Child language input does not reflect word frequency: Typical and atypical feature description across development

Anonymous CogSci submission

Abstract

Language provides children a powerful source of information about the world. From language alone, simple distributional learning models can recover enough information to perform comparably to non-native college applicants on the TOEFL (Landauer & Dumais, 1997). Blind children learn the same kinds of relationships among visual categories as sighted children, without any of the relevant visual input (Landau & Gleitman, 1985). However, language does not perfectly reflect the world: the most typical features of natural kinds may often go unremarked. For instance, adults rarely describe the color of an orange carrot, as world knowledge makes this description redundant. Given children's nascent world knowledge, does parents' speech to children follow this pattern? From longitudinal corpus data of parent-child communication (Goldin-Meadow et al., 2014) between 14–58 months, we extracted usage data for 684 high-frequency concrete nouns and co-occurring adjectives. Independent raters coded the typicality of over 2,000 unique adjective–noun pairs on a 7-point Likert scale. If language statistics reflect world statistics, description should be dominated by the typical (strong negative skew); however, across all ages, we see descriptors concentrated in the atypical range (positive skewness = 0.65). However, parents were reliably more likely to use typical descriptors when talking to younger copmared with older children. Overall, child language input reflects notable more than typical features, but increased description of typical features early in development may provide a foothold for young learners.

Keywords: language input, language acquisition, child-directed speech

Children learn a tremendous amount about the structure of the world around them in just a few short years, from the rules that govern the movement of physical objects to the hierarchical structure of natural categories and even relational structures among social and cultural groups (Baillargeon, 1994; Legare & Harris, 2016; Rogers & McClelland, 2004). Where does the information for this rapid acquisition come from? Undoubtedly, a sizeable component comes from direct experience observing and interacting with the world (Sloutsky & Fisher, 2004; Stahl & Feigenson, 2015). But another important source of information comes from the language people use to talk about the world (Landauer & Dumais, 1997; Rhodes, Leslie, & Tworek, 2012). For many of the things children learn about–like the roundness of the earth–the information must come from language (Harris & Koenig, 2006). How similar is the information available from children's direct experience to the information available in the language children hear?

Two lines of work suggest that they may be surprisingly similar. One compelling area of work is the comparison of semantic structures learned by congentinally blind children to those of their sighted peers. In several domains that would at first blush rely heavily on visual information, such as color terms or verbs of visual perception (e.g. *look*, *see*), blind children's semantic similarity judgments are quite similar to those of sighted children (Landau, Gleitman, & Landau, 2009). Further, blind adults' judgments of visual perception verbs are sensitive to highly detailed information like variation in intensity (e.g. blaze vs. glow), just like sighted adults (Bedny, Koster-Hale, Elli, Yazzolino, & Saxe, 2019). A second line of evidence supporting the simliarity of information in perception and language is the broad success of statistical models trained on language alone in approximating human judgments across a variety of domains (Devlin, Chang, Lee, & Toutanova, 2018; Landauer & Dumais, 1997; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Even more compellingly, models trained on both linguistic usage and perceptual features for some words can infer the perceptual features of linguistically related words entirely from the covariation of language and perception (Johns & Jones, 2012).

Still, there is reason to believe that some semantic features may be harder to learn from language than these data suggest. While the co-occurrence structure of language may provide strong clues that carrots are like tomatoes, carrots are vegetables, and carrots are eaten for dinner, they may provide very little evidence that carrots are orange (Willits, Sussman, & Amato, 2008). This is because people rarely use language merely to provide running comentary on the world around them. Instead, we use language to talk about things we find interesting–things that are surprising, exciting, or otherwise divergent from our conversational partners' expectations (Grice, 1975).

