions, and did so in pilot testing, a one-way ANOVA on total time elapsed from the beginning of Russian training or the filler task to the end of the Russian test phase or the completion of the filler task, as derived from E-prime data, indicates that there was a significant difference in the amount of time spent between the first and second rounds of English naming for the control and experimental groups (F (4, 68) = 13.29, $p < 0.001$), with a planned comparison indicating that the control group spent significantly less time (M = 8.52 minutes, SD = 0.51, range = 7.84 - 13.05) than the experimental groups (M = 15.85, SD = 4.59, range = 5.60 - 28.47), t (68) < 0.001. Even when comparing only the experimental group that took the least amount of time to complete the training runs (the group that received feedback and an intermixed presentation order; M = 13.70 minutes, SD = 3.17) with the control condition, there was

still a significant difference in the amount of time between the first and second English rounds, t (27) = 6.04, $p < 0.001$.

However, correlations among the experimental groups make this alternate explanation less likely. Specifically, the amount of time of time spent between the two rounds of English naming was not significantly correlated with participants' consistency scores (range = 60.00 – 93.75), r = -0.15, p = 0.27. Similarly, consistency was not significantly correlated with the number of training runs (range = 1 – 8) participants needed to reach the criterion to advance, r = -0.20, p = 0.12. These non-significant correlations, along with the non-significant differences for only the test items, suggest that it is the Russian training on specific objects—and not how long participants are spending on the Russian training—that leads to differences in participants' English lexical patterns after training.

Among the four experimental groups, there were no immediately discernible differences in English consistency scores, although the original predictions stated that participants in the blocked conditions and those who received feedback would remain more consistent than those in the intermixed conditions and who did not receive feedback. Specifically, in terms of English consistency for the training items (those which participants had thought of in terms of the Russian pattern most often), analyses of consistency scores (measured as the sum of scores for each training object divided by 35 and then multiplied by 100) showed no main effects of either feedback or presentation order, nor a significant interaction (see Table 6 for statistical values). The same analysis was also done for only the test items (not including the training items that were presented

during the test phase). There were no main effects of feedback or presentation order, nor a significant interaction.

Consistency scores were also analyzed separately for the discrepant items and for the compatible items. See Table 7 for mean consistency scores for these items and Table 8 for statistical values of the analyses. For the discrepant items there was a significant effect of presentation order. Participants in the blocked conditions were more consistent in their English naming choices (M = 81.82%, SD = 7.53) than participants in the intermixed conditions (76.10%, SD = 9.68). There was no main effect of feedback nor a significant interaction. For the compatible items, there were no significant differences of feedback or presentation order, nor a significant interaction. The discrepant and compatible items were also broken down into training and testing items (see Table 9 for the statistical values for the analyses of the discrepant items). Among the discrepant training items, there was a significant effect of presentation order. Participants in the blocked condition were more consistent (M = 79.17%, SD = 9.37) for these items than participants in the intermixed conditions (M = 73.23%, SD = 10.01). There was a non-significant effect of feedback and a non-significant interaction. For the discrepant testing items, there was a marginally significant effect of presentation order. Participants in the blocked conditions were marginally more consistent (M = 87.50%, SD = 8.61) than those in the intermixed conditions (M = 82.26%, SD = 13.15). The medium effect size of this finding again suggests that the experiment may have lacked the statistical power needed due to the small sample size. There was a non-significant effect of feedback and a non-significant interaction. Among the compatible training items (see Table 10 for statistical values for these analyses), there were no significant effects of feedback or presentation

order, nor was there a significant interaction. Among the compatible testing items, there were no significant effects of feedback or presentation order, nor was there a significant interaction. The original predictions for presentation order were only supported for the discrepant items, then, while the predictions for feedback were not supported at all.

To determine what objects were least likely to be consistent from pretest to posttest, the English consistency data were explored in a bit more detail. The following analyses took only the experimental groups into account. The number of people who choose the modal name for the object in the pre-test (a measure of pre-training consensus for the object name) was significantly correlated with the number of people who remained consistent in their choice for that object among all the objects ($r = 0.85$, $p < 0.00001$), the training items alone ($r = 0.85$, $p = < 0.00001$), and the testing items alone ($r = 0.87$, $p < 0.00001$), suggesting that people were more likely to change their answer for the objects which had previously had low consensus, regardless of the item name's discrepant or compatible status with the correct Russian word.

To determine what names participants were switching to and from, a median split was used to separate the 40 objects that were most likely to receive a consistent name (M = 91.36%, SD = 5.44) from the 40 objects that were least likely to receive a consistent name (M = 66.36%, SD = 7.32). If the second-most common name for an object was chosen by at least 10 people and the least common name was chosen by no more than 5 people, that object was considered be roughly equally split between the first and second most-common names. Using these criteria, 21 objects were split between the names *cup* and *mug*, 9 objects were split between *cup* and *glass*, and 1 object was split between *mug* and *glass*. Additionally, 9 of these objects did not meet these criteria to be considered

equally split between two names, either because the second most common name was chosen by fewer than 10 people or because the object was split among all three names. Of the 21 *cup*/*mug* items, 20 were called *cup* more often at post-test than at pre-test. Of the 9 *cup*/*glass* items, 4 were called *cup* more often at post-test than at pre-test, while 5 were more often called *glass*. The 1 *mug*/*glass* item was called *mug* more often at post-test than at pre-test. Furthermore, the modal name for each object in the pre-test was compared to the modal name for each object in the post-test. The modal name did change for four objects (three *mug* to *cup*; 1 *cup* to *glass*), but each of these objects had low consensus during the pre-test. These measures together suggest that people may not be switching to reflect the Russian categories per se, but that the Russian training makes them less sure of their English choice. The likelihood of an object receiving the modal name (measured as the number of people giving the object the modal name over the total number of people and multiplied by 100 to create a percentage; see Table 11 for the means and standard deviations) did not differ from pre-test to post-test either for all the objects (t (79) = 1.06, p = 0.30), for the training items alone (t (44) = 1.05, p = 0.30), or for the testing items alone, t (34) = 0.35, p = 0.73.

Feature Listing Task

The results of the feature listing task were briefly examined to determine what types of features participants were listing and, specifically, if participants were generally noting features of *stakan* that were broader than those of *glass* and/or features of *chashka* that were narrower than those of *cup*. The most common feature listed for *stakan* was that these objects did not have handles, which is not a feature that generally distinguishes the Russian word from *glass* as glasses generally also do not handles. Due to this lack of

relevant data from these preliminary analyses, the results of this task will not be discussed further here.

Discussion

The weak interconnection hypothesis predicts that participants would have fewer training runs, fewer errors in the initial training runs, greater Russian accuracy scores, and higher English consistency scores in the feedback conditions and in the blocked presentation order conditions. For feedback, the prediction was not borne out at all, as evidenced by the lack of any significant effects of feedback in any analysis. Although feedback has been found to be helpful in the categorization literature (Ashby et al., 2002), it did not make any difference for the participants here for either Russian accuracy or English consistency. The presentation order of the exemplars, however, did affect participants' performance. Although there were no significant differences between the blocked and intermixed conditions for the overall Russian accuracy scores, participants in the intermixed conditions did make fewer errors on their first run of Russian training and took fewer trials to reach criterion. These results for the more sensitive measures of learning suggest that an intermixed presentation order may be beneficial if the goal of learning is to quickly and accurately learn an L2 pattern. In terms of English consistency, however, the significant differences that were found were in the opposite direction—that is, participants in the blocked conditions were more consistent than those in the intermixed conditions, for the discrepant items and the training subset of the discrepant items. If the goal of learning is to accurately learn an L2 pattern while maintaining, as much as possible, the L1 pattern, then, the blocked presentation order may be more beneficial. Because these results were only found for the discrepant items, this may

support the idea that the more participants gain in the Russian pattern, the more they lose the sense of their original English pattern.

