Real-World Processing of Word-Referent Pairs in Adults

Laura Cacciamani

Carnegie Mellon University

Abstract

The process of pairing words with their referents in the real world is often ambiguous; the same word may refer to different objects, the same object may be referred to by different words, and there may just be insufficient environmental cues to assist us in the process. Thus, how is it that adults are able to go about this, particularly when the word-referent pairings are not constant? Three different experiments were completed which tested the ability of adults to pair up novel words and shapes after being exposed to a learning phase where the words referred to the correct object either 50% or 75% of the time, in order to simulate the ambiguity of real-world referents and determine to what extent that ambiguity impairs their ability to correctly match them. The results showed that although all adults were able to determine word boundaries within the stream of speech, the difference in ability between the two conditions on pairing the words with their referents was not significantly different.

Real-World processing of Word-Referent Pairs in Adults

In our daily lives, we are constantly matching up words with their referents in order to better understand our world. This word-referent pairing, however, is not a simple task; sometimes one word may refer to multiple objects (e.g. "ball" might mean baseball or soccer ball), an object might be able to be defined by multiple words (e.g. the device you use to turn on a television might be called a remote control or a clicker), and sometimes the object in reference is not even present (e.g. it is in the next room). These are just a few of the difficulties that arise in matching up words with objects as we go about our lives.

Because the process of word-referent pairing is so difficult and often ambiguous, how is it that we do it so easily on a daily basis without even a second thought? Developmental psychologists have looked into this question with the idea that understanding how the process occurs in infants may give us insight into how we go about matching up words with referents so easily as adults. In the process of language learning in infants, it is first required that words are able to be parsed from a stream of speech. Multiple theories have been thrown around about how this process works, and whether or not the parsing of words from sentences is based on experience. Crain (1991) suggested that language tasks such as determining word boundaries are not experience-driven due to the fact that the language that infants are exposed to is often incomplete, grammatically incorrect, or skewed from what normal speech sounds like due to the infant-directed nature of talking to children. Others, however, have argued that experience may be more pertinent in language learning than previously assumed. Saffran, Aslin, and Newport (1996) argued for the idea that word boundaries are defined by being repeatedly exposed to speech such that statistical regularities arise. They found that upon giving infants a stream of nonsense speech comprised of novel words, 8-month-olds were able to differentiate between words and part-words during the test trials by utilizing the transitional probabilities that arose during habituation; syllables that comprised a word had a transitional probability of 1.0, whereas the probability between words was 0.33 (Saffran et. al., 1996), which infants were able to pick up on as determined by their looking times. Thus, it can be assumed from this research that infants first learn to parse words in a sentence by utilizing probabilities, which is an experience-driven process.

Now that it has been shown that infants are able to determine word boundaries, we can extend this further by asking how infants use these words to refer to an object. One study relating to determining word-object pairs was conducted by Gogate and Bahrick (1998). In this study, 7-month-old infants were presented with objects on the screen, and each object had an arbitrary vowel that was spoken in tandem with it. The conditions were defined by either a moving object or a static object, with the vowels being either synchronous or asynchronous with the object. The results indicated that infants were best able to detect word-object pairings during the moving-synchronous condition, which is the condition that can be defined by the most redundancy. Thus, it can be implied that intermodal redundancy helps to make word-object relationships more salient and allows infants of this age to be capable of knowing what word (or vowel, in this case) refers to which object. Other studies have had similar findings, stating that presentation of stimuli using an intermodal approach is pertinent (Kuhl and Meltzoff, 1984; Bahrick, 1994; MacKain et. al. 1983) and synchrony between the auditory and visual stimuli enables the pairings to be made (Lyons-Ruth, 1977).

The above research has shown that infants are capable of determining word boundaries and finding word-object relations, but how does this apply to adults? Studies have shown that like infants, adults have a better performance rate when bimodal rather than unimodal cues are used and when redundancy is present as well, particularly in word-object relation tasks (Miller, 1982), but applicable to spatial localization tasks also (Neil et. al., 2006). Another, more recent adult study conducted by Yu and Smith (2007) looked at the ability of adults to pair up multiple spoken words with multiple pictures over various trials. Participants were not given specific instructions to learn the word-objects pairs, but rather were just presented with both stimuli simultaneously. Despite this fact, results showed that participants were still able to match up the word-referent pairs accurately. It was also found that participants learned the word-referent pairs better when there are more pairs to be learned, indicating that the size of the data set is more important than the number of repetitions of the word-referent pairs, which is an important factor to keep in mind.