Thus, speakers are informative in relation to common knowledge amongst interlocutors and available information in the environment, not redundant with these other sources of knowledge. Language production tasks in the lab largely bolster this claim, finding that people are sensitive to informativity in formulating and interpreting utterances (Mangold & Pobel, 1988; MORE CITES). In particular, speakers are informative in relation to both the current context of objects in the environment and the typical features of those objects (Mitchell et al., 2013; Westerbeek et al., 2015; Rubio-Fernandez, 2016). While speakers overwhelmingly refer to an object that is typical of its category with a bare noun—e.g., calling a yellow banana "a banana"—they often describe more about the object's features when they are atypical, rarely referring to a green or blue banana without specifying its color. Listeners are similarly affected by color description, expecting a color adjective to be used when an object's color is atypical (Sedivy, 2003). **BOLD PART IS SUPER JARGONY**

For things like carrots–that children learn about both from perception and from language–this issue may be resolved by intergrating both sources of information: Likely almost all of the carrots they see are orange even if no one comments on them. But for things for which they lack perceptual access–zoo animals, other social groups, etc.–the structure of language alone may lead them to have distorted categories. This may be especially problematic because children's understanding of the pragmatics of adjective use change significantly over development (Horowitz & Frank, 2016).

Should we expect children's categories to be biased in this way, with typical features underrepresented and atypical features overrepresented? We examine the typicality of adjectives in a large, diverse corpus of parent-child ineractions recorded in children's homes to ask whether parents talking to their children–just like adults speaking and writing–tend to use adjectives predominantly to mark atypical features. However, we also find that parents' use adjectives differently over the course of children's development, noting typical features more often for younger children. We then ask whether the co-occurrence structure of language nonetheless captures typicality information, and find that relatively little of it is represented. Thus, children must either have distorted representations of categories learned only through languae or else learn typicality through some means.

Corpus Analysis

We first analyze caregiver speech to extract the usage statistics for all co-occuring adjective-noun pairs within an utterance. After subsetting the data to the concrete concepts (Brysbaert et al., 2014), human raters judged the typicality of each adjective-noun pair. Combining our usage data across the developmental range (ages 14-58 months) with the typicality judgements, we can examine how information is distributed in caregiver descriptions.

Corpus Method We used data from the Language Development Project– a large-scale, longitudinal corpus of parent child-interaction in the home with families who are representative of the Chicago community in socio-economic and racial diversity (Goldin-Meadow et al., 2014). Recordings were taken in the home every 4-months from when the child was 14-months-old until they were 58-months-old, resulting in 12 timepoints. Recordings were 90 minute sessions, and participants were given no instructions. The Language Development Project corpus contains transcription of all speech and communicative gestures produced by children and their caregivers over the course of the 90-minute home recordings.

To determine the extant information structure in children's language input, we subset the corpus to only caregiver speech, yielding a set of 828,392 distinct utterances. While it is important to consider children's own productions as an integral part their language environments (CITATION), description of a concept relies heavily on linguisitic and world knowledge, and so it is necessary to analyze caregiver speech separately. For example, at our earliest timepoint (14-months), children are not producing multiword utterances with adjectives and nouns, or even reliably producing many adjectives at all.

Based on part-of-speech tags, we extracted usage rates for all nouns in the corpus. To ensure that the set of nouns we examined were relatively constant across our developmental samples, we subset all nouns that caregivers produced to only include nouns that were produced at least once every 3 sessions (i.e. per developmental year). This yielded a set of some 2,000 potential target nouns.

We next tagged every utterance that included both one of the target nouns and an adjecitve. Any possible resulting pairs was counted seperately (i.e. utterances with one noun and multiple adjectives were coded as multiple pairs) Many resulting high-frequency pairs proved difficult to classify on our typicality schema (e.g., "good"– "job"; "little" – "bit"). To identify potentially problematic pairs, we joined human judgments of concreteness for both the noun and the adjective (Brysbaert). We set our threshold at the top 25% of the concreteness ratings, and excluded any pair where either the adjective or noun was not rated above that threshold. Finally, all pairs were given to 7 human coders to judge whether the pair was "incoherent or unrelated" and we thus excluded a further XXX pairs from the sample (e.g., incoherent pairs such as "wet" – "brown", and presumably errorfully coded pairs such as "orange" – "orange").

