



HHS Public Access

Author manuscript

J Child Lang. Author manuscript; available in PMC 2017 May 01.

Published in final edited form as:

J Child Lang. 2017 May ; 44(3): 734–751. doi:10.1017/S030500091600009X.

Prosodic disambiguation of noun/verb homophones in child-directed speech*

ERIN CONWELL

North Dakota State University

Abstract

One strategy that children might use to sort words into grammatical categories such as NOUN and VERB is DISTRIBUTIONAL BOOTSTRAPPING, in which local co-occurrence information is used to distinguish between categories. Words that can be used in more than one grammatical category could be problematic for this approach. Using naturalistic corpus data, this study asks whether noun and verb uses of ambiguous words might differ prosodically as a function of their grammatical category in child-directed speech. The results show that noun and verb uses of ambiguous words in sentence-medial positions do differ from one another in terms of duration, vowel duration, pitch change, and vowel quality measures. However, sentence-final tokens are not different as a function of the category in which they were used. The availability of prosodic cues to category in natural child-directed speech could allow learners using a distributional bootstrapping approach to avoid conflating grammatical categories.

INTRODUCTION

As children learn their first language, they must arrive at a set of representations that are at once general enough to produce and process the potentially limitless number of utterances in that language, but also sufficiently constrained to avoid producing the equally limitless number of utterances that are not permitted in that language. One aspect of language that supports both productivity and constraint is the availability of grammatical categories, such as noun and verb. These categories encode the syntactic contexts in which a particular word can and cannot be used. How children sort words into such categories is a major question for language acquisition research.

If noun and verb categories were solely characterized by semantic properties (i.e. if the set of nouns really were all words that referred to people, places, and things), then the process of sorting words into categories would be relatively straightforward. A child who had learned the meaning of a word would also have perfect information about its grammatical

[*]This research was supported by Grant 1R15HD077519-01 to the author from the National Institute of Child Health and Human Development, which is part of the National Institutes of Health (NIH). The contents of this paper are the sole responsibility of the author and do not necessarily represent the official views of NICHD or NIH. Additionally, I thank Brenden Melvie, Katelyn Tallas, Matthew Kramer, Felix Pichardo, Cheyenne Brady, Adrienne MacDonald, Elisabeth Dukowitz, and Alexandra Howatt for their assistance with the token extraction and measurement, Alejandrina Cristia for sharing her PRAAT scripts, and two anonymous reviewers for their helpful comments on the manuscript.

Address for correspondence: Erin Conwell, North Dakota State University, NDSU Dept. 2765, P.O. Box 6050, Fargo, ND 58108-6050. tel: 1-701-231-6123; Erin.R.Conwell@ndsu.edu.

category. Unfortunately for both learners and researchers, the situation is not so straightforward. The grammatical category of a word is defined by its relationship to other grammatical categories. Nouns are words that appear in noun contexts: as the heads of noun phrases, as the objects of verbs and prepositions, etc. Likewise, verbs are words that occur in verb contexts: as the heads of verb phrases, with noun phrase arguments, with adverbial modifiers, etc. Known as the BOOTSTRAPPING PROBLEM, the inherent circularity of these definitions makes the learning process challenging to describe (Gleitman & Wanner, 1982; Pinker, 1989). Further complicating the learning process is the fact that in English, as well as many other languages, a number of words are AMBICATEGORICAL, that is, they can be used in more than one major syntactic category. For example, the word *walk*, which describes an action, may be used in both noun and verb contexts.

Language acquisition researchers have posited a range of solutions to the problem of sorting words into grammatical categories. While some proposals leverage the imperfect correlation between meaning and grammatical category (e.g. SEMANTIC BOOTSTRAPPING; Pinker, 1989), one weakness of these proposals is that infants have access to only a moderate number of word meanings, yet they demonstrate sensitivity to grammatical violations created by swapping nouns and verbs within a sentence (Soderstrom, White, Conwell & Morgan, 2007) and they are able to use the distribution of nouns in a sentence to guess at the meanings of novel verbs (Gertner, Fisher & Eisengart, 2006; Yuan & Fisher, 2009; *inter alia*). This suggests that very young children may base their early grammatical categorizations on something other than the meanings of the words. In particular, DISTRIBUTIONAL BOOTSTRAPPING suggests that infants may use local co-occurrence information to begin lexical categorization (Maratsos & Chalkley, 1980).

The premise of distributional bootstrapping is that local information could provide adequate support for first-pass categorization, thus avoiding the problems that the self-referential nature of formal category definitions might create. For example, nouns and adjectives frequently appear immediately after the word *the*, but verbs never do. A number of empirical studies indicate that such cues are useful for categorizing nouns and verbs, and that infants will use such information to categorize words (Höhle, Weissenborn, Kiefer, Schulz & Schmitz, 2004; Maratsos & Chalkley, 1980; Mintz, 2003; Mintz, Newport & Bever, 2002; Monaghan, Chater & Christiansen, 2005; Redington, Chater & Finch, 1998; Shi & Melançon, 2010). These results provide support for the idea that children's earliest lexical categories might be based on local co-occurrence information. However, the computational work in support of distributional bootstrapping (e.g. Mintz, 2003; Redington *et al.*, 1998) typically uses the homogeneity of the resultant categories as the measure of accuracy of categorization, and does not articulate how ambicategorical words would be treated in this process (but see Cartwright & Brent, 1997, for an exception).

Distributional bootstrapping runs up against a potential problem when it comes to ambicategorical words. Such words should, in theory, significantly confound children who are trying to sort words into categories. As Pinker (1987) points out, encountering sentences such as (1a–c) should lead a learner to conclude that (1d) is also a grammatical sentence in English. After all, the words *fish* and *rabbits* have previously appeared in the same grammatical contexts and it would be reasonable for a learner using only distributional cues

to assume that they can, therefore, appear in all the same contexts. Such examples are used to argue against the very possibility of distributional bootstrapping because such errors are only rarely observed in children's speech.