All of the previously discussed studies have shown that both infants and adults are successful at matching up words with their referents when the two are matched up 100% of the time. However, in the real world, words often do not refer to the same object all the time, which is an important factor to take in consideration when trying to generalize. Thus, the current study asks the question of just how good adults are at matching up words with their referents when they are not paired together 100% of the time. Two different conditions (50% matched and 75% matched) were used to determine if a threshold exists where individuals are no longer able to match the words with their correct referents past a certain point. Since it has already been shown that individuals can match up words and referents when they are paired all the time, it is hypothesized in this study that participants will be significantly more accurate at pairing in the 75% condition compared to the 50% condition, suggesting that the threshold for word-referent pairing exists between the two percentages.

Experiment 1

Method

Participants.

Participants consisted of 55 Carnegie Mellon University undergraduates (29 males, 26 females). All participants were recruited through the psychology experiment system and were given course credit for participating.

Stimuli and Design.

Stimuli during the learning phase consisted of 4 different objects—a white cross, a green diamond, a purple heart, and a yellow hexagon—which appeared in a looming manner on the computer screen. Each object had a 3-syllable nonsense word that was paired with it—*padoti*, *bidaku*, *tupiro*, and *golabu*, respectively—which was spoken in a monotone computerized female voice over headphones. Each object appeared on the screen 18 times, making for a total of 72 items in the learning phase that appeared in random but consecutive order. The frequency of correct versus incorrect pairing of the words with the objects, however, differed depending on condition. In condition A, 50% of the objects (36 objects) were paired with the correct nonsense word, and the other 50% were paired with one of the other words (12 times per each of the other 3 words, totaling 36). In condition B, 75% of the objects (54 objects) were paired with the correct nonsense word, and the other 25% were paired with one of the other words (6 times per each of the other 3 words, totaling 18).

Procedure.

Participants were sat down at a desk in front of a 12" portable DVD player, which is what the experiment DVD was played on. They were given instructions to simply watch the screen and listen over the headphones during the 72-item learning phase; no specific instructions were given to look for word/object pairings. Following the learning phase, participants completed two 16-item questionnaires: the first questionnaire, the word-segmentation task, presented the participant with one of the words (e.g. *bidaku*) and one part-word (e.g. *ku-gola*) and asked them to determine which one was a word they had heard during the learning phase by circling the correct item on the response sheet. Each word was presented 4 times, totaling 16 questions. The second questionnaire, the word-object pairing task, gave the participant a word (e.g. *bidaku*) and 4 choices of what object goes with that word (e.g. white cross, green diamond, purple heart, or yellow hexagon), and the participant was asked to circle his or her answer on the response sheet. This was done 4 times for each word, totaling 16 questions.

Results

The results of each participant's questionnaires were entered into an excel spreadsheet for further analysis, and the two questionnaires were analyzed separately. A one-sample T-test was run on the number of correct answers on the word-segmentation task. Across both conditions, the average number of correct answers out of 16 was 10.68 with a standard deviation of 3.02. This is significantly greater than chance, $t(43)=5.880$, $p<.001$. These results imply that no matter whether the words were correctly paired 50% or 75% of the time, all participants were able to parse the words from the stream of speech. If we break down the scores and look at the two conditions separately with an independent-samples T-test, we see means of 10.27 and 11.09 and standard deviations of 3.25 and 2.79 in conditions A and B, respectively. The differences between these two averages, however, is not significant, $t(42)=-.895$, $p=.376$. This implies that the frequency of correct word/object pairings, whether it's 50% or 75%, has no effect on participants' ability to parse words from the stream of speech.

