1. Introduction

Successfully differentiating questions and statements early in the acquisition process is critical in the acquisition of many languages. Questions constitute a large percentage of infant-directed speech and can differ from statements in surface word order and syntactic structure. For example, in English, polar interrogatives and wh-questions make up 23% and 21% of infant-directed speech respectively (Newport, 1977) and involve do-support, subject-auxiliary inversion and related structural changes. An inability to distinguish these sentence types from statements with a canonical word order would lead infants to make erroneous generalizations about the grammatical structure of their language (Pinker, 1984; Slobin & Bever, 1982).

Prior studies suggest that prosodic cues might be available that would allow children to distinguish statements and questions. For example, a number of prosodic cues correlate with different sentence types in adult speech. Prominently, pitch cues and duration in the final syllables of a sentence type are frequently associated with certain sentence types in English, French, and Dutch. In English, the final syllables of polar interrogatives and declarative questions frequently have rising pitch intonations, while statements have falling pitch intonations (Bartels, 1999; Hedberg, Sosa & Fadden, 2004; Ladd, 2008). In French, rising pitch intonations are common in the final syllables of polar interrogatives. Speakers in an experimental context have been able to identify polar interrogatives and statements using this cue (Vion & Colas, 2006). In Dutch, questions are often spoken with faster speech rates than statements (van Heuven & van Zanten, 2005). If infants could use prosodic cues to differentiate statements and questions at the earliest stages of syntax acquisition, this would allow them to associate different syntactic properties to the two sentence types. Indeed, infants learning European Portuguese appear to be able to do so (Frota, Butler & Vigário, 2014).

Geffen and Mintz (2017) investigated whether pitch, duration and intensity in the final two syllables of polar interrogatives, statements, and wh-questions in
English infant-directed speech could be used to provide an initial basis for distinguishing sentence types. The study found that while the prosodic cues in the final two syllables provided sufficient evidence to distinguish polar interrogatives and statements at above chance rates, these cues tended to group wh-questions with statements. Thus, these cues were not sufficient to differentiate questions and statements generally, leaving open the question of whether other prosodic cues could be used.

Previous work has shown that some prosodic information correlated with wh-questions can also occur in areas other than the final two syllables of the sentence. Maxwell and Fletcher (2011) found that Bengali-English and Kannada-English speakers often have higher accent scaling for the sentence-initial syllables of wh-questions when compared to statements. Pragmatics and discourse function of a question have also been found to influence the nuclear contours of a wh-question both at the beginning and at the end of the sentence in North American English. The characteristic low pitch intonations at the end of wh-questions are correlated with the contexts in which these sentences occur (Hedberg, Sosa, Görgülü, & Mameni, 2010). The pitch accents of wh-questions and some type of polar interrogatives have also been found to vary by context (Hedberg & Sosa, 2002). Hence, the current study investigates whether the prosodic cues in sentence-initial syllables could provide a discrimination cue to further distinguish polar interrogatives, wh-questions, and statements.

2. Methods

2.1. Corpus Features

The Brent Corpus (Brent & Siskind, 2001) from the CHILDES database (MacWhinney, 2000) was analyzed for this study. The corpus contains audio recordings of the interactions between 16 mother-infant dyads. Interactions of the pairs occurred at home in a naturalistic setting. This study analyzed the infant-directed speech of three of the mothers from this corpus—s1, s2, and v1.