- (i).
- a. John likes fish.
 - b. John likes rabbits.
 - c. John can fish.
 - d. *John can rabbits.

However, an increasing body of work suggests that ambicategorical words need not pose any problem for learners. Although noun/verb homophones are present in speech to children (Conwell & Morgan, 2012; Lippeveld & Oshima-Takane, 2014; Nelson, 1995), words used as both noun and verb may be perceptually distinct depending on the category of use. In general, disyllabic noun/verb homophones in English are distinguished by changes in syllabic stress (e.g. PROject and proJECT; Kelly & Bock, 1988; Sereno & Jongman, 1995), and adult speakers use this information to guide their hypotheses about the meanings of nonsense words (Kelly, 1988). Monosyllabic ambicategorical words are longer when used as nouns than when used as verbs, but this effect disappears when sentence position is controlled for (Sorensen, Cooper & Paccia, 1978). Conwell & Barta (unpublished observations) find that sentence-medial noun uses of monosyllabic ambicategorical words are longer than sentence-medial verb uses of the same words, but sentence-final tokens do not exhibit such a difference. English sentence prosody creates lengthening and enhanced pitch contours at the ends of sentences (Shattuck-Hufnagel & Turk, 1996). Sentence-final prosody, therefore, may overwhelm the prosodic differences in noun/verb homophones. Even though these words are measurably different, those differences may not be perceived by listeners. However, adult listeners presented with isolated noun and verb tokens of real ambicategorical words show different event-related potential (ERP) responses to those words as a function of category (Conwell, 2015). Tokens of nonsense words produced in the same syntactic contexts do not elicit similar differences in ERP. If noun and verb uses of the same words are, in fact, prosodically distinct, then children might be able to maintain separate lexical entries for the noun form and the verb form of ambicategorical words during the earliest stages of development.

The findings from adult-directed speech do not necessarily translate to child-directed speech, as child-directed speech differs significantly from adult-directed speech (see Soderstrom, 2007, for a recent review of these differences). In speech to children, vowels and pauses both tend to be elongated (Bernstein Ratner, 1984, 1986; Fernald, Taeschner, Dunn, Papousek, de Boysson-Bardies & Fukui, 1989; Fisher & Tokura, 1996). Child-directed speech also contains exaggerated pitch contours relative to adult-directed speech (Ferguson, 1964; Fernald, 1989), and the vowel formant space in speech to children tends to be exaggerated as well (Bernstein Ratner, 1984; Cristia & Seidl, 2014; Kuhl *et al.*, 1997). These features make child-directed speech prosodically and phonetically distinct from adult-directed speech, which may affect the extent to which perceptual cues to the grammatical category of uses of

ambicategorical words are available in speech to children. The exaggerated nature of child-directed speech could enhance the differences between noun and verb tokens of the same words or, alternatively, that exaggeration could apply to all tokens, thus masking those cues. Further, most studies of prosodic cues to the category of noun/verb homophones have involved speech produced by reading, and the prosody of read speech is distinct from that of spontaneous speech (Howell & Kadi-Hanifi, 1991).

Previous work on noun/verb homophones in child-directed speech suggests that at least some perceptual cues are available to infants. Canadian-French-speaking mothers reliably disambiguate noun and verb uses of disyllabic nonsense words when reading to their infants (Shi & Moisan, 2008). Specifically, they increase both pitch and duration of the second vowel more for noun uses than for verb uses. A similar study of Mandarin-speaking mothers also found differences in pitch and the ratio of the first and second vowel durations as a function of lexical category (Li, Shi & Hua, 2010). However, both of those studies used disyllabic nonsense words, making their results hard to generalize to the case of English, as a preponderance of the ambiguous word types in speech to English-learning children are monosyllabic (Conwell, 2009). In an analysis of the stimuli used in their infant speech perception study, Conwell and Morgan (2012) reported reliable prosodic differences between noun and verb uses of English monosyllables taken from the speech of one mother to her child. Specifically, noun tokens were longer than verb tokens of the same words. They also reported higher first vowel formants in noun tokens than in verb tokens, resulting in a larger formant ratio for verbs than for nouns. Unfortunately, their study used only a small number of tokens from one mother and did not consider sentence position as a factor. Therefore, the prevalence of such cues in child-directed English remains unknown.

The study presented in this paper asks whether prosodic cues to lexical category are available for noun/verb homophones in child-directed speech. Using a naturalistic corpus of child-directed speech, it examines the relative contributions of grammatical category and sentence position to the realization of differences in the pronunciation of noun/verb homophones, as well as the roles of child age and category frequency in those effects. The features that characterize child-directed speech could have one of two possible effects on the differential production of noun and verb uses of ambicategorical words. One possibility is that the exaggerated nature of child-directed prosody will exaggerate these differences. An alternative, however, is that just as sentence-final prosody overwhelms the more subtle differences between noun and verb uses of words in adult-directed speech, the extreme prosody of child-directed speech may mask these prosodic distinctions. If the latter is the case, then prosodic cues to category are not helpful in disambiguating information for children. If, however, the former is the case, then these cues may allow children to maintain separate representations of the noun use and the verb use of ambicategorical words.

METHOD

Corpus

This study examines the maternal speech in the providence corpus (Demuth, Culbertson & Alter, 2006). This longitudinal corpus contains interactions of six mother-child dyads recorded for approximately 1 hour every 1–2 weeks beginning with the emergence of the

child's first words (ages 00;11–01;04) and continuing for two to three years. The dyads were recorded in their homes to obtain interactions that were as natural as possible. This corpus contains 364 hours of recordings and was selected because audio-recordings are available for all sessions. Details of the number of recordings and the ages of the children are presented in Table 1.

