Statistical Word Learning and Object Categorization: A Cross-Linguistic Study in English and Mandarin

Chi-hsin Chen, Chen Yu ({chen75, chenyu}@indiana.edu)
Cognitive Science Program, Indiana University, Bloomington, IN 47405 USA
Chih-Yi Wu, Hintat Cheung
Graduate Institute of Linguistics, National Taiwan University, Taipei, 10617 Taiwan

Abstract

This study investigates whether, in addition to learning the first-order regularities between individual words and objects under a cross-situational learning context, adult learners are sensitive to the second-order correlations between the phonological features of labels and the visuo-perceptual features of objects and whether they can use such features as a cue in categorizing novel objects. Two experiments were designed and conducted to examine whether native speakers of English and Mandarin Chinese performed differently when they encountered different types of training data that either reflected or were inconsistent with the linguistic features in their native language. We found that when the training stimuli reflected the linguistic structures of their native language, English speakers were able to use the phonological features of labels as a cue in object categorization just as their Mandarin-speaking counterparts did. Moreover, our results also suggest bi-directional real-time interactions between learning the first-order word-referent mappings and the higher-order mappings between phonological features in labels and perceptual features in visual objects.

Keywords: word learning; statistical learning, object categorization

Introduction

The ability to categorize enables us to efficiently deal with the enormous amount of information provided by the environment. Previous studies have shown that object categorization and early word learning bootstrap each other (e.g., Yoshida & Smith, 2005; Yu, 2005). The cues that can be used in word learning and categorization includes visuo-perceptual cues (e.g., Soja *et al*., 1991; Colunga & Smith, 2003), social-pragmatic cues (e.g., Baldwin *et al*. 1996; Tomasello & Akhtar 1995), and linguistic cues (e.g., Colunga & Smith, 2003; Yu, 2005). Among these cues, linguistic labels play an important role in guiding categorization (e.g., Colunga & Smith, 2003; Plunkett *et al*., 2008; Sloutsky *et al*., 2001, Waxman & Braun, 2005).

Colunga and Smith (2003) proposed that there are three steps in the process of early categorization. In the first step, a child learns the association between labels and individual instances without realizing that some referents might share the same name. For example, a child may use the word "ball" to refer to both a rubber ball and a basketball and the word "cup" for a big white cup and a smaller red one. However, he does not realize that some of the objects share the same label. In the second phase, the child notices that each linguistic label refers to objects that have something in common, in this case, their shapes. For example, the objects with the name "ball" are all round and the things with the name "cup" are all cup-shaped. After that, he may generalize that every solid object has a certain shape. For example, from the patterns that he learned, he may then predict that forks are all fork-shaped.

In previous research on object categorization (e.g., Sloutsky *et al*., 2001), the subjects were often taught the names for novel objects and then asked whether two objects belonged to the same category or different categories. The testing objects might either share the same label or have different names. The subjects were expected to group together the objects with the same names. However, giving identical labels to the same-category objects corresponds to the condition of giving a basic or super-ordinate label to its members. It is unclear whether learners are able to use more fine-grained cues, such as the phonological features of the labels, in object categorization.

Unlike derivational and inflectional morphology in Indo-European languages, compounding plays a major role in Chinese word formation (Li and Thompson, 1981). In modern Mandarin, names of objects are often disyllabic compounds. Items belonging to a category often contain the name of that category, with the first syllable as the modifier and the second syllable as the head of a compound. In other words, items belonging to the same category often share a final syllable signifying the label of the category. For example, in English, the words "pork", "beef", and "chicken" do not share any constituent in their forms. In Mandarin, however, they share a final syllable, *rou4* (meat). The second syllables of the words *zhu1rou4* (pork), *niu2rou4* (beef), and *ji1rou4* (chicken) indicate that they belong to the category of "meat". Given that words ending with the same syllable are likely to belong to the same category, one strategy is to form categories based on the phonological forms of the labels. For this reason, Mandarin speakers might be more likely to use the linguistic features of labels in object categorization than their English-speaking counterparts.

