Quantifying Lexical Ambiguity in Speech To and From English-Learning Children

Stephan C. Meylan*(smeylan@mit.edu)
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology
Department of Psychology and Neuroscience, Duke University
Jessica Mankewitz* (jmank@stanford.edu)
Department of Psychology, Stanford University
Sammy Floyd (sfloyd@princeton.edu)
Department of Psychology, Princeton University
Hugh Rabagliati (hugh.rabagliati@ed.ac.uk)
School of Philosophy, Psychology and Language Sciences, University of Edinburgh
Mahesh Srinivasan (srinivasan@berkeley.edu)
Department of Psychology, University of California, Berkeley

Abstract

Because words have multiple meanings, language users must often choose appropriate meanings according to the context of use. How this potential ambiguity affects first language learning, especially word learning, is unknown. Here, we present the first large-scale study of how children are exposed to, and themselves use, ambiguous words in their language learning environments. We tag 180,000 words in two longitudinal child language corpora with word senses from WordNet, focusing between 9 and 51 months and limiting to words from a popular parental vocabulary report. We then compare the diversity of sense usage in adult speech around children to that observed in a sample of adult language, as well as the diversity of sense usage in children's own productions. To accomplish this we use a Bayesian model-based estimate of sense entropy, a measure of diversity that takes into account uncertainty inherent in small sample sizes. This reveals that sense diversity in caregivers' speech to children is similar to that observed in a sample of adult-directed written material, and that children's use of nouns—but not verbs—is similarly diverse to that of adults. Finally, we show that sense entropy is a significant predictor of vocabulary development: children begin to produce words with a higher diversity of adult sense usage at later ages. We discuss the implications of our findings for theories of word learning.

Keywords: word learning; lexical semantics; corpus annotation; child language acquisition; polysemy; homonymy; WordNet

Introduction

When adults talk to one another, they use language that is rife with ambiguous words. The vast majority of frequent word forms (*e.g., nail, line, bottle, hold*) are associated with more than one meaning (Zipf, 1935). How do children come to learn the various meanings associated with each word form?

Both classic and contemporary theories of word learning hold that children assume that a single word will carry a single meaning, as this bias is believed to facilitate lexical development (Markman, 1989; Trueswell et al., 2013). These theories thus predict that children should struggle to learn multiple meanings for words, and may only learn these word meanings after an initial, simplifying one-to-one assumption is abandoned. Contrary to these predictions, recent work suggests that by the early preschool years, children have learned multiple meanings for many familiar words (Srinivasan & Snedeker, 2011; Rabagliati et al., 2010; Floyd et al., 2020). Moreover, by 3 to 4 years of age, children can simultaneously learn and retain multiple meanings for novel words (Srinivasan et al., 2019; Floyd & Goldberg, 2020). However, these laboratory experiments do not inform us about the real-life conditions under which children learn ambiguous words.

For instance, we currently know little about the amount and types of lexical ambiguity that children encounter in their language environments, or how this might change over development. Is ambiguity widespread in speech to children, as it is in adult-directed speech? Or might caregivers avoid using ambiguous words when speaking to children, similar to how they employ simpler words (Golinkoff et al., 2015; Soderstrom, 2007)? If the input to children contains minimal ambiguity, then it may not pose a problem for contemporary single-meaning theories of word learning. Alternatively, children may hear extensive ambiguity from the outset of development, which would constitute a challenge for existing models (*e.g.,* (Trueswell et al., 2013; Stevens et al., 2017).

Laboratory experiments also tell us little about how much ambiguity children use themselves in naturalistic environments. For example, we do not know when children first start to spontaneously use words with multiple meanings, or whether they are able to learn to use words with the wide range of meanings that adults use. Moreover, we do not know whether words that are used with more meanings–which appear in more diverse contexts–are more or less challenging for children to learn (Roy et al., 2015 vs. Hills et al., 2010). The questions highlighted above — which have broad implications not only for the acquisition of ambiguous words but for theories of word learning more broadly — can only be answered using observational, naturalistic methods. The present study is an initial step toward that goal.

Characterizing and quantifying the ambiguity in language to and from children presents a number of logistical and computational challenges, from the labor that is required to tag corpora (since words are not typically annotated for their specific meanings), to the development of analytic procedures that can cope with the sparsity within these data. The existence of these challenges can perhaps explain why this topic has rarely been addressed. To our knowledge, there have only

*These authors contributed equally to this work.