Thus, our final sample included 2,220 unique adjective-noun pairs drawn from over 6,000 utterances. The pairs were combinations of 684 distinct concrete nouns and 118 distinct concrete adjectives. We compiled these pairs and collected human judgments on Amazon Mechanical Turk for each pair, as described below. Table 1 contains example utterances from the final set and typicality judgments from our human raters.

Ratings Method

To evaluate the typicality of the adjective–noun pairs pulled from the corpus, we asked participants on Amazon Mechanical Turk to rate each pair. Participants were presented with a question of the form "How common is it for a cow to be a brown cow?" and asked to provide a rating on a seven-point scale: (1) never, (2) rarely, (3) sometimes, (4) about half the time, (5) often, (6) almost always, (7) always. Each participant rated 20 pairs, and each pair was rated by four participants; we used Dallinger [CITE], a tool for automating complex recruitment on Amazon Mechanical Turk, to balance recruitment. Overall, we recruited 444 participants to rate 2200 adjective–noun pairs. After exclusions using an attention check, we retained 8580 judgments, with each adjective–noun pair retaining at least two judgments.

Utterance	Pair	Judge_1	Judge_2	Judge_3	Judge_4	Mean_Typicality
especially with wooden shoes	wooden-shoe	2.00	2.00	3.00	2.00	2.75
some of our green bananas too	green-banana	4.00	3.00	5.00	[illegible]	3.75
are bananas yellow?	yellow-banana	5.00	5.00	7.0[illegible]	[illegible]	5.75

