Children's overextension as communication by multimodal chaining

Renato Ferreira Pinto Junior (renato@cs.toronto.edu)
Department of Computer Science
University of Toronto

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science
Cognitive Science Program
University of Toronto

Abstract

Young children often stretch terms to novel objects when they lack the proper adult words—a phenomenon known as overextension. Psychologists have proposed that overextension relies on the formation of a chain complex, such that new objects may be linked to existing referents of a word based on a diverse set of relations including taxonomic, analogical, and predicate-based knowledge. We build on these ideas by proposing a computational framework that creates chain complexes by multimodal fusion of resources from linguistics, deep learning networks, and psychological experiments. We test our models in a communicative scenario that simulates linguistic production and comprehension between a child and a caretaker. Our results show that the multimodal semantic space accounts for substantial variation in children's overextension in the literature, and our framework predicts overextension strategies. This work provides a formal approach to characterizing linguistic creativity of word sense extension in early childhood.

Keywords: language acquisition; linguistic creativity; overextension; word sense extension; multimodality; chaining; communication

Young children often stretch terms to describe novel objects when they lack the proper adult words, a phenomenon known as overextension (Clark, 1978). Overextension is a communicative strategy that draws on knowledge of diverse relations in the world. For instance, a child may use "dog" to refer to a *squirrel*, "ball" to refer to a *balloon*, or "key" to refer to a *door*. This creative use of words toward novel meanings, or *word sense extension*, is not only attested in child language acquisition, but it is also reflected in historical meaning change, e.g., we extended the meaning of "mouse" from a rodent to a computer device. We explore the origin of word sense extension by asking how the cognitive capacity of overextension in childhood can be characterized formally.

Early work by Vygotsky (1962) suggests that overextension relies on "chain complex", a critical element of concept formation in childhood. He demonstrated chain complex by a series of overextension cases from a child who extended the meanings of "quah" to wide-ranging things including a duck, water in a pond, liquids in general, an eagle on a coin, and any coin-like objects. Vygotsky's account resonates with work from philosophy and cognitive linguistics that suggest the complex structure of word meanings (e.g. Wittgenstein, 1953) is formed possibly due to a process of chaining (Lakoff, 1987), where one referent is linked to another forming a chain-like structure. More recent work has shown that chaining predicts word sense extension in the history of English (Ramiro, Srinivasan, Malt, & Xu, 2018) and other languages (Xu, Regier, & Malt, 2016). However, these works did not offer a formal account of how one might represent the rich knowledge in a chain complex.

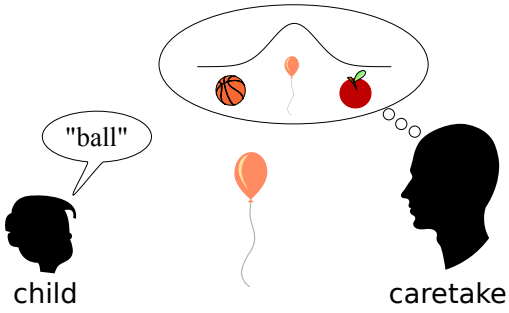


Figure 1: Overextension in child-caretaker communication.

Empirical work from Rescorla (1980) provided clues to the knowledge underlying children's overextension. Specifically, she identified three main types of relations between core and overextended meanings of a word, summarized as 1) *categorical* relation: overextension by linking objects within a taxonomy, e.g., "dog"→*squirrel*, 2) *analogy* or visual analogy: overextension by linking objects with shared perceptual properties, e.g., "ball"→*balloon*, and 3) *predicate-based* relation: overextension by linking objects that co-occur frequently in the environment, e.g., "key"→*door*. An open question we address in this work is how to combine these types of relations to predict overextension strategies in early childhood.

