To appear in Cognition.

A computational theory of child overextension

Renato Ferreira Pinto, Jr.[a], Yang Xu[b,*]

[a]*Department of Computer Science, University of Toronto*
[b]*Department of Computer Science, Cognitive Science Program, University of Toronto*

Abstract

Overextension—the phenomenon that children extend known words to describe referents outside their vocabulary—is a hallmark of lexical innovation in early childhood. Overextension is a subject of extensive inquiry in linguistics and developmental psychology, but there exists no coherent formal account of this phenomenon. We develop a general computational framework that captures important properties of overextension reported separately in the previous literature. We operationalize overextension as probabilistic inference over a conceptual space that draws on a fusion of knowledge from lexical semantics, deep neural networks, and psychological experiments to support both production and comprehension. We show how this minimally parameterized framework explains overextension in young children over a comprehensive set of noun-referent pairs previously reported in child speech, and it also predicts the behavioral asymmetry in children's overextensional production and comprehension reported in lab settings. Our work offers a computational theory for the origins of word meaning extension and supports a single-system view of language production and comprehension.

Keywords: overextension; lexical innovation; word meaning extension; multimodality; production-comprehension asymmetry

1. Introduction

Young children often extend known words to referents outside their vocabulary, a phenomenon known as overextension [1]. For example, children might extend *dog* to

*Corresponding author
Email address: yangxu@cs.toronto.edu (Yang Xu)

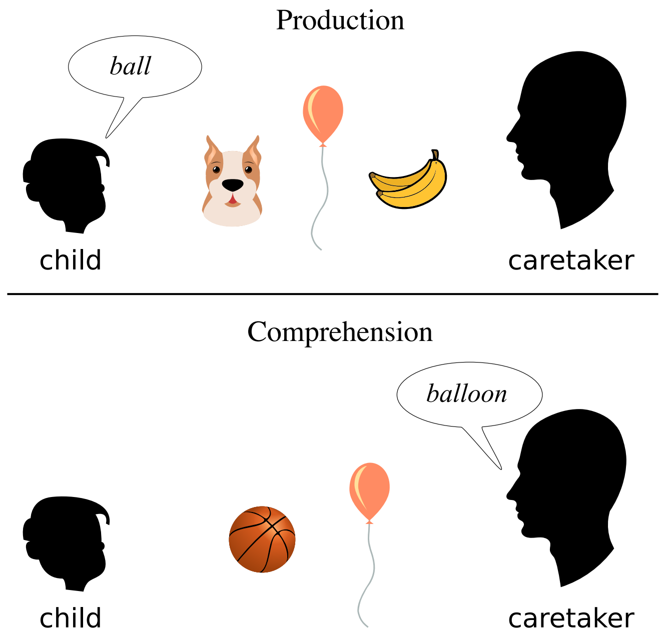

Figure 1: Illustration of overextension in child-caretaker communication. Production: The child chooses to extend the meaning of a known word—*ball* in this scenario—to refer to the object balloon, word for which has not yet entered the child's vocabulary. Comprehension: The child, as a listener, must infer the meaning of the caretaker's utterance—*balloon*—given possible confounding referents in the environment (e.g., a ball).

refer to a squirrel, *ball* to refer to a balloon, or *key* to refer to a door. Overextension takes place typically between 1 and 2.5 years in child development [2] and evidences early capacity for lexical innovation under communicative and cognitive pressures. Work in linguistics and developmental psychology has made important discoveries about overextension [3, 2, 4, 1, 5], but to our knowledge there exists no formal coherent account that synthesizes these ideas to explain the wide array of behaviors in overextension, both in terms of children's production and comprehension (see Figure 1 for an illustration). Here we present a computational framework for characterizing the origins of word meaning extension that connects different findings about overextension in the literature.

Vygotsky [3] describes overextension as a crucial stage of early concept formation. In his classic example, a child first uttered *quah* to refer to a duck in a pond, then to bodies of water, to liquids in general, including milk in a bottle, as well as to a coin with an eagle imprinted on it, and subsequently other round, coin-like objects. Vygotsky's work provides an anecdotal account of overextension and resonates with work in philosophy and cognitive linguistics suggesting how word meanings involve rich but perplexing semantic relations [e.g., 6, 7]. However, this account does not specify the conceptual basis and mechanism that give rise to the word choices that children produce in overextension.

A study by Rescorla [5] extends the early work by suggesting that children's lexical production of overextension relies on rich conceptual knowledge. In her diary study of six children, Rescorla has identified three main types of semantic relations that connect conventional and overextended referents of a word, described as 1) *categorical* relation: overextension by linking objects within a taxonomy (e.g., *dog* referring to a squirrel); 2) *analogical* relation: overextension by linking objects with shared visual or other perceptual properties (e.g., *ball* referring to an apple); and 3) *predicate-based* relation: overextension by linking objects that co-occur frequently in the environment (e.g., *key* referring to a door).

Separate from the literature that documents children's overextension from the perspective of lexical production [e.g., 8, 2, 5], several studies have shown that children's lexical comprehension also exhibits the property of overextension, and there are important behavioral differences in terms of overextensional production and comprehension. In particular, children tend to misintepret the meaning of a word by overextending to other (related or confounding) referents in the environment [9, 10, 11]. The extent that overextension behavior in comprehension mirrors that in production has been a subject of controversy [9, 12, 10], but one observation persists [13, 14]: children often overextend in production even when they correctly infer the appropriate adult word in comprehension, i.e., there exists a production-comprehension asymmetry such that comprehension tends to mature earlier than production in development. For example, Rescorla [5] reports a child who consistently identified the correct referent upon hearing the word *strawberry*, but would still overextend the word *apple* to refer to strawberries in production. This asymmetry reflects the general trend that comprehension leads production in language development [15], but it remains debated whether comprehension and production rely on two separate systems or a single system [16].

Although several hypotheses have been proposed to explain both the mechanisms behind overextension as well as the relationship between production and comprehension, existing views are mixed as to the explanation of overextension in terms of: 1) incomplete conceptual system [2, 17, 18], 2) pragmatic choice under limited vocabulary [4], and 3) retrieval error [12, 19, 20, 11]. The first view poses children's immature conceptual development as the root of overextension, suggesting that children overextend words because their developing conceptual system cannot yet distinguish concepts to the extent that adult words do. This explanation addresses the semantic aspect of lexical innovation, but not the production-comprehension asymmetries, since incomplete conceptual knowledge alone could not explain words being correctly understood but not produced. The other two views focus on this latter aspect by suggesting that children overextend words either as a communicative strategy when they lack the proper vocabulary and thus rely on an approximation to accomplish their communicative goals, or due to performance errors caused by the cognitive effort of retrieving unfamiliar words. However, these theories do not propose a formal model to explain the conceptual leaps that children make when they do overextend words in production or comprehension.

