4pSCb21. Lexical bias and prosodic cues: An eye-tracking study of compound/phrase disambiguation

Jessica L. Gamache*

*Corresponding author's address: Department of Linguistics and Languages, Michigan State University, East Lansing, MI 48824, gamache1@msu.edu

Despite compounds and phrases exhibiting distinct prosodic cues (e.g., pitch and duration), adults fail to use these cues in certain contexts. In two eye-tracking experiments, I examined English speakers' attention to prosody in adjective-noun string disambiguation to explore a lexical bias previously reported (e.g., 'hot dog' is interpreted as the food, regardless of prosody.) In experiment 1, 24 participants were presented with pictures of the compound and phrasal representations for possible compounds (20 known, 10 novel) accompanied by an audio presentation of either a phrasal or compound production. Replicating previous findings, participants sometimes ignored prosody and exhibited biases towards compound pictures for common compounds and towards phrasal pictures for novel compounds. The fixation data reveals similar patterns in which participants appeared to not consider the alternative picture, even when it matched available prosodic cues. In experiment 2, the phrasal/compound pictures were decoupled and placed with unrelated distractors and participants were asked whether either picture matched the audio presentation of a phrasal or compound production. This experiment helps to determine whether the exhibited lexical bias is an artifact of experimental designs using minimal pairs or whether prosody is still ignored when frequency and novelty are not directly competing.
INTRODUCTION

In order to be a competent speaker of a language, speakers must be able to use prosodic cues to resolve structural ambiguities. Consider the sentence *The boy saw the man with the telescope*. This has two readings: one where the man has a telescope and one where the boy uses a telescope to see the man. This ambiguity can be resolved by prosodic features, namely the placement of a short pause either after *saw*, signaling that the man had the telescope, or after *man*, signaling that the boy used the telescope to see the man. Many studies have looked at children’s and adults’ ability to use prosodic cues in ambiguous domains, including adjunct attachment, as in the telescope example above (Carlson et al., 2001; Snedeker and Yuan, 2008; Snedeker and Casserly, 2010), compounds vs. lists, as in “fruit-salad and oranges” vs. “fruit, salad, and oranges” (Wells et al., 2004; Good, 2008), and adjective-noun strings, as in ‘yellow-jacket’ as a compound (insect) vs. a phrase (jacket) (Vogel and Raimy, 2002).

Adults generally use prosody effectively in disambiguation tasks. However, when faced with a conflict between prosodic cues and lexical cues, the prosodic cues are often ignored (Good, 2008; Vogel and Raimy, 2002). For example, Good (2008) found that when adults were presented with two nouns with compound stress they would prefer a phrasal interpretation if the two nouns did not readily lend themselves to a compound interpretation (e.g., an object called a ‘nail-key’). Similarly, Vogel and Raimy (2002) found that given a novel adjective-noun compound (‘red-cup’ as a name for a red flower) participants would choose a phrasal interpretation (a cup that is red) over a compound interpretation (a red flower).

Children are generally reported to not use prosody in an adultlike manner in disambiguation tasks until after age 10 (Crystal, 1978; Cruttenden, 1985; Wells et al., 2004; Vogel and Raimy, 2002; Good, 2008). There are some exceptions to these findings, however. Good (2008) shows that children are actually better than adults at using prosody in cases where the prosodic and lexical information conflict. That is, children are more likely than adults to give ‘nail-key’ with compound stress a compound interpretation, even though it is not obvious what this object would be. Given that children’s lexical and world knowledge is weak relative to adults’ knowledge, children are less biased by lexical information and therefore use prosody more effectively in these situations (Good, 2008). Moreover, Snedeker and Yuan (2008) and Zhou et al. (2012) find that children’s eye movements mirror adults’ in on-line prosodic disambiguation tasks, even when their responses do not reflect the prosodic information. These findings suggest that children’s ability to use prosody is likely more proficient than previously reported, but that task demands may interfere with their offline responses.