To examine participants' sensitivity to the phonological features of labels in word learning process, we adopted the cross-situational word learning paradigm, a statistical learning paradigm proposed by Yu & Smith (2007). In their study, the authors presented adults with a set of learning trials, with multiple words and multiple pictures within each trial. To successfully solve the word-referent mapping problem, the subjects needed to track the co-occurring statistics between words and pictures across trials. They found that adults could rapidly learn the word-referent mappings in highly ambiguous learning situations (see also Smith & Yu, 2008, for an infant version). This experimental

paradigm provides an ideal learning context to examine whether, in addition to learning the first-order correlations between individual labels and objects, participants are also able to extract higher-order regularities between the linguistic features of labels and certain shared visual features of category members.

The goal of this study was to investigate whether adults were sensitive to the phonological features of labels under statistical learning conditions in which the final syllables (Experiment 1) or the first syllables (Experiment 2) of labels were correlated with certain visual features of the objects. Furthermore, we examined whether native speakers of English and Mandarin performed differently in the learning task.

Experiment 1

In this experiment, we investigated whether adults were able to use the final syllables of labels in object categorization under a cross-situational learning context (Yu & Smith, 2007), in which the subjects had to keep track of word-object co-occurrences across trials in order to find the correct word-referent mappings. In addition, we tested whether native speakers of English and Mandarin Chinese performed differently when they encountered different types of linguistic data, the English-like pseudo-words and the Mandarin-like pseudo-words. The reason for this comparison is that in Mandarin the high frequency and the head-final feature of compound nouns (i.e., the final component of a compound noun determines its semantic category) allow the speakers to identify category membership from the final syllable of a word, while in English, the final syllables of object labels is not a reliable cue to category membership. Therefore, speakers of these two languages might perform differently in the categorization task. More specifically, Mandarin-like pseudo-words might cue Mandarin speakers about the compound noun patterns in their native language and this might facilitate learning. In contrast, English speakers might not benefit from it.

Participants

Participants were 48 English-speaking undergraduates at Indiana University who received course credit for volunteering and 46 native speakers of Mandarin recruited from National Taiwan University, Taiwan.

Design and Stimuli

We used a 2*2 between-subject design with the native language of the participants (English vs. Mandarin) and the types of linguistic stimuli (English-like pseudo-words vs. Mandarin pseudo-words) as between-subject factors.

The participants were trained under a cross-situational learning paradigm in which they were presented with 4 novel objects and 4 novel labels in each trial. However, they were not provided with the information about which object was matched to which label. In order to find the correct mappings, they had to keep track of the co-occurrence regularities between objects and labels across different trials. There were a total of 18 object and label mappings to learn. Over the training procedure, there were 12 repetitions per instance, yielding a total of 54 trials (18 pairs x 12 repetitions / 4 pairs per trial).

The objects used in the training were divided into three different categories, with six objects in each category. Members in a category had an attached part that looked similar. Take the objects in Figure 1 as an example, items in this category all had an attached spiral part that spreads at the end.

Objects in a category had labels with identical final syllables. For example, objects in Figure 1 and other members in that same category all had a name that ended with the syllable *–ti* in the English stimuli (and the syllable *–ti2* in the Mandarin stimuli).