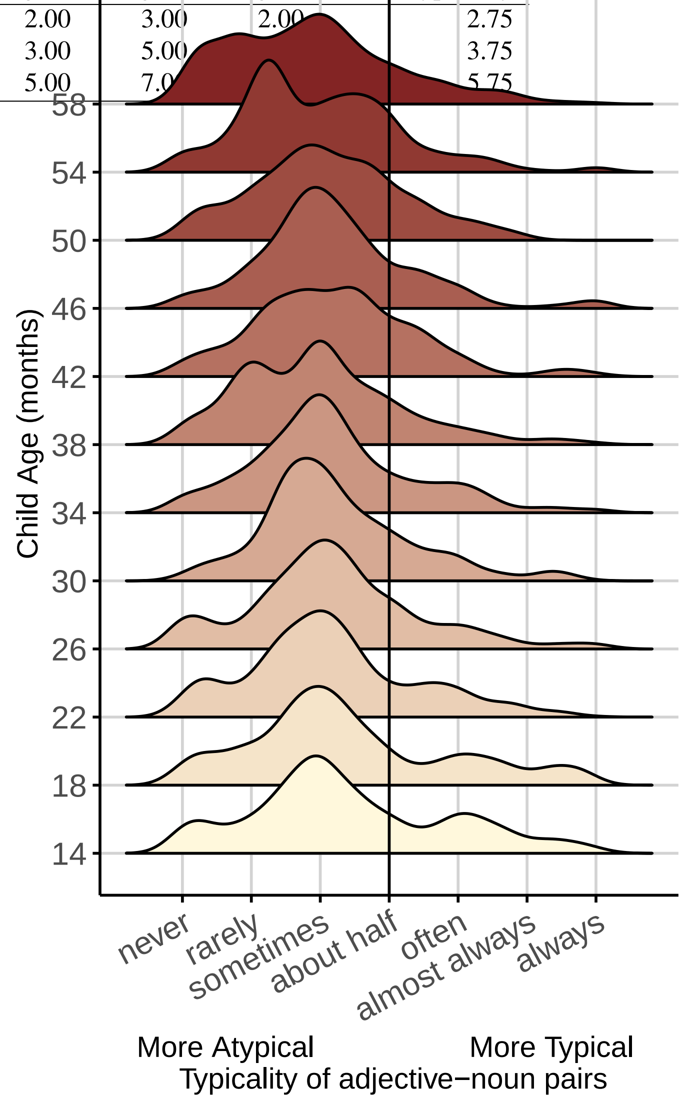
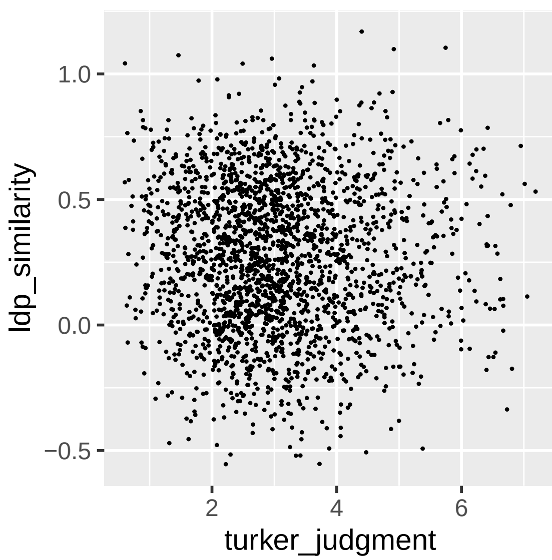
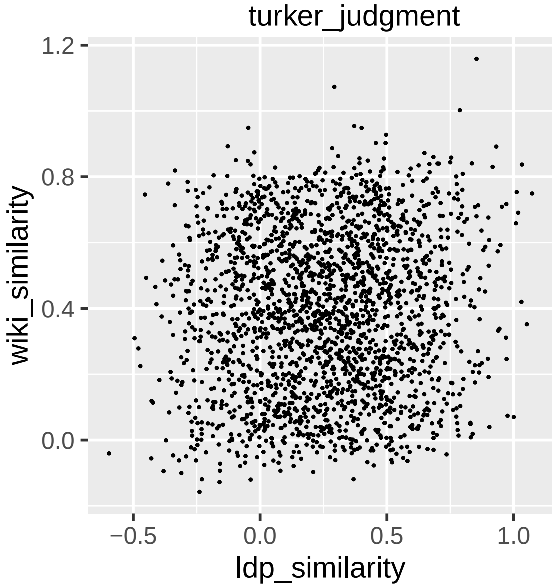
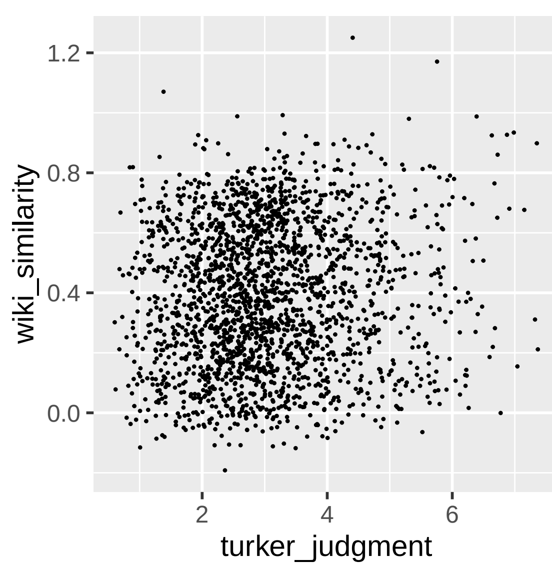


Figure 1: Denisity plots showing the usage amount at each timepoint based on the typicality of the adj-noun pair.







Results

If description in child-directed speech mirrors adult-directed speech, we should see that caregiver descripiton is dominated by modifiers that are sometimes true of the noun they modify. If instead child-directed speech privledges redundant or assumed information, caregiver description would yield a distinct distribution dominated by highly typical modifiers. As can be seen in figure 1, there is remarkable developmental consistency such that caregiver description largely focuses on features that are only sometimes true of the concept.