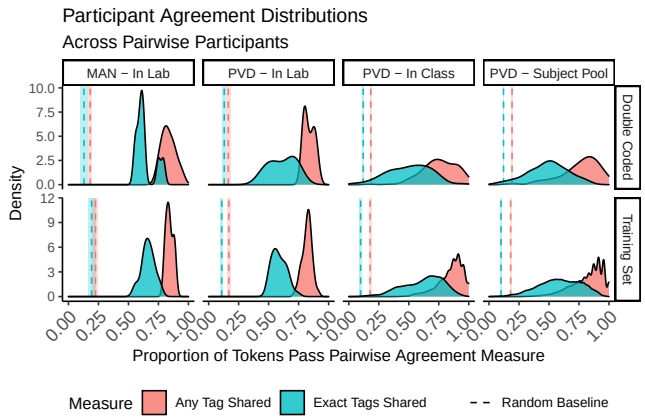

Figure 1: Pairwise agreement between annotators of each participant type (In Lab vs In Class vs Subject Pool) and corpus (Manchester (MAN) vs Providence (PVD)).

been two prior large-scale investigations of ambiguity in child language (Casas et al., 2018, 2019), but neither tagged words with meanings in context.

Below, we describe our solutions to these challenges, including newly sense-annotated corpora and a set of novel analytic tools. These allow us to answer, for the first time, three basic questions about the diversity of word meanings in speech to and from children: whether speech to children reflects fewer senses than language directed to adults, how children's use of word senses compares with that of adults, and whether greater diversity of meanings in the input facilitates or inhibits word learning.

Methods

Corpora

We tagged two English language corpora with WordNet 3.0 (Miller, 1995) senses: Providence (Demuth et al., 2006) and Manchester (Theakston et al., 2001). They were chosen as they provide longitudinal coverage at a critical age range (9-51 months in Providence; 18-39 months in Manchester). Additionally, these two corpora offer two varieties of English: British English through the Manchester corpus and American English through the Providence corpus. We used WordNet because it is the most commonly used ontology for word sense tagging, and allows comparison to adult corpora (Miller et al., 1993). WordNet should only be taken as a rough model of the range of word meanings entertained by people, in that the discrete set of senses for each word may be more or less granular than those used in an individual's lexicon (Gangemi et al., 2001), and that it overlooks richly structured relationships between different meanings (Nair et al., 2020). Despite these shortcomings, the WordNet ontology allows us to characterize meanings of a very broad range of words with respect to a reasonable first-pass characterization of the adult language.

Because tagging *all* words with their word senses would be prohibitively resource-intensive, we restricted the set of word tokens receiving tags to the set of word types that are also present in the Communicative Development Index (CDI), a common measure of child vocabulary from parental report (Fenson et al., 2007).* Additionally, we downsampled highly frequent words: For each word type, we tagged a sample of up to 50 tokens in each 3-month interval of child age for each speaker role (child or parent).

Annotation

The annotators comprised research assistants, undergraduate students, and subject pool participants, drawn from three different institutions (Tab. 1). To account for word use differences between the two varieties of English, UK based annotators primarily tagged transcripts from the Manchester corpora while US based annotators were assigned transcripts from the Providence corpora. They used a web app built on top of the data architecture of childes-db, version 2018.1 (Sanchez et al., 2019), a database mirror of transcripts in CHILDES (MacWhinney, 2000). Upon clicking on a word for transcription, a panel is populated with possible tags from WordNet (Miller, 1995), taking into account the lemma and the broad part of speech tag (Fig. 2).

For each token, the annotator was instructed to select *all* of the WordNet senses that could reasonably apply, given the surrounding transcript. We allowed annotators to select multiple options motivated by the fact that many WordNet senses are similar, and that it might be difficult to discern which sense is used in a given context. To address cases where the meaning of the word was unclear to the annotator, or the sense was not attested in the WordNet inventory, the annotators could also choose "I don't know" or "Other meanings (none of the below)." Additionally, annotators had the option to flag a token as "Wrong Part of Speech" in cases where the CHILDES part of speech tag caused the system to present inappropriate sense options (5.0% of words). Though videos exist for the Providence corpus, annotators were presented with only the transcript.

Hired, in-lab annotators were assigned a training transcript according to their assigned corpus (597 tokens for Providence annotators and 208 tokens for Manchester annotators). All undergraduate student and subject pool participants were assigned one of five 25-token-long segments from a Providence in-lab training transcript.