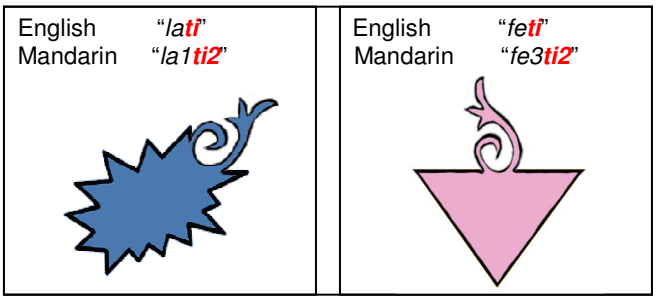

Figure 1 Sample objects and labels used in the study

To avoid the problem that some final syllables might be more attractive than others, we used two sets of stimuli in each language. Set A and Set B were English-like pseudo-words while Set C and Set D were Mandarin-like pseudo-words. The stimuli in Set A and Set B were matched in their first syllables while different final syllables were used. The same design was applied to the stimuli in Set C and Set D. More importantly, under the constraint of using legal syllables (i.e., following the phonological rules of either English or Mandarin), the stimuli in Set C were closely matched to their counterparts in Set A in their phonetic features while the stimuli in Set D were matched to those in Set B. Thus, Set C and Set D could be viewed as the Mandarin version of Set A and Set B with additional tonal information, though the tonal information did not serve any contrastive function in the current study[1]. The auditory input was based on natural human voices generated by AT&T Natural Voices (for English-like pseudo-words) and InterPhonic by iFLYTEK (for Mandarin pseudo-words) with modification of the speed, pitch, loudness, and lengths of individual syllables.

Procedure

There were two phases in this experiment, the Training phase and the Testing phase. In the Training phase, the

[1] We also tried to have the tones of the Mandarin-like pseudo-words distributed evenly. Among these six categories of labels (3 categories * 2 sets), one category had the final syllable in Tone 1, two categories in Tone 2, another one in Tone 3, and two in Tone 4. As to the first syllables of the labels, in each linguistic set, four words had the first syllable in Tone 1 and Tone 3, respectively, and five words in Tone 2 and Tone 4, respectively. To avoid the tone *sandhi* problem, none of the Mandarin pseudo-words had the Tone 3+ Tone 3 combination.

participants were presented with picture slides, each containing 4 objects, and auditory labels. However, the labels in each trial were presented in a random order. Therefore, the learners could not tell the label of an object within one single trial. Nevertheless, the probability of an object given its label was always 1.0. As long as the learners keep track of the co-occurrences between objects and labels, they should be able to find out the correct object-label mappings. Moreover, they might notice a higher-level statistical regularity between object-parts and the final syllables of labels. The category-syllable association might in turn facilitate the word and meaning mapping.

After training, there were two tasks in the Testing session, the word-object Mapping task and the Generalization task. The Mapping task tested how well the participants learned the names of the training objects. The participants were presented with one word at a time and were instructed to select its referent from 4 alternatives. There were 18 trials in the Mapping task.

In the Generalization task, the participants heard one novel word on each trial and were asked to select its referent from three alternatives, each containing the object-part that corresponded to the particular feature of one category. For example, a participant might be presented with a novel label "*joti*," which rhymed with the labels of the training objects with a spiral part, like those in Figure 1, and asked to select its referent from three novel objects, one of which had the spiral part. If the participant chose that object, this indicates that he might have acquired the relationships between the final syllables of labels and the corresponding perceptual features of the categories. There were 9 trials in the Generalization task (3 for each category).

Results

As illustrated on the left-hand side of Figure 2, participants learned more word-referent mappings than expected by chance. Even the group with the lowest accuracy, English speakers tested with Mandarin stimuli, performed significantly above chance ($t(23) = 2.41$, $p = .024$). We performed a 2*2 ANOVA with the native language of the participants and the language of the stimuli as between-subject variables. We found a significant interaction ($F(1, 90) = 5.204$, $p = .025$). Participants performed better when they were trained with the pseudo-words that followed the phonological features of their own native language. On the other hand, we did not find any main effect of the participants' native language ($F(1, 90) = .026$, *n.s.*) or of the language of the stimuli ($F(1, 90) = 2.384$, *n.s.*). Post-hoc analyses using Bonferroni correction show that English speakers trained with English-like pseudo-words (i.e., the E-E group) performed significantly better than their English-speaking counterparts trained with Mandarin-like stimuli (the E-M group, $p = .045$). Even though Mandarin speakers trained with English stimuli (the M-E group) did not learn as well as those trained with Mandarin stimuli (the M-M group), the difference did not reach statistical significance.