We propose a computational framework that considers overextension as a communicative game between a child and a caretaker, illustrated in Figure 1. The game involves a child and a caretaker in a situation where the child needs to refer to an out-of-vocabulary novel object. In this context, the child faces a *production problem*, where the goal is to extend a word from the existing vocabulary (e.g., "ball") to the novel object (e.g., a balloon). The caretaker instead faces a *comprehension problem*, where the goal is to guess the intended referent based on the child's utterance. Since "ball" does not typically map to balloons, we wish to reconstruct the cognitive processes that could have given rise to successful communication between the child and the caretaker in common cases of overextension. As such, our framework should support both strategic word choices for the child and prediction of intended referents for the caretaker.

Our communication-based framework relates to earlier work in overextension from the developmental literature. For example, Bloom (1973) argued that overextension is a performance error caused by vocabulary limitations, whereby a child may consciously use an incorrect word (from the adult perspective) as a strategy to convey the desired referent meaning. A related hypothesis poses overextension as a retrieval error (Fremgen & Fay, 1980; Gershkoff-Stowe, 2001; Huttenlocher, 1974; Thomson & Chapman, 1977), suggesting that children may overextend an earlier acquired word even if the correct adult word has been partially acquired (e.g., understood in comprehension), because the latter may be more difficult to produce. We explore evidence for a retrieval error hypothesis by evaluating whether children may favour words with higher usage frequencies in their overextended word choices. Another extensive line of research suggests that overextension arises from children's incomplete conceptual knowledge of the semantic features underlying different categories (Clark, 1973; Kay & Anglin, 1982; Mervis, 1987). While we do not directly test claims about children's conceptual space in this work, we show that a combination of semantic relations helps to explain overextension strategies, and may play an integral role in characterizing the mechanisms that subserve children's early word learning.

Our framework also draws on a multimodal space of semantic relations (cf. Rescorla, 1980) that serves as the knowledge engine for creating chain complexes in overextension. The notion of multimodality is motivated in part by work on visually-grounded word learning (e.g. Lazaridou, Chrupała, Fernández, & Baroni, 2016; Roy & Pentland, 2002; Yu, 2005), which shows that perceptual features play an important role in children's acquistion of core (or conventional) word meanings. Our focus is on investigating how by integrating diverse semantic relations one might account for word usage beyond the core meanings that children normally acquire. Our work thus differs from the extensive literature on cross-situational word learning (Fazly, Alishahi, & Stevenson, 2010; Frank, Goodman, & Tenenbaum, 2009; Kachergis, Yu, & Shiffrin, 2017; Siskind, 1996), where the emphasis has been typically on modeling children's behaviour in learning conventional word meanings, but not on how they extend existing terms to describe novel objects. Our work also extends existing computational studies that explore overextension in specific domains such as color terms (Beekhuizen & Stevenson, 2016) to more general cases of overextension that involve mappings across domain boundaries, e.g., "ball"→*balloon*.

Computational framework

We present our computational framework for overextension following two steps: 1) Specification of a probabilistic model that simulates child's word choices (production) and caretaker's inference of intended referents (comprehension); 2) Construction of a semantic space that supports multimodal chaining of word meanings, encapsulated in the same model.

For this work, we focus on overextension of nouns, but the general framework that we present can be used to explore other types of overextension (e.g., in verbs and adjectives).

Probabilistic formulation

We formulate overextension as communication between a child and a caretaker. In particular, the child wishes to refer to a novel object c in an environment E. The child does so by choosing (and stretching) a word w from her vocabulary V. We assume that the correct term for the novel object is not yet acquired by the child, hence $c \neq w$ and $c \notin V$. Based on the child's utterance w, the caretaker wishes to infer the referent c among possible referents in E. We then model the child's behaviour by a *production model* and pair it with a *comprehension model* for the caretaker's behaviour.