*We used the North American English CDI forms for the Manchester corpus because the North American English forms were systematically more common the corresponding ones on the British CDI (*e.g., truck* vs. *lorry*)

Table 1: Contributions from different annotator types.

Type	Institution	Taggers	N. Tokens	% Tokens
RA Staff	Edinburgh	7	67,770	37.78%
RA Staff	Berkeley	11	36,983	20.61%
Subj Pool	Berkeley	506	39,395	21.96%
Subj Pool	Princeton	276	33,588	18.72%
Subj Pool	Edinburgh	18	2,252	1.26%
Class	Berkeley	208	27,807	15.5%

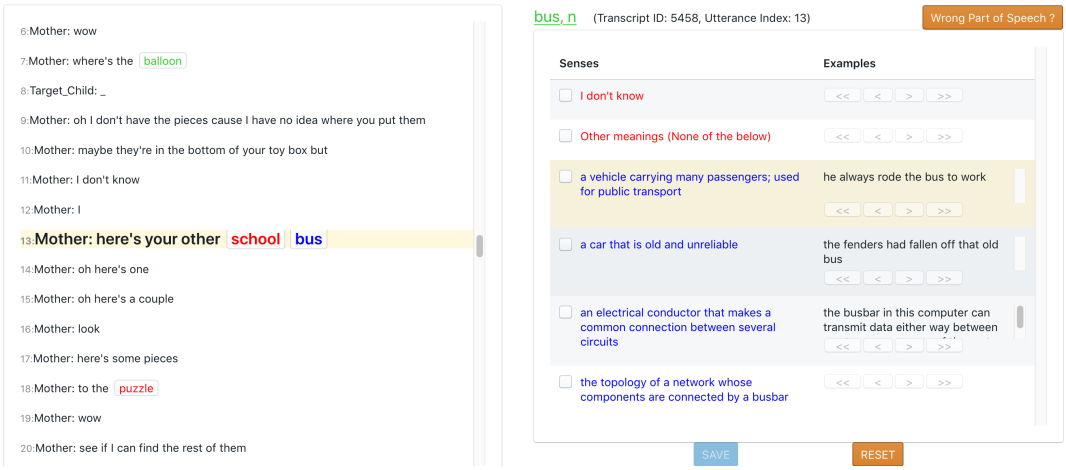

Figure 2: Annotation interface. Participants tag selected words in CHILDES corpora (left) with senses from WordNet (right).

Agreement and reliability

To assess inter-annotator agreement, we compared results for the annotators' assigned training segments, as well as their performance on the 21.28% (38,022) of tokens that we ensured were double coded. We computed pairwise inter-annotator agreement in two ways: a maximally charitable version where participants "agreed" if they shared at least one common tag for a token, and a minimally charitable version where participants only "agreed" if they provided the exact same set of tags for a token (see Fig. 1 for pairwise agreement distributions). An individual participant's agreement score was computed as the average agreement with all other participants of the same participant type. Participants who had an agreement score two standard deviations below the mean agreement for their training segment were dropped from further analysis (n=82, n=887 remaining).

This procedure often yielded multiple tags for a given word token. For analyses requiring a single tag for each token, we took the majority tag or sampled one of the highest-frequency tags when there was no majority. The resulting majority-tag dataset consists of 61,888 child and 112,802 adult word tokens, covering 701 and 719 unique word lemma+part of speech combinations, respectively, from 18 child-caregiver pairs from the Providence and Manchester corpora. This collection of tags reflects approximately 2,700 hours of annotation time.

Analyses

We conduct three analyses to answer basic questions about the diversity of word meanings in speech to and from children and their consequences for word learning. The annotation interface code, coding manual, and analysis code can be accessed at `https://osf.io/9uqrv`. The complete dataset will be shared upon the completion of annotation.

Two central questions are whether child-directed language reflects less diverse sense usage than adult-directed language (Analysis 1a), and whether children use word senses in similar ways to adult caregivers (Analysis 1b). To responsibly compare the diversity of sense usage across language samples, we need a measure that takes into account both the total number of senses as well as the probability of each of those senses for a given word. Further, the method needs to take into account the amount of data available: small sample sizes should be reflected in uncertainty in the estimate of sense diversity. Here we use the measure of *sense entropy*.