Procedure

To obtain a list of possible noun/verb homophones, the Brown Corpus (Francis & Kucera, 1982) was used to create a list of all words that were used at least once as a noun and at least once as a verb. This list contained 2,075 word types. The maternal speech in the Providence Corpus was searched for uses of all of these word types in sentence-medial and sentence-final position. To ensure adequate tokens to meaningfully compare the effects of both category and sentence position, only those word types that a mother used at least ten times in the middle of an utterance and at least ten times at the end of an utterance were targeted. Using the *kwal* function in CLAN (MacWhinney, 2000), every utterance containing one of the target words was pulled from the corpus and each use was categorized by hand as either a noun use or a verb use, based on the syntax of the utterance. Words that were used in isolation were not categorized and therefore not included in the analysis. Words that were used in only one category by mothers were not included in further analysis.

Following this categorization process, tokens of the eighty-one noun/verb homophone types used in both categories and in both utterance positions by at least one mother were extracted from the audio-recordings of the Providence Corpus. (See 'Appendix' for a complete list of word types.) A total of 2,775 tokens were included in the final analysis: 941 medial noun uses, 930 medial verb uses, 457 final noun uses, and 447 final verb uses. Trained research assistants measured these tokens using the PRAAT program (Boersma & Weenink, 2014). Boundaries were placed at the beginning and end of each token as well as at the beginning and end of the vowel. In the case of disyllabic words, vowel measurements were taken from the vowel in the stressed syllable. Placement of boundaries was based on a combination of auditory and visual examination of the waveforms and spectrograms. A PRAAT script was used to extract the duration of the token and of the vowel (in seconds), the mean, maximum, and minimum pitch in the token (in Hertz), and the first and second formant frequencies at the midpoint of the stressed vowel. Rate of pitch change was computed by taking the total change in pitch over the course of the token (maximum pitch–minimum pitch) and dividing by token duration (in seconds). The second formant frequency was divided by the first formant frequency to obtain the formant ratio.

In addition to category of use and sentence position, a number of covariates were included in the model. Phrase position, which was coded as either medial or final, was treated as a fixed factor. To control for possible effects of vowel type, vowel type was also included as a fixed factor. Because the frequency of a homophone meaning affects duration in adult-directed speech (Gahl, 2008), the relative frequency of use in each lexical category was included in the analysis as well. For each word type, the category of use was coded as either higher frequency or lower frequency, based on counts in the CELEX database (Baayen, Piepenbrock & Gulikers, 1995). To examine whether mothers adjust their use of prosodic

cues to noun and verb category as the child becomes more linguistically proficient, the child's age in months at the time the token was produced was also included in the analysis. Speaker identity was included as a random effect to account for individual differences in prosody. Word type was included to control for differences in length in phonemes.

The data were analyzed in R (R Core Team, 2015) using linear mixed models (*lme4*; Bates, Maechler, Bolker & Walker, 2015) with category, sentence position, phrase position, vowel type, child age, and relative frequency of category as fixed effects, and word type and speaker as random effects. Statistical significance for each factor was determined using likelihood ratio tests comparing the full model to a model without the factor under consideration. The interaction of category and sentence position was also included in the analysis.

RESULTS

Based on previous research, token and vowel duration were predicted to be affected by both category of use (conwell & Barta, unpublished observations; Li *et al.*, 2010; Shi & Moisan, 2008; Sorensen, *et al.*, 1978) and sentence position (Shattuck-Hufnagel & Turk, 1996). The results of the mixed model analyses for the durational measures (token and vowel duration) are presented in Table 2. First, we consider the random effects. Because the word types included in this analysis varied in number of phonemes, word type significantly affected token duration ($\chi^2(1) = 233.62$, $p < .001$) and vowel duration ($\chi^2(1) = 342.8$, $p < .001$). Speaker identity also produced significant differences in token duration ($\chi^2(1) = 65.55$, $p < .001$) and vowel duration ($\chi^2(1) = 45.22$, $p < .001$), likely due to individual differences in speaking rate.

Turning to the fixed effects, token duration showed significant effects of both category of use ($\chi^2(2) = 17.63$, $p < .001$) and sentence position ($\chi^2(2) = 188.38$, $p < .001$), as well as a significant interaction of category and sentence position ($\chi^2(1) = 11.83$, $p < .001$). Noun uses were significantly longer than verb uses in sentence-medial positions ($t(1816) = 7.876$, $p < .001$, Cohen's $d = .358$), but not in sentence-final positions ($t(8p6) = 0.277$, $p = .78$, Cohen's $d = .018$). Vowel duration showed a pattern similar to token duration, with significant effects of category ($\chi^2(2) = 13.5$, $p = .001$) and sentence position ($\chi^2(2) = 52.89$, $p < .001$), as well as a marginal interaction of category and sentence position ($\chi^2(1) = 3.01$, $p = .083$). Noun uses have significantly longer vowels than verb uses in medial positions ($t(1793) = 6.465$, $p < .001$, Cohen's $d = .295$), but not in final positions ($t(902) = 0.444$, $p = .657$, Cohen's $d = .029$). Token and vowel duration also showed significant effects of phrase position ($\chi^2(1) = 63.31$, $p < .001$; $\chi^2(1) = 58.26$, $p < .001$, respectively), while vowel duration was significantly affected by vowel type ($\chi^2(15) = 51.86$, $p < .001$), and token duration was marginally affected by vowel type ($\chi^2(15) = 24.44$, $p = .058$). Token and vowel duration were both affected by the relative frequency of the category of use, marginally in the case of token duration ($\chi^2(1) = 2.98$, $p = .085$) and significantly in the case of vowel duration ($\chi^2(1) = 3.95$, $p = .047$). These analyses show no effect of child age on either of the durational measures (both $\chi^2 < 2.2$, both $p > .1$). The durational data are presented in Figure 1.