Although one of the variables in this experiment did lead to some significant differences among the experimental groups, which would suggest the possibility of the weak interconnection hypothesis being more accurate, this hypothesis could be only be true if at least one of the experimental groups was as consistent in their English naming choices as the control group. The strong interconnection hypothesis predicts that participants in the control condition would remain more consistent in terms of their English lexical choices than those in the experimental conditions. The results of these analyses support this prediction for all of the items tested as well as the training items alone, although not for the testing items alone. Because the training items were more strongly affected than the testing items, this suggests that participants—and possibly real-world L2 learners—may show greater shifts in their L1 patterns for those objects which they are most familiar with in the L2 context. Further, despite the significant differences of presentation order, the results of these consistency analyses suggest that the strong interconnection hypothesis is more accurate than the weak interconnection hypothesis, as the control group was more consistent in their English lexical choices than the experimental groups.

Experiment 2

The second experiment was designed to relate to immersion experiences. For many immigrants, entering an immersion context results in almost exclusive L2 usage. These immigrants are often students or young professionals who come to an L2 context alone and only occasionally speak in their first language. Many other immigrants,

however, move to a new country with their family or move to an area with a large immigrant population. These immigrants may continue to use their L1 on a regular basis when interacting with their family members or other members of the immigrant population. However, they will also use their L2 frequently when interacting with native speakers at work or school or within the larger community. These immigrants switch back and forth between their two languages across the course of a day.

This continual switching between languages is likely to increase the cross-talk between a bilingual speaker's two languages (Wolff & Ventura, 2009). A key question of this experiment was to what extent this switching leads to a decreased ability to either master a new lexical pattern or maintain a previously-learned lexical pattern.

In order to answer this question, participants in the second experiment were trained on the Russian lexical pattern either with or without English trials mixed in. Because the effect of continual updating was predicted to play a role in participants' abilities to learn the Russian lexical pattern and to maintain the English pattern, the performance of participants who received English naming trials mixed with Russian was expected to be poorer than the performance of participants who did not receive the English naming trials. The weak interconnection hypothesis predicts that participants in the Russian-only condition would perform better than those in the mixed-language condition, due to the effects of continually updating the feature maps. As participants learned the Russian pattern, their feature networks should update to reflect the Russian lexical pattern. However, if the word *glass* had recently been activated, the updating of the word *stakan* may have also affected the features the participant associated with the word *glass*. Feature updating may be unable to effectively update the word in only one

language, especially if the direct link between translation "equivalents" is strong. If the strong interconnection hypothesis is correct, no differences between groups would be expected.

If the weak interconnection hypothesis is true, another variable that could affect the role of this updating is metalinguistic knowledge. If participants know that the Russian lexical pattern differs from the English lexical pattern, they may be able to actively prevent feature updating in the language that they are not currently using. This metalinguistic knowledge is a variable that is likely to naturally occur among immigrants in an L2 environment, so half of the participants in this experiment were explicitly told that the lexical patterns across the two languages differed while half were given no such information. The manipulations of language switching and metalinguistic knowledge resulted in a 2 x 2 design. While metalinguistic knowledge could still affect participants' English consistency scores if the strong interconnection hypothesis is true, that hypothesis does predict that the experimental groups' consistency scores will still be lower than the control group's consistency scores.

Main effects of both switching between languages and metalinguistic knowledge were expected in the measures of Russian learning and English consistency. Specifically, as mentioned above, the weak interconnection hypothesis predicts that participants who were presented with Russian-only trials would be more accurate in their Russian naming and more consistent in their English naming than participants who received mixed-language trials. Similarly, the weak interconnection hypothesis also predicts that participants who received metalinguistic knowledge would name objects more accurately in Russian and more consistently in English than participants who did not receive this

knowledge. The strong interconnection hypothesis predicts that any differences found among the experimental groups for either measure would still result in higher English consistency scores for the control group than for any of the experimental groups. Although no interactions were expected under either hypothesis, the two variables were combined into a single experiment in order to make the design more efficient and use fewer participants while maintaining a relatively high level of statistical power.

Additionally, the English naming consistency of the experimental groups in Experiment 2 were also compared to the control group described under the Experiment 1 heading with a one-way ANOVA. As with Experiment 1, if the weak interconnection hypothesis is correct, participants in the experimental conditions should not differ in terms of their English consistency from the control group. If the strong interconnection hypothesis is correct, differences between the groups would be expected, with the control group expected to be more consistent with their English naming than the other groups.

Method

Participants

Participants were 56 monolingual English speakers from the participant pool at Lehigh University, with 14 participants in all conditions. In addition, the 14 participants who completed the control condition of Experiment 1 also served as the control condition for Experiment 2.

Materials

The same Language History Questionnaire and photographs from Experiment 1 were used in Experiment 2.

Procedure

The procedure for Experiment 2 was identical to the procedure for Experiment 1 with the exception of the training phase.

Training phase. After the first round of English baseline naming, participants were asked to imagine that they were in Russia and needed to focus on learning how to speak Russian in order to communicate with the people around them. In this phase, participants viewed the 45 photographs in the training set labeled with their Russian names. None of the participants in this experiment were given the dictionary translations of the Russian words. In the instructions for this phase, half of the participants were given explicit metalinguistic knowledge that Russian words often are not exactly equivalent to English words while half of the participants were given no such knowledge. All participants then saw the 45 photographs from the training set labeled with their Russian names. The photographs in this experiment were presented in random order, as it is unlikely that real-world bilingual speakers in an immersion context would be presented with a blocked presentation of any one category. After seeing each of the labeled photographs, participants were asked to label each of the 45 photographs with the correct Russian word. None of the participants received feedback in this experiment, as many real-world bilingual speakers often do not receive feedback if their intended meaning can be understood (Malt, 2008). After labeling the objects in Russian, half of the participants labeled the objects again in English. Only those 45 photographs from the training set were labeled in English in this portion of the experiment. Those participants who were in the Russian-only condition completed a short filler task designed to take approximately the same amount of time as the English naming. This filler task consisted of viewing each of the 45 pictures from the training set again and making a height judgment (deciding if

each object was at least four inches tall or shorter than four inches). This task was designed to give participants the same amount of exposure to each object as participants in the mixed-language condition while not requiring a linguistic judgment or decision in either language. Each round of viewing the photographs, labeling the photographs, and completing the English naming or the filler task made up a training run. Participants in both conditions completed their training runs until a criterion of 90% accuracy on the Russian labeling was reached.

Test phase and posttest. Once this criterion was reached, participants in both conditions were asked to generalize their Russian knowledge to novel objects (which were all objects that are called *chashka, stakan, kruzhka, riumka,* or *fuzher*; participants were limited to a choice of only these names) and then to label the objects from the testing set in English again. They then completed the feature listing task and were debriefed and dismissed. See Appendix C for the instructions used throughout this experiment.

Control group. The control group described in Experiment 1 also served as a control group for Experiment 2.

Results

A 2 (metalinguistic knowledge vs. no metalinguistic knowledge) by 2 (language training: mixed-language vs. Russian-only) fully between-subjects ANOVA was used for all analyses among the experimental groups.

As with Experiment 1, both consistency in English naming and accuracy in Russian naming (based on naming choices obtained from native Russian speakers) were analyzed.