Production. We cast the production problem as probabilistic inference over existing words in the child's vocabulary given the probe novel object c, via Bayes' rule:

$$p_{\text{prod}}(w|c) \propto p_{\text{prod}}(c|w)p(w) \tag{1}$$

We define the prior $p(w)$ proportional to the logarithmic usage frequency of a word with add-one smoothing $p(w) \propto \log(1+\text{freq}(w))$. This formulation is consistent with the frequency effect found in the study of overextension in color terms (Beekhuizen & Stevenson, 2016). It captures the intuition that all things being equal, the child is more likely to choose a common word versus a rare word for overextension. We define the likelihood function $p_{\text{prod}}(c|w)$ by a meta similarity measure that encapsulates the three types of semantic relations reported by Rescorla (1980) which the novel referent c can bear with the existing referent c_w of word w:

$$\begin{aligned} p_{\text{prod}}(c|w) &\propto \text{sim}(c, c_w) \\ &= \exp\left(-\frac{d_c(c,c_w)+d_v(c,c_w)+d_p(c,c_w)}{h}\right) \end{aligned} \tag{2}$$

We take the exponential-decay form from the generalized context model (GCM) or exemplar model of categorization (Nosofsky, 1986), where the influence of each relational type is proportional to how similar c and c_w are under that relation. We represent similarity by inverse distance, where d_c, d_v, and d_p represent distances measured according to categorical relation, visual analogy, and predicate-based relation, respectively. We describe the construction of each of these relational features in the next section. To control for model sensitivity to these distance functions, we use a single parameter h that we estimate empirically from data. The magnitude of h determines how slowly the meta similarity or the likelihood function decreases with respect to the distance measured in the multimodal relations.

Comprehension. We pair the child's production model with a comprehension model for the caretaker. Specifically, the caretaker solves the inverse inference problem as the child by a probability distribution over the space of intended referents based on the child's utterance w:

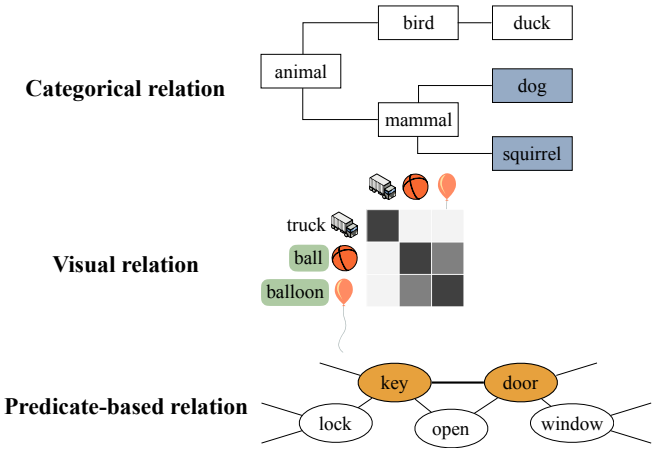


Figure 2: Types of semantic relations in multimodal space.

$$p_{\text{comp}}(c|w) \propto p_{\text{comp}}(w|c)p(c) \tag{3}$$

We consider the space of c to be all referents that appear in the communicative environment E where the child and the caretaker are situated in. We also assume a uniform prior $p(c)$ on possible referents in this environment, although it may be possible to enrich this prior by considering perceptual salience, eye gaze, pointing, and other pragmatic cues, which we do not model or have explicit access to in this work.

We define the likelihood function identical to the formulation used in the production model, under the assumption that both the child and the caretaker have knowledge of the multimodal relations:

$$p_{\text{comp}}(w|c) \propto \text{sim}(c_w, c) = \text{sim}(c, c_w) \tag{4}$$

Although it is possible to model perspective taking in a recursive way $p_{\text{comp}}(w|c) = p_{\text{prod}}(w|c) = p_{\text{prod}}(c|w)p(w)$ under the assumption that the caretaker takes into account the child's word choice in guessing the intended referent, e.g., similar to the rational speech act model (Goodman & Frank, 2016), we choose to work with the simplest version of this model that does not make any recursive assumption in the caretaker. We show in *Results* that our framework accounts for data well even without this assumption.

Multimodal semantic space

We define a multimodal semantic space that captures the three types of relational features in Rescorla (1980): categorical relation, visual analogy, and predicate-based relation. We construct these relational features using a fusion of resources drawn from linguistics, deep learning networks, and psychological experiments, as illustrated in Figure 2.