Pitch information was predicted to be affected by sentence position and phrase position (Shattuck-Hufnagel & Turk, 1996) and by the age of the child (Soderstrom, 2007). The results of the mixed model analyses for the pitch measures are presented in Table 3. Mean pitch and rate of pitch change over the token were analyzed in separate mixed models. Speaker identity affected mean pitch ($\chi^2(1) = 25.33$, $p < .001$) as well as pitch change ($\chi^2(1) = 97.1$, $p < .001$), reflecting individual differences in speakers' use of pitch. There was also a significant effect of word type on both mean pitch ($\chi^2(1) = 26.7$, $p < .001$) and pitch change ($\chi^2(1) = 9.37$, $p = .002$).

When considering fixed effects, mean pitch and pitch change were expected to vary as a function of sentence and phrase position, but not necessarily as a function of lexical category. For mean pitch, this was the case, with a main effect of sentence position ($\chi^2(2) = 6.86$, $p = .032$), but neither a significant effect of lexical category nor a significant interaction (both $p > .2$). Pitch change, however, exhibited no main effect of lexical category, nor of utterance position, and no significant interaction (all $p > .17$). Pitch change was significantly affected by phrase position, with phrase-final tokens having greater pitch change than phrase-medial tokens ($\chi^2(1) = 6.6$, $p = .01$). Phrase position did not affect mean pitch ($\chi^2(1) = 1.36$, $p = .24$). Vowel type did not affect either pitch measure (both $\chi^2(15) < 19$, both $p > .2$). The age of the child also significantly affected both mean pitch ($\chi^2(1) = 42.74$, $p < .001$) and pitch change ($\chi^2(1) = 13.67$, $p < .001$), as the prosody of child-directed speech became less pronounced as the children aged. The relative frequency of category of use affected mean pitch ($\chi^2(1) = 8.75$, $p = .003$) as well as pitch change ($\chi^2(1) = 4.866$, $p = .027$). The pitch data are presented in Figure 2.

Vowel quality is affected by lexical category in disyllabic words (Kelly, 1988; Shi & Moisan, 2008). The formant ratio showed a significant random effect of word type ($\chi^2(1) = 195.36$, $p < .001$). Speaker identity also significantly affected the formant ratio ($\chi^2(1) = 28.43$, $p < .001$). The formant ratio of the vowel showed significant main effects of both category ($\chi^2(2) = 6.454$, $p = .04$) and position ($\chi^2(2) = 6.65$, $p = .036$), as well as a significant interaction ($\chi^2(1) = 6.338$, $p = .012$). Verb uses had more exaggerated formant ratios than noun uses in sentence-medial position ($t(1865) = 2.357$, $p = .018$, Cohen's $d = .109$), but not in sentence-final position ($t(901) = 0.926$, $p = .354$, Cohen's $d = .0620$). As anticipated, vowel type significantly affected the formant ratio ($\chi^2(15) = 116.77$, $p < .001$), but phrase position, child age, and relative frequency of category use did not (all $\chi(1) < 1.9$, all $p > .17$). The results of the mixed model analysis of formant ratio are presented in Table 4. The formant ratio data are presented in Figure 3.

In summary, this analysis found several differences between noun and verb tokens of ambicategorical words in natural child-directed speech, but only for sentence-medial uses. Noun tokens are longer than verb tokens and contain longer vowels. Noun tokens also exhibited greater pitch change than verb tokens, while verbs have more exaggerated formant ratios. Although these effects were only found for sentence-medial uses, they show prosodic differentiation of noun and verb uses of ambiguous words in children's natural experience.

DISCUSSION

This study asked whether the speech that children hear contains acoustic or prosodic information that would allow them to distinguish noun and verb uses of ambicategorical words. The results show that verb tokens are distinct from noun tokens in terms of their duration, vowel duration, pitch change, and formant ratio. In sentence-medial positions, noun tokens are longer and exhibit greater pitch change, and verb tokens have larger formant ratios. No differences between noun and verb uses were found in any of the measurements of tokens in sentence-final position. Although speakers show individual variation in speaking rate, and word type affects the acoustic cues as well, child age does not affect pronunciation of these words as a function of lexical category, and relative frequency of use in a category shows only small or marginal effects on duration, but larger effects on pitch and pitch range.

The results regarding duration are congruent with previous findings regarding the differentiation of noun and verb tokens of ambiguous words in adult-directed speech. Sorensen *et al.* (1978) reported differences in the duration of tokens sentence-medially, but not when sentence position was controlled for by placing the tokens at the end of the sentence. Conwell and Barta (unpublished observations) manipulated both phrase and sentence position and found that medial noun and verb tokens differed from one another along durational, pitch, and vowel quality measures, but that final tokens did not exhibit such differences. In that study, the lack of distinction between noun and verb tokens in final positions was attributed to overwhelming effects of sentence-/phrase-final prosody. Similarly, in this study, differences in duration and pitch created by sentence prosody are much larger than any category-level differences, thus sentence-level prosodic effects may mask the differences in noun and verb productions in sentence-final position in child-directed speech as they do in adult-directed speech. A more continuous measure of sentence position might reveal a more gradual change in the presence of acoustic differences over the course of an utterance as the availability of syntactic information for disambiguation improves. Naturalistic utterances, however, vary greatly in length and include sufficient disfluency and repetition that analyzing such an effect in these data is not feasible. An experimental study may be needed to examine this possibility. Still, the results presented here support the hypothesis that adults differentiate noun and verb uses of ambicategorical words when speaking to children.