In the Generalization task, only Mandarin speakers trained with Mandarin-like pseudo-words performed significantly above chance ($t(21) = 3.341$, $p = .003$). Nevertheless, compared to chance level, the scores of the English speakers trained with the English stimuli and the Mandarin speakers trained with the English stimuli both yielded near-significant differences (E-E: $p = .065$, M-M: $p = .103$).

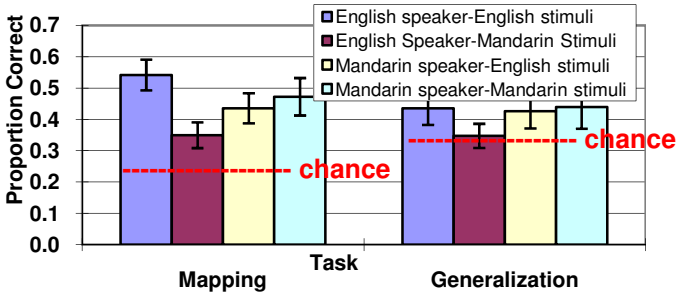

Figure 2: Proportion of accurate response in Experiment 1

To the question of whether the numbers of words learned correlate with how likely the participants use final syllables as a cue in generalization, we found significant correlations in all subject-stimuli conditions (E-E: $r = .558$; E-M: $r = .569$; M-E: $r = .416$; M-M: $r = .876$, $ps < .05$).

Discussion

Like previous studies on cross-situational learning, participants in this study were able to learn the first-order regularities between labels and objects (i.e., individual word-referent mappings), regardless of the language in which they were trained. Even though the labels in each category only differed in their first syllables and had the second syllables in common, the participants were still able to perform above chance after 10 minutes of training. In addition to learning the mappings between individual labels and pictures, Mandarin speakers trained with Mandarin-like pseudo-words, were also able to learn the second-order correlations, in which the final syllables of labels correlated with certain visual features of objects. In order to find the second-order correlations, the learners had to figure out several correct word-referent mappings before they could see the higher-order patterns.

In general, participants learned better with the stimuli that followed the phonological rules of their own native language. This suggests that the familiarity of the phonological structures of the pseudo-words used in the training might influence word learning. Those trained with a set of stimuli that followed the phonological rules of a foreign language might need more time to get familiarized with the stimuli before they could start learning the mappings between objects and their names.

As to the performance of generalization, only Mandarin speakers trained with Mandarin-like pseudo-words (i.e., the M-M group) performed significantly above chance. Meanwhile, the M-M group was not the group that performed the best in the Mapping task. Thus, the native language of the participants seemed to be a key factor that influenced the participants' use of final syllables in the Generalization task. In Mandarin, as compared to English, the final syllable of an object label is a more reliable cue to

category membership. Mandarin speakers might therefore pay more attention to the forms of labels than their English counterparts did. This might facilitate their use of this feature in categorizing novel objects. However, it is not to say that the native language of the participants was the sole factor that influenced their use of such cue. The number of words learned seemed to be another factor that influenced their generalization. From the correlation analyses, the numbers of correct answers the participants provided in the Mapping task were positively correlated with the numbers of correct responses that they provided in the Generalization task.

It is worth-mentioning that from the results of this experiment, we were not able to tell whether the participants took the final syllables as category markers or whether the participants only associated those syllables with object-parts. If final syllables were viewed as category markers, they might then directly guide categorization. On the other hand, if the learners only formed associations between shared syllables and object-parts, this association might influence categorization indirectly, in that shared sounds high-lighted the similarities among certain objects. These shared phonological features might then help the participants pay attention to the right visual properties of objects in the Generalization task.

As we predicted, English speakers did not use final syllables as a cue in the Generalization task and this might result from the fact that final syllables of labels is not a reliable cue for judging category membership in English. But what would they do if the training data reflect the linguistic characteristics in their native language? Even though in English, compounding is not as frequent as it is in Mandarin, both languages have similar noun phrase structures in that an adjective in a noun phrase is placed before the head noun (e.g., "*red car").* Moreover, in phrases like *red car, red hat,* and *red book*, the first syllables (i.e., the first words in this case) indicate that these objects have something in common. Similar structures can be found in Mandarin. For example, the first syllables in the phrases, *hong2-che1* (red car), *hong2-mao4* (red hat), and *hong2-shu1* (red book), indicate that these things are same in colors. Therefore, if, instead of using final syllables, we put the category markers in the first syllables to simulate the noun phrase structures in these languages, would English speakers still perform differently from their Mandarin-speaking counterparts?