We present a formal approach to child overextension that is aimed at explaining the various findings about this phenomenon under a coherent view. We propose a general computational framework that models child overextension both in terms of production and comprehension, and we evaluate this framework rigorously against empirical findings reported previously from naturalistic and lab settings.

We focus on modeling the overextension of nouns which represent a broad class of concepts in the lexicon. We contribute a new dataset of 236 noun pairs (i.e., noun-referent) collected from the literature which we have made publicly available (see *Supplementary Material*). We show that our computational framework not only explains children's overextended word choices over different semantic modalities, but with no further modification it also replicates the empirical findings about production and comprehension from independent psychological experiments. Our framework shows that overextension in both production and comprehension can be explained by inferential processes on common conceptual knowledge, thus providing support for the single-system view of language production and comprehension.

Although we focus our experiments and discussion on presenting a unified model that explains overextension in production and comprehension, it is worth highlighting that our work can also be seen as a more general framework of reference from which overextension follows. We elaborate on this view in our second set of experiments, which shows that our model can predict both correct and overextended production and comprehension following empirical findings, and in *Supplementary Material*, in which we show how a longitudinal simulation of our model can suggest developmental trajectories from early overextension to adult concept formation and reference.

2. Relations to existing computational work

Our work extends the broad literature on computational modeling of word learning and language acquisition.

A prominent line of research emphasizes modelling cross-situational word learning, which posits that children infer the conventional meanings of words by leveraging the statistical regularities in natural utterances across different situations [21, 22, 23, 24]. Cross-situational word learning has been tackled by several methodological approaches, including symbolic [21], associative [25, 26, 27], and Bayesian [28, 29, 30] models; independent research has also proposed connectionist accounts of word learning algorithms [31, 32, 33, 34, 35]. Differing from this rich area of research, our work instead focuses on the innovative aspects of the lexicon on the path toward the acquisition of proper or conventional language.

Our framework draws on a multimodal semantic representational space that is inspired partly by recent work on visually grounded word learning [36, 37, 38]. This line of research uses visual features in the environment to model word learning as a process grounded in visual perception. Our work employs similar techniques to account for overextension patterns based on visual analogy but also goes beyond by incorporating semantic relations of other types, including taxonomic and predicate relations.

Although computational approaches to child lexical innovation are still in their infancy, some recent research has explored particular aspects of this problem. For example, Alishahi and Stevenson [39, 40] developed a probabilistic model of early argument structure acquisition that simulates a transient period of overgeneralized verb argument structure (e.g., *Mary fall toy*). Related work has studied the relationship between cross-linguistic variation in lexicalization and child overextension of spatial prepositions and color terms [41, 42]. This line of research has suggested that both word frequency and implicit cognitive biases inferred from cross-linguistic tendencies play a role in predicting children's overextension patterns in these individual semantic domains. Our approach here offers a general way of constructing semantic relations that approximates children's conceptual structure in overextension, and we show how these relations can be integrated to reproduce overextension behavior across (as opposed to within) domains. We also show that our models predict the differences between production and comprehension observed in child overextension without additional parameter tuning.

3. Computational formulation of theory

We first present three theoretical hypotheses we explore in our computational approach to overextension. We then formulate overextension as probabilistic inference during communication in which a child, in production, wishes to refer to a novel object given vocabulary and cognitive constraints, and, in the opposite comprehension scenario, needs to infer the intended meaning of an utterance given available referents in the environment (see Figure 1). We describe our framework in terms of two main components: 1) a generic probabilistic process of overextension for production and comprehension, and 2) the construction of a multimodal semantic space that supports probabilistic inference.

3.1. Theoretical hypotheses

We posit three hypotheses under our framework:

1. *Multimodality*: a combination of multiple types of semantic relations should better predict children's overt strategies of word choices in overextension than features treated in isolation;

2. *Effort-saving production (or frequency effect)*: words that occur more frequently in children's linguistic environment are favored over less common words in overextended production;
3. *Production-comprehension asymmetry*: by reflecting task differences between production and comprehension, a single framework should predict the empirical observations on child behavior in production and comprehension including the reported asymmetry.

Each of these hypotheses is grounded in the previous findings about overextension: the first hypothesis integrates the idea that a developing conceptual system forms the basis of children's overextension [2, 17, 18, 3] with the observations of Rescorla [5] on the multimodal nature of the semantic relations underlying individual word choices; the second hypothesis represents the view of cognitive difficulty in retrieving unfamiliar or recently-learned words as a cause of overextended word choices in production [12, 19, 20, 11]; and the third hypothesis materializes the proposal of Thomson & Chapman [11] that task differences may be the key to combining early conceptual organization and retrieval difficulty into a general model of overextension. In this respect, our model does not make new discoveries. However, an important distinction between our work and the previous studies is that we provide a single account of overextension that coherently explains these empirical findings reported previously in separation, whereas the existing literature has not proposed or evaluated a general formal theory that specifies how the different findings may be explained coherently. We test the validity of each of our hypotheses through computational experiments with a large meta dataset of child overextension in production as reported in an array of previous studies, as well as independent behavioral data of production and comprehension collected from lab experiments.

3.2. Probabilistic framework

Production. Consider a child with limited vocabulary V who wishes to refer to some concept c in the environment (e.g., a balloon), where the adult word for c may not be in the child's existing vocabulary. Given a candidate word $w \in V$ for production (e.g., *ball*), we specify the following probabilistic model of word choice in overextension:

$$p_{\text{prod}}(w|c) = \frac{p(c|w)p(w)}{\sum_{w' \in V} p(c|w')p(w')} \tag{1}$$

The likelihood term $p(c|w)$ measures the appropriateness of referring to (or categorizing) concept c with word w, and is defined as a density function (specified later) that depends on the semantic similarity between c and c_w, or the concept that word w signifies conventionally, e.g., *ball* for "ball":

$$p(c|w) = f_{\text{sim}}(c|c_w) \tag{2}$$

The prior $p(w)$ encodes the notion of cognitive effort, that is, some words are easier to retrieve than others. Following previous work showing the effect of word frequency on overextension [42], we define $p(w)$ as a frequency-based word prior:

$$p(w) = \frac{F(w)}{\sum_{w' \in V} F(w')} \tag{3}$$

where $F(w)$ is the total frequency of word w in a representative corpus of children's linguistic environment. In this account, frequent words are more likely to be chosen for overextension, and we test this assumption rigorously against the lexical choices that children were reported to make in overextension.