The pitch and vowel quality results were only partly predicted from previous findings. That pitch and pitch change are affected by sentence position is unsurprising, as pitch accents tend to occur at the ends of sentences (Shattuck-Hufnagel & Turk, 1996). The greater pitch change in the noun tokens was unexpected, but may be an effect of these words receiving greater stress or being presented in didactic contexts in child-directed speech. The result that formant ratios are larger in verb tokens than in noun tokens was also unexpected. Formant ratios were included in this study as an indicator of vowel neutralization, which might be predicted in unstressed or shorter tokens. Somewhat surprisingly, verb tokens showed greater formant ratios than noun tokens, suggesting that vowels in ambiguous words used as verbs are less neutralized than the vowels in their noun counterparts. Why this might be the

case is unclear, but it is consistent with data from Conwell and Morgan's (2012) analysis of their stimuli.

One question that arises from these findings is how the effects seen in child-directed speech might differ (or not) from those in adult-directed spontaneous speech. To test this, one would wish to compare these data with measurements of noun/verb homophones from an adult-directed corpus. However, the child-directed speech corpora differ from the existing adult-directed speech corpora in a number of ways (e.g. familiarity of the speakers; presence in the same physical space; smaller numbers of word types) that make finding the right adult-directed control quite difficult. The results in the present study follow a pattern similar to those previously reported for adult-directed speech, but with somewhat greater differences in duration for noun and verb tokens. In adult-directed speech, noun tokens in sentence-medial position are longer than verb tokens by 20–30 ms; in the present study, the average difference between nouns and verbs for sentence-medial tokens was 42 ms. However, previous data on noun/verb differentiation in adult-directed productions come from studies of speech produced by reading, in order to parameterize factors such as length in phonemes, vowel type, phrasal position, and type frequency, so these differences should be interpreted with caution, as read speech has very different prosody from spontaneous speech (Howell & Kadi-Hanifi, 1991). What these data do show is that child-directed prosody does not mask the differences between noun/verb homophones that are present in adult-directed speech. If anything, it may exaggerate them.

Given that the differences between noun and verb tokens seem to be available only in sentence-medial positions and not in sentence-final positions, one might wonder how useful these cues are for child language learners. This is an important question and there are a few issues to consider when answering it. First, most of the (non-isolated) words that children hear are utterance-medial, which means that even information that is only available in sentence-medial cases is information that children have frequent access to. Second, in general, there is an asymmetry in the distribution of nouns and verbs across sentence positions. Nouns are more likely than verbs to appear in sentence-final positions and therefore to be subject to sentence-final lengthening (Shattuck-Hufnagel & Turk, 1996). Therefore, when sentence position is not parameterized (and it is not in natural speech), nouns overall will be distinct from verb tokens. That is, in children's natural experience, noun tokens will be longer than verb tokens sentence-medially and there will simply be more noun tokens than verb tokens that are subject to sentence-final lengthening. Finally, it is worth considering how sentence position affects word learning. Golinkoff and Alioto (1995) reported that adult participants show better segmentation and recall of words in an unfamiliar language when those words are in sentence-final positions in infant-directed speech, which would suggest that cues confined to sentence-medial position might not be especially useful for learners. However, infants are able to segment and recall sentence-medial words that co-occur with high-frequency lexical items (Bortfeld, Morgan, Golinkoff & Rathbun, 2005), and they might, therefore, be able to use the differences in sentence-medial tokens to maintain separate categories. Future research could consider how these factors interact in a word-learning task.

Of course, acoustic and prosodic information is only one possible source of evidence for children who are categorizing tokens of ambicategorical words. Another important cue could be the semantic context in which such words are used. Cases of spontaneous denominalization and deverbalization by children have been documented and constitute evidence that preschool-aged children have an understanding that some words can be used as both nouns and verbs (Bowerman, 1982; Clark, 1982; Lippeveld & Oshima-Takane, 2014; Oshima-Takane, Barner, Elsabbagh & Guerriero, 2001). Two-and-a-half-year-old French-learning children are able to comprehend deverbal uses of novel nouns (e.g. *vop* to mean the action done with that object), provided that their mothers produce a large number of cross-category uses in spontaneous speech (Lippeveld & Oshima-Takane, 2014). By three years of age, French-learners are adept at interpreting novel denominal verbs (Lippeveld & Oshima-Takane, 2015). However, these studies examined only instrumental verb uses of novel words. Older children (aged 4 and 5 years) often interpret even familiar source verbs (e.g. *milk*) as goal verbs, suggesting that the concepts necessary for correctly selecting the appropriate meaning of a denominal verb are still developing at those ages (Srinivasan & Barner, 2013). Likewise, Bushnell and Maratsos (1984) reported that seven-year-olds performed better at appropriately producing novel verb uses of familiar nouns (e.g. *spooning* to mean using a spoon to transfer an object) than five-year-olds did. Taken together, the studies described here indicate that semantic information is undoubtedly important and useful to older children who are learning about polysemous noun/verb pairs, but that such information may not be available to younger children for reasons related to conceptual development.

Based on the findings in this and previous papers, children likely use both bottom-up information, such as the prosodic cues described here and elsewhere, and top-down information, such as lexical semantics and world knowledge, to learn about ambicategorical words. These two types of information are available to different degrees across development. The prosodic information described here is perceived by infants (Conwell & Morgan, 2012), while children may not have the ability to use world knowledge to interpret cross-category uses until late in their preschool years (Srinivasan & Barner, 2013). The bottom-up cues described here may be most useful to children who are engaged in initial attempts at categorizing words based on their distributions (e.g. Mintz, 2003). For those children, *kiss* the noun and *kiss* the verb may sound sufficiently distinct to be treated as two different items, and thus not confound the process of sorting words into distributional categories. As children mature, they may be able to use their developing conceptual structures to fine-tune those categories with semantic information.