This research addresses the relationship between prosodic and lexical information in perception, specifically when the two types of information are in conflict. To examine the relationship between prosodic and lexical information, we performed two experiments with adjective-noun string disambiguation tasks. Lexical knowledge and prosody could conflict in two ways for adjective-noun stings. First, frequent meanings of some compounds like ‘hot-dog’ may be so strong that the string ‘hot dog’ will be interpreted as a compound regardless of its prosody. On the other end of the spectrum, novel compounds, which are by definition infrequent, could be interpreted as phrases regardless of their prosody due to the lacking lexical knowledge associated with these forms. All else being equal, one could hypothesize that prosodic information is strong enough to create new lexical items on the fly, thereby turning novel compounds into known compounds on the first encounter (i.e., fast mapping (Carey, 1978b,a,)). While adjective-noun compounds are non-compositional (e.g., a ‘yellow-jacket’ is an insect, not a jacket), they are nevertheless are semi-productive and have compound prosody that reliably differentiates them from their phrasal counterparts (Morrill, 2012).

Experiment 1 examined listeners’ use of prosody when presented with images of both the compound of the phrasal interpretation of an adjective-noun string. Given the findings that more fine-grained online measures can give insight into the processing of prosodic information, accuracy data, response times, and eye fixations were all recorded. We predicted that response data would replicate that of previous findings, showing a strong lexical bias for novel items (Vogel and Raimy, 2002) but that the on-line measures of response times and fixation data should reflect speakers’ use of prosody. Experiment 2 examined whether or not a mismatch in prosody would cause speakers to reject images that were otherwise matched for lexical description (i.e., would speakers reject an image of the compound interpretation if they heard the phrasal prosody.) Given that this study removed the explicit lexical bias of having the compound and phrasal images side-by-side, we predicted that speakers would accept more interpretations in this experiment relative to Experiment 1.
EXPERIMENT 1

Method

Participants

Twenty-four undergraduate students from Michigan State University participated in Experiment 1 in exchange for course credit. Due to a software error, eye tracking data is only reported for 23 participants.

Materials

Twenty known and ten novel adjective-noun compounds were selected for Experiment 1 (see Appendix for list.) For each of the 30 items, two images were drawn, one for the compound interpretation and one for the phrasal interpretation. The compound and phrasal images for each item were presented together on the screen (see Figure 1.) Sixty unrelated filler items were also included in the experiment. Audio stimuli for both types of prosody were recorded by the author for each item. Compound and phrasal presentations differed in timing, pitch pattern, and stress pattern.

![Figure 1: A sample screen shot for yellow jacket from Experiment 1. Each item was presented with the compound (yellow jacket left) and phrasal (yellow jacket right) images.](image)

Apparatus

The experiment was presented on a 48-cm. cathode ray tube monitor with 1024 × 768 pixels resolution. The experiment was controlled by Experiment Builder software (SR Research). A head mounted EyeLink II eye tracker was used to record eye movements at a sampling rate of 500 Hz. A Microsoft SideWinder Game Pad was used to collect button presses.

Design

The variables Novelty of Compound (Known, Novel) and Prosody of Audio (Compound, Phrasal) were varied within participants, creating four types of items: Known-Compound, Known-Phrasal, Novel-Compound, and Novel-Phrasal. All 30 test items were presented to each participant in either their compound or phrasal form. Prosody was counterbalanced such that there were even numbers of participants receiving phrasal and compound prosody for each item. No participant heard both prosodies for any item. The 60 filler items were placed such that no two test items were adjacent. Two pseudorandom orders were generated and inverted for a total of four item list orders.

Procedure

Participants were randomly assigned to one of four order lists. Participants read the instructions and were calibrated in the eye tracker. Each trial started with the presentation of a fixation cross, followed by a
5000 ms familiarization interval with the target images. The familiarization interval was intended to expose participants to the stimuli and diminish any effects of novelty in the images. After familiarization a fixation cross reappeared for 1000 ms and subsequently the target images were displayed along with the audio presentation “Look at the ≪target≫ now.” Using the game pad, participants clicked either the left or right trigger buttons to convey which image on the screen matched the audio presentation.