Categorical relation. We define categorical relation between two referents via a standard distance measure d_c in natural language processing by Wu and Palmer (1994), based on taxonomic similarity. Concretely, for two concepts c_1 and c_2 under a taxonomy T (i.e., a tree), the distance is:

$$d_c(c_1, c_2) = 1 - \frac{2N_{\text{LCS}}}{N_1 + N_2} \tag{5}$$

N_{LCS} denotes the number of shared parent nodes of the two concepts in the taxonomy. N_1 and N_2 denote the depths of the two concepts in the taxonomy. This distance measure is effectively the negated taxonomic similarity between c_1 and c_2, and is bounded between 0 and 1. Under this measure, concepts from the same semantic domain (such as *dog* and *squirrel*) should yield a lower distance than those from across domains (such as *ball* and *balloon*). To derive the categorical features, we took the taxonomy from WordNet (Miller, 1995) and annotated words by their corresponding *synset*'s in the database. We used the *NLTK* package (Bird & Loper, 2004) to calculate similarities between referents for this feature.

Visual analogical relation. We define visual analogical relation by cosine distance between vector representations of referents in visual embedding space. In particular, we extracted the visual embeddings from convolutional neural networks—VGG-19 (Simonyan & Zisserman, 2015), a state-of-the-art convolutional image classifier pre-trained on the ImageNet database (Deng et al., 2009)—following procedures from work on visually-grounded word learning (Lazaridou et al., 2016). Under this measure, concepts that share visual features (such as *ball* and *balloon*, both of which are round objects) should yield a relatively low distance even if they are remotely related in the taxonomy. To obtain a robust visual representation for each concept c, we sampled a collection of images $I_1, \ldots, I_k$ up to a maximum of 512 images from ImageNet. With each image I_j processed by the neural network, we extracted the corresponding visual feature vector from the first fully-connected layer after all convolutions: v_j^c. We then averaged the sampled k feature vectors to obtain an expected vector v^c for the visual vector representation of c.

Predicate-based relation. We define predicate-based relation by leveraging the psychological measure of word association. We assume that two referents that frequently co-occur together should also be highly associable, e.g., *key* and *door*. Specifically, we followed the procedures in De Deyne, Navarro, Perfors, Brysbaert, and Storms (2018) and took the "random walk" approach to derive vector representations of referents in a word association probability matrix. This procedure generates word vectors based on the positive pointwise mutual information from word association probabilities propagated over multiple leaps in the associative network. As a result, concepts that share a common neighbourhood of associates are more likely to end up closer together in the vector space. De Deyne et al. (2018) showed that this measure yields superior correlations with human semantic similarity judgements in comparison to other measures of association. We used word association data from the English portion of the Small World of Words project (De Deyne et al., 2018). The data is stored as a matrix of cue-target association probabilities for a total of 12292 cue words. We used the implementation provided by the authors (`https://github.com/SimonDeDeyne/SWOWEN-2018`)

to compute vector representations from the association probability matrix. We used cosine distance to compute predicate-based distances between pairs of referent vectors.

To ensure that the three types of relational features provide complementary information, we calculated their intercorrelations based on 66 concept pairs that we used for our analyses. Although correlations were significant ($p < .001$), all coefficients were low (category & visual: 0.179; category & predicate: 0.186; visual & predicate: 0.274).

Data

We collected linguistic data from three sources: 1) Metadata of child overextension from the literature; 2) Vocabulary of early childhood; 3) Text corpora of child-caretaker speech.

Metadata of child overextension. We performed a meta survey of 12 representative studies from developmental psychology and collected a total of 86 overextension example word-referent pairs. Each pair consists of an overextended word and the novel referent that word has been extended to. We kept word-referent pairs that overlapped with the available data from the three features we described, resulting in a total of 66 word-referent pairs. Table 1 shows examples from this meta dataset and their sources from the literature.

While the data we used for analysis may not constitute an unbiased sample of child overextension, two factors help to alleviate this concern. First, we followed a systematic approach in data collection by recording every utterance-referent pair in which both constituents could be denoted by one noun. Second, the diversity of the sources that we examined reduces the possibility of biasing our sample from any individual study.

Table 1: Examples of overextension data.