Developmental Consistency. A mixed effects model predicting typicality from usage confirms that caregiver speech is signficantly biased toward description of atypical speech (B XXX, $p <$ YYY). The full mixed effects model included random effects of noun and rater, specified as follows `lmer(typicality ~ log(age) + (1|noun) + (1|rater)` using an offset of 4 which is the midpoint of our scale. Rather than focusing on typical information, caregiver language tends on average be about less typical features (i.e. ‘sometimes’, ‘rarely’ or ‘never’).

Examining usage data as a function of typicality (see figure 1), we see evidence of a positive skew (0.65). Data from every time point from 14-58 months seems to show a similar pattern (skews 0.23 - 0.82). These skews provide further evidence that the the bulk of caregiver language reflects lower-typicality adjective-noun pairs.

Our mixed effects model also yields some evidence of developmental change. There is a negative effect of age on typicality (B XXX, $p <$ YYY), suggesting that children’s language environments become increasingly focused on atypicial features as they get older. This result is discussed further below.

Adult-Directed Speech. Next, we briefly confirm that adult-directed speech shows the similar usage pattern using an identical analysis framework on usage data from the Corpus of Contemporary American English (COCA). Using the set of adjective-noun pairs for which we have judgments from our analysis of caregiver speech, we repeat our analysis of usage frequencies for a set of 1,357 distinct adjective-noun pairs.

As predicted, an identical mixed effects model predicting typicality from usage shows that adult-directed speech speech is signficantly biased toward description of atypical features (B XXX, $p <$ YYY). ADS also shows a similar a positive skew (0.68), such that the bulk of language reflects adjective-noun pairs rated < 4 on typicality (i.e. ‘sometimes’, ‘rarely’ or ‘never’).

The usage distributions in adult-directed speech seem qualitatively similar to the distributions we found in child-directed speech. In sum, these data suggest that even when talking with very young children, caregiver speech structures information similarly to adult-directed speech. We next briefly consider how children’s own productions are structured.

The difference of settings makes it difficult to compare these data directly with our data from the Language Devleopment Project corpus. Four of the COCA datasets are drawn from written texts, which puts in place distinct language pressures and removes visual common ground. While there is one subset of the corpus drawn from spoken data, those utterances come from TV and radio programs. All of these settings are likely quite different environements for language than the naturalistic in-home setting that our Language Development Project corpus draws from. In light of these differences, any observed similarity in usage seems remarkable.

Child Speech. What kind of information is contained in children’s own speech? By analyzing children’s productions, we can determine when children come to use description in a way that looks like caregiver speech. Are children mirroring adult-like uses of description even from a young age, or are they choosing to describe other features of the world?

The Language Development Corpus contains 442,048 child utterances. Using the set of adjective-noun pairs for which we have judgments from our analysis of caregiver speech, we repeat our analysis on usage data for a set of 533 distinct adjective-noun pairs.

While preliminary, an identical mixed effects model predicting typicality from usage shows that adult-directed speech speech is signficantly biased toward description of atypical features (B XXX, $p <$ YYY). As before, children’s productions shows a positive skew (0.61, comapre with skewness = 0.65 seen in the adults), such that the bulk of language reflects adjective-noun pairs rated < 4 on typicality (i.e. ‘sometimes’, ‘rarely’ or ‘never’).

Talk about the Highly-Typical. Despite of the striking consistency of description in caregiver speech across development, our mixed effects model revealed a negative effect of age. In line with our hypotheses, it seems that caregivers are more likely to provide descrpiption of typical features for their young children, compared with older children.

Indeed, when we look at the proportion of all description that is about highly-typical features (i.e. features that are ‘often’, ‘almost always’, or ‘always’ true), we see a significant negative correlation with age (r = -0.8231762, p $<$ 0.01). While children at all ages hear more talk about what is atypically true (see figure 1), younger children hear relatively more talk about what is typically true than older children (figure 2).