Comprehension. In the case of comprehension, the child hears word w and estimates probabilistically that it refers to some concept c in the referential environment. The comprehension model recovers the similarity-based measure used above in its probabilistic formulation:

$$p_{\text{comp}}(c|w) = \frac{p(w|c)p(c)}{\sum_{c' \in E} p(w|c')p(c')} \tag{4}$$

The likelihood term $p(w|c)$ measures the appropriateness of word w to refer to concept c, and is defined by the multimodal similarity function: $p(w|c) = f_{\text{sim}}(c_w|c)$. The prior $p(c)$ is set to the uniform distribution over the set of possible referents E in the child's environment, reflecting the assumption that referents in the environment are equally likely to be chosen as the target referent *a priori*. This choice also reflects the

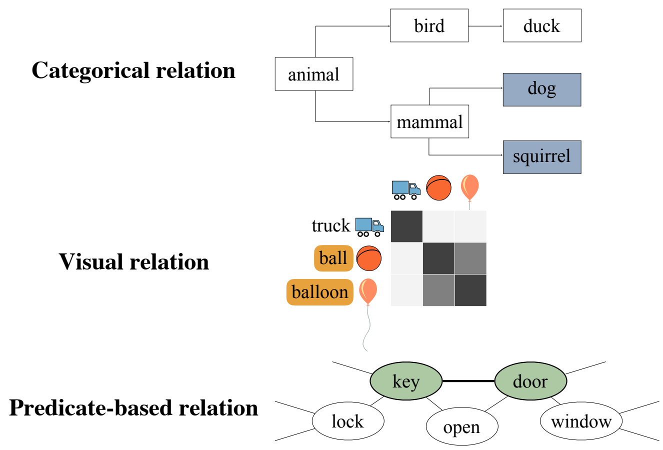

Figure 2: Types of semantic relations in multimodal semantic space.

hypothesis that cognitive effort plays a larger role in overextension in production than in comprehension [12, 19, 20, 11]. However, we show in the *Supplementary Material* that under a frequency-based prior (as in the case of the production model), our models also capture important components of the production-comprehension asymmetry reported in the literature, and hence elucidating the contribution of both the prior and likelihood components in our models.

Multimodal semantic space. We define a multimodal semantic space that captures the three types of relational features described by Rescorla [5]: categorical relation, visual analogy,[1] and predicate-based relation. We construct these relational features using a fusion of resources drawn from lexical semantics, deep learning networks, and psychological experiments, as illustrated in Figure 2 and specified as the following.

1) *Categorical relation*. We define the categorical relation between two referents via a standard distance measure d_c in natural language processing by Wu and Palmer [43], based on taxonomic similarity. Concretely, for two concepts c_1 and c_2 under a taxonomy

[1]While Rescorla defined analogy to include broader perceptual features, such as auditory, we restrict this investigation to visual features in the interest of data availability for a large-scale study.

T (i.e., a hierarchy), the distance is:

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

N_{LCS} denotes the depth of the least common subsumer of c_1 and c_2 in the taxonomy, and N_1 and N_2 denote the depths of the two concepts. This distance measure is bounded between 0 and 1, and is larger for concepts that are more distantly related (i.e., share fewer common ancestors) in the taxonomy. 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 [44] and annotated words by their corresponding *synsets* in the database. We used the *NLTK* package [45] to calculate similarities between referents for this feature.

2) *Visual analogical relation.* We define the 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 [46], a state-of-the-art convolutional image classifier pre-trained on the ImageNet database [47]—following procedures from work on visually-grounded word learning [36]. 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 256 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.

3) *Predicate-based relation.* We define the predicate-based relation by leveraging the psychological measure of word association. Word associations reflect many kinds of semantic relationships, and importantly some of these relationships are predicate-based that are not captured by either the "categorical" or the "visual" component of the model, e.g., in the case of key and door. We assumed that two referents that co-occur together

frequently should also be highly associable, and we followed the procedures in [48] by taking 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 point-wise mutual information from word association probabilities propagated over multiple leaps in the associative network. As a result, concepts that share a common neighborhood of associates are more likely to end up closer together in the vector space. De Deyne et al. [48] showed that this measure yields superior correlations with human semantic similarity judgments in comparison to other measures of association. We used word association data from the English portion of the Small World of Words project [48]. The data is stored as a matrix of cue-target association probabilities for a total of 12,292 cue words. We used the implementation provided by the authors[2] to compute vector representations from the association probability matrix. We used cosine distance to compute predicate-based distances between pairs of referent vectors.

To complete our formulation of the multimodal semantic space, we integrate the three types of semantic relations specified above into a density function based on conceptual similarity that measures the likelihood of concepts being associated by overextension in the probabilistic framework.

We take the Gaussian-Euclidean form of the generalized context model (GCM) or exemplar model of categorization [49], which defines the similarity between two concepts c_1 and c_2 as a decaying function of the distance separating them in psychological space. First, the model computes the distance between the concepts as the Euclidean norm over the distance components in each psychological dimension:

$$d(c_1,c_2) = \left[d_c(c_1,c_2)^2 + d_v(c_1,c_2)^2 + d_p(c_1,c_2)^2\right]^{1/2} \tag{6}$$

Under this formulation, the psychological dimensions correspond to the three types of multimodal relations: categorical distance d_c, visual analogical distance d_v, and predicate-based distance d_p. Then, a Gaussian kernel computes concept similarity as a

[2] https://github.com/SimonDeDeyne/SWOWEN-2018

decaying function of psychological distance:

$$\mathrm{sim}(c_1, c_2) = \exp\left(-\frac{d(c_1, c_2)^2}{h}\right) \tag{7}$$

This similarity measure is modulated by a single *kernel width* parameter h, which controls the sensitivity of the model to the distance function. The magnitude of h determines how slowly the similarity measure decreases with respect to distance in the multimodal relations. We empirically estimate the value of h from data in the experiments and provide a simulation of the model based on a range of values for the sensitivity parameter in *Supplementary Material*.

To formulate a parsimonious model, we use a single kernel width parameter to modulate all three unmodified distance measures (instead of three separate parameters). While further refinements such as normalization strategies may be valuable to explore, we found this simple formulation to be sufficient for our empirical evaluations and theoretical inquiries. Furthermore, we show in *Supplementary Material* that allowing independent kernel width parameters to act on each psychological dimension does not change the conclusions from our experiments.

In practice, this similarity measure readily yields the density function required by the production and comprehension models; formally, it must be normalized to form a proper density function:

$$f_{\mathrm{sim}}(c_1|c_2) = \frac{\mathrm{sim}(c_1, c_2)}{Z_h} \tag{8}$$

where Z_h depends only on h,[3] and thus need not be explicitly computed in the models.

To ensure that the three types of relational features provide complementary information, we calculated their inter-correlations based on the 236 concept pairs that we used for our analyses. Although correlations were significant ($p < .001$), coefficients were low or moderate (Spearman's ρ; category *vs* visual: 0.238; category *vs* predicate: 0.445; visual *vs* predicate: 0.421), suggesting that each feature contributes to information encoded in the multimodal semantic space. We further verify the contribution of each

[3]Concretely, $Z_h = \int \exp\left(-\frac{x^2}{h}\right) \mathrm{d}x$

individual feature in a predictive task on overextension (see Section 6).