We treat the choice of word sense for a given word type as a discrete random variable (*i.e.,* a many-sided weighted die), over which we can compute entropy, or the average level of uncertainty over the word's possible meanings in the absence of context. Sense entropy $H(X)$ is calculated as:

$$H(X) = -\sum_{i=1}^{n} P(x_i) log P(x_i), \quad (1)$$

where $P(x_i)$ is the normalized tag frequency of the ith sense for that word type. Intuitively, both more biased uses of a word type (*i.e.,* heavy preference for one sense) as well as the word type having fewer senses overall both result in lower

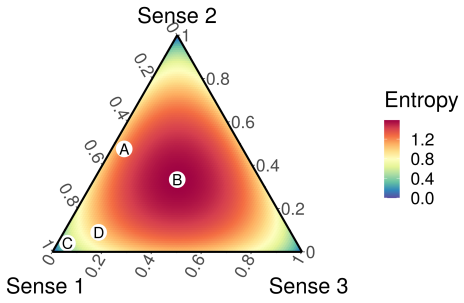

Figure 3: Possible sense entropy values for a word type with 1-3 senses. The three sides correspond to the probabilities of the three senses. A: A word type where Sense 1 and 2 are equiprobable and sense 3 unused (1 bit). B: A word type with maximum entropy for a word with 3 senses (all three senses are equiprobable, $\approx$ 1.58 bits). C: a word type with minimal entropy, where only Sense 1 is observed (0 bits). D: Example of a sense usage that is approximately 1 bit though all three senses have nonzero probability (compare with A).

sense entropy. Only observed senses contribute to the entropy estimate, in that unobserved senses carry no weight ($P(x_i) = 0$). This contrasts with the approach of computing a proportion, where the denominator would reflect the number of possible senses for that word in WordNet (*i.e.,* adult-specific and rare uses).

The maximum likelihood estimate (*i.e.,* point estimate) of sense entropy is likely to be a significant *under*estimate whenever a word type has a small number of tagged uses. This is because a small sample is unlikely to contain low frequency senses for rare words. For this reason, we treat the "true" sense probabilities of a word type ($\theta := p(x_1), ..., p(x_k), \sum\theta = 1$) as a latent variable, and estimate it for each partition using a Bayesian model, specifically a Dirichlet-multinomial model. Taking advantage of the fact that 55% of adult productions correspond to the first listed sense in WordNet, we assume a weak asymmetric Dirichlet prior that assigns an average 55% of the probability to the first-listed sense and approximately evenly distributes the remaining probability among the other senses ($\alpha_1 = 1$, $\alpha_{2,...,k} = \frac{k-1}{.8}$). To ensure that the prior provides the same contribution to the posteriors for both samples, we draw a subset of tags without replacement from the dataset with more observations for each word type.

Fitting the Dirichlet-multinomial model with Gibbs sampling in JAGS (Plummer, 2003), we sample from the posterior over sense probabilities, and compute entropy directly from θ in each Markov chain Monte Carlo sample. When there are few tags and many senses, the estimates of sense probabilities (and hence sense entropy) captured in the MCMC chains have high variance, such that entropy estimates are minimally constrained. On the other hand, if a word type has many tags, then its sense entropy estimate is much more constrained.

For null hypothesis significance testing, we fit the Dirichlet-multinomial models to the two samples and compute the difference between entropy estimates. We then check whether 99% of the resulting distribution lies above or below 0, and report the number of word types reaching significance.

Analysis 1a: Child-directed vs. adult-directed language

How does the ambiguity in child-directed language compare to the ambiguity in language directed to adults? To evaluate this, we compare model-based entropy estimates for adult speech in our child language corpus (reflecting largely child-directed speech, though also some overheard speech between adults or speech directed at older siblings) with adult-directed written language in Semcor (Miller et al., 1993).

Results Figure 5 shows that speech to children exhibits comparable diversity in sense usage compared to a sample of adult-directed written material. Specifically, for the large majority of words examined, child-directed sense entropy was similar to adult-directed sense entropy. This suggests that adults do not systematically avoid ambiguity when talking to children. However, there were large differences in how the words were *used* across the two environments. For example, *fix* as in "restore by replacing a part" is the most common meaning of the word in child environments, while other uses such as "fix a meal" (*cook.v.02*) are common in adult environments. By contrast, the sense for *full* in "a full glass" is more common in adult written material than other senses, such as "full from dinner" (Fig. 7A). Intuitively, many of these differences in word usage stem from contextual differences between adult written text and child-oriented home recordings.