Discussion

In sum, language is used to discuss atypical features of the world, more so than typical features. Description in caregiver speech seems to largely recapitulate the usage patterns that we observed in adult-directed speech, suggesting that these patterns arise from general pressures on how communication is used. Indeed even children’s own productions show a similar usage pattern, with more description of atypical features of world as early as we can measure. A

It should be noted that children’s utterances come from naturalistic conversations with caregivers, and the pattern of de-

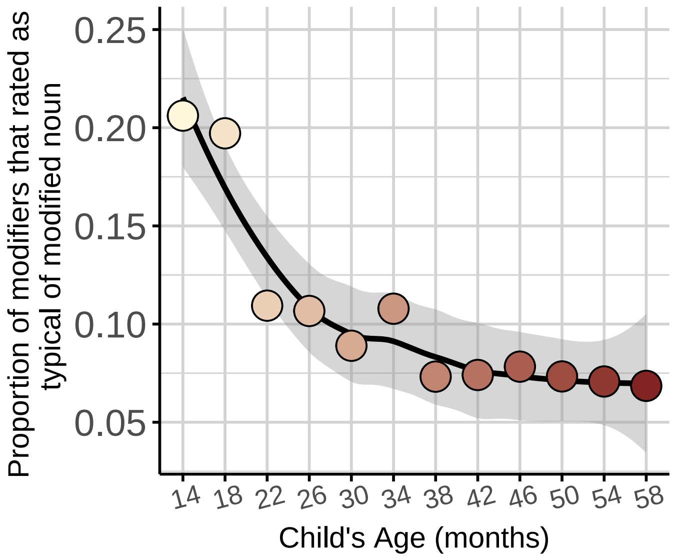


Figure 2: This plot shows the proportion of caregiver description that is about typically-true features, as a function of age.

scription is likely highly dependent on that caregiver. That is, if a parent chooses to describe the *purpleness* of a cat in book, the child may well respond by asking about that same feature. Thus, it is possible that some of the children's productions captured take place within a parent-led discourse. Future analyses would need to better disentangle the extent to which children's productions are imitatitve of caregivers.

While children's own descriptions largely mirror adults', we do see some evidence that children's linguistic environment changes across development, becoming increasingly focused on atypical features. The higher prevalance of typical descriptors in early development may help young learners; however, even at the earliest point we measured, the vast majority of language input describes atypical features.

Across adult, parent, and child language corpora, we find robust evidence that language use systematically overerpresents atypical features. This usage aligns with the idea that language is used informatively in relation to background knowledge about the world. It may pose a problem, however, for young language learners with still-developing world knowledge. If language does not transparently convey the typical features of objects, and instead (perhaps misleadingly) notes the atypical ones, how might children come to learn what objects are typically like? One possibility is that information about typical features is captured in regularities across many utterances. If this is true, language may still be an important source of information about typicality as children may be able to extract more accurate typicality information by tracking second-order co-occurence.

Other Ways of Extracting Structure from Language?

Second-order Co-occurence

Kim et al. (2019) demonstrate blind people's striking convergence with sighted people's feature judgments about animals, and posit that blind people may use animal taxonomy knowledge to selectively generalize features between species. In a response, Lewis et al. (2019) note that complex inferential machinery is not necessary to make these nuanced generalizations, and in fact much of this feature information is captured by simpler associations between words in the language one hears. We take a similar approach to ask whether typical feature information is expressed in the structure of the language children hear.

Information in language goes beyond what can be learned from any one utterance. Though one might never hear the words /kumquat/ and /grapefruit/ in the same sentence, their common contexts—other words like /eat/, /rind/, /tree/, /tart/, and /seeds/—are clues that they have some similarities. Models of distributional semantics capitalize on these patterns by representing words using their contexts, and judging two words to be similar if they are surrounded by similar sets of words. Though we found that children get more information about atypical than typical features of objects on the utterance level, perhaps patterns of language use across many utterances could be used to extract typical feature information.

Word2Vec To test this possibility, we trained word2vec, a model that predicts words using their contexts, on the same corpus of child-directed speech used in our first set of analyses. Our model is a continuous-bag-of-words word2vec model trained using the package gensim [CITE]. If the model captures information about the typical features of objects, we should see that the model's word pair similarities are correlated with the typicality ratings we elicited from human raters.