Uttered word → *Referent*	Source
"banana" → *moon*	Behrend, D. A. (1988)
"car" → *truck*	Fremgen, A., & Fay, D. (1980)
"apple" → *orange juice*	Rescorla, L. A. (1981)
"ball" → *bead*	Barrett, M. D. (1978)
"fly" → *toad*	Clark, E. V. (1973)
"cow" → *horse*	Gruendel, J. M. (1977)
"apple" → *egg*	Rescorla, L. A. (1980)

Vocabulary from early childhood. To approximate children's vocabulary in early childhood, we collected nouns reported to be produced by children of up to 30 months of age from the American English subset of the Wordbank database (Frank, Braginsky, Yurovsky, & Marchman, 2017). Because overextension has been typically reported to occur between 1;1 and 2;6 years (Clark, 1973) (that covers the range in Wordbank), we constructed a vocabulary V using all the nouns from Wordbank for which we could obtain the required semantic features. The resulting vocabulary includes 316 out of the 322 nouns from the database.

Corpora of child-caretaker speech. To evaluate our models in a realistic communicative context, we collected a large set of child-caretaker speech transcripts from the CHILDES database (MacWhinney, 2014), for child Eve (age 1;6 to 2;3) from the Brown corpus (Brown, 1973), Peter (1;9 to 3;1) from the Bloom70 corpus (Bloom, Hood, & Lightbown, 1974), and Nina (1;11 to 3;3) from the Suppes corpus (Suppes, 1974). We chose these children's data because their ages closely match the typical overextension period reported in child development. We considered each transcript as forming a communicative environment, and from each environment, we collected the set of all nouns uttered by the child and the caretaker for the analyses detailed in the next section. In total, we obtained 1586 communicative environments with a median of 139 distinct nouns per context.

Results

We assess our proposed framework in three aspects: 1) model accuracy in reconstructing child and caretaker strategies in overextension; 2) evidence for multimodal chaining in overextension; 3) model generation of chain complex.

Model reconstruction of overextension strategies

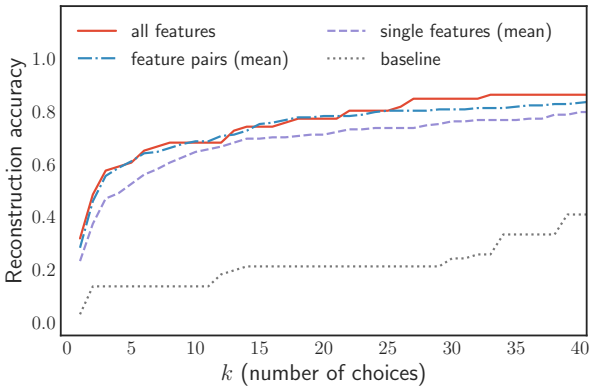
Production. We evaluated the child production model against the curated set of overextension word-referent pairs, $O = \{(w_i, c_i)\}$, with respect to all words in the child vocabulary V. For each pair, the model chooses the target word based on the given overextended sense c_i by assigning a probability distribution over words w in V. We assessed the model by finding the maximum *a posteriori* probability (MAP) of all the overextension pairs under the single sensitivity parameter h, which we optimized to the MAP objective function via standard stochastic gradient descent:

$$\max_h \prod_i p_{\text{prod}}(w_i|c_i;h,V) = \max_h \prod_i \frac{p_{\text{prod}}(c_i|w_i;h)p(w_i)}{\sum_{w\in V} p_{\text{prod}}(c_i|w;h)p(w)} \tag{6}$$

To assess the contribution of the three relational features, we tested this production model under single features and all possible combinations of features in pairs and triplets. We also compared these models under the frequency-based prior versus those under a uniform prior, along with a baseline model that chooses words only based on the prior distribution. We evaluated all models under two metrics: the Bayesian information criterion (BIC), which is a standard measure for probabilistic models that considers both degree of fit to data (i.e., likelihood) and model complexity (i.e., number of free parameters); a performance curve that measures model accuracy at different values of k, similar to the standard receiver-operating curve (ROC), where we assessed the predictive accuracy of each model from its choice of top k words for different levels of k, or the proportion of overextension pairs (w_i, c_i) for which the model ranks the correct production w_i among its top k predictions for referent c_i.