2.2. Utterance features

Following the procedures outlined by Geffen and Mintz (2017), polar interrogatives, wh-questions, and statement sound files were selected based on several criteria. The selected polar interrogatives included subject and auxiliary inversion. The selected wh-questions included auxiliary inversions as well as a wh-word in the utterance initial position (e.g. who, what, where, when, why, how). The selected declarative statements had transitive word order. Four fragments were also included in the analysis. Context was used to decide the inclusion of three of these fragments, which were determined to be declarative statements. The fourth was a polar interrogative fragment (e.g. Are you?). Sound files were excluded if extraneous noise impacted the subsequent feature extraction.
2.3. Feature Extraction

The initial two syllables of 103 statements, 81 polar interrogatives, and 87 wh-questions from the corpus were segmented and analyzed. For each syllable average pitch, lowest pitch, highest pitch, duration, and average intensity were extracted. All pitch, duration, and intensity values were extracted using the ProsodyPro script (Xu, 2012). Pitch excursion was subsequently calculated using the values of the lowest and highest pitch of the first and second syllable. The extreme values for the syllables were identified and subtracted from one another to produce a positive or negative value reflecting a rising or falling pitch. To reduce the dimensionality of variables, a principal component analysis (PCA) was conducted. The first four principle components accounted for 90% of the variance between the measured variables. Component 1 was a moderate loading on all variables excluding duration of the first syllable. Component 2 was primarily a function of the first syllable’s duration. Component 3 was a moderate loading on all variables excluding the maximum pitch of the first syllable, the duration of the second syllable, and the mean pitch of the first syllable. Component 4 was primarily a function of the duration of the second syllable (see Figure 1 in the Appendix).

2.4. Supervised Analyses

Three logistic regressions were subsequently conducted using these four factors as the predictors, and sentence type as the dependent variable. The first regression compared polar interrogatives and statements. The second regression compared wh-questions to statements. The third regression compared wh-questions to polar interrogatives. Critically, for each regression we set the regression weights based on half the utterances of the relevant types, and used them to predict the category of the other half.

2.5. Unsupervised Analyses

Follow-up analyses were conducted. An unsupervised method, k-means clustering, was also used to analyze the data. The data of the extracted values from ProsodyPro were used. Clusters of two and three were used for this analysis. Two for the categories question and statements, and three for the categories polar interrogative, wh-question, and statement.

3. Results
3.1. Supervised Analyses

The first regression categorized polar interrogatives and statements with 72% accuracy. Inaccurate categorizations in this group were evenly split among the two sentence types. The second regression categorized wh-questions and statements with 59% accuracy. Among the inaccurate categorizations, statements
were more likely to be erroneously categorized as wh-questions than vice-versa. Therefore, both questions types were discriminated from statements at above-chance levels. The third regression was basically at chance in categorizing polar interrogatives and wh-questions, thus the prosodic cues were similar across the two types of questions, just as both sentence types are similar with respect to their distributional and structural differences from statements.

3.2. Unsupervised Analysis

The k-means analyses for both two and three clusters yielded low ratios of the between sum of squares and total sum of squares. The ratio for the two cluster analysis was .239; the ratio for the three cluster analysis was .362, with 1.0 being the highest possible. Thus, these values indicate a poor fit.

4. Discussion

The results of the study provide support for previous findings that prosody can be used to differentiate sentence types and demonstrate that utterance-initial prosodic cues could provide additional information for a listener. Similar to the results of the Geffen and Mintz (2017) study, the findings show that prosody can be a source of information to distinguish polar interrogatives from statements, as the utterance-initial prosodic cues in polar interrogatives patterned differently from statements. The results from this study also suggest that utterance-initial prosody patterns similarly between wh-questions and polar interrogatives, and differently from statements. While the categorization is not perfect, these results nevertheless demonstrate that sentence-initial prosody is correlated with sentence type and available as a source of information to distinguish questions and statements.

However, it is important to note that the manner in which the regression weights were set in the analyses are supervised. The results for the unsupervised method indicate that the prosodic information was not reliable enough to differentiate questions from statements in this context. It may be the case that cues other than the utterance-initial prosodic cues are necessary for children to initially distinguish the two sentence types.