The data presented here show that mothers disambiguate noun and verb uses of ambicategorical words when speaking to their children, but only when those uses are in the middle of utterances. This effect mirrors previous findings on adult-directed speech, in which noun and verb uses of the same word types are differentiated sentence-medially, but sentence-final prosody eliminates these differences (Conwell & Barta, unpublished observations; Sorensen *et al.*, 1978). The presence of prosodic cues to lexical category in child-directed speech could prevent language learners from conflating noun and verb uses of ambicategorical words, allowing them to sort words into grammatical categories while sidestepping the potential pitfalls of ambiguity.

Appendix: word types included in analysis

back	guess	ride
bit	hammer	ring
bite	hand	rock
bow	help	run
break	hug	saw
brush	kiss	scoop
building	leaves	set
call	line	sleep
catch	look	slide
check	love	snap
circle	match	snow
clean	matter	sound
color	mess	spin
count	mix	stick
crash	name	sticks
cut	pack	stuff
dance	page	swing
drawing	paint	tie
dress	painting	toast
drink	plant	train
end	play	try
face	pop	turn
fall	press	walk
fit	puzzle	watch
flower	rain	water
fly	reading	wonder
go	rest	work

REFERENCES

- Baayen, HR., Piepenbrock, R., Gulikers, L. The CELEX lexical database. 1995. Online: <<http://celex.mpi.nl/>>
- Bates, D., Maechler, M., Bolker, BM., Walker, S. Fitting linear mixed-effects models using lme4. ArXiv e-print. 2015. online: <<http://arxiv.org/abs/1406.5823>>
- Bernstein Ratner N. Patterns of vowel modification in mother-child speech. *Journal of Child Language*. 1984; 11:557-78. [PubMed: 6501465]
- Bernstein Ratner N. Durational cues which mark clause boundaries in mother-child speech. *Phonetics*. 1986; 14:303-9.
- Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 5.3.672014. online: <<http://www.praat.org/>>
- Bortfeld H, Morgan J, Golinkoff RM, Rathbun K. Mommy and me: familiar names help launch babies into speech stream segmentation. *Psychological Science*. 2005; 16:298-304. [PubMed: 15828977]
- Bowerman M. Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni di Semantica*. 1982; 3:5-66.

- Bushnell EW, Maratsos MP. 'Spooning' and 'basketing': children's dealing with accidental gaps in the lexicon. *Child Development*. 1984; 55:893–902. [PubMed: 6734324]
- Cartwright TA, Brent MR. Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cognition*. 1997; 63:121–70. [PubMed: 9233082]
- Clark, EV. The young word maker: a case study of innovation in the child's lexicon. In: Wanner, E., Gleitman, LR., editors. *Language acquisition: the state of the art*. Cambridge University Press; Cambridge: 1982. p. 309–426.
- Conwell, E. Unpublished doctoral dissertation. Brown University; Providence, RI: 2009. Resolving ambicategoricity in language acquisition: the role of perceptual cues.
- Conwell E. Neural responses to category ambiguous words. *Neuropsychologia*. 2015; 69:85–92. [PubMed: 25637057]
- Conwell E, Morgan JL. Is it a noun or is it a verb? Resolving the ambicategoricity problem. *Language Learning and Development*. 2012; 8:87–112.
- Cristia A, Seidl A. The hyperarticulation hypothesis of child-directed speech. *Journal of Child Language*. 2014; 41:913–34. [PubMed: 23406830]
- Demuth K, Culbertson J, Alter J. Word-minimality, epenthesis and coda licensing in the acquisition of English. *Language and Speech*. 2006; 49:137–74. [PubMed: 17037120]
- Ferguson CA. Baby talk in six languages. *American Anthropologist*. 1964; 66:103–14.
- Fernald A. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*. 1989; 60:1497–510. [PubMed: 2612255]
- Fernald A, Taeschner T, Dunn J, Papousek M, de Boysson-Bardies B, Fukui I. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*. 1989; 16:477–501. [PubMed: 2808569]
- Fisher C, Tokura H. Acoustic cues to grammatical structure in infant-directed speech: crosslinguistic evidence. *Child Development*. 1996; 67:3192–218. [PubMed: 9071777]
- Francis, WN., Kucera, H. *Frequency analysis of English usage: lexicon & grammar*. Houghton Mifflin; Boston, MA: 1982.
- Gahl S. *Time and thyme* are not homophones: the effect of lemma frequency on word durations in spontaneous speech. *Language*. 2008; 84:474–96.
- Gertner Y, Fisher C, Eisengart J. Learning words and rules: abstract knowledge of word order in early sentence comprehension. *Psychological Science*. 2006; 17:684–91. [PubMed: 16913951]
- Gleitman, LR., Wanner, E. *Language acquisition: the state of the state of the art*. In: Wanner, E., Gleitman, LR., editors. *Language acquisition: the state of the art*. Cambridge University Press; Cambridge: 1982. p. 3–49.
- Golinkoff RM, Alioto A. Infant-directed speech facilitates lexical learning in adults hearing Chinese: implications for language acquisition. *Journal of Child Language*. 1995; 22:703–26. [PubMed: 8789520]
- Höhle B, Weissenborn J, Kiefer D, Schulz A, Schmitz M. Functional elements in infants' speech processing: the role of determiners in syntactic categorization of lexical elements. *Infancy*. 2004; 5:341–53.
- Howell P, Kadi-Hanifi K. Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*. 1991; 10:163–9.
- Kelly MH. Phonological biases in grammatical category shifts. *Journal of Memory and Language*. 1988; 27:343–58.
- Kelly MH, Bock JK. Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*. 1988; 14:389–403. [PubMed: 2971769]
- Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina V, et al. Cross-language analysis of phonetic units in language addressed to infants. *Science*. 1997; 277:684–6. [PubMed: 9235890]
- Li A, Shi R, Hua W. Prosodic cues to noun and verb categories in infant-directed Mandarin speech. *Speech Prosody*. 2010; 100088:1–4.
- Lippeveld M, Oshima-Takane Y. The effect of input on children's cross-categorical use of polysemous noun–verb pairs. *Language Acquisition*. 2014; 22:209–39.