Measures of Russian Accuracy and Learning

Accuracy of Russian naming was measured as described in Experiment 1. See Table 12 for mean accuracy scores. Participants who received metalinguistic knowledge and those who were in the Russian-only conditions were expected to be more accurate in their Russian naming than those who did not receive metalinguistic knowledge and those who were in the mixed-language conditions. Accuracy scores of the training and test items were analyzed separately (see Table 13 for statistical values). For the training items, there was a marginally significant effect of metalinguistic knowledge in the opposite direction as originally predicted, with participants who did not receive metalinguistic knowledge having marginally higher accuracy scores (M = 94.90%, SD = 3.65) than participants who did receive metalinguistic knowledge (M = 92.86%, SD = 4.58). Based on the medium effect size found for this measure, the sample size may have been too small to detect the effect. There was no significant effect of language training and no significant interaction. For the primary measure of interest, novel test items alone, there was no significant effect of either metalinguistic knowledge or language training, nor a significant interaction. The original predictions were therefore unsupported.

Analyses of variables related to the speed of learning also failed to indicate significant effects of either metalinguistic knowledge or language training (see Table 14 for statistical values), although it was originally predicted that participants who received metalinguistic knowledge and those who completed the Russian-only trials would learn the Russian pattern more quickly than those who did not receive metalinguistic knowledge and those who completed the mixed-language trials. For the number of trials it took participants to reach the 90% criterion, there were no effects of metalinguistic

knowledge, no significant effect of language training, and no interaction. For the number of errors made on the first run of Russian naming, there were no significant effects of metalinguistic knowledge, no significant effects of language training, and no interaction. Again, then, the original predictions were not supported.

M**easures of English Consistency**

English consistency scores were also obtained as described for Experiment 1. The strong interconnection hypothesis predicts that the control group would have higher consistency scores than any of the experimental groups. The scores of the control group were therefore compared to those of the experimental groups (see Table 15 for statistical values). As predicted, among all items tested, there was a significant effect of condition, with a planned comparison indicating that the control group was significantly more consistent (M = 87.14%, SD = 6.38) than the experimental groups (M = 78.75%, SD = 9.73). This effect was also found for the training items alone, with a planned comparison indicating that the control group was again more consistent (M = 85.56%, SD = 6.09) than the experimental groups (M = 77.22%, SD = 10.64). For the test items alone, there was a marginal effect of condition, which could be due to the small sample size; based on the significant results for all items and for the training items alone, it is likely that a larger sample size would result in a significant result for the test items alone as well. A planned comparison indicated that the control group was significantly more consistent (M = 89.18%, SD = 8.47) than the experimental groups (M = 80.71%, SD = 10.95).

Consistency scores for the discrepant and compatible items were analyzed separately (see Table 15 for statistical values). For the discrepant items, there was no significant difference between groups. For the compatible items, however, there was a

significant difference, with a planned comparison indicating that the control group was significantly more consistent (M = 88.89%, SD = 7.54) than the experimental groups (M = 77.08%, 12.14). The discrepant and compatible items were also broken down into training and testing items. The only significant difference among groups in these four analyses was for the compatible training items, with a follow-up comparison indicating that the control group (M = 89.52%, SD = 6.25) was significantly more consistent than the experimental groups (M = 75.83%, SD = 12.75). As with Experiment 1, the predictions were generally supported, with the control group being more consistent than the experimental groups overall as well as for most of the subsets of the data.

Again, however, another possible explanation of the differences between groups is that participants in the control group spent less time between the rounds of English naming than did participants in the experimental groups, (F (4, 65) = 10.43, $p < 0.001$), with a planned comparison indicating that the control group spent significantly less time (M = 8.52 minutes, SD = 0.51, range = 7.84 – 13.05) on the filler task than the experimental groups spent on the Russian training (M = 17.65 minutes, SD = 5.46, range = 6.39 – 32.18), t (65) = 6.22, $p < 0.001$. As in Experiment 1, though, the correlations between English consistency and training run times suggest that this explanation is unlikely. Specifically, English consistency (range = 53.75 – 92.50) was not significantly correlated with either how long participants spent on the Russian training (r = -0.11, p = 0.44) or how many trials participants needed to reach the criterion to advance, range = 1 – 7, r = -0.24, p = 0.08.

The English consistency scores were also analyzed for only the experimental groups, with the original predictions stating that participants who received metalinguistic

knowledge and those completing the Russian-only trials would be more consistent than those who did not receive metalinguistic knowledge and those who completed the mixed-language trials. See Table 16 for mean consistency scores for all groups and Table 17 for statistical values for the following analyses. Among only the training items, there was a marginally significant effect of language training. Participants in the mixed-language conditions were marginally more consistent (M = 79.92%, SD = 9.78) than those in the Russian-only conditions (M = 74.52%, SD = 10.94). The medium effect size found for this result indicates that the sample size may have been too small. There was no significant effect of metalinguistic knowledge nor a significant interaction. For only the test items, there were no significant differences of metalinguistic knowledge or language training, nor was there a significant interaction.

Consistency scores were also analyzed separately for the discrepant items and the compatible items. See Table 18 for consistency scores for these items and Table 19 for the statistical values for the following analyses. For the discrepant items, there was a marginally significant effect of language training (with a medium effect size) with participants in the mixed-language condition marginally more consistent (M = 87.79%, SD = 9.08) than participants in the Russian-only condition (M = 77.43%, SD = 11.08). There was no main effect of metalinguistic knowledge nor a significant interaction. For the compatible items, there were no significant effects of metalinguistic knowledge or language training, nor a significant interaction. As in Experiment 1, the discrepant and compatible items were also broken down individually into the training and testing items (see Table 20 for the statistical values for the analyses of the discrepant items and Table 21 for the statistical values for the analyses of the compatible items). Among the

discrepant training items, there was a significant effect of language training. Participants in the mixed-language condition were more consistent (M = 81.07%, SD = 10.27) than participants in the Russian-only condition (M = 74.76%, SD = 12.62). There was no significant effect of metalinguistic knowledge nor a significant interaction. There were no significant effects for the discrepant testing items, the compatible training items, or the compatible testing items. In sum, the predictions for metalinguistic knowledge were not supported at all, while those for the Russian-only vs. mixed-language condition were generally marginal and in the opposite direction as predicted.

Analyses of the English consistency data indicated that, as in Experiment 1, there was a strong correlation between the number of people who chose the modal name for an object during the pre-test and how many people remained consistent with their choice, whether the correlation was conducted for all objects ($r = 0.84$, $p < 0.00001$), the training items alone ($r = 0.82$, $p < 0.00001$), or the testing items alone ($r = 0.87$, $p < 0.00001$) again suggesting that people were more likely to change the names for items that had low initial agreement, regardless of an object's discrepant or compatible status with the Russian word.

As in Experiment 1, a median split was used to separate the 40 objects that were most likely to receive a consistent name (M = 88.08%, SD = 6.37) from the 40 objects that were least likely to receive a consistent name (M = 68.39%, SD = 6.21). Using the same criteria as Experiment 1, 20 objects were split between the names *cup* and *mug*, 13 objects were split between *cup* and *glass*, and 1 object was split between *mug* and *glass*. Additionally, 6 objects either received the second-most common name from fewer than 10 people or were split among all three names. Of the 20 *cup*/*mug* items, 19 were called

cup as often or more often at post-test than at pre-test. Of the 13 *cup*/*glass* items, 11 were called *cup* as often or more often at post-test than at pre-test. The 1 *mug*/*glass* item was called *mug* more often at post-test than at pre-test. The modal name changed for three objects (1 *glass* to *cup*, 1 *mug* to *cup*, and 1 *cup* to *glass*). Additionally, one object that was called *mug* most often at pre-test was called *cup* and *mug* equally often at post-test. Together these results again suggest that participants are not changing their English pattern to look more like the Russian pattern, but that the Russian training makes them less sure of their English choice. A paired-samples *t*-test indicated that objects were more likely to receive the modal name during the post-test than during the pre-test (see Table 22 for means and standard deviations) for all objects (t (79) = -2.97, p = 0.004), for the training items alone (t (44) = -2.04, p = 0.05), or for the testing items alone, t (34) = -2.50, p = 0.02.