The left two columns of Table 2 summarize the BIC scores of the family of production models. We made three observations. First, models that incorporate features performed

(a) Production model



(b) Comprehension model

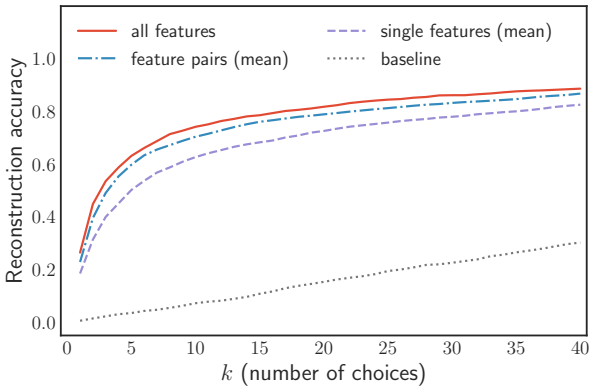


Figure 3: Performance curves for production and comprehension models.

Table 2: Bayesian Information Criterion (BIC) scores for production and comprehension (comp.) models.

	Production		**Comp.**
Model likelihood	freq. prior	unif. prior	unif. prior
Baseline	695	760	13268
category (cat.)	502	583	9663
visual (vis.)	496	574	10344
predicate (pred.)	499	582	9890
cat. + vis.	454	537	8949
cat. + pred.	457	544	8885
vis. + pred.	461	546	9261
all features	**439**	**526**	**8594**

better than the baseline (i.e., lower in BIC scores), suggesting that children overextend words by making explicit use of the semantic relations we considered. Second, models with the frequency-based prior performed dominantly better than those with the uniform prior, suggesting that children jointly consider word usage frequency (or effort) and semantic relations in overextension. Third, models with featural integration performed better than those with isolated features (i.e., all features<features pairs<single features in BIC score), suggesting that children rely on multiple kinds of semantic relations in overextensional word choices. Figure 3a further confirms these findings in performance curves that show the average predictive performance under the full range of k in top k modelled word choices: all features>features pairs>single features>baseline in the area under curves.

Comprehension. We next assessed the caretaker comprehension model by asking whether the model can retreive the intended referent from an uttered overextended word. Because we do not have the actual records of caretakers' inferences, we simulated a dataset for model evaluation by 1) identifying child-caretaker speech scripts that contain the overextended referents $\{c_i\}$ from our curated data; 2) replacing the correct word for a referent c_i (in the script) with the overextended word w_i reported in the literature. We then examined if the model is able to retrieve the correct referent c_i based on w_i among all other competing nouns in the communicative context of a script. As an example, knowing that "ball" has been reported to be overextended to *balloon*, we would identify child speech scripts that contain the word "balloon" and replace that word with "ball". We would then run our comprehension model and check if the top referents recovered by the model contain "balloon" among other nouns that appeared as context in that given script.

Similar to the case of production, we assessed the model by maximizing the posterior comprehension probability over all curated referents based on their appearances in the CHILDES transcripts. We optimized the MAP objective function under the sensitivity parameter h using stochastic gradient descent:

$$\max_h \prod_i p_{\text{comp}}(c_i|w_i;h,E_i) = \max_h \prod_i \frac{p_{\text{comp}}(w_i|c_i;h)p(c_i)}{\sum_{c\in E_i} p_{\text{comp}}(w_i|c;h)p(c)} \quad (7)$$

We used the same two metrics to evaluate the family of comprehension models and summarized the BIC-based results in the third column of Table 2 and ROC-based results in Figure 3b. We observed that results are qualitatively similar to those obtained in the production model: the rank order of performance among baseline model, models with single features, feature pairs, and all features, remains unchanged.