- Lippeveld M, Oshima-Takane Y. Nouns to verbs and verbs to nouns: When do children acquire class-extension rules for deverbal nouns and denominal verbs? *Applied Psycholinguistics*. 2015; 36:559–88.
- MacWhinney, BJ. *The CHILDES project: tools for analyzing talk*. 3rd ed. Erlbaum; Mahwah, NJ: 2000.
- Maratsos, MP., Chalkley, MA. The internal language of children's syntax: the ontogenesis and representation of syntactic categories. In: Nelson, K., editor. *Children's language*. Vol. Vol. 2. Gardner Press; New York: 1980. p. 127-214.
- Mintz TH. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*. 2003; 90:91–117. [PubMed: 14597271]
- Mintz TH, Newport EL, Bever TG. The distributional structure of grammatical categories in speech to young children. *Cognitive Science*. 2002; 26:393–424.
- Monaghan P, Chater N, Christiansen MH. The differential role of phonological and distributional cues in grammatical categorization. *Cognition*. 2005; 96:143–82. [PubMed: 15925574]
- Nelson, K. The dual category problem in the acquisition of action words. In: Tomasello, M., Merriman, WE., editors. *Beyond names for things: young children's acquisition of verbs*. Erlbaum; Mahwah, NJ: 1995. p. 223-250.
- Oshima-Takane, Y., Barner, D., Elsabbagh, M., Guerriero, AMS. Learning of deverbal nouns. In: Almgren, M, Barreña, A, Ezeizabarrena, M-J, Idiazabal, I., MacWhinney, B., editors. *Research in language acquisition: proceedings of the 8th congress of the International Association for the Study of Child Language*. Cascadilla Press; Somerville, MA: 2001. p. 1154-1170.
- Pinker, S. The bootstrapping problem in language acquisition. In: MacWhinney, B., editor. *Mechanisms of language acquisition*. Erlbaum; Hillsdale, NJ: 1987. p. 399-442.
- Pinker, S. *Learnability and cognition: the acquisition of argument structure*. MIT Press; Cambridge, MA: 1989.
- R Core Team. R Foundation for Statistical Computing. Vienna: 2015. R: A language and environment for statistical computing. Online: <<http://www.R-project.org/>>
- Redington M, Chater N, Finch S. Distributional information: a powerful cue for acquiring syntactic categories. *Cognitive Science*. 1998; 22:425–69.
- Sereno JA, Jongman A. Acoustic correlates of grammatical class. *Language and Speech*. 1995; 38:57–76.
- Shattuck-Hufnagel S, Turk AE. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*. 1996; 25:193–247. [PubMed: 8667297]
- Shi R, Melançon A. Syntactic categorization in French-learning infants. *Infancy*. 2010; 15:517–33.
- Shi, R., Moisan, A. Prosodic cues to noun and verb categories in infant-directed speech. In: Chan, H, Jacob, H., Kapia, E., editors. *Proceedings of the 32nd Annual Boston University Conference on Language Development*; Somerville, MA: Cascadilla Press; 2008. p. 450-461.
- Soderstrom M. Beyond babytalk: re-evaluating the nature and content of speech input to pre-linguistic infants. *Developmental Review*. 2007; 27:501–32.
- Soderstrom M, White KS, Conwell E, Morgan JL. Receptive grammatical knowledge of familiar content words and inflection in 16-month-old infants. *Infancy*. 2007; 12:1–29.
- Sorensen JM, Cooper WE, Paccia JM. Speech timing of grammatical categories. *Cognition*. 1978; 6:135–53. [PubMed: 679646]
- Srinivasan M, Barner D. The Amelia Bedelia effect: world knowledge and the goal bias in language acquisition. *Cognition*. 2013; 128:431–50. [PubMed: 23787375]
- Yuan S, Fisher C. “Really? She blicked the baby?” Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*. 2009; 20:619–26. [PubMed: 19476591]

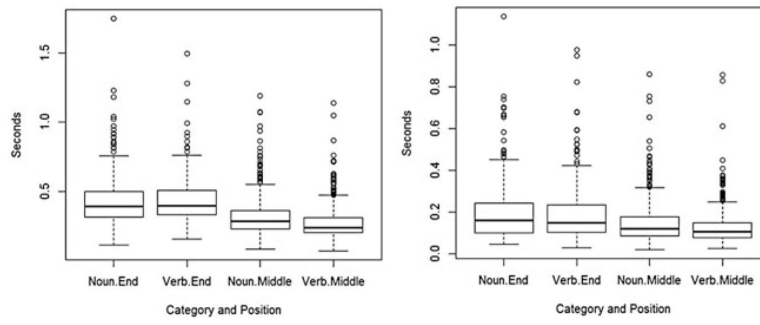


Fig. 1. Token duration (left) and vowel duration (right) in seconds of noun/verb homophones by category and sentence position.