Experiment 2

Experiment 1 revealed that in general English speakers did not take final syllables of labels as a cue in categorizing novel objects. One possible explanation is that the phonological forms of labels is not a reliable cue for categorization in their language environment and this leads to their insensitivity to the relationships between the final syllables of labels and the visual features of objects in the training data. In the second experiment, we designed a learning condition that that simulated a linguistic structure similar in both languages, in which the category markers were placed in the first syllables of labels. In this experiment, members of a category had a same object-part in their appearances and the same first syllable in their labels.

Participants

Participants were 48 English-speaking adults and 41 Mandarin-speaking adults. None of them participated in Experiment 1.

Design and Stimuli

The visual stimuli in Experiment 2 were identical to those in Experiment 1. New sets of linguistic stimuli were used in which the category markers were placed in the first syllables of labels. For example, objects in one category might all had their labels beginning with the syllable *jo-* while the members of another category might have labels starting with the syllable *che-*. Again, the English-like pseudo-words and the Mandarin-like pseudo-words were closely matched in their phonetic features. In addition, there were two sets of stimuli in each language to avoid the problem that some syllables might be more attractive than others.

Procedure

The procedure was identical to that in Experiment 1. The participants were trained under a cross-situational learning paradigm using a 4*4 presentation, with a total of 18 word-picture pairs to learn. After the training, the participants were first tested with the mappings between the trained objects and labels. Following the Mapping task, they were then presented with one novel label at a time and were asked to choose its referent from 3 novel objects, each containing the object-part of a particular category.

Results

Again, the participants performed significantly above chance in the Mapping task (proportion correct: E-E: .60; E-M: .49; M-E: .56; M-M: .53). We conducted a 2*2 ANOVA with the native language of the participants and the language of the stimuli as between-subject variables. We did not find any main effect or interaction.

In the Generalization task, the Mandarin speakers performed significantly above chance, regardless of their training data (M-E: $t(20) = 4.480$ $p < .001$, M-M: $t(19) = 4.703$, $p < .001$). The English speakers trained with the English stimuli also performed significantly above chance ($t(23) = 5.839$, $p < 001$). Compared to chance, the difference scores of the E-M group (English speakers trained with Mandarin stimuli) were near-significant ($t(23) = 1.869$, $p = .074$). A two-way ANOVA using the native language of the participants and the language of the stimuli as between-subject variables indicated that there was no significant difference among the participants in different conditions in the Generalization task.

We also found strong correlations between the numbers of words learned and the numbers of responses using the first syllables of labels as a cue in the Generalization task. This suggests that the more one learned, the more likely the person was to use the first syllable as a cue in categorizing novel objects. (English speakers: $r = .667$, $p < .001$; Mandarin speakers: $r = .631$, $p < .001$).

Comparing the performance of the participants in Experiment 1 and Experiment 2 in the Mapping task, we used a 3-way ANOVA with experiment (Experiment 1 vs. Experiment 2), native language of the participants (English vs. Mandarin), and the language of the stimuli (English vs. Mandarin) as between-subject variables. We found a significant main effect of experiment ($F(1,175) = 5.636$, $p = .019$) and significant interaction between the native language of the participants and the language of the stimuli ($F(1, 175) = 4.061$, $p = .045$).As shown in Figure 3, participants in Experiment 2 learned better than those in Experiment 1, which was true for participants in all conditions. Moreover, learners trained with their own native language tended to perform better than those trained with a foreign language.