Feature Listing Task

As in Experiment 1, preliminary analyses of the feature listing task did not lead to relevant data, so the results of that task will not be discussed further here.

Discussion

The specific experimental predictions of this experiment, if the weak interconnection hypothesis is correct, were that the participants who received metalinguistic knowledge and those in the Russian-only condition would perform better in terms of Russian accuracy and English consistency than those who did not receive metalinguistic knowledge and those in the mixed-language condition. However, the results generally suggest that neither variable made a large impact in participants' scores. The non-significant trends that were found, as well as the one significant result (an effect

of language training for consistency scores on only the discrepant training items) were all in the opposite direction as predicted. While the overall lack of differences by itself does not eliminate the weak interconnection hypothesis, as it is possible that variables other than the ones tested here could lead to differences in Russian accuracy or English consistency, the results of the control group as compared to the experimental groups suggest that the strong interconnection hypothesis is more accurate.

As in Experiment 1, the strong interconnection hypothesis predicts that the control group would have higher English consistency scores than the experimental groups. This expected result was found, strongly suggesting that the strong interconnection hypothesis is likely more accurate than the weaker version.

General Discussion

The primary goal of these experiments was to distinguish between the strong and weak interconnection hypotheses. Under the strong interconnection hypothesis, bilingual speakers' lexical patterns will inevitably differ from the monolingual patterns of either language due to unavoidable cross-connections between the two lexicons. Under the weak interconnection hypothesis, the differences that have been observed in real-world bilingual speakers are due to learning conditions as well as to the structure of the lexical network. This is not because the interconnections do not exist but because specific learning conditions allow for those interconnections to affect one another in terms of the lexical choices of bilingual speakers, while other types of learning conditions may prevent these effects from occurring. If this hypothesis is true, bilingual speakers can obtain monolingual-like lexicons in each of their languages if they learn each language

under optimal conditions. Taken together, the results of both experiments indicate that the strong interconnection hypothesis is likely more accurate than the weaker version.

Although participants in these experiments were able to accurately learn the L2 Russian pattern—which has often not been found for real-world bilinguals (e.g., Malt & Sloman, 2003)—they were unable to maintain the original L1 pattern while doing so, at least for the items on which they were directly trained, suggesting that even brief exposure (15-18 minutes, on average) to an alternate lexical pattern may influence speakers' L1.

Furthermore, the various learning conditions across both experiments made little difference in the participants' ability to learn the Russian patterns or to remain consistent in the English patterns, suggesting that the commonalities among the experimental conditions across both experiments—repeated exposure to multiple exemplars of a limited number of names—made it possible for participants to learn the Russian pattern, rather than any specific learning condition.

The one experimental variable that made a consistent difference in participants' scores—blocked vs. intermixed presentations of the exemplars in Experiment 1—was in the opposite direction as originally predicted for learning the Russian pattern, but in the predicted direction for measures of English consistency, suggesting that different presentation orders may be differentially beneficial for L2 learners depending on what the learner's goal is. Originally, the blocked presentation order was thought likely to be more effective in helping participants learn the L2 pattern, as it would help participants see the commonalities of each category (Carvalho & Goldstone, 2011). However, as the intermixed category presentation actually led to fewer mistakes in the first round of

Russian naming and fewer trials needed to reach the 90% criterion to advance to the testing stage of the experiment, it seems that being able to see the differences across the Russian categories was beneficial for participants, a finding which fits with general findings from the categorization literature (e.g., Lavis & Mitchell, 2006). These results may also have been influenced by the format of the test phase, which was also presented in an intermixed fashion and could therefore have led to benefits due encoding specificity. The blocked presentation order of the Russian lexical categories did, though, lead to higher English consistency scores, so seeing the commonalities within the categories may have helped participants understand how they were different from the English lexical categories and therefore may have led to less of an influence of L2 on L1.

Although the English consistency scores of the control group, as compared to the experimental groups, do indicate that the brief L2 training led to a difference in participants' L1 lexical patterns, the differences are not simple shifts to more closely match the L2 pattern. That is, participants are not, for example, simply more likely to call tall plastic items *glass* in order to more closely fit the native Russian usage of *stakan*. Instead, it seems that the decision-making may simply become looser in a sense. The lack of one-to-one shifts may also suggest that the typicality of an item for a particular name may have an effect on how likely the name is to shift (Pavlenko & Malt, 2011). Pavlenko and Malt suggest that the items most likely to shift names may be either those with the lowest L1 typicality or those with the highest L2 typicality. Because the objects with the highest typicality for a Russian name may not be the same as the objects with the highest typicality for the English name, these two possibilities may compete with one another and lead to non-direct influences from the L2 pattern to the L1 pattern.

Because the changes in consistency were generally stronger for the training items than for the testing items across both experiments, this may also indicate that the lexical network does not immediately generalize across all objects upon exposure to a second language. That is, at least some of the connections in the network may be connected to specific exemplars and it is only these connections that shift, at least initially. It is also possible that the connections that have shifted due to the training may still be quite weak, and when participants are presented with exemplars that they were not trained on, the connections are more easily able to revert back to their original pattern. It is possible that, over time, bilingual speakers' lexical networks may generalize more fully, leading to changes in naming patterns even for specific exemplars to which they had not been exposed in an L2 context, and that the relatively short amount of time between the baseline and posttest measures of English naming did not provide the opportunity for the participants' lexical networks to generalize, but more testing would be needed to explore this possibility fully.

One unexpected disconnect between the two experiments is that in Experiment 1, objects were equally likely to receive the modal English name during both the pre-test and the post-test, while in Experiment 2, objects were more likely to receive the modal name during the post-test than during the pre-test. While the result for Experiment 2 may seem counter-intuitive at first, it is possible that the Russian training led participants to focus more closely on specific features of objects and therefore led to higher consensus for the object names. If this explanation is correct, it raises the question of why the difference was not found for Experiment 1.

Future Directions

The participants in these experiments were able to learn the Russian pattern quickly and accurately, which has not been found with real-world bilingual speakers, some of whom differ significantly from monolingual norms even after 10 or more years in an immersion environment (Malt & Sloman, 2003). The ceiling effects of Russian accuracy that were found are perhaps not surprising, given the 90% criterion that participants were required to meet before being tested on novel objects, although the speed at which these ceiling levels were reached was unexpected. This quick learning and the ceiling effects do suggest that the language learning here may differ in important ways from the language learning of real-world L2 learners. One possible way to address this concern would be to give all participants an equal amount of exposure to the L2 lexical pattern by requiring all participants to complete the same number of training runs, regardless of their accuracy levels on those runs. This would likely lead to greater variability in the Russian accuracy scores, potentially allowing for more significant differences among the experimental variables. This measure would be a way of exploring how these variables might affect learning of a new pattern, and how that new pattern influences the old pattern, before participants reach ceiling levels on the new pattern. This would, though, eliminate the possibility of using trials to criterion as a measure of Russian learning.

Many of the findings of this experiment were only marginally significant, but measures of effect size indicated that the findings were in the medium range. It is therefore likely that increasing the level of statistical power in these experiments by increasing the number of participants would lead to a greater number of significant effects.