This may be especially relevant to wh-questions, as prosodic cues could vary by wh-question type. There are cases in which this seems to occur. For example, House (2005) found that phrase-final focal accent of Swedish speakers differed by the wh-word in the utterance-initial position. In order to investigate the possibility that different wh-question types could have different prosodic cues, future studies would need to analyze a larger amount of utterances and perhaps a larger corpus. Geffen, Burkinshaw and Curtin (in prep) are currently analyzing a corpus of approximately 5000 sentences to evaluate the prosodic cues that may be available to infants in the first two words of infant-directed statements and wh-questions. Hopefully, this will provide further support for the availability of utterance-initial cues for distinguishing statements and questions.
**Figure 1:** Full loadings of individual factors for principal component analysis

<table>
<thead>
<tr>
<th></th>
<th>Comp.1</th>
<th>Comp.2</th>
<th>Comp.3</th>
<th>Comp.4</th>
<th>Comp.5</th>
<th>Comp.6</th>
<th>Comp.7</th>
<th>Comp.8</th>
<th>Comp.9</th>
<th>Comp.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>-0.21</td>
<td>-0.26</td>
<td>0.35</td>
<td>0.00</td>
<td>0.14</td>
<td>0.00</td>
<td>-0.57</td>
<td>0.00</td>
<td>-0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>Max</td>
<td>-0.42</td>
<td>0.00</td>
<td>-0.25</td>
<td>0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.57</td>
<td>0.18</td>
<td>0.00</td>
<td>0.18</td>
</tr>
<tr>
<td>excursion</td>
<td>-0.21</td>
<td>0.30</td>
<td>-0.60</td>
<td>0.21</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.33</td>
<td>0.10</td>
</tr>
<tr>
<td>minf01</td>
<td>-0.29</td>
<td>-0.23</td>
<td>0.21</td>
<td>-0.20</td>
<td>-0.28</td>
<td>-0.61</td>
<td>0.00</td>
<td>0.35</td>
<td>0.41</td>
<td>-0.13</td>
</tr>
<tr>
<td>minf02</td>
<td>-0.24</td>
<td>-0.21</td>
<td>0.26</td>
<td>0.00</td>
<td>0.35</td>
<td>0.50</td>
<td>0.12</td>
<td>-0.12</td>
<td>0.63</td>
<td>-0.12</td>
</tr>
<tr>
<td>maxf01</td>
<td>-0.41</td>
<td>0.00</td>
<td>0.00</td>
<td>0.15</td>
<td>-0.43</td>
<td>0.42</td>
<td>0.21</td>
<td>0.29</td>
<td>-0.36</td>
<td>-0.44</td>
</tr>
<tr>
<td>maxf02</td>
<td>-0.38</td>
<td>0.00</td>
<td>-0.22</td>
<td>0.00</td>
<td>0.37</td>
<td>-0.35</td>
<td>0.00</td>
<td>-0.50</td>
<td>-0.21</td>
<td>-0.43</td>
</tr>
<tr>
<td>duration1</td>
<td>-0.12</td>
<td>0.63</td>
<td>0.53</td>
<td>0.52</td>
<td>0.00</td>
<td>-0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>duration2</td>
<td>-0.22</td>
<td>0.56</td>
<td>0.00</td>
<td>-0.77</td>
<td>0.00</td>
<td>0.15</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>meanf01</td>
<td>-0.36</td>
<td>-0.12</td>
<td>0.11</td>
<td>0.00</td>
<td>-0.48</td>
<td>0.00</td>
<td>0.14</td>
<td>-0.61</td>
<td>0.00</td>
<td>0.45</td>
</tr>
<tr>
<td>meanf02</td>
<td>-0.31</td>
<td>-0.13</td>
<td>0.00</td>
<td>0.00</td>
<td>0.46</td>
<td>0.00</td>
<td>0.51</td>
<td>0.32</td>
<td>-0.27</td>
<td>0.39</td>
</tr>
<tr>
<td>meanintensity1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.35</td>
</tr>
<tr>
<td>meanintensity2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.27</td>
</tr>
</tbody>
</table>
References


Brent, Michael R., & Siskind, Jeffrey M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* 31-44.