Analysis 1b: Child-produced vs. adult-produced speech

Do young children use a variety of word senses in their own speech, and how might this compare to the ways in which caregivers use words? Because most of our target words are relatively low frequency, a comparison of sense usage *within* a single child's family yields low precision estimates of any differences. Instead, we make a "megachild" (and analogous"megacaregiver") assumption: sense usage from all children (and caregivers) can be combined as representative of an "average" English-learning child (and caregiver; note we made the same simplifying assumption for child-directed speech in Analysis 1).

Results Sense entropy estimates from the Dirichlet-multinomial models (Fig. 6) reveal that children use most words with comparable sense diversity to adults (*i.e.,* few words are above the diagonal, and child sense entropies are nonzero). However, adults use a more diverse range of senses for verbs and adverbs, such as *cry* (esp. using it to mean *to*

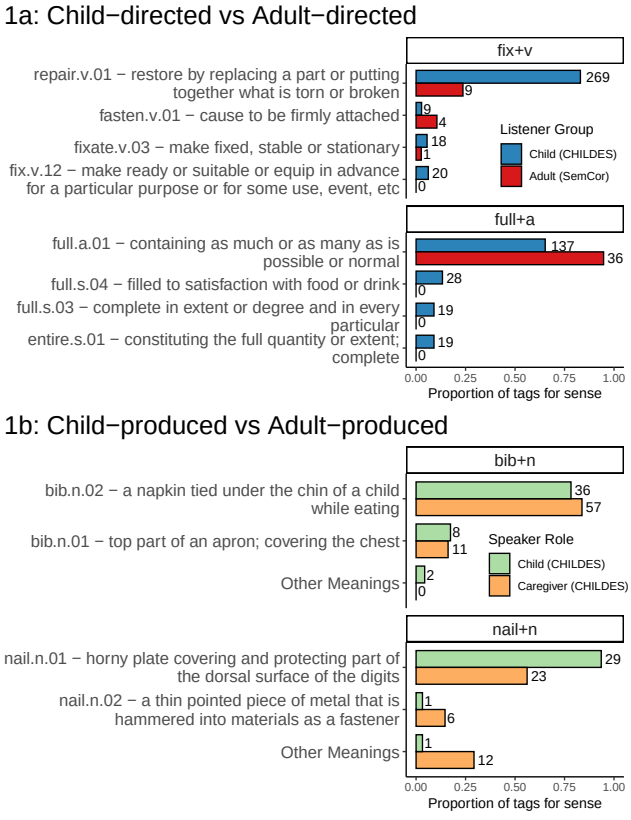

Figure 4: Example empirical sense distributions.

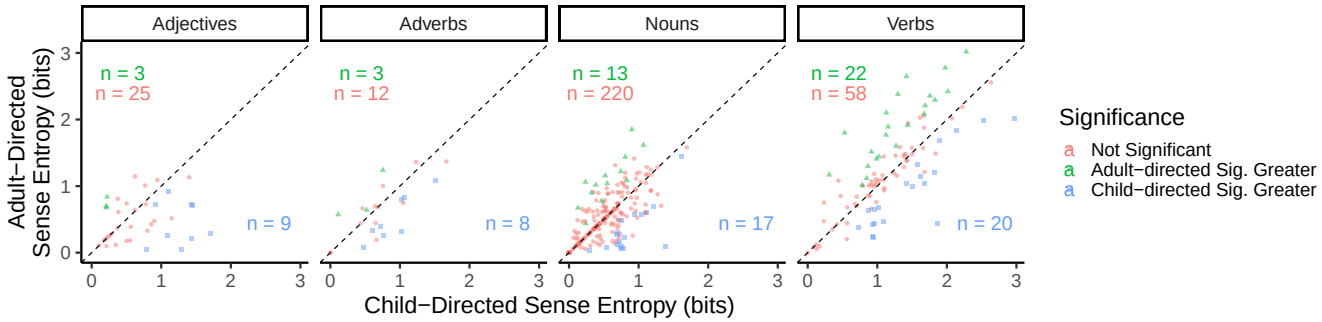

Figure 5: Sense entropy in adult speech to children vs. adult written language (**Analysis 1a**). Numbers (n =) indicate the number of word types where the difference in sense entropy is significant in favor of adult-directed speech, in favor of child-available speech, or not significant. Child-directed speech has comparable sense entropy to a sample of adult written material (mostly newspaper articles).

exclaim), *sit* (esp. to be located somewhere), and *hear* (esp. *to become aware of*). Adults also demonstrate a longer tail of low frequency uses of these verbs, *e.g., take* in *to take a picture*. Finally, a detailed analysis of nouns suggested *why* children use some of these words with a greater diversity of senses than adults: because they are more likely to use meanings outside of the WordNet sense ontology (*e.g.,* children employed a high number of other meanings for *sun*, *head*, and *bib*). Verbs have higher sense entropies than other parts of speech (Fig. 6).