One concern regarding the materials for these experiments is that there was only a single novel exemplar of *fuzher* and of *riumka* included in the testing set. Although the original full testing set of 88 objects included several objects that were thought likely to receive one of those two names, based on similarity to the objects that had received those names in Pavlenko and Malt's (2011), the native Russian speakers called many of the objects that were thought likely to be called *fuzher* by the name *bokal* instead while many of the objects thought likely to be *riumka* were most often called by that name, but by fewer than 7 of the 11 speakers. Future research should address this limitation by including more objects in the final testing set that are called *fuzher* and *riumka* by native speakers, as this would help ensure that participants are truly generalizing their knowledge to novel objects. With only one *fuzher* and one *riumka* in the testing set, both of which may be highly typical exemplars (the lack of agreement from the native Russian speakers on the objects thought likely to be called *fuzher* or *riumka* indicates that the objects that did receive one of those names may be high in typicality) it is impossible to know how well participants are truly generalizing the Russian lexical pattern to objects that may be less typical.

Future research should also ensure that the filler task used in the control condition more closely matches the time spent on the L2 training in the experimental conditions to eliminate the possibility that the differences in English consistency between the control and the experimental groups found in these experiments were due to the shortened amount of time in between the two rounds of English naming in the control condition. If a filler task that took the same amount of time as the Russian training resulted in English consistency scores that were not significantly different from one another across the

control and experimental groups, this would indicate that the Russian training does not, in fact, have a greater influence on the English lexical patterns than would be expected from the inconsistency that occurs naturally in everyday language use.

Conclusions

The results of these experiments suggest that the preliminary conclusion from the pilot study (Jobe, 2012)—that the weak interconnection hypothesis was more accurate than the stronger interconnection hypothesis—was likely incorrect. Although L2 learners may be able to learn a new lexical pattern to match monolingual norms given an adequate learning environment, they may not be able to do so while simultaneously maintaining their original L1 pattern. Overall, then, the results here, particularly those comparing the control group to the experimental groups, indicate that learning an L2 lexical pattern may affect participants' L1 patterns even with less than 20 minutes of exposure to the new pattern, suggesting that bilingual lexical interactions are likely inevitable.

Table 1

Experiment 1 Russian Accuracy Means (and Standard Deviations)

	Training Items			Testing Items		
	Feedback	No Feedback	Total	Feedback	No Feedback	Total
Blocked	94.92 (6.36)	94.87 (3.05)	94.90 (5.00)	87.81 (9.74)	88.31 (7.92)	88.02 (8.85)
Intermixed	92.06 (5.23)	92.86 (5.22)	92.47 (5.18)	87.81 (8.26)	90.36 (6.16)	89.12 (7.24)
Total	93.49 (5.93)	93.76 (4.43)		87.81 (8.88)	89.52 (6.86)	

Table 2

Experiment 1 Analysis of Variance for Russian Accuracy Scores

Source	*df*	*F*	Partial η^2	*p*
		Training Objects		
Feedback	1	0.08	0.001	0.78
Presentation Order	1	3.24	0.06	0.08
Interaction	1	0.10	0.002	0.76
Error	55	(26.88)		
		Testing Objects		
Feedback	1	0.49	0.009	0.49
Presentation Order	1	0.22	0.004	0.64
Interaction	1	0.22	0.004	0.64
Error	53	(95.69)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 3

Experiment 1 Analysis of Variance for Number of Trials to Criterion and Errors Made on First Training Run

Source	*df*	*F*	Partial η^2	*p*
		Trials to Criterion		
Feedback	1	1.34	0.03	0.24
Presentation Order	1	8.10**	0.13	0.01
Interaction	1	1.92	0.03	0.17
Error	55	(1.18)		
		Errors Made on First Training Run		
Feedback	1	0.28	0.01	0.60
Presentation Order	1	8.12**	0.13	0.01
Interaction	1	1.67	0.03	0.20
Error	55	(52.04)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 4

Experiment 1 English Consistency Score Means (and Standard Deviations)

	All Items			Training Items			Test Items		
	Feedback	No Feedback	Total	Feedback	No Feedback	Total	Feedback	No Feedback	Total
Blocked	79.92 (6.45)	81.15 (8.55)	80.49 (7.38)	78.96 (7.83)	77.61 (8.29)	78.33 (7.92)	81.52 (9.53)	84.83 (8.68)	83.06 (9.13)
Intermixed	77.08 (8.64)	78.98 (7.84)	78.06 (7.54)	74.07 (8.43)	75.56 (8.07)	74.84 (8.14)	85.33 (8.14)	86.25 (9.91)	85.81 (8.96)
Total	78.50 (7.38)	79.96 (7.42)		76.52 (8.37)	76.47 (8.09)		83.43 (8.92)	85.62 (9.24)	

Table 5

Experiment 1 Analysis of Variance and Planned Comparisons for English Consistency Scores (Control and Experimental Groups)

	df (one-way ANOVA)	*F*	*p*	*df* (follow-up *t*-test)	Follow-up *t*-value	*p*	*r*
All Items	4	4.37**	0.003	68	-3.35**	0.001	0.40
Training	4	4.76**	0.002	68	-3.91**	<0.001	0.43
Testing	4	0.88	0.48				
Discrepant	4	3.52*	0.01	68	-2.69**	0.01	0.31
Compatible	4	3.63*	0.01	68	-3.41**	0.001	0.38
Discrepant Training	4	3.11*	0.02	68	-2.04*	0.05	0.24
Compatible Training	4	3.93**	0.01	68	-3.79**	<0.001	0.42
Discrepant Testing	4	0.80	0.53				
Compatible Testing	4	2.70*	0.04	68	-1.75	0.09	0.21
Error	68						

Note. *$p < 0.05$. **$p < 0.01$.

Table 6

Experiment 1 Analysis of Variance for English Consistency Scores

Source	*df*	*F*	Partial η^2	*p*
		Training Items		
Feedback	1	0.001	<0.001	0.98
Presentation Order	1	2.66	0.05	0.11
Interaction	1	0.44	0.01	0.51
Error	55	(66.44)		
		Test Items		
Feedback	1	.079	0.01	0.38
Presentation Order	1	1.20	0.02	0.28
Interaction	1	0.25	0.01	0.62
Error	55	(83.22)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 7

Experiment 1 English Consistency Score Means (and Standard Deviations) For Discrepant and Compatible Items

	Discrepant Items			Compatible Items		
	Feedback	No Feedback	Total	Feedback	No Feedback	Total
Blocked	82.42 (8.08)	81.12 (7.09)	81.82 (7.53)	76.85 (8.04)	79.27 (13.30)	77.98 (10.66)
Intermixed	75.30 (10.79)	76.85 (8.79)	76.10 (9.68)	79.26 (9.38)	81.60 (8.54)	80.47 (8.89)
Total	78.86 (10.04)	78.76 (8.22)		78.06 (8.67)	80.56 (10.78)	

Table 8

Experiment 1 Analysis of Variance for English Consistency of Discrepant and Compatible Items

Source	*df*	*F*	Partial η^2	*p*
		Discrepant Items		
Feedback	1	0.003	<0.001	0.96
Presentation Order	1	6.08*	0.10	0.02
Interaction	1	0.38	0.01	0.54
Error	55	(78.31)		
		Compatible Items		
Feedback	1	0.85	0.02	0.36
Presentation Order	1	0.84	0.02	0.36
Interaction	1	<0.001	<0.001	0.99
Error	55	(97.37)		

Table 9

Experiment 1 Analysis of Variance for the Training and Testing Subsets of the Discrepant Items

Source	*df*	*F*	Partial η^2	*p*
		Training Items		
Feedback	1	0.27	0.01	0.61
Presentation Order	1	5.28*	0.09	0.03
Interaction	1	0.89	0.02	0.35
Error	55	(94.30)		
		Testing Items		
Feedback	1	1.16	0.29	0.29
Presentation Order	1	3.37	0.06	0.07
Interaction	1	0.05	0.001	0.83
Error	55	(128.02)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 10