As to the participants' performance in the Generalization task, we conducted another 3-way ANOVA using experiment (Experiment 1 vs. Experiment 2), native language of the participants (English vs. Mandarin) and the language of the stimuli (English vs. Mandarin) as between-subject variables. Participants in Experiment 2 were more likely to use the phonological features of labels in generalization than their counterparts in Experiment 1 ($F(1, 175) = 4.216$, $p = .042$). However, we did not find any significant main effect of the participants' native language or the language of the stimuli, nor did we find any interaction. As Figure 4 illustrates, participants in Experiment 2 performed better than those in Experiment 1.

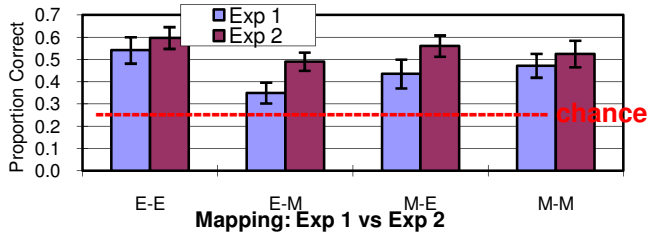

Figure 3 Comparison of participants' performance in the Mapping task

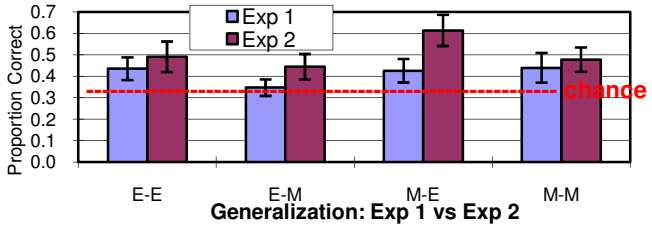

Figure 4 Comparison of participants' performance in the Generalization task

Discussion

The results from Experiment 2 showed that participants were able to learn the mappings between labels and pictures under a highly ambiguous learning context. Moreover, both English-speakers, especially those trained with English-like pseudo-words, and Mandarin-speakers were able to use the first syllables of labels as a cue in object categorization. This indicated that changing the stimuli to simulate the linguistic features in the participants' environment might influence the participants' use of such features in various tasks. Since both English and Mandarin have adjectives placed before the head nouns in noun phrases, the similarity in the noun phrase structures might lead to similar performance of speakers of these two languages.

Similar to the pattern seen in Experiment 1, the number of words learned played an important role in the participants' performance in the Generalization task. The more a person learned, the more likely he/she was to use the phonological features of labels in categorizing novel objects. This makes sense since learners need a certain number of correct word-referent pairs to infer the relationship between the phonological features of labels and the visual features of objects. Moreover, discovering the relationship between these cues might also help word learning. Taken together, there might be a bi-directional feedback loop between the mapping task and the categorization task.

Participants in Experiment 2 performed significantly better in both Mapping and Generalization tasks than their counterparts in Experiment 1 did. There are two possible explanations to this pattern and both predict that the gain from (partial) knowledge about categories help word-referent mappings. The first one is that the adjective-like first syllables facilitate the learners' attention to the visual features of objects to a greater extent than the final syllables did. On the other hand, it has been suggested that first syllables are more salient in speech processing (e.g., Jusczyk *et al.*, 1999). Therefore, the second possibility is that it might be easier to learn category membership when the category markers lie in the first syllables. From the current study, we are not able to determine which of the explanations better accounts for the data. It is also possible that the effects found in the study resulted from both processes.

General Discussion

This study shows that in addition to learning the first-order statistical regularities under a highly ambiguous learning context, adult learners were also able to infer the second-order correlations between the phonological features of labels and the visual features of objects. That is, in addition to learning the individual mappings between labels and objects, the subject were also able to detect that members of a category had a shared phonological feature in their labels and a same object-part. Furthermore, the more words a participant learned, the more likely he/she was to use the phonological features as a cue in the Generalization task. In addition to pure co-occurrence statistics, prior linguistic experiences also seemed to have an impact on one's sensitivity to different aspects of statistical regularities in the same training data. The fact that the participants in Experiment 2 performed better in both the Mapping and Generalization tasks indicates that word learning and object categorization might bootstrap each other. That is, finding the mappings between some word-object pairs helps the learners notice that objects with labels that share the same first or second syllables have something in common in their visual appearances. This just acquired knowledge might in

turn help their learning of other words, in which they could use this information to rule out certain candidates or to decide what a potential referent of a word should look like or what a potential label of an object should sound like.