If we look at how both conditions performed on the word-object pairing task with a one-sample T-test, we see that the average number of correct answers out of 16 is 4.52 with a standard deviation of 2.96. This number, however, is not significantly greater than chance, $t(43)=1.171$, $p=.248$, which would be 4 out of 16. This result suggests that participants in both conditions had difficulty in pairing up the word with the correct object. To break it down further, an independent-samples T-test between conditions on this data shows that the means are 3.95 and 5.09 with standard deviations of 2.44 and 3.36 for condition A and B, respectively. These means are not significantly different from each other, $t(42)=-1.282$, $p=.207$, which implies that the condition that the participant was in did not have an effect on their ability to pair up the words with the correct objects.

Because few significant results were found through this experiment despite the large number of participants, we decided to adjust the learning phase in hopes of creating more significant differences between the two conditions. Many participants in this first experiment stated that they were unable to pick up on the trend (what words were paired with what objects) until the very end of the learning phase; thus, in experiment 2, we tried doubling the length of the learning phase.

Experiment 2

In order to attempt to create a significant difference between conditions A and B, a second experiment was run with 24 participants, where the only difference from Experiment 1 was that the stimuli in the learning phase was doubled. Thus, the participants were presented with 144 objects rather than 72, but the questionnaires to follow remained the same. The same results were found, with no significant difference between conditions A and B.

Because doubling the length of the learning phase proved unhelpful, we decided that the next step would be to adjust the actual stimuli in hopes of achieving more significant results.

Experiment 3

Method

Participants.

Participants were 34 Carnegie Mellon University undergraduates (18 males and 16 females). All participants were recruited through the psychology experiment system and were given course credit for participating.

Stimuli and Design.

The stimuli for this experiment were identical to Experiment 1, utilizing the 4 looming shapes with the same 3-syllable words paired up with them in a 72-item learning phase. Likewise, the objects and words in condition A were paired up correctly 50% of the time, and the objects and words in condition B were paired up correctly 75% of the time. The only difference in this experiment was that for the other percentage of the time (50% and 25% for condition A and B, respectfully), a plain black screen was shown on the screen rather than one of the objects that contradicted the word being said.

Procedure.

The procedure for this experiment was identical to Experiment 1.

Results

Results from Experiment 3 were achieved similarly to Experiment 1, using an Excel spreadsheet to record the number of answers the participant answered correctly for each of the two response sections. A one-sample T-test was run on the results of the word-segmentation task, which showed that across both conditions, the average number correct was 10.68 out of 16, with a standard deviation of 2.42. This is significantly greater than chance, $t(33)=25.7$, $p<.001$, indicating that all individuals were able to parse the nonsense words from the stream of speech despite how often the words and objects were paired together correctly.

If we break this down even more and look at the two conditions separately, an independent-samples T-test shows that the means of condition A and B, respectively, were 10.94 and 10.41 answers correct. The difference between these two averages is nowhere near significant, $t(32)=.632$, $p=.532$, which indicates that the 75% condition was not any better than the 50% condition, and rather both conditions were equally good at being able to segment the nonsense words from the stream of speech.

Moving on to the results of the second questionnaire, the word-object pairing task, a one-sample T-test shows that the average number correct across both conditions is 5.82 with a standard deviation of 4.36. This mean is significantly greater than chance, $t(33)=7.78$, $p<.001$, which would be 4 correct out of 16, which implies that both conditions were able to successfully match up words with their correct objects a significant portion of the time. Knowing this information, it is important to look at how the conditions compare to each other, so an independent-samples T-test was run on the results from the word-object matching task. The means of conditions A and B were 5.71 and 5.94, respectively, with standard deviations of 4.98 and 3.79. Although the overall mean was significantly above chance, the difference between the means of these two conditions was not significantly different, $t(32)=.155$, $p=.878$. This shows that although all participants were able to correctly match the word with its referent in Experiment 3, the 75% condition did not perform significantly better than the 50% condition, which goes against the stated hypothesis.

Discussion

The current study took the previous research on word-referent pairing, applied it more to adults, and took it one step further by not having the words and objects paired up 100% of the time in order to relate more to real-world situations. In the two conditions, the words and objects in the learning phase were either paired correctly 50% or 75% of the time in order to determine if a threshold exists where individuals begin to lack in their ability to pair accurately. The two subsequent tasks, the word segmentation task and the word-object pairing task, were used to test the participants on their learning.