Experiment 1 Analysis of Variance for the Training and Testing Subsets of the Compatible Items

Source	*df*	*F*	Partial η^2	*p*
		Training Items		
Feedback	1	0.02	<0.001	0.88
Presentation Order	1	1.16	0.02	0.29
Interaction	1	0.29	0.01	0.59
Error	55	(159.47)		
		Testing Items		
Feedback	1	1.45	0.03	0.23
Presentation Order	1	0.24	0.004	0.62
Interaction	1	0.19	0.003	0.67
Error	55	(139.70)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 11

Experiment 1 Means (and Standard Deviations) for Likelihood of Objects Receiving Modal Names

All Items		Training Items		Testing Items	
Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
82.65 (14.05)	81.37 (13.94)	81.13 (13.84)	79.32 (14.47)	84.60 (14.28)	84.02 (12.95)

Table 12

Experiment 2 Russian Accuracy Means (and Standard Deviations)

	All Items			Novel Items		
	Metalinguistic Knowledge	No Metalinguistic Knowledge	Total	Metalinguistic Knowledge	No Metalinguistic Knowledge	Total
Mixed Language	89.16 (3.47)	88.39 (7.20)	88.78 (5.56)	87.55 (4.83)	85.71 (11.15)	86.63 (8.48)
Russian Only	89.41 (3.09)	88.39 (8.57)	88.90 (6.34)	86.53 (4.11)	84.49 (13.76)	85.51 (10.02)
Total	89.29 (3.22)	88.39 (7.77)		87.04 (4.43)	85.10 (12.30)	

Table 13
Experiment 2 Analysis of Variance for Russian Accuracy Scores

Source	*df*	*F*	Partial η^2	*p*
		Training Objects		
Metalinguistic Knowledge	1	3.47	0.06	0.07
Language Training	1	1.54	0.03	0.22
Interaction	1	1.54	0.03	0.22
Error	52	(16.82)		
		Testing Objects		
Metalinguistic Knowledge	1	0.60	0.01	0.44
Language Training	1	0.20	0.004	0.66
Interaction	1	0.002	<0.001	0.97
Error	52	(88.46)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 14

Experiment 2 Analysis of Variance for Number of Trials to Criterion and Errors Made on First Training Run

Source	*df*	*F*	Partial η^2	*p*
		Trials to Criterion		
Metalinguistic Knowledge	1	0.01	<0.001	0.91
Language Training	1	0.58	0.01	0.45
Interaction	1	0.30	0.01	0.59
Error	52	(1.51)		
		Errors Made on First Training Run		
Metalinguistic Knowledge	1	0.34	0.01	0.56
Language Training	1	0.34	0.01	0.56
Interaction	1	1.49	0.03	0.23
Error	52	(50.65)		

Table 15

Experiment 2 Analysis of Variance and Planned Comparisons for English Consistency Scores (Control and Experimental Groups)

	df (one-way ANOVA)	*F*	*p*	*df* (follow-up *t*-test)	Follow-up *t*-value	*p*	*r*
All Items	4	3.31*	0.02	65	-3.08**	0.003	0.36
Training	4	3.20*	0.02	65	-2.84**	0.01	0.33
Testing	4	2.34	0.07	65	-2.68**	0.01	0.32
Discrepant	4	1.99	0.11				
Compatible	4	3.39*	0.01	65	-3.43**	0.001	0.39
Discrepant Training	4	1.70	0.16				
Compatible Training	4	4.01**	0.01	65	-3.84**	<0.001	0.43
Discrepant Testing	4	1.18	0.33				
Compatible Testing	4	1.73	0.16				
Error	65						

Note. *$p < 0.05$. **$p < 0.01$.

Table 16

Experiment 2 English Consistency Means (and Standard Deviations)

	All Items			Training Items			Test Items		
	Meta-linguistic Knowledge	No Meta-linguistic Knowledge	Total	Meta-linguistic Knowledge	No Meta-linguistic Knowledge	Total	Meta-linguistic Knowledge	No Meta-linguistic Knowledge	Total
Mixed Language	81.33 (9.88)	80.89 (9.70)	81.11 (9.61)	80.63 (10.33)	79.21 (9.53)	79.92 (9.78)	82.25 (12.23)	83.06 (11.45)	82.65 (11.64)
Russian Only	76.52 (8.43)	76.25 (10.63)	76.38 (9.42)	75.56 (10.39)	73.49 (11.77)	74.52 (10.94)	77.76 (8.55)	79.80 (11.62)	78.78 (10.06)
Total	78.93 (9.35)	78.57 (10.26)		78.10 (10.49)	76.35 (10.91)		80.00 (10.61)	81.43 (11.44)	

Table 17

Experiment 2 Analysis of Variance for English Consistency Scores

Source	*df*	*F*	Partial η^2	*p*
		Training Objects		
Metalinguistic Knowledge	1	0.38	0.01	0.54
Language Training	1	3.67	0.07	0.06
Interaction	1	0.01	<0.001	0.91
Error	52	(111.02)		
		Testing Objects		
Metalinguistic Knowledge	1	0.23	0.004	0.63
Language Training	1	1.72	0.03	0.20
Interaction	1	0.04	0.001	0.84
Error	52	(122.23)		

Table 18
Experiment 2 English Consistency Score Means (and Standard Deviations) For Discrepant and Compatible Items

	Discrepant Items			Compatible Items		
	Metalinguistic Knowledge	No Metalinguistic Knowledge	Total	Metalinguistic Knowledge	No Metalinguistic Knowledge	Total
Mixed Language	82.63 (10.50)	82.95 (7.81)	82.79 (9.08)	79.76 (11.35)	78.37 (14.58)	79.07 (12.84)
Russian Only	77.92 (11.04)	76.95 (11.51)	77.44 (11.08)	74.80 (10.70)	75.40 (12.22)	75.10 (11.28)
Total	80.28 (10.84)	79.95 (10.13)		77.28 (11.11)	76.88 (13.29)	

Table 19

Experiment 2 Analysis of Variance for English Consistency of Discrepant and Compatible Items

Source	*df*	*F*	Partial η^2	*p*
		Discrepant Objects		
Metalinguistic Knowledge	1	0.01	<0.001	0.91
Language Training	1	3.78	0.07	0.06
Interaction	1	0.06	0.001	0.82
Error	52	(106.41)		
		Compatible Objects		
Metalinguistic Knowledge	1	0.02	<0.001	0.90
Language Training	1	1.46	0.03	0.23
Interaction	1	0.09	0.002	0.76
Error	52	(151.33)		

Table 20

Experiment 2 Analysis of Variance for the Training and Testing Subsets of the Discrepant Items

Source	*df*	*F*	Partial η^2	*p*
		Training Items		
Metalinguistic Knowledge	1	0.92	0.02	0.34
Language Training	1	4.13*	0.07	0.05
Interaction	1	0.01	<0.001	0.91
Error	52	(134.966)		
		Testing Items		
Metalinguistic Knowledge	1	2.61	0.05	0.11
Language Training	1	1.00	0.02	0.32
Interaction	1	0.15	0.003	0.70
Error	52	(153.83)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 21

Experiment 2 Analysis of Variance for the Training and Testing Subsets of the Compatible Items

Source	*df*	*F*	Partial η^2	*p*
		Training Items		
Metalinguistic Knowledge	1	0.04	0.001	0.84
Language Training	1	1.06	0.02	0.31
Interaction	1	0.01	<0.001	0.95
Error	52	(168.32)		
		Testing Items		
Metalinguistic Knowledge	1	0.10	0.002	0.76
Language Training	1	1.25	0.02	0.27
Interaction	1	0.24	0.01	0.62
Error	52	(201.87)		

Note. Values in parentheses represent mean square errors.