One potential limitation of our construction of multimodal space is that some of the data sources, namely taxonomy and word association, come from adult-based knowledge (taxonomy) or from experiments performed with adult participants (word association); child-specific sources of similar data are scarce for the purposes of our large-scale experiments. While we acknowledge that features obtained from these data might not perfectly correspond to children's mental representations, we expect these extensively tested data sources to provide useful signal to our experiments, which we confirm by corroborating developmental psychologists' hypotheses in a formal setting. Future work can explore the representational and predictive effects of using child-specific semantic features if they become available at scale, either by collecting such data or by attempting to degrade the adult-level features in a systematic way.

4. Meta data of child overextension

One important evaluation of our framework involves testing our model against a comprehensive array of word-referent pairs comprising children's overextensional production as reported in the child language literature. We collected this meta dataset by performing a meta survey of 8 representative studies from the literature and collected a total of 323 examples of overextension noun-referent pairs. We selected studies containing the most examples of overextended noun-referent pairs as recorded in one of the following conditions: diary records, videotaped play sessions, or picture naming activities. Most (51%) overextension entries for our analyses came from Rescorla's diary studies [50, 51, 5], and the remaining sources complemented this extensive resource.

Each entry in our dataset consisted of an overextended noun and the novel referent that noun has been extended to. We kept word-referent pairs of nouns that overlapped with the available data from the three feature resources we described, resulting in a total of 236 word-referent pairs from 8 different sources. Table 1 shows some examples from this meta dataset and their sources from the literature, and we have made the entire meta dataset available to the community.

While the data we used for analysis may not constitute an exhaustive range of

Table 1: Examples of overextension data, one for each source included in this study.

Uttered word	→	Referent	Source empirical study
ball	→	onion	[11]
car	→	truck	[12]
apple	→	orange juice	[50]
ball	→	marble	[8]
fly	→	toad	[2]
cow	→	horse	[52]
apple	→	egg	[51]
truck	→	bus	[5]

child overextension, we followed a thorough procedure in data collection by recording every word-referent pair in which both constituents could be denoted by one noun. Furthermore, we collected a diverse set of overextensional cases from multiple sources surveyed from the literature as opposed to an individual study. Future empirical efforts to collect larger and systematic records of children's overextension could provide a valuable addition to our work, and we believe that the models we propose here can be applied to those records.

5. Materials and methods

5.1. Data and code availability

Data and code for replication, including a demonstration, are deposited at: `https://github.com/r4ferrei/computational-theory-overextension`.

5.2. 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 [53]. This database is based on the MacArthur-Bates Communicative Development Inventories [54] and aggregates average age of acquisition for over 680 English words. Because overextension has been typically reported to occur between 1 and 2.5 years [2] (that covers the range in Wordbank), we constructed a vocabulary V using all the nouns from Wordbank for which the required semantic features could be obtained. The resulting vocabulary includes 317 out of the 322 nouns from the database (see *Supplementary Material*, Table S4 for a complete list).

5.3. Word frequencies in child-caretaker speech

To capture the distribution of nouns in young children's linguistic environment, we collected a large set of child-caretaker speech transcripts from the CHILDES database [55]. Specifically, we worked with all transcripts from studies performed in naturalistic child-caretaker situations for children between 1 and 2.5 years (the typical overextension period), resulting in 1,713 transcripts with over 200K noun tokens in total.[4]

We measured the relative frequency of each noun by dividing its total number of token occurrences across all transcripts by the total number of noun tokens. Then, to alleviate the effect of minor spelling differences or variability in child versions of adult words (e.g., *mama/mommy/mom*), we counted the frequency of each entry in the overextension dataset as the total frequency of the lemma variations of its *synset* in the WordNet database.

5.4. Model optimization and evaluation methods

In Section 6.1, we evaluated our probabilistic models against the meta set of overextension word-referent pairs, $O = \{(w_i, c_i)\}$, with respect to all words in the child vocabulary 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:

$$\max_h \prod_i p_{\text{prod}}(w_i|c_i;h) = \max_h \prod_i \frac{p(c_i|w_i;h)p(w_i)}{\sum_{w\in V} p(c_i|w;h)p(w)} \tag{9}$$

We maintained this value of h for all other experiments in this paper.

[4]Specifically, we collected transcripts from the studies in [56, 57, 58, 4, 59, 60, 61, 62, 63, 64, 65, 66, 67, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78].

For our likelihood-based evaluations, we used the Bayesian Information Criterion (BIC), a standard measure for probabilistic models that considers both degree of fit to data and model complexity. The score is defined as $\mathrm{BIC} = \log(n)k - 2\log(\hat{L})$, where n is the number of data points, $\hat{L}$ is the maximized likelihood of the model, and k is the number of free parameters (here, $k = 0$ for the prior-only baselines and $k = 1$ for all other models, which are parameterized by the kernel width h).

6. Results

We evaluate the proposed computational framework following two steps: 1) we test model accuracy in predicting children's overextended word choices, as reported from the literature; and 2) we use the same model from step 1) with no parameter tuning to assess its explanatory power on explaining behavior differences in production and comprehension under an independent set of lab experiments, also as reported from the literature.

6.1. Explaining word choices in overextension

To assess how well the model captures children's word choices in overextension, we first evaluated the production model against the meta set of overextension word-referent pairs, $O = \{(w_i, c_i)\}$, with respect to all words in the child vocabulary. 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 the vocabulary.

To assess the contribution of the three features, we considered all possible restrictions of the multimodal space, and thus tested the production model under single features and all possible combinations of feature pairs, along with the full multimodal model consisting of categorical, visual analogical, and predicate-based relations. We also compared these models under the frequency-based prior $p(w)$ versus those under a uniform prior, as well as a baseline model that chooses words only based on the prior distribution.

We evaluated all models under two metrics: Bayesian Information Criterion (BIC) and performance curves similar to receiver operating characteristics. The BIC 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). All of our models that incorporate semantic features contain a single parameter, the kernel width h, and baseline models do not contain any free parameters. Under the second criterion, we assessed model performance curves that measure predictive accuracy at different numbers of allowed model predictions m. Concretely, for each level of m, we measured the predictive accuracy of the model from its choice of top m words in the vocabulary, which is equivalent to the proportion of overextension pairs (w_i, c_i) for which the model ranks the correct production w_i among its top m predictions for referent c_i. Since the dataset for this experiment focuses on overextended word-referent pairs, we similarly limited the word choices available to the model in each prediction by removing the appropriate word from the set of candidates for that concept.