Analysis 2: Predicting age of first production

Does ambiguity help or hinder word learning? Many features of words are known to help predict when English-learning children first produce a word, including overall frequency in the input, word-final frequency in the input ("final frequency"), frequency as sole word in an utterance ("solo frequency"), concreteness, part of speech, and "babiness" (see Braginsky et al., 2019 for a review). We use a linear mixed effects model following Braginsky et al. (2019) to ask whether sense entropy of a word might predict when a word is first produced, over and above these variables. We restrict the analysis to n=211 CDI items where entropy estimates are sufficiently constrained (99% HPD for sense entropy less than .5 from Analysis 1b).

Sense entropy is strongly correlated with two existing predictors, a) frequency (Pearson's $r = .47$), likely because words with more meanings may appear more often, and b) concreteness (r = .41), likely because more senses typically include metaphorical uses in addition to concrete ones. We residualize both with respect to sense entropy. A complete set of correlations between sense entropy and predictors used in Braginsky et al. (2019) is presented in Fig. 8.

We note two contradictory predictions from the literature regarding the effect of sense entropy on age of first production of a word. Hills et al. (2010) find that words that are used in a broader variety of lexical contexts are produced earlier. They interpret this as evidence that more diverse associative structure promotes faster word learning. The contrasting proposal, set forth in (Roy et al., 2015), is that words that are more broadly distributed across high-level linguistic contexts (*e.g.,* mealtime vs. playtime vs. storytime) are learned later. By this account, words with *lower* sense diversity may appear in more stable contexts and thus be easier to master.

Results The regression model reveals a similar pattern of effects to Braginsky et al. (2019), though with some differences (e.g., we find a *negative* main effect of the number of phonemes, and a positive coefficient for the interaction of the number of phonemes and child age, as opposed to the reverse pattern). Beyond these previously-documented effects, we also find a statistically significant *negative* main effect of sense entropy: children are *less* likely to produce words with higher sense entropy in the caregiver input (Fig 9). Importantly, word frequency and concreteness were still significant predictors, even with sense entropy partialed out, suggesting the importance of their contribution besides that captured by sense entropy.

Discussion

Using a newly-annotated collection of word meanings, we provide a new quantitative characterization of lexical ambiguity in speech to and from children, with three implications for theories of word learning.

First, contrary to the idea that adults avoid ambiguity when speaking to children, we find that the sense diversity in adult speech to children is actually comparable to the sense diversity in a sample of adult-directed written language. This finding is important because it suggests that children do, in fact, have to contend with ambiguity from the earliest stages of language acquisition; thus, theories of word learning cannot assume that the dominant situation of word learning is one in which children are exposed to an unambiguous word.

Second, we find that children's use of words is actually comparable in its sense diversity to that of adults (outside of the case of verbs); this is interesting because it suggests that children do not go through a long, protracted stage in which they only use a word with a single meaning, as dominant theories of word learning might predict.

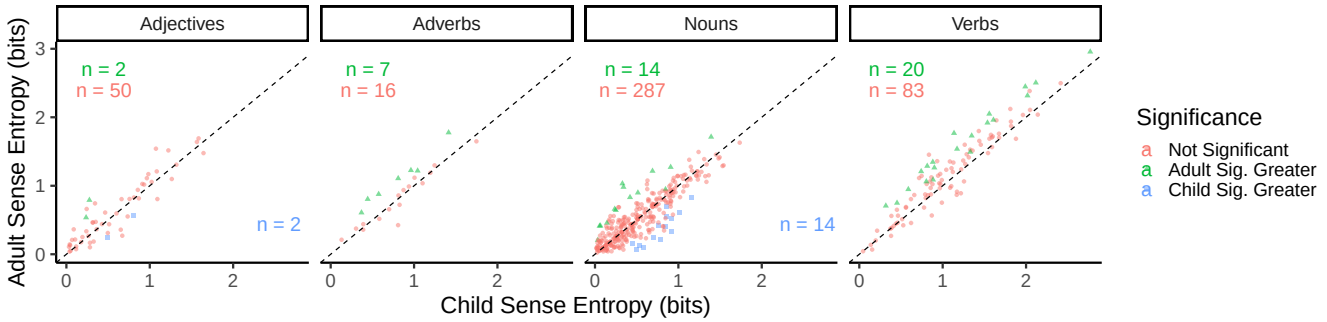

Figure 6: Caregiver vs. child sense entropy stratified by part of speech (**Analysis 1b**). Dashed line indicates a perfect correlation between adult and child sense entropy, consistent with the null hypothesis. Numbers (n =) indicate the number of word types where the difference in sense entropy is significant in favor of adult speech, in favor of child speech, or not significant.