Experiment 1 was the most basic of the three, consisting of a 72-item learning phase which had objects (shapes) associated with each word that was said. It was found that no matter the condition, all individuals were able to segment the words from speech, even though in condition B the words and objects were only paired up correctly half of the time. These results actually proved true for experiments 2 and 3 as well, with all participants being able to parse the words from the speech despite condition. A possible explanation for this is the fact that just *having* a shape that looms on the screen at the same time as a word, no matter what the shape or word is, allows the words to be extracted from the speech later. This goes along with the idea that bimodal stimuli allow for more salience such that just the presence of a shape presented in unison with a word allows it to be recalled later (Miller, 1982).

Although individuals across conditions in all three experiments were able to segment the words from the speech, the same is not true for being able to match the words with the objects in the second task. In experiment 1, individuals performed slightly but not significantly better than chance on this task, getting only around 4-5 questions correct out of 16 (which could occur with guessing). An explanation for why individuals were unable to learn the word-referent pairings could be because the learning phase was too short to pick up the pattern. It was often the case that upon completing the study and being debriefed on how it worked, the participant would claim to have only begun to pick up on the word-referent pairings towards the very end of the learning phase, thus not having enough time to fully encode it. In order to try to correct for this, experiment 2 was created, which consisted of a doubled learning phase (144 items rather than 72) in order to account for the possibility of participants being unable to pick up on the pattern quick enough. After running a preliminary trial of this, however, the results were identical to that of experiment 1, indicating that the length of the learning phase may not have been the reason for the inability of participants to accurately pair words with referents. This makes sense, since Yu and Smith (2007) found that the amount that word-referent pairs are repeated does not have an impact on the ability of people to pair them. Thus, a different explanation for these results is necessary.

Another possible reason for the inability of participants to correctly match the words with their referents in experiment 1 could have been because upon watching the stimuli, participants may have formed a hypothesis on which word went with which object, but upon seeing contradictory stimuli in the other 50% or 25% of the time (for conditions A and B, respectively), their hypothesis was disproven. Thus, it would almost be as if a hypothesis on the word-referent pairings was never even formulated, hence why they did not perform better than chance on the later task. Experiment 3 attempted to account for this possibility by not having the contradicting stimuli in the other 50% or 25% of the time, but rather a plain black screen presented with the spoken word instead. Interestingly, the results of this manipulation in the experiment showed that all participants across conditions *were* able to accurately match the word with the correct object, indicating that the above explanation for why participants originally struggled with the task in experiment 1 is feasible.

An interesting aspect of the results of these three experiments is that the differences between the conditions were never significant, which goes against the original hypothesis that the 75% condition would perform better. A reason for this may be that the threshold of where individuals begin to lack in their word-referent pairing ability lies in a different place, such as below 50%. If this were the case, it would explain why both the 50% and 75% conditions performed equally, because they are both above the ability threshold. It is also possible that there are different thresholds for segmenting words from speech and for matching words with objects, since it seems that participants had more difficulty with the latter.

Using this idea, it is important for future research to look at how participants respond to other percentages too, such as less than 50% and more than 75%, and perhaps another condition in the middle of the two as well. That way, it is possible to gain a more accurate idea of where the threshold, if any, lies. If finding a threshold still proves difficult, then perhaps priming the participant to let them know what to pay attention to might improve their accuracy. In addition, Yu and Smith (2007) found that larger data sets improve the encoding of them, as mentioned earlier, so in the future it might be interesting to add more word-referent pairs rather than just 4 in order to see if an effect exists in this case. The order in which the tasks are laid out in the experiment may also have made a difference during this study, so in the future it would be important to consider order effects and counterbalancing the two tasks.

References

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology*. *6*(2), 111-123.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*. *14*(4), 597-650.

Gogate, L. J., & Bahrick, L. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*. *69*(2), 133-149.

Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior & Development*. *7*(3), 361-381.

Lyons-Ruth, K. (1977). Bimodal perception in infancy: Response to auditory-visual incongruity. *Child Development*. *48*(3), 820-827.

MacKain, K., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left-hemisphere function. *Science*. *219*(4590), 1347-1349.

Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*. *14*(2), 247-279.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*. *9*(5), 454-464.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*. *274*(5294), 1926-1928.

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