Table 2 summarizes the BIC scores of the family of production models. We make three observations. First, models that incorporate features performed 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. This confirms the first theoretical hypothesis that we presented. Second, models with the frequency-based prior performed dominantly better than those with the uniform prior, suggesting that word usage frequency or cognitive effort and semantic relations jointly affect children's word choices in overextension. This confirms our second hypothesis. Third, models with featural integration performed better than those with isolated features (i.e., all features $<$ feature pairs $<$ single features in BIC score), suggesting that children rely on multiple kinds of semantic relations in overextensional word choices. This provides further evidence for our first hypothesis.

Figure 3 further confirms these findings in performance curves that show average predictive performance under a range of m possible word choices: all features $>$ feature pairs $>$ single features $>$ baseline in the area under curves. Although Figure 3 shows a large range of possible word choices to clearly contrast the performance trends of each family of models, note that predictive performance is reasonable even within a smaller, more plausible number of possible word choices: the full multimodal model correctly predicts 55% of the overextension data in its top 5 word choices (compared to

Table 2: Bayesian Information Criterion (BIC) scores for production models with respect to overextension dataset ($N = 236$). A lower BIC score indicates a better model.

Model	**BIC score**	
	frequency prior	**uniform prior**
baseline	2471	2717
categorical (cat.)	1863	2093
visual (vis.)	1817	2041
predicate (pred.)	1853	2072
vis. + pred.	1732	1947
cat. + vis.	1682	1904
cat. + pred.	1646	1871
all features	**1592**	**1812**

12% accuracy of the frequency baseline model).

It could be argued that, when very few word choices are allowed (e.g., under 5), the model accuracy is still relatively low in absolute terms, and limitations of our current model and data sources could help explain this result—for example, differences between children's knowledge and the adult knowledge comprising our conceptual space, and discrepancies between idealized model inference and the actual inferential processes performed by children in word selection could both be factors limiting the performance of our models. However, we also emphasize that the overextension dataset cannot be taken as an exhaustive account of all possible overextensions that children produce. For instance, the following model predictions are counted as incorrect because the dataset does not contain such word-referent pairs: *tuna* for fish, *tiger* for jaguar, and *orange* for peach. These examples show that many incorrect predictions are still closely related to the target referents and capture the kind of semantic relationship displayed by typical cases of overextension. *Supplementary Material*, Table S3 provides more sample model outputs for both correct and incorrect predictions.

We further evaluated the ability of the three features in our multimodal space to

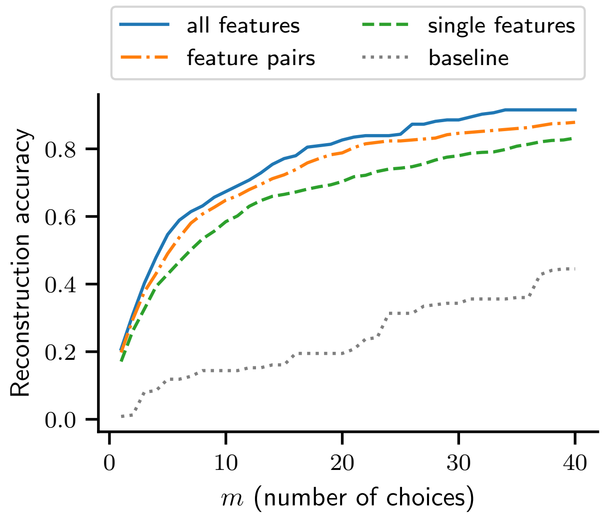

Figure 3: Performance curves for production models showing cross-validated model accuracies in reconstructing word choices ($N = 236$). Aggregated results (single features and feature pairs) show mean accuracy over individual models; see *Supplementary Material*, Figure S2 for a fine-grained comparison of all models.

capture the diversity of semantic relations present in children's overextension in a logistic regression analysis that achieved 84% accuracy in distinguishing the true overextension word pairs in our dataset from randomized control pairs (see more details of this analysis in *Supplementary Material*). Figure 4 shows the estimated contribution of each semantic relation toward characterizing the overextension dataset, along with some examples best explained by each multimodal feature that illustrate how the model captures the different types of semantic relations on which children rely in overextension.

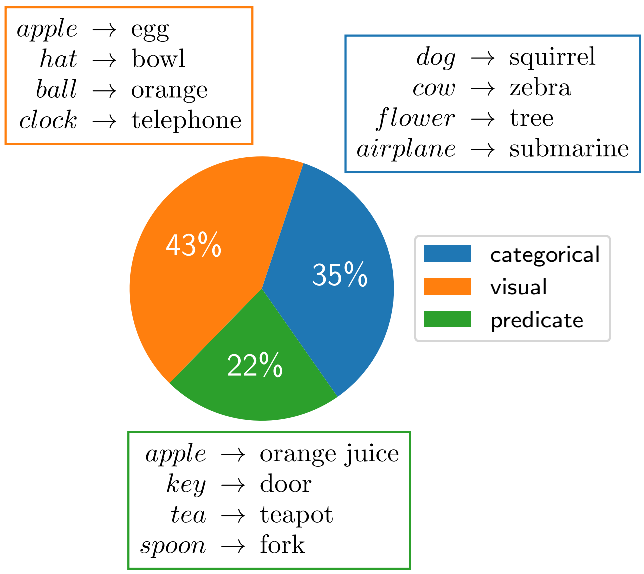

Figure 4: Percentage shares and overextension examples explained by the three types of features in the collected meta dataset ($N = 236$).

6.2. *Explaining production-comprehension behavioral differences*

To assess whether the same modeling framework also accounts for the overextension behaviors in production and comprehension, we performed a set of replication analyses based on the independent empirical study conducted by McDonough [79]. That study analyzed children's performance in production and comprehension with respect to a set of nouns and corresponding visual stimuli in four domains: animals, food, vehicles, and clothes. The 30 nouns were split into two groups by age of acquisition (16 early and 14 late nouns) to test the hypothesis that items typically learned early in development would suffer overextension less frequently than those learned later in development.

In the production task, children were shown the stimuli in sequence and asked to name them. In the comprehension task, in each trial, experimenters showed a triplet of stimuli, uttered a target word, and asked the child to find the stimulus corresponding to the target word. The comprehension task included trials in two conditions: high contrast, in which the two distractors belonged to a different domain than the target stimulus, and low contrast, in which one of the distractors belonged to the same domain as the target stimulus (see Figure 5). Table 3 shows the stimuli triplets and conditions.