*$p < 0.05$. **$p < 0.01$.

Table 22

Experiment 2 Means (and Standard Deviations) for Likelihood of Objects Receiving Modal Names

All Items		Training Items		Testing Items	
Pre-Test	Post-Test	Pre-Test	Post-Test	Pre-Test	Post-Test
78.04 (13.15)	80.69 (13.94)	76.87 (12.95)	79.76 (14.41)	79.54 (13.43)	81.89 (13.42)

Figure 1. Sample pictures of *chashka*, *kruzhka*, and *stakan*.

References

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language, 60,* 270-290.

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53,* 60-80.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory and Cognition, 30,* 666-677.

Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly, 32,* 233-262.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carvalho, P. F., & Goldstone, R. L. (2011). Sequential similarity and comparison effects in category learning.

Cook, V. (2003). Introduction: The changing L1 in the L2 user's mind. In V. Cook (Ed.), *Effects of the Second Language on the First* (pp.1-18). Clevedon, UK: Multilingual Matters Ltd.

De Groot, A. M. B. (1993). Word-type effects in bilingual processing tasks: Support for a mixed-representational system. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 27-51). Amsterdam: John Benjamins Publishing Company.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford, England: Oxford University Press.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory and Cognition, 33*, 1220-1234.

Guttentag, R. E., Haith, M. M., Goodman, G. S., & Hauch, J. (1984). Semantic processing of unattended words by bilinguals: A test of the input switch mechanism. *Journal of Verbal Learning and Verbal Behavior, 23,* 178-188.

Jobe, R. (2012). Bilingual lexical interactions: Inevitable or malleable? Unpublished manuscript.

Kronenfeld, D. B., Armstrong, J. D., & Wilmoth, S. (1985). Exploring the internal structure of linguistic categories: An extensionist semantic view. In J. W. D. Dougherty (Ed.), *Directions in cognitive anthropology* (pp. 91-113). Urbana: University of Illinois Press.

Lavis, Y., & Mitchell, C. (2006). Effects of preexposure on stimulus discrimination: An investigation of the mechanisms responsible for human perceptual learning. *The Quarterly Journal of Experimental Psychology, 59,* 2083-2101.

Macnamara, J. T. (1967). The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior, 19,* 485-502.

Malt, B. C. (2008). Grant proposal.

Malt, B. C., & Li, P. (2011). Grant proposal.

Malt, B. C. & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition, 6*, 47-67.

Malt, B. C., Sloman, S. A., Gennari, S., Shi, M., & Wang, Y. (1999). Knowing versus naming: Similarity and the linguistic categorization of artifacts. *Journal of Memory and Language, 40*, 230-262.

Pavlenko, A., & Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and first-language object naming by Russian-English bilinguals. *Bilingualism: Language and Cognition, 14*, 19-45.

Penfield, W., & Roberts, L. (1959). *Speech and Brain-Mechanisms.* Princeton, NJ: Princeton University Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Sahlin, B. H., Harding, M. G., & Seamon, J. G. (2005). When do false memories cross language boundaries in English-Spanish bilinguals? *Memory and Cognition, 33*, 1414-1421.

Schneider, H. D., & Hopp, J. P. (2011). The use of the Bilingual Aphasia Test for assessment and transcranial direct current stimulation to modulate language acquisition in minimally verbal children with autism. *Clinical Linguistics & Phonetics, 25*, 640-654.

Schwanenflugel, P. J., & Rey, M. (1986). Interlingual semantic facilitation: Evidence for a common representational system in the bilingual lexicon. *Journal of Memory and Language, 25*, 605-618.

Appendix A

Language History and Use Questionnaire

1. Age

2. Sex

3. Native language

4. Where did you grow up? (City or town, state, and country) List all places where you lived before coming to Lehigh. If more than one location, please list them all and how many years you spent in each.

5. If you have lived in any non-English speaking country, where was it and for how long?

6. Have you learned any other languages besides English? If so, please list each language and the age at which you began to learn it, and rate your overall proficiency (reading, writing, speaking, listening combined) using the 1 to 7 scale.

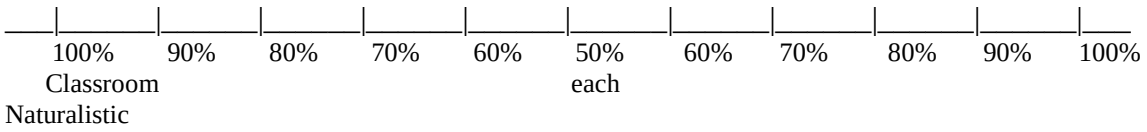
____1_____2_____3_____4_____5_____6_____7_____

minimal intermediate completely proficient;

entirely native-like

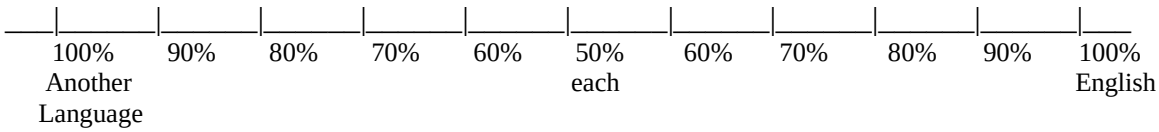
	Age	Overall Proficiency
6a1. Language 1: _____	6a2. _____	6a3. _____
6b1. Language 2: _____	6b2. _____	6b3. _____
6c1. Language 3: _____	6c2. _____	6c3. _____

7. For any languages you speak other than English, what percentage of your learning occurred in a classroom and what percentage occurred in more naturalistic ways (i.e., speaking with native speakers of that language, spending time in a country where that language is spoken, reading for pleasure books or magazines written in that language, etc.)?

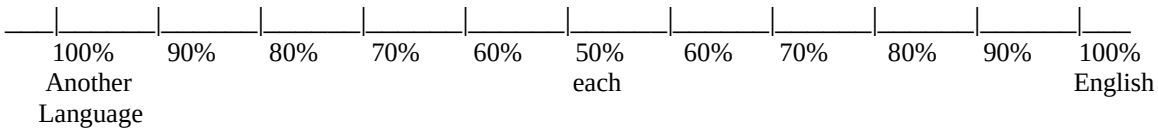


Please consider the following activities within the time frame of the past year.

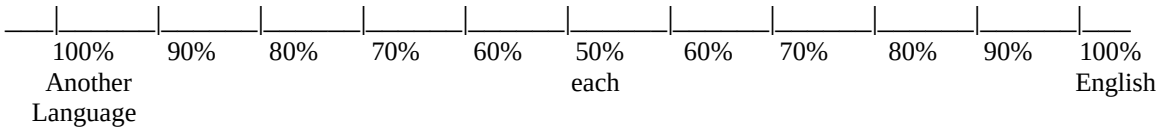
8. For all daily activities combined, what is the relative amount of English vs. another language that you use?



9. For all daily activities combined EXCLUDING formal second-language classes (i.e. a Spanish or French class you might take here at Lehigh), what is the relative amount of English vs. another language that you use?



10. When you voluntarily (not for a class assignment) complete leisure activities such as speaking, reading, or watching TV, what percentage of those activities are in English vs. another language?



Appendix B

Experiment 1 Instructions

Baseline Testing

Welcome to the experiment!

There are several stages to this experiment. For the first stage, you will see pictures of common objects on the screen.