**Results**

Analyses of Variance (ANOVAs) were performed on the proportion of correct responses, as well as the response times and proportion looks to target image.

**Correct Responses**

Figure 2 displays mean accuracy rates as a function of Novelty and Prosody. Accuracy was significantly higher for the Known items than the Novel items \( (F(1,23)=17.798, \ p<.001) \) and significantly higher for Phrasal prosody than for Compound prosody \( (F(1,23)=74.360, \ p<.001) \). For Novel items, accuracy was significantly higher for Phrasal items than Compound items, while for Known items there was no significant difference in accuracy, leading to a significant interaction between Novelty and Prosody \( (F(1,23)=165.020, \ p<.001) \). Paired-samples t-tests confirm the significant difference in accuracy for Novel items \( (t(23)=-14.303, \ p<.001) \), but not Known items \( (p=.183) \).

![Figure 2: Mean accuracy rates by Novelty and Prosody. Error bars are +/− 1 standard error.](image)

**Response Times**

Figure 3 displays mean response times (RTs) as a function of Novelty and Prosody. RTs were significantly faster for Known items than for Novel items \( (F(1,23)=71.133, \ p<.001) \) and significantly faster for Compound prosody than for Phrasal prosody \( (F(1,23)=5.496, \ p=.028) \). For Known items, accuracy was significantly faster for Compound items than for Phrasal items, while for Novel items there was no significant difference in RTs, leading to a significant interaction between Novelty and Prosody \( (F(1,23)=6.023, \ p=.022) \). Paired-samples t-tests confirm the significant difference in RTs for Known items \( (t(23)=-3.867, \ p=.001) \), but not for Novel items \( (p=.611) \).

**Fixations**

Figure 4 displays mean proportion of looks to target image as a function of Novelty and Prosody. The target image was looked at significantly more for Known items than for Novel items \( (F(1,22)=19.150, \ p<.001) \), but not significantly more for Compound prosody than for Phrasal prosody \( (p=.084) \). For Novel items, participants looked significantly more to the target image for Phrasal items than for Compound items, while for Known
items there was no difference in looks to the target, leading to a significant interaction between Novelty and Prosody ($F(1,22)=10.046, p=.004$). Paired-samples $t$-tests confirm the significant difference in RTs for Novel items ($t(22)=-3.025, p =.006$), but not for Known items ($p=.401$).

**Experiment 1 Summary**

Experiment 1 tested speakers’ ability to use prosodic information to choose between compound and phrasal images of adjective-noun word strings for both known and novel compounds. The results replicated previous findings (Vogel and Raimy, 2002): while adult speakers are good at using prosody for known compounds, they fail to acknowledge prosody in cases where there is a more readily available phrasal interpretation. This trend is echoed in all three types of data collected. For accuracy rates, participants correctly chose the compound interpretation of novel compounds only $\sim 20\%$ of the time, while accuracy rates were above 60\% in all other conditions. For RTs, participants were significantly slower for Novel items than for Known items. And finally, the fixation data shows that participants looked significantly less at the novel compound targets, suggesting that they may not have even considered them a viable choice for the auditory stimuli. In other words, given that a phrasal interpretation of an adjective-noun string is always available, when faced with a novel compound, participants used the lexical information over the prosodic information.
EXPERIMENT 2

The goal of Experiment 2 was to examine participants' ability to use prosody when lexical bias was not explicitly present in the experiment. There were two critical changes from Experiment 1: (i) images were placed with distracters instead of their minimal pairs and (ii) participant could either answer that the audio matched the left image, the right image, or neither image. Given the changes to Experiment 2, we considered two possible outcomes. First, decoupling the prosodic minimal pairs could remove the lexical bias (for Known compounds and Novel phrases) and increase participants' accuracy on the task. The alternative outcome we considered was that decoupling the prosodic minimal pairs could remove the cue to the importance of prosody for the task (i.e., that different prosody changed the interpretation) and participants would make their decisions based solely on the lexical information.

Method

Participants

Thirty-two undergraduate students from Michigan State University participated in Experiment 2 in exchange for course credit.