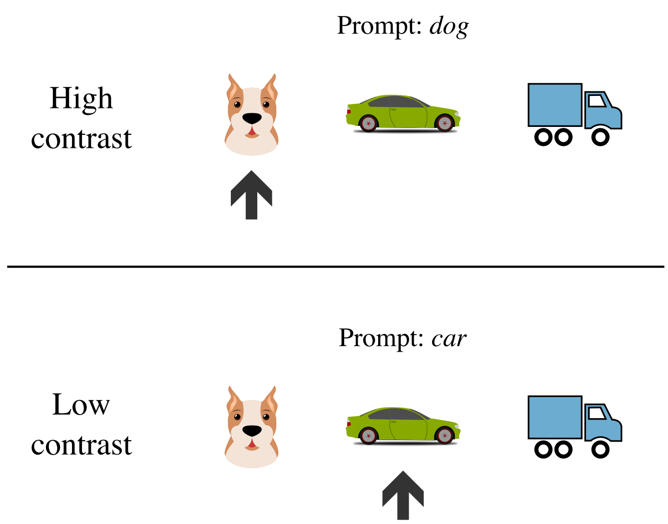

Figure 5: Two conditions in comprehension experiment devised by McDonough [79].

We replicated these experiments with our computational framework. For the production experiment, we presented the production model based on Equation 1 with each stimulus referent c, and measured the probability of correct (target word) production, $p_{\text{prod}}(w|c)$, versus all other words in the child vocabulary. For the comprehension experiment, we presented the model based on Equation 4 with each target word w, and computed the probability of the target referent versus the two distractors in the triplet, $p_{\text{comp}}(c|w)$ (with $E =$ the triplet of stimuli in Equation 4).

The empirical data on the left panel of Figure 6 demonstrates the behavioral asymmetry between production and comprehension. The drop in performance from comprehension to production is particularly striking for late nouns, but even among early nouns, children performed better in the high-contrast condition of the comprehension task than in the production task.

The right panel of Figure 6 shows the results from our model reproduction in terms of the predicted proportion of correct responses per task (production, and comprehension in

Table 3: Experimental stimuli from McDonough [79]. Each row shows one triplet as presented in the comprehension experiment, and columns organize them into high and low contrast selections, as well as early and late items. The bottom section shows triplets omitted from this experiment due to lack of feature data for the stimuli marked by asterisks.

Early noun (High contrast)	**Early noun** (Low contrast)	**Late noun** (Low contrast)
pig	train	bus
cow	pants	shorts
orange	bicycle	motorcycle
dog	car	truck
apple	shirt	vest
cat	dress	sweater
egg	airplane	rocket
shirt	pig	hippo
bicycle	cow	moose
boat	carrot	celery
pants	orange	beet
dress	dog	fox
car	apple	strawberry
train	cat	raccoon
carrot	shoe	*sandal
airplane	cake	*pie

low and high contrast conditions) and per noun group (early and late nouns). We observe that these results replicate the trends from empirical data: in the comprehension task, low contrast trials elicited higher rates of overextension than high contrast trials, and there was no difference between early and late items in comprehension (e.g., *pig* overextended to hippo and *hippo* overextended to pig at similar rates). Welch's t-tests confirmed these results: over the 14 triplets of stimuli, the proportion of correct comprehension in the high contrast, early noun condition ($M = 1.0$) was significantly higher than in the low contrast, early noun condition ($M = 0.92$): $t(13) = 3.05$, $p < 0.01$; and there was no significant difference in the proportion of correct comprehension between the low contrast, early noun condition ($M = 0.92$) and the low contrast, late noun condition ($M = 0.92$): $t(25) = 0.01$, $p = 0.995$. Although the model predicts lower rates of overextension than empirical results, it is worth highlighting that we did not re-tune the parameter h in any way from the previous experiment, and thus the qualitative match shows that the model is able to predict patterns of overextension in comprehension without any exposure to such data beforehand. In the production task, correct labels were produced for early items ($n = 16$, $M = 0.68$) more often than for late items ($n = 14$, $M = 0.30$), and the difference between the two groups was significant (Welch's $t(23) = 6.08$, $p < 0.001$).

To ensure that our results were not tainted by the overlap between overextension data from the previous experiment and the stimuli from the computational replication described here, we repeated this experiment with a model parameter h that was tuned only on overextension pairs in which neither the produced word nor the referent appear in the data from Table 3. This procedure removed 111 out of the 236 overextension pairs from the training data. We observed no relevant changes to our experiment results: all significance values reported above were maintained, as were the relative performance values shown in Figure 6.

Comparing the results in the two tasks, we make two observations. First, the semantic space and probabilistic formulation enable the model to make predictions that recapitulate empirical findings in both production and comprehension, suggesting that the framework captures relevant features of young children's linguistic abilities. Second, the model predicted the asymmetry between production and comprehension without

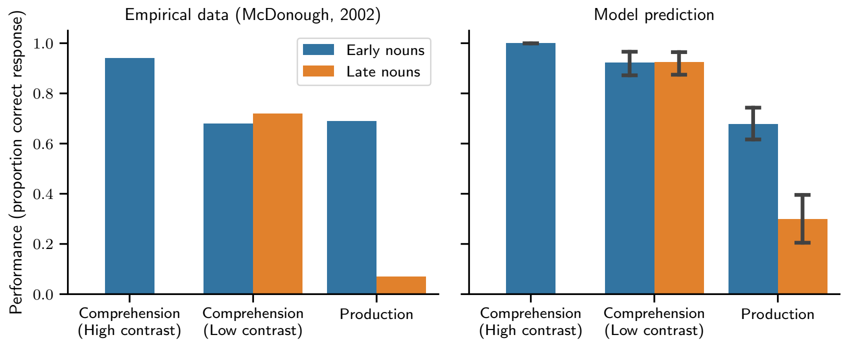

Figure 6: Results of comprehension and production experiments from empirical data of McDonough [79] and from model reproduction. Each bar shows the proportion of correct responses (referent selection in comprehension, and word utterance in production). Comprehension bars show performance over 14 triplets of stimuli, and production bars show performance over 16 early nouns and 14 late nouns. Error bars represent bootstrap 95% confidence intervals.

any modification, showing that a single system can explain the common phenomenon in developmental psychology: that children often overextend words even when they seem to correctly understand the appropriate adult words in comprehension. Our modeling framework reveals that incorporating the task demands of production and comprehension into a probabilistic process grounded in the same representational knowledge is sufficient to capture this asymmetry. Together this set of results confirms the third hypothesis that we proposed.

We highlight that, whereas our first experiment focused on predicting overextended word choices only, our second experiment allowed the model to predict both correct and overextended behaviour in both word choice (production) and referent selection (comprehension), as evidenced by our comparisons of the rates of overextension observed in empirical data and predicted by our model. This observation indicates that our model not only explains overextension in production and comprehension, but also serves as a more general framework of reference from which both overextension and appropriate word usage might follow. We demonstrate this possibility in a longitudinal simulation of our model in *Supplementary Material*.

7. Discussion and conclusion

We have presented a formal computational account of children's overextension. We formulated the problem of overextension in production and comprehension under a probabilistic framework and showed that a shared set of multimodal semantic relations between production and comprehension (combining categorical, visual analogical, and predicate-based features) and a minimally-parameterized model can explain substantial variation in children's overextended word choices from the developmental literature. Furthermore, we showed how the same framework leveraging cognitive effort in word retrieval, specified as a frequency-based prior, enhances model predictability of word choices in production while helping to explain the asymmetry between production and comprehension. Our computational analyses have confirmed the three theoretical hypotheses that we presented initially, and we have provided support for an integrated view of production and comprehension [16], such that production and comprehension in overextension both rely on a single system that supports probabilistic inference over a shared set of representations in a single conceptual space.