Results We find that similarities in the model have near zero correlation with human adjective–noun typicality ratings (r = 0.018). This is in spite of better correlations with large sets of human similarity judgments between different kinds of word pairs (correlation with wordsim353, 0.37; correlation with simlex, 0.15). This suggests that statistical patterns in child-directed speech are likely insufficient to encode information about the typical features of objects, despite encoding at least some information about word meaning more broadly. However, the corpus on which we trained this model was small; perhaps our model did not get enough language to draw out the patterns that would reflect the typical features of objects. To test this possibility, we asked whether word vectors trained on a much larger corpus—English Wikipedia—strongly correlate with typicality ratings. We find that while the correlation between similarities in the Wikipedia–trained model and human noun–adjective typicality ratings is stronger, it is still fairly weak at r = 0.24. Overall, these results suggest that models of distributional se-

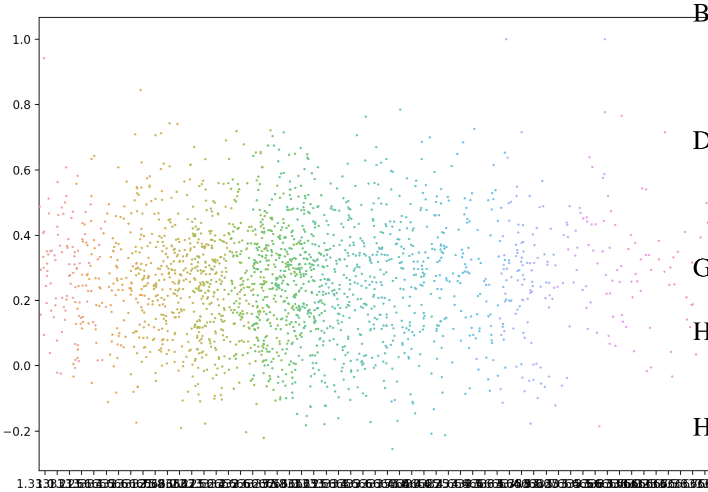


Figure 3: Correlation of vector distances from word2vec (trained on the LDP corpus) and human-rated typicality judgments.

mantics fail to extract typical feature information from language in which atypical features are more often described.

Syntax

General Discussion

importance of conversational pressures in determining information in language

> AL: "Why would adults provide so little typical feature information (or at least less than we might think)? Are they just avoiding over-informativity, even with presumably more ignorant conversational partners? Maybe the way they talk is actually important for kids to learn pragmatics / how to talk about features?"

learning problem for kids. maybe the prototypicals is a foothold for young learners. . . .

> AL: Do you think there is enough of the prototypicals? given that even the youngest kids only hear like 20% typical adjectives – but maybe that's already a lot (?)

more thoughts on potential ways kids extract more information

limitations. . . abstract language?

Acknowledgements

Place acknowledgments (including funding information) in a section at the end of the paper.

References

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*(5), 133–140.

Bedny, M., Koster-Hale, J., Elli, G., Yazzolino, L., & Saxe, R. (2019). There's more to "sparkle" than meets the eye: Knowledge of vision and light verbs among congenitally blind and sighted individuals. *Cognition*, *189*, 105–115.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Preprint arXiv:1810.04805*.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*(3), 505–524.

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*(3), 807–819.

Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, *4*(1), 103–120.

Landau, B., Gleitman, L. R., & Landau, B. (2009). *Language and experience: Evidence from the blind child* (Vol. 8). Harvard University Press.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Legare, C. H., & Harris, P. L. (2016). The ontogeny of cultural learning. *Child Development*, *87*(3), 633–642.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Rhodes, M., Leslie, S.-J., & Tworek, C. M. (2012). Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences*, *109*(34), 13526–13531.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research*, *32*(1), 3–23.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Willits, J. A., Sussman, R. S., & Amato, M. S. (2008). Event knowledge vs. Verb knowledge. In *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2227–2232).