Materials

The same 30 compounds from Experiment 1 were used. For each trial in Experiment 2, however, the target item no longer occurred with its prosodic minimal pair. Target images were paired with random distractors and each participant only saw one representation of each item (i.e., only the compound image or only the phrasal image.)

Apparatus

The experiment was presented on a 52-cm. flat screen monitor with $1024 \times 768$ pixels resolution. The experiment was controlled by Experiment Builder software (SR Research). A desk mounted EyeLink 1000 eye tracker was used to record eye movements at a sampling rate of 1000 Hz. A Microsoft SideWinder Game Pad was used to collect button presses.

Design

In addition to the two variable from Experiment 1 (Novelty (Known, Novel) and Prosody (Compound, Phrasal)), Match (Match, Mismatch) was added. Match reflects whether or not the audio stimulus matches the image that the participants see on the screen. These three variables created eight types of trials:

<table>
<thead>
<tr>
<th>Trial Type</th>
<th>Prosody</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known Compound Match</td>
<td>Compound</td>
<td>Compound</td>
</tr>
<tr>
<td>Known Compound Mismatch</td>
<td>Compound</td>
<td>Phrasal</td>
</tr>
<tr>
<td>Known Phrasal Match</td>
<td>Phrasal</td>
<td>Phrasal</td>
</tr>
<tr>
<td>Known Phrasal Mismatch</td>
<td>Phrasal</td>
<td>Compound</td>
</tr>
<tr>
<td>Novel Compound Match</td>
<td>Compound</td>
<td>Compound</td>
</tr>
<tr>
<td>Novel Compound Mismatch</td>
<td>Compound</td>
<td>Phrasal</td>
</tr>
<tr>
<td>Novel Phrasal Match</td>
<td>Phrasal</td>
<td>Phrasal</td>
</tr>
<tr>
<td>Novel Phrasal Mismatch</td>
<td>Phrasal</td>
<td>Compound</td>
</tr>
</tbody>
</table>

All 30 test items were presented to each participants, with Prosody and Match counterbalanced in the audio presentations such that there were even numbers of participants receiving each type of Prosody and
each type of Match for each item. No participant heard both types of prosody for any item. The 60 filler items were placed such that no two test items were adjacent. Two pseudorandom orders were generated and inverted for a total of four item list orders.

Procedure

The same procedure from Experiment 1 was used.

Results

Analyses of Variance (ANOVAs) were performed on the proportion of correct responses and response times. Eye tracking data is not be presented for this experiment.

Correct Responses

Figure 5 displays mean accuracy rates as a function of Novelty, Prosody, and Match. Accuracy was significantly higher for Match trials than for Mismatch trials ($F_{(1,31)}=364.109, p<.001$), reflecting participants' incorrect acceptance of most Non-Match trials as being matches. Accuracy was also significantly higher for Phrasal trials than for Compound trials ($F_{(1,31)}=94.111, p<.001$). Accuracy did not differ significantly by Novelty overall ($p=.712$). However, accuracy was significantly higher for Known Match items than Novel Match items and significantly lower for Known Mismatch trials than for Novel mismatch trials, leading to a significant interaction between Match and Novelty ($F_{(1,31)}=48.099, p<.001$). This interaction reflects participants' acceptance of only $\sim40\%$ of the novel compounds as being compounds. The interactions between Match and Prosody ($p=.173$) and Match, Prosody and Novelty ($p=.710$) were not significant.

Response Times

Figure 6 displays mean RTs as a function of Novelty, Prosody, and Match. RTs were significantly faster for Known trials than for Novel trials ($F_{(1,31)}=364.109, p<.001$). RTs did not differ significantly overall by Match ($p=.395$) or Prosody ($p=.593$). However, Compound trials were significantly faster for Known trials than for Novel trials, while Phrasal items did not differ by Novelty, leading to a significant interaction between Prosody and Novelty ($F_{(1,31)}=364.109, p<.001$). The interaction between Novelty and Prosody was also significant ($F_{(1,31)}=6.823, p=.014$), reflecting the significantly longer RTs for Match trials as compared to Mismatch trials for Phrasal Prosody. The interaction between Prosody, Match, and Novelty was also significant ($F_{(1,31)}=20.957, p<.001$). The interactions between Match and Novelty was not significant ($p=.195$).