Our computational approach also offers a synthesis of the previous psychological findings about overextension. By expressing children's conceptual knowledge via multimodal semantic relations; their lexical choices via a probabilistic process that can overextend in-vocabulary words to out-of-vocabulary referents under communicative need; and cognitive effort in word retrieval as a probabilistic process in which the correct word competes for retrieval with other words in the vocabulary, our framework integrates these ideas into a general account of overextension that explains a broad range of data ranging from naturalistic settings to lab experiments.

Our work adds to an extensive body of computational studies that model word learning in children. While previous research has made substantial progress in modeling the acquisition of conventional language use [25, 28, 27, 80, 81, 82], there is relatively little work on modelling how children innovatively use words to bypass their linguistic limitations for naming out-of-vocabulary referents. Our framework helps to elucidate the computational processes of early word meaning extension and extends related work on modeling overextension within individual domains [39, 40, 41, 42] to modeling meaning extension across domains.

It is important to acknowledge that overextension is a general phenomenon that applies to word classes beyond nouns [2]. Psychologists have observed that children also systematically overextend a variety of other classes of words, for example: antonym pairs related to quantity [*less*/*more* 83] and time [*before*/*after* 84]; dimensional terms such as *big* for more specialized properties including *tall* and *high* [85]; verbs such as *ask* and *tell* [86]; kinship terms such as *brother* and *sister* [87]; spatial terms, with one general purpose term standing in for a variety of spatial relations [1], among others. A challenge remains as how to formalize semantic knowledge more generally that would be applicable to overextension in these other word classes.

A comprehensive formal account of lexical acquisition should also specify a mechanism to capture the phenomenon of vocabulary growth over time. One way to model this process would be to integrate word learning and overextension strategies into a coherent model. Future work should explore this possibility of combining the mechanisms of overextension and word learning to account for child behavior under naturalistic environment.

We have offered a computational account of child overextension that incorporates theories from developmental psychology and supports probabilistic construction and inference of innovative word usages that resemble those described in classical work [3]. Our framework along with the meta dataset that we have collected will pave the way for a formal and scalable characterization of children's lexical innovation. Our work sheds light on the computational basis of word meaning extension as a manifestation of human lexical creativity in early childhood.

Acknowledgments

We are grateful to Leslie Rescorla for sharing the data of child overextension from her dissertation work and providing feedback on the manuscript. We would like to thank Yu B Xia for helping with the collection of the overextension data. We thank Barbara C. Malt, Charles Kemp, Mahesh Srinivasan, and Suzanne Stevenson for their constructive comments on this work. We also thank the members of the Computational Linguistics group at the University of Toronto and Stanford Linguistics for helpful suggestions, and the anonymous reviewers for comments and suggestions that improved this work. This research is supported by a UofT entrance scholarship to RFP Jr. from UofT, a Connaught New Researcher Award, a NSERC Discovery Grant RGPIN-2018-05872, and a SSHRC Insight Grant #435190272 to YX.

References

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

[2] E. V. Clark, What's in a word? on the child's acquisition of semantics in his first language, in: Cognitive development and acquisition of language, 1973.

[3] L. S. Vygotsky, Language and thought, MIT Press, 1962.

[4] L. Bloom, One word at a time: the use of single word utterances before syntax, Mouton, 1973.

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

[6] L. Wittgenstein, Philosophical investigations, Prentice Hall, 1953.

[7] G. Lakoff, Women, fire, and dangerous things: What categories reveal about the mind, U Chicago Press, 1987.

[8] M. D. Barrett, Lexical development and overextension in child language, Journal of Child Language 5 (2) (1978) 205–219.

[9] R. S. Chapman, J. Thomson, What is the source of overextension errors in comprehension testing of two-year-olds? a response to fremgen and fay, Journal of Child Language 7 (3) (1980) 575–578.

[10] C. B. Mervis, K. Canada, On the existence of competence errors in early comprehension: a reply to fremgen & fay and chapman & thomson, Journal of Child Language 10 (2) (1983) 431–440.

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

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

[13] D. A. Behrend, Overextensions in early language comprehension: evidence from a signal detection approach, Journal of Child Language 15 (01) (1988) 63.

[14] L. G. Naigles, S. A. Gelman, Overextensions in comprehension and production revisited: preferential-looking in a study of dog, cat, and cow, Journal of Child Language 22 (1) (1995) 19–46.

[15] E. V. Clark, B. F. Hecht, Comprehension, production, and language acquisition, Annual Review of Psychology 34 (1) (1983) 325–349.

[16] M. J. Pickering, S. Garrod, An integrated theory of language production and comprehension, Behavioral and Brain Sciences 36 (4) (2013) 329–347.

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

[18] C. B. Mervis, Child-basic object categories and early lexical development, in: Concepts and conceptual development: Ecological and intellectual factors in categorization, 1987.

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

[20] J. Huttenlocher, The origins of language comprehension, in: Theories in cognitive psychology: The Loyola Symposium, 1974.

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

[22] C. Fisher, D. G. Hall, S. Rakowitz, L. Gleitman, When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth, Lingua 92 (1994) 333–375.

[23] L. Gleitman, The structural sources of verb meanings, Language Acquisition 1 (1) (1990) 3–55.

[24] S. Pinker, Language learnability and language acquisition, Cambridge, MA: Harvard University Press, 1984.

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

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

[27] C. Yu, D. H. Ballard, A unified model of early word learning: Integrating statistical and social cues, Neurocomputing 70 (13-15) (2007) 2149–2165.

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

[29] M. C. Frank, D. Ichinco, R. Saxe, Cross-situational word learning respects mutual exclusivity, in: CogSci 31, 2009.

[30] N. Goodman, J. B. Tenenbaum, M. J. Black, A bayesian framework for cross-situational word-learning, in: Advances in Neural Information Processing Systems, 2008, pp. 457–464.

[31] M. H. Davis, Connectionist modelling of lexical segmentation and vocabulary acquisition, in: Connectionist models of development: Developmental processes in real and artificial neural networks, 2003.

[32] P. Li, I. Farkas, B. MacWhinney, Early lexical development in a self-organizing neural network, Neural Networks 17 (8-9) (2004) 1345–1362.

[33] P. Li, X. Zhao, B. Mac Whinney, Dynamic self-organization and early lexical development in children, Cognitive Science 31 (4) (2007) 581–612.