These objects are usually called "cup," "mug," or "glass" by English speakers. Please indicate which of these names you think you would use for these objects by pressing the "c" key for cup, the "m" key for mug, or the "g" key for glass.

You can work at your own pace for this section of the experiment.

Please hit the spacebar to begin the experiment.

Training Phase

Now, please imagine that you will soon be leaving to study abroad in Russia and that you need to begin learning the Russian language. In order to help you with this task, you will see the same objects that you previously named in English. First, you will see each object labeled with its Russian name and then you will try to produce the names on your own.

There are only five names that you need to learn: kruzhka, stakan, chashka, fuzher, and riumka. Your goal is to learn which objects are called by each name.

Press the spacebar for more instructions.

Training Phase Screen 2

You will now see each object presented with its Russian name for three seconds. Just try to focus on learning which objects are called by each name. After you see each

object once, you will be asked to name the objects as kruzhka, stakan, chashka, fuzher, or riumka. As you see each picture, please say the Russian word out loud. You will repeat this task (both seeing the objects and producing your labels) until you label at least 90% of the objects correctly.

"Kruzhka" is typically translated as "mug," "stakan" is typically translated as "glass," and "chashka" is typically translated as "cup." You will also see objects called "fuzher" and "riumka." These words do not have good English translations.

Press the spacebar to continue.

Training Phase Screen 3

You will now see the same objects you have just seen. Please indicate what you believe the correct Russian name for each object is by pressing the "k" key for kruzhka, the "s" key for stakan, the "c" key for chashka, the "f" key for fuzher, or the "r" key for riumka. As you make your choice, please say the word out loud. After you give your answer, you will receive feedback on your response. Please use this information to help you learn.

You can move at your own pace for this section of the experiment.

Press the spacebar to continue.

Training Phase: Criterion Not Met

You will now see the objects labeled with the correct Russian names again. Keep learning the names and you will be tested again.

Press the spacebar to continue.

Training Phase: Criterion Met

Congratulations! You've now reached the criterion to move on to the next phase of the experiment. Press the spacebar to continue.

Test Phase

You will now see more pictures of objects. Some of these will be pictures you've already been tested on in on Russian, while some will be new. Please indicate, by pressing the "k," "s," "c," "f," or "r" keys, if you think the object would be called a kruzhka, a stakan, a chashka, a fuzher, or a riumka. As you make your choice, please say the word out loud. [This time, you will not receive feedback on your choice.]

Press the spacebar to begin.

Posttest

In this final phase of the experiment, you will see each of the objects again. This time, please indicate what you would call each object in English by pressing "c" for cup, "m" for mug, or "g" for glass.

Press the spacebar to continue.

Final Screen

Thank you for participating. Please let the experimenter know that you are finished with the experiment.

Appendix C

Experiment 2 Instructions

Baseline Testing

Welcome to the experiment!

There are several stages to this experiment. For the first stage, you will see pictures of common objects on the screen.

These objects are usually called "cup," "mug," or "glass" by English speakers. Please indicate which of these names you think you would use for these objects by pressing the "c" key for cup, the "m" key for mug, or the "g" key for glass.

You can work at your own pace for this section of the experiment.

Please hit the spacebar to begin the experiment.

Training Phase

Now, please imagine that you have moved to Russia and need to be able to speak the Russian language. In order to help you with this, you will see the same objects that you previously named in English. First, you will see each object labeled with its Russian name and then you will try to produce the names on your own.

There are only five names that you need to learn: kruzhka, stakan, chashka, fuzher, and riumka. Your goal is to learn which objects are called by each name.

Press the spacebar for more instructions.

Training Phase Screen 2

You will now see each object presented with its Russian name for three seconds. Just try to focus on learning which objects are called by each name. After you see each

object once, you will be asked to name the objects as kruzhka, stakan, chashka, fuzher, or riumka. As you see each picture, please say the Russian word out loud.

Once you name these objects, you will be given instructions about another task to complete. You will repeat this cycle (seeing the objects, producing your labels, and completing the additional task) until you label at least 90% of the objects correctly in Russian.

[Please note that the Russian words are not exactly equivalent to any English words.]

Press the spacebar to continue.

Training Phase Screen 3

You will now see the same objects you have just seen. Please indicate what you believe the correct Russian name for each object is by pressing the "k" key for kruzhka, the "s" key for stakan, the "c" key for chashka, the "f" key for fuzher, or the "r" key for riumka. As you make your choice, please say the word out loud.

You can move at your own pace for this section of the experiment.

Press the spacebar to continue.

Training Phase Screen 4 (Mixed Languages Conditions): English Naming

Now, you will see each of the objects you just labeled again. This time, you will label each of the objects in English as either "cup," "mug," or "glass" by selecting "c," "m," or "g" key. As you make your choice, please say the word out loud.

Press the spacebar to continue.

Training Phase Screen 4 (Russian Only Conditions): Filler Task

Now, you will see the same pictures you've just seen. Using the ruler below the object as a scale, your job is to determine, as accurately as possible within a time limit of 2 seconds, if the object is more or less than four inches tall. If the object is four inches or taller, press the orange key; if it is shorter than four inches, press the green key.

Press the spacebar to continue.

Training Phase: Criterion Not Met

You will now see the objects labeled with the correct Russian names again. Keep learning the names and you will be tested again.

Press the spacebar to continue.

Training Phase: Criterion Met

Congratulations! You've now reached the criterion to move on to the next phase of the experiment.

Press the spacebar to continue.

Test Phase

You will now see more pictures of objects. Some of these will be pictures you've already been tested on in on Russian, while some will be new. Please indicate, by pressing the "k," "s," "c," "f," or "r" keys, if you think the object would be called a kruzhka, a stakan, a chashka, a fuzher, or a riumka. As you make your choice, please say the word out loud.

Press the spacebar to continue.

Posttest

In this final phase of the experiment, you will see each of the objects again. This time, please indicate what you would call each object in English by pressing "c" for cup, "m" for mug, or "g" for glass.

Press the spacebar to continue.

Final Screen

Thank you for participating. Please let the experimenter know that you are finished with the experiment.

Rachel Jobe

Department of Psychology, Lehigh University
17 Memorial Drive East
Bethlehem, PA 18015
Cell phone: (434) 942-9441
Email: rlj311@lehigh.edu
Email: rachellynnjobe@gmail.com

Education

September 2013: **Master of Science in Psychology, Lehigh University, Bethlehem, PA**
Thesis: Bilingual Lexical Interactions: Inevitable or Malleable?
Advisor: Barbara C. Malt, Ph.D.

May 2010: **Bachelor of Science in Psychology and English, University of Mary Washington, Fredericksburg, VA**
Magna Cum Laude

Academic Employment

June 2012-June 2013: Research assistant for Professor Barbara Malt: collect experimental data, input data, data analysis, perform other duties as asked, Bethlehem, PA

January-June 2012: Teaching assistant for Professor Kiri Lee (half-time), Introduction to Linguistics: graded exams and assignments, met with students regarding questions on material, scheduling, etc., Bethlehem, PA

January-June 2012: Teaching assistant for Professor Aladdin Yaqub (half-time), Introduction to Cognitive Science: graded exams and assignments, met with students regarding questions on material, scheduling, etc., Bethlehem, PA

August-December 2011: Teaching assistant for Professor Jessecae Marsh, Introduction to Cognitive Psychology: graded exams and assignments, met with students regarding questions on material, scheduling, etc., Bethlehem, PA

January-May 2010: Computer lab aide, University of Mary Washington: enforced computer lab code of conduct for students, updated and maintained computer software and supplies as needed, provided basic Microsoft office troubleshooting assistance for other students, Fredericksburg, VA

Professional Memberships

Virginia Psychological Association, student member

Women in Cognitive Science