Evidence for multimodal chaining

To examine directly how the multimodal semantic space we constructed accounts for variation in the overextension data, we performed a logistic regression analysis. In particular, we considered two sets of data: the *attested set* overextension word-referent pairs, and a *control set* that shuffles the word-referent mappings from the attested set. We then performed a binary classification task via logistic regression to assess whether the attested pairs can be detected from the control pairs, given the same three relational features that we used for our previous analyses. The logistic model achieved 83% accuracy, compared to 50% chance. We also trained models on subsets of the feature space, achieving best feature pair performance of 82% and best single feature performance of 80%. This suggests that semantic relations provide significant predictability of concepts that might undergo overextension.

Figure 4 shows the distribution of dominant features across the 66 overextension pairs (we labelled each pair according to the top-scoring feature in the logistic regression model), along with a few examples that are best explained by each relational type. We observed that the contributions of these features are roughly even, providing support for the view that children rely on a combination of modalities in overextension.

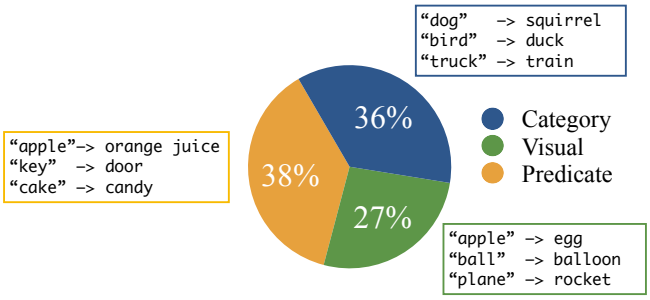


Figure 4: Percentage shares and examples explained by the three types of features from the curated overextension dataset.

Model simulation of chain complex

To illustrate how our model might simulate a chain complex, we applied an iterative scheme to sample a chain of concepts from the multimodal semantic space. Specifically, we began the chaining process with a seed word w_0 and initial chain $C^0 = \{w_0\}$. In the j-th iteration, we sampled word w uniformly from C^{j-1}, and word w' from children's vocabulary V according to probability distribution $p(w'|w) \propto \text{sim}(c_{w'}, c_w)$, where the similarity function is defined in Equation 2. We then added w' to the chain complex by linking it to w, hence extending the chain to $C^j = C^{j-1} \cup \{w'\}$. Figure 5 shows a chain complex sampled from seed concept "door". Similar to Vygotsky's "quah" example, it features referent-to-referent extensions that involve different types of relations, illustrating the thought processes that could have given rise to the diverse overextension patterns attested in young children.

While exploratory in nature, our simulation demonstrates the potential of a multimodal approach to capture the formation of chain complexes in child overextension. Future work should explore this generative aspect of the framework in more rigorous terms.

Discussion

We have presented a formal framework for characterizing children's overextension. We have shown that this framework yields good accuracies in reconstructing child-caretaker communication based on a relatively large set of overextension examples we curated from the developmental literature. Our results indicate that the diverse range of overextension patterns can be explained by our framework that encapsulates a multimodal representation of semantic relations with categorical, visual, and predicate-based features.

With respect to earlier work from developmental psychology, our results support the view that children's overextended

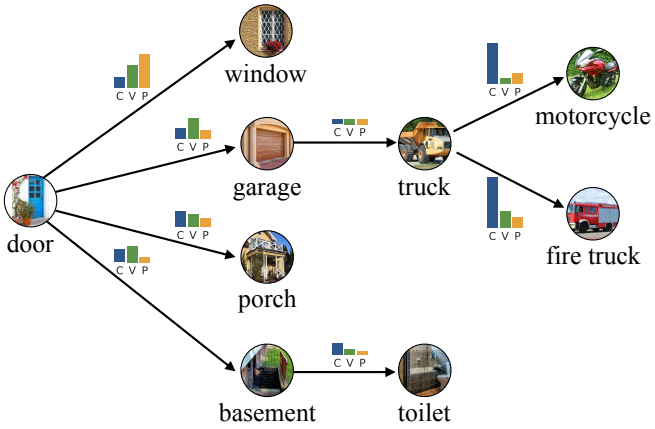


Figure 5: Chain complex sampled from the multimodal semantic space, and contributions of categorical (C), visual (V), and predicate-based (P) relations to chaining probabilities.

word choices reflect a communicative strategy under a limited vocabulary. Moreover, we have shown that children tend to favour high-frequency words in overextension, which provides evidence for the retrieval-error view of overextension. Future work should explore whether the current framework can explain overextension in children's language comprehension, as well as account for the later convergence to adult word usage.