The results from the current study are consistent with the proposal of the Attentional Learning Account (e.g., Colunga & Smith, 2008). According to the Attentional Learning Account, statistical regularities in the environment direct learners' attention to certain predictive cues. Furthermore, this weighted attention then influences or guides future learning. Yet, the first-order regularity in the input is not the only information that the learners attend to. They also pay attention to higher-order correlations in *the whole system,* as it is evident from the performance differences between the English- and Mandarin-speakers in the current study. However, it does not mean that learning is determined by what is in the environment. Momentary information processed in real-time learning also causes "contextually cued dynamic shifts in attention" (Colunga & Smith, 2008, p. 197). As can be seen from the current study, the positive correlations between the numbers of words learned and the numbers of responses using the phonological features in generalization indicate that the statistical regularities in the experimental setting might also induce the English speakers' use of final syllables in categorization, even though it is not a reliable cue from their language experiences.

As the statistical regularities were the same for all sets of stimuli in both experiments and one could only distinguish the members of one category solely from one syllable in the labels (i.e., first syllable in Experiment 1 and second syllable in Experiment 2), participants in different conditions did not perform the same. The design of the current study, an extension of the standard cross-situational paradigm to different linguistic stimuli and with speakers from different linguistic environments, also shows that statistical regularity is not the only factor that influences learning. The factors such as the familiarity of the phonological features of the pseudo-words and how the training data reflect the regularities in participants' language experiences might all influence how one learns those regularities. Even though the English-like and the Mandarin-like stimuli were closely matched in phonetic features, the Mandarin-like stimuli actually provides additional tonal information that could be used in distinguishing each individual labels. In spite of the fact that the tonal information did not serve any contrastive function in the study and that they actually provided additional information, the unfamiliarity to the phonological features of Mandarin stimuli might have a negative effect on the English speakers' learning of individual words and this in turn affects their performance in the Generalization task. Also, we are not able to determine the factor that leads to significant increase in correct responses in Experiment 2 Further investigation is needed to address these issues.

Acknowledgments

This research was supported by National Institute of Health Grant R01HD056029, National Science Foundation Grant BCS 0544995, and National Science Council Grant NSC96-2411-H-002-068-MY3. The first author was also supported by a scholarship from the National Science Council of Taiwan (NSC-095-SAF-I-564-077-TMS).

References

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M. & Tidball, G. (1996). Infant's reliance on a *social* criterion for establishing word–object relations. *Child Development, 67*, 3135-3153.

Colunga, E. and Smith, L. B. (2003). The emergence of abstract ideas: Evidence from networks and babies. *Philosophical Transactions, 358*, 1205-1214.

Colunga, E. and Smith, L. B. (2008). Knowledge embedded in process: the self-organization of skilled noun learning. *Developmental Science, 11(2)*, 195-203.

Jusczyk, P. W., Houston, D., & Newsome, M. (1999). The beginning of word segmentation in English-speaking infants. *Cognitive Psychology, 39*, 159-207.

Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.

Plunkett, K., Hu, J.-F., Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. Cognition, 106, 665-681.

Sloutsky, V. M., Lo, Y-F., & Fisher, A. V. (2001). How much does a shared name make things similar? Linguistic labels, similarity, and the development of inductive inference. *Child Development, 72(6)*, 1695-1709.

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., and Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13* (1), 13-19.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition, 106, 1558-1568.

Soja, N. N., Carey, S., and Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition, 38*, 179-211.

Tomasello, M. & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development,* 10, 201-224.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. Cognition, 95, B59-B68.

Yoshida, H. and Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science, 16 (2)*, 90-95.

Yu, C. (2005). The emergence of links between lexical acquisition and object categorization: a computational study. *Connection science, 17*, 381-397.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414-420.