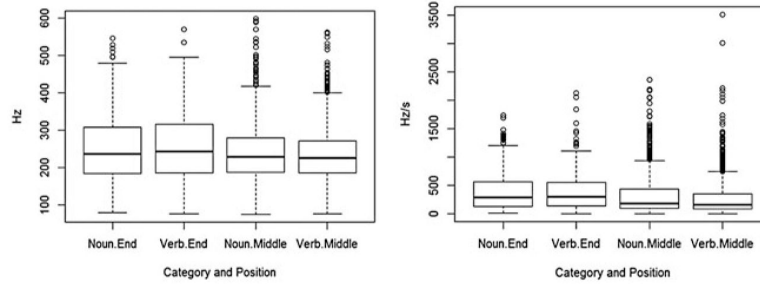


Fig. 2. Mean pitch in Hz (left) and pitch change over the token (right) of noun/verb homophones by category and sentence position.

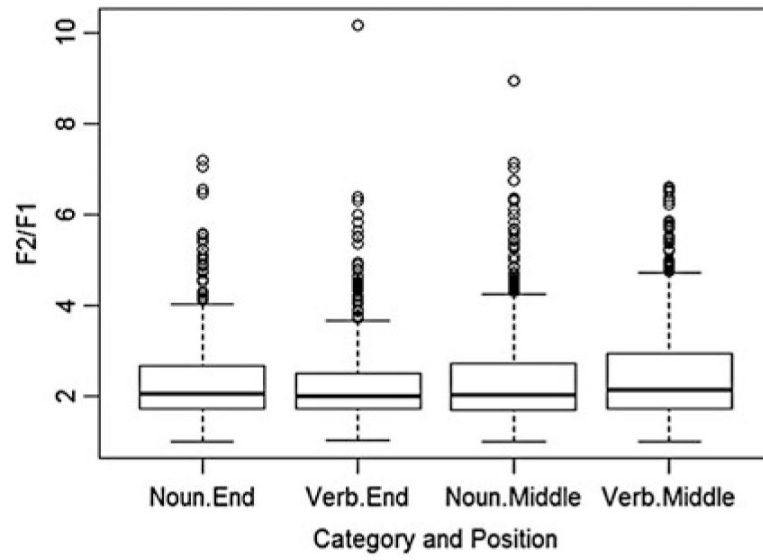


Fig. 3. Ratio of first and second vowel formants at the midpoint of the stressed vowel by category and sentence position.

TABLE 1

Number of recordings and ages for each dyad in the corpus

Child	Ages	Total recordings
Alex	1;04–3;05	51
Ethan	0;11–2;11	50
Lily	1;01–4;00	80
Naima	0;11–3;10	87
Violet	1;02–3;11	54
William	1;04–3;04	44

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 2

Results of the linear mixed effects models for token and vowel duration. Statistical significance was determined using likelihood ratio tests

Effect	Parameter	Token duration		Vowel duration	
		Estimate	SE	Estimate	SE
<i>Fixed effects</i>					
Intercept	β	.42649	.04173	.23299	.02573
Age of child	β	-.00017	-.00032	-.0003	-.0002
Relative frequency of category	β	-.00781	-.00452	-.00562 *	-.00283
Vowel type	β	-.03036	-.058	-.0371 ***	-.03619
Phrase position	β	-.05566 ***	-.00695	-.03337 ***	-.00435
Category	β	-.00457 ***	-.00784	-.00301 **	-.0049
Sentence position	β	-.09144 ***	-.00801	-.02792 ***	-.00501
Category \times sentence position	β	-.02035	-.00973	-.01057	-.00609
<i>Random effects</i>					
Word type	σ^2	.00391 ***	.0069	.00152 ***	.0043
Speaker	σ^2	.00076 ***	.01124	.0002 ***	.0058
Residual	σ^2	.0135	.00221	.00528	.02968

* NOTES: $p < .05$,

**
 $p < .01$,

 $p < .001$.

TABLE 3

Results of the linear mixed effects models for mean pitch and pitch change. Statistical significance was determined using likelihood ratio tests

Effect	Parameter	Pitch		Pitch change	
		Estimate	SE	Estimate	SE
<i>Fixed effects</i>					
Intercept	β	279.13	16.58	410.6	64.381
Age	β	1.449 ^{***}	0.2204	3.514 ^{***}	0.948
Relative frequency of category	β	9.149 ^{**}	3.09	29.517 [*]	13.375
Vowel type	β	17.651	21.63	69.45	78.208
Phrase position	β	5.533	4.736	52.542 [*]	20.435
Category	β	2.634	5.357	9.366	23.182
Sentence position	β	8.461 [*]	5.407	23.218	23.254
Category \times sentence position	β	7.789	6.658	22.965	28.81
<i>Random effects</i>					
Word type	σ^2	305 ^{***}	1.928	2333 ^{**}	5.334
Speaker	σ^2	184.2 ^{***}	5.54	4578 ^{***}	27.62
Residual	σ^2	6258.1	1.509	117327	6.535

* NOTES: $p < .05$,

** $p < .01$,

*** $p < .001$.

TABLE 4

Results of the linear mixed effect model for the ratio of first and second vowel formants at the midpoint of the stressed vowel. Statistical significance was determined using likelihood ratio tests

Effect	Parameter	F_2/F_1	
		Estimate	SE
<i>Fixed effects</i>			
Intercept	β	1.693	.1934
Age	β	0.001	.002
Relative frequency of category	β	0.0202	.0283
Vowel type	β	0.7955***	.2677
Phrase position	β	0.0599	.0435
Category	β	0.0875*	.0491
Sentence position	β	0.0816*	.0499
Category \times sentence position	β	0.1537*	.061
<i>Random effects</i>			
Word type	σ^2	0.0694***	.0291
Speaker	σ^2	0.0127***	.06
Residual	σ^2	0.5302	.0138

* NOTES: $p < .05$,

** $p < .01$,

*** $p < .001$.