We have shown the initial promise of a multimodal representational scheme toward a better characterization of the generative capacity for word sense extension in early childhood. Future work could explore the generality of this framework in accounting for overextension beyond nouns, as well as historical changes of word meaning.

Acknowledgements

We would like to thank Yu B Xia for helping with collection of child overextension data, Charles Kemp for reference to Vygotsky's work, and Suzanne Stevenson for constructive comments on the draft. We are also thankful to the members of the Computational Linguistics group at the University of Toronto for comments on an early version of this work. This research is supported by an NSERC DG grant and a Connaught New Researcher Award to YX.

References

Beekhuizen, B., & Stevenson, S. (2016). Modeling developmental and linguistic relativity effects in color term acquisition. In *CogSci 38*.

Bird, S., & Loper, E. (2004). Nltk: the natural language toolkit. In *ACL 42*.

Bloom, L. (1973). *One word at a time: the use of single word utterances before syntax*. Mouton.

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when, and why. *Cognitive Psychology*, *6*(3), 380–420.

Brown, R. (1973). *A first language: The early stages*. Harvard U. Press.

Clark, E. V. (1973). What's in a word? on the child's acquisition of semantics in his first language. In *Cognitive development and acquisition of language*.

Clark, E. V. (1978). Strategies for communicating. *Child Development*, 953–959.

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2018). The small world of words english word association norms for over 12,000 cue words. *Behavior Research Methods*, 1–20.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *CVPR 2009*.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, *34*(6), 1017–1063.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *J Child Lang*, *44*(3), 677–694.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychol Sci*, *20*(5), 578–585.

Fremgen, A., & Fay, D. (1980). Overextensions in production and comprehension: A methodological clarification. *J Child Lang*, *7*(1), 205–211.

Gershkoff-Stowe, L. (2001). The course of children's naming errors in early word learning. *Journal of Cognition and Development*, *2*(2), 131–155.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Huttenlocher, J. (1974). The origins of language comprehension. In *Theories in cognitive psychology: The loyola symposium* (pp. xi, 386–xi, 386).

Kachergis, G., Yu, C., & Shiffrin, R. M. (2017). A bootstrapping model of frequency and context effects in word learning. *Cognitive Science*, *41*(3), 590–622.

Kay, D. A., & Anglin, J. M. (1982). Overextension and underextension in the child's expressive and receptive speech*. *Journal of Child Language*, *9*(1), 83–98.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. U Chicago Press.

Lazaridou, A., Chrupała, G., Fernández, R., & Baroni, M. (2016). Multimodal semantic learning from child-directed input. In *NAACL-HLT 15*.

MacWhinney, B. (2014). *The childes project: Tools for analyzing talk, volume ii: The database*. Psychology Press.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 201–233).

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

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences*, *115*(10), 2323–2328.

Rescorla, L. A. (1980). Overextension in early language development. *J Child Lang*, *7*(2), 321–335.

Roy, D. K., & Pentland, A. P. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, *26*(1), 113–146.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *ICLR 2015*.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

Suppes, P. (1974). The semantics of children's language. *American Psychologist*, *29*(2), 103.

Thomson, J. R., & Chapman, R. S. (1977). Who is daddy revisited: the status of two-year-olds' over-extended words in use and comprehension. *Journal of Child Language*, *4*(3), 359–375.

Vygotsky, L. S. (1962). *Language and thought*. MIT Press.

Wittgenstein, L. (1953). *Philosophical investigations* (G. Anscombe, Trans.). Prentice Hall.

Wu, Z., & Palmer, M. (1994). Verbs semantics and lexical selection. In *ACL 32*.

Xu, Y., Regier, T., & Malt, B. C. (2016). Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, *40*(8), 2081–2094.

Yu, C. (2005). The emergence of links between lexical acquisition and object categorization: A computational study. *Connection Science*, *17*(3-4), 381–397.