Third, we find that the diversity of sense usage in the adult input is a significant predictor of when a word is first pro-

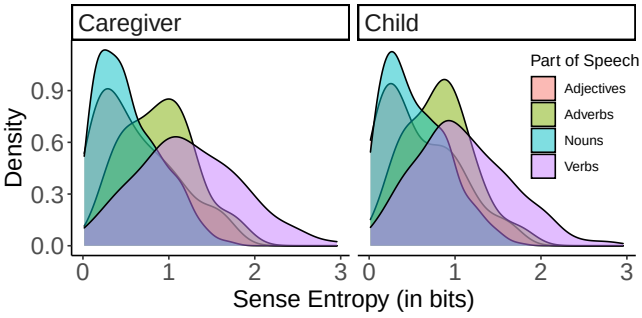

Figure 7: Sense entropy distributions by speaker and part of speech (cf. Fig. 6).

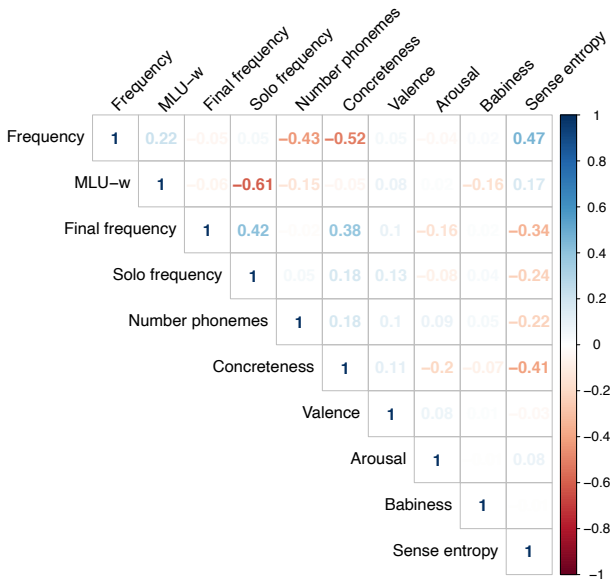

Figure 8: Correlations between common predictors of vocabulary acquisition from (Braginsky et al., 2019) and sense entropy for $n = 211$ word types with well-constrained sense entropy estimates.

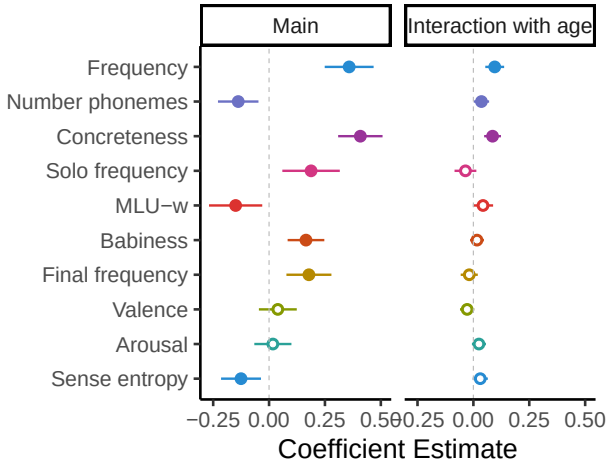

Figure 9: Coefficients for vocabulary production model (**Analysis 2**). Children begin to produce words with more diverse sense usage later. Filled circles indicate statistical significance, $p < .05$.

duced by children, adding to decades of prior work on what predicts word production. Words with more diverse sense usage are produced later, consonant with the finding of Roy et al. (2015) that words that are more closely linked to a specific communicative context are learned faster. We note, however, that although ambiguity might make it difficult for children to initially learn to use a word at all, it is possible that having learned one meaning of a word helps children learn additional meanings for that word, particularly when they are related (as in cases of polysemy). This rapid extension of words to new, related meanings is supported by experimental studies (Floyd & Goldberg, 2020; Srinivasan et al., 2019).