Experiment 2 Summary

Experiment 2 examined whether or not speakers would reject an image whose description differed only in prosody from the audio stimulus. The answer is a resounding “no”. For the known items, participants...
generally accepted all of the trials as matches, even though 50% were mismatch trials. For novel items, the results were split: half of the compounds were treated like known items and were accepted across the board, for both match and mismatch trials, while the other half of the novel compounds were rejected across the board. In other words, without any cue to prosodic information being important, participants will associate an adjective-noun string to either a phrasal or a compound interpretation as long as that adjective-known string is associated to a given meaning (whether compositionally or as a stored compound meaning.) For example, participants know that ‘yellow-jacket’ with compound prosody is an insect, they also allow the phrasal prosody of ‘yellow jacket’ to be associated with the insect. On the other hand, speakers reject any adjective-noun compound for which they do not have an associated compound meaning. That is, if participants do not know that ‘red-cup’ with compound prosody is a red flower, they reject the flower image regardless of the prosody of ‘red-cup’ being phrasal or compound. The only interpretation available for speakers in this case is the phrasal one (i.e., a cup that is red.)

**DISCUSSION**

This research examined speakers' usage of prosodic and lexical information in perception. Specifically, we examined how speakers weighed prosodic information against lexical information when the two were in conflict. To do this, we performed two experiments using adjective-noun strings, which are ambiguous between compounds and phrases. The first experiment examined speakers’ use of prosody when both a compound and a phrasal alternative were present and the second experiment examined speakers’ use of prosody when only one alternative was present and the prosody either did or did not match the image. Experiment 1 showed that it was not the case that prosodic information could override lexical information and create new lexical items; known items were reliably treated as compound when presented with compound prosody and as phrases when presented with phrasal prosody. Novel items, on the other hand, were generally treated as phrases, regardless of the type of prosody. Contrary to predictions, the eye tracking data also did not show any sign of the speakers considering the prosodic information for novel items. Experiment 2, which was predicted to lessen the strength of lexical information, only served to support the results of Experiment 1; if an interpretation of an image could match the auditory stimulus, the participants accepted the trial as a match, regardless of the prosody being a match or mismatch. If participants did not have an interpretation that could match the auditory stimulus, they would reject the trial as a mismatch, regardless of the prosody being a match or mismatch. These results have implications for both word learning and the effect of lexicalization on speech processing.

**ACKNOWLEDGEMENTS**

I would like to thank Cristina Schmitt, Alan Munn, Mark Becker, Tuuli Morrill and the members of the MSU Language Acquisition Lab for their help with this project.
REFERENCES


APPENDIX

<table>
<thead>
<tr>
<th>Known Compounds</th>
<th>Novel Compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>big bird*</td>
<td>big wind (a strong fan)</td>
</tr>
<tr>
<td>big wig</td>
<td>blue banner* (a fish with a blue stripe)</td>
</tr>
<tr>
<td>black board*</td>
<td>flat bridge* (a type of oar)</td>
</tr>
<tr>
<td>black top</td>
<td>green rope (a green snake)</td>
</tr>
<tr>
<td>blue book</td>
<td>long neck (a bottle with a long neck)</td>
</tr>
<tr>
<td>blue jay*</td>
<td>orange belly (a frog with an orange stomach)</td>
</tr>
<tr>
<td>cheap skate</td>
<td>red cup* (a red flower)</td>
</tr>
<tr>
<td>dark room</td>
<td>tall back (a chair with a tall back)</td>
</tr>
<tr>
<td>green house*</td>
<td>white knife (a white pointy shark)</td>
</tr>
<tr>
<td>hard drive</td>
<td>yellow face (a bird with a yellow face)</td>
</tr>
</tbody>
</table>

* used in Vogel and Raimy (2002)