[34] K. Plunkett, C. Sinha, M. F. Møller, O. Strandsby, Symbol grounding or the emergence of symbols? vocabulary growth in children and a connectionist net, Connection Science 4 (3-4) (1992) 293–312.

[35] T. Regier, The emergence of words: Attentional learning in form and meaning, Cognitive Science 29 (6) (2005) 819–865.

[36] A. Lazaridou, G. Chrupała, R. Fernández, M. Baroni, Multimodal semantic learning from child-directed input, in: NAACL-HLT 15, 2016.

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

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

[39] A. Alishahi, S. Stevenson, A probabilistic model of early argument structure acquisition, in: CogSci 27, 2005.

[40] A. Alishahi, S. Stevenson, A computational model of early argument structure acquisition, Cognitive Science 32 (5) (2008) 789–834.

[41] B. Beekhuizen, A. Fazly, S. Stevenson, Learning meaning without primitives: Typology predicts developmental patterns, in: CogSci 36, 2014.

[42] B. Beekhuizen, S. Stevenson, Modeling developmental and linguistic relativity effects in color term acquisition, in: CogSci 38, 2016.

[43] Z. Wu, M. Palmer, Verbs semantics and lexical selection, in: ACL 32, Association for Computational Linguistics, 1994.

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

[45] S. Bird, E. Loper, Nltk: the natural language toolkit, in: ACL 42, Association for Computational Linguistics, 2004.

[46] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: ICLR 2015, 2015.

[47] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: CVPR 2009, Ieee, 2009.

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

[49] R. M. Nosofsky, Attention, similarity, and the identification–categorization relationship, Journal of Experimental Psychology: General 115 (1) (1986) 39–57.

[50] L. A. Rescorla, Category development in early language, Journal of Child Language 8 (2) (1981) 225–238.

[51] L. A. Rescorla, Concept formation in word learning, Ph.D. thesis, Yale University (1976).

[52] J. M. Gruendel, Referential extension in early language development, Child Development (1977) 1567–1576.

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

[54] L. Fenson, V. Marchman, D. Thal, P. Dale, J. Reznick, E. Bates, MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual, 2nd Edition, Baltimore, MD: Brookes, 2006.

[55] B. MacWhinney, The CHILDES project: Tools for analyzing talk, Volume II: The database, Psychology Press, 2014.

[56] E. Bates, I. Bretherton, L. S. Snyder, From first words to grammar: Individual differences and dissociable mechanisms, Cambridge, MA: Cambridge University Press, 1988.

[57] N. Bernstein-Ratner, Dissociations between vowel durations and formant frequency characteristics, Journal of Speech, Language, and Hearing Research 28 (2) (1985) 255–264.

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

[59] S. R. Braunwald, Mother-child communication: the function of maternal-language input, Word 27 (1-3) (1971) 28–50.

[60] M. R. Brent, J. M. Siskind, The role of exposure to isolated words in early vocabulary development, Cognition 81 (2) (2001) B33–B44.

[61] R. Brown, A first language: The early stages, Harvard U. Press, 1973.

[62] D. P. Hayes, M. G. Ahrens, Vocabulary simplification for children: A special case of 'motherese'?, Journal of Child Language 15 (2) (1988) 395–410.

[63] M. J.-A. Demetras, Working parents' conversational responses to their two-year-old sons, Ph.D. thesis, The University of Arizona (1986).

[64] A. Feldman, L. Menn, Up close and personal: A case study of the development of three english fillers, Journal of Child Language 30 (4) (2003) 735–768.

[65] E. F. Masur, J. B. Gleason, Parent-child interaction and the acquisition of lexical information during play, Developmental Psychology 16 (5) (1980) 404–409.

[66] R. P. Higginson, Fixing: Assimilation in language acquisition, Ph.D. thesis, Washington State University (1985).

[67] S. A. Kuczaj II, The acquisition of regular and irregular past tense forms, Journal of Verbal Learning and Verbal Behavior 16 (5) (1977) 589–600.

[68] L. McCune, A normative study of representational play in the transition to language, Developmental Psychology 31 (2) (1995) 198–206.

[69] J. McMillan, Corpus collected and transcribed by Julie McMillan, `https://childes.talkbank.org/access/Eng-NA/McMillan.html`, accessed: 2019-06-26 (2004).

[70] C. E. Morisset, K. E. Barnard, C. L. Booth, Toddlers' language development: Sex differences within social risk, Developmental Psychology 31 (5) (1995) 851–865.

[71] A. Ninio, C. E. Snow, B. A. Pan, P. R. Rollins, Classifying communicative acts in children's interactions, Journal of Communication Disorders 27 (2) (1994) 157–187.

[72] R. S. Newman, M. L. Rowe, N. B. Ratner, Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development, Journal of Child Language 43 (5) (2016) 1158–1173.

[73] M. J. Demetras, K. N. Post, C. E. Snow, Feedback to first language learners: The role of repetitions and clarification questions, Journal of Child Language 13 (2) (1986) 275–292.

[74] P. R. Rollins, Caregiver contingent comments and subsequent vocabulary comprehension, Applied Psycholinguistics 24 (2003) 221–234.

[75] J. Sachs, Talking about the there and then: The emergence of displaced reference in parent-child discourse, in: K. E. Nelson (Ed.), Children's Language, Volume 4, 1983.

[76] P. Suppes, The semantics of children's language, American Psychologist 29 (2) (1974) 103–114.

[77] V. Valian, Syntactic subjects in the early speech of American and Italian children, Cognition 40 (1991) 21–81.

[78] A. Leubecker-Warren, J. N. Bohannon III, Intonation patterns in child-directed speech: Mother-father differences., Child Development 55 (4) (1984) 1379–1385.

[79] L. McDonough, Basic-level nouns: first learned but misunderstood, Journal of Child Language 29 (2) (2002) 357–377.

[80] F. Xu, J. B. Tenenbaum, Word learning as bayesian inference, Psychological Review 114 (2) (2007) 245–272.

[81] O. Abend, T. Kwiatkowski, N. J. Smith, S. Goldwater, M. Steedman, Bootstrapping language acquisition, Cognition 164 (2017) 116–143.

[82] S. Niyogi, Bayesian learning at the syntax-semantics interface, in: CogSci 24, 2002.

[83] M. Donaldson, G. Balfour, Less is more: A study of language comprehension in children, British Journal of Psychology 59 (4) (1968) 461–471.

[84] E. V. Clark, On the acquisition of the meaning of before and after, Journal of Verbal Learning and Verbal Behavior 10 (3) (1971) 266–275.

[85] E. V. Clark, On the child's acquisition of antonyms in two semantic fields, Journal of Verbal Learning and Verbal Behavior 11 (6) (1972) 750–758.

[86] C. Chomsky, The acquisition of syntax in children from 5 to 10, MIT Press, 1969.

[87] J. Piaget, Judgment and reasoning in the child, London: Kegan Paul, Trench, Trubner, 1928.