Our new dataset, which we plan to publicly release once complete, will open the door to a range of new questions about lexical ambiguity, and provide new answers for classic questions about semantic development. For example, our dataset will allow us to quantify the degree and manner in which children make "semantic overgeneralizations" by using words with non-conventional meanings (see Analysis 1b).

Such an analysis would allow us a far more rigorous exploration of this phenomenon than was previously possible.

Several limitations temper the strength of the conclusions, and invite further empirical work and theoretical refinement. Most prominently, word meanings are reduced to the discretized, unstructured sense ontology of WordNet. On the first point, many of the senses in WordNet overlap. On the second point, WordNet does not distinguish between homonymous (accidental collisions in wordforms) and polysemous (motivated re-use of wordforms) relationships between senses. While the choice of the WordNet sense ontology was motivated here by a desire for compatibility with existing work, we hope that our data collection strategy — allowing annotators to note when none of the senses in WordNet were appropriate and to report multiple senses for a token — will allow for the development of new word sense representations that overcome these limitations.

Further, our choice of adult-produced and adult-directed corpora is limited to existing semantically tagged corpora, which is primarily composed of written text. The absence of adult directed conversational corpora may under-represent more embodied usages of word meanings such as *full* (as in "full from dinner"). These meanings are common in child-directed conversational corpora, but relatively rare in written text. We hope to annotate additional, more representative adult corpora to create stronger comparison datasets.

A final limitation concerns the "megachild" assumption that we made for the current analyses, by collapsing sense usage across families. This modeling assumption was necessary in the present work because of sparsity in the data (rare words have rare senses), but sense usage may vary widely across families. In future work, we will test variants of the Dirchlet-multinomial model presented above which can pool evidence of sense usage across families when data is sparse, and model family-specific usage in instances where there is sufficient data.

Conclusion

We present the first large-scale study of ambiguity in the speech to and from English learning children. Our analyses of new, longitudinal sense-tagged corpora reveal several basic facts about variation in word meanings in speech to children, as well as children's own productions. These analyses and the dataset itself will enable new research into word meanings, above and beyond word forms.

References

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Casas, B., Català, N., Ferrer-i Cancho, R., Hernández-Fernández, A., & Baixeries, J. (2018). The polysemy of the words that children learn over time. *Interaction Studies*, *19*(3), 389–426.

Casas, B., Hernández-Fernández, A., Català, N., Ferrer-i Cancho, R., & Baixeries, J. (2019). Polysemy and brevity versus frequency in language. *Computer Speech & Language*, *58*, 19–50.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Lang Speech*, *49*(2), 137–174.

Fenson, L., et al. (2007). *MacArthur-Bates communicative development inventories*. Paul H. Brookes Publishing Company Baltimore, MD.

Floyd, S., & Goldberg, A. E. (2020). Children make use of relationships across meanings in word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Floyd, S., Goldberg, A. E., & Lew-Williams, C. (2020). Toddlers recognize multiple meanings of polysemous words. In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.*

Gangemi, A., Guarino, N., & Oltramari, A. (2001). Conceptual analysis of lexical taxonomies: The case of WordNet top-level. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001* (pp. 285–296).

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010, Oct). The Associative Structure of Language: Contextual Diversity in Early Word Learning. *J Mem Lang*, *63*(3), 259–273.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. transcription format and programs* (Vol. 1). Psychology Press.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. MIT Press.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Miller, G. A., Leacock, C., Tengi, R., & Bunker, R. T. (1993). A Semantic Concordance. In *Proceedings of the workshop on Human Language Technology* (pp. 303–308).

Nair, S., Srinivasan, M., & Meylan, S. (2020). Contextualized word embeddings encode aspects of human-like word sense knowledge. In *Proceedings of the Workshop on the Cognitive Aspects of the Lexicon* (pp. 129–141).

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.*

Rabagliati, H., Marcus, G. F., & Pylkkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition*, *117*(1), 17–37.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668.

Sanchez, A., Meylan, S., Braginsky, M., MacDonald, K., Yurovsky, D., & Frank, M. (2019). childes-db: A flexible and reproducible interface to the child language data exchange system. *Behavior Research Methods*, *51*(4), 1928–1941.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

Srinivasan, M., Berner, C., & Rabagliati, H. (2019). Children use polysemy to structure new word meanings. *J Exp Psychol Gen*, *148*(5), 926–942.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, *62*(4), 245–272.

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*, 638–676.

Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language*, *28*(1), 127–152.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Zipf, G. (1935). *The Psychobiology of Language*. Houghton-Mifflin.