A Computational Theory of Children's Overextension

by

Renato Ferreira Pinto Junior

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Computer Science
University of Toronto

© Copyright 2019 by Renato Ferreira Pinto Junior

Abstract

A Computational Theory of Children's Overextension

Renato Ferreira Pinto Junior

Master of Science

Graduate Department of Computer Science

University of Toronto

2019

Young children extend known words to refer to novel objects, a phenomenon commonly known as overextension. This ability evidences linguistic creativity in early childhood and depends on a critical process of complexive thinking (Vygotsky, 1962). Psychological research has indicated that overextension relies on diverse semantic relations that link existing and overextended referents of a word, drawing on taxonomic, analogical, and predicate-based knowledge. Previous work has also observed asymmetrical behaviour in overextension between production and comprehension. However, no existing research has provided a unified framework to account for the overextension strategies across a comprehensive array of words and the asymmetries in overextension behaviour. I propose a computational theory that explains children's overextension under minimal parameterization. My framework constructs overextension behaviour with a multimodal fusion of knowledge derived from lexical semantics, deep neural networks, and psychological experiments. The framework captures the asymmetries between production and comprehension with an effort-based prior and suggests how children's word usage might converge to conventional usage through development. I tested my approach against a novel meta dataset curated from the developmental literature that includes 236 reported cases of children's overextension. My model reproduced overextended word-refrent pairs with 84% accuracy, captured 55% of children's word choices in the top-5 predictions, and replicated the empirical patterns of production and comprehension in overextension. This

work provides a formal approach to characterizing linguistic creativity of word meaning extension in early childhood.

Contents

Chapter 1

The problem of overextension

Young children creatively stretch known words to refer to novel objects, a phenomenon known as overextension (Clark, 1978). For example, children might say *dog* to refer to a squirrel, *ball* to refer to a balloon, or *key* to refer to a door. This intriguing phenomenon, which mostly takes places between 1 and 2.5 years in development (Clark, 1973), showcases children's early capacity for creative use of language under cognitive and communicative pressures (Bloom, 1973), as illustrated in Figure 1.1. 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. In this thesis, I explore the origin of sense extension by asking how the cognitive capacity for overextension in childhood can be characterized formally.

Children's overextended word choices rely on their nascent semantic representations of concepts, and how those relate to adult word meanings (Clark, 1973). Critical to the present work, Rescorla (1980) showed that children draw from diverse semantic relations in their lexical innovations. In her diary study of six children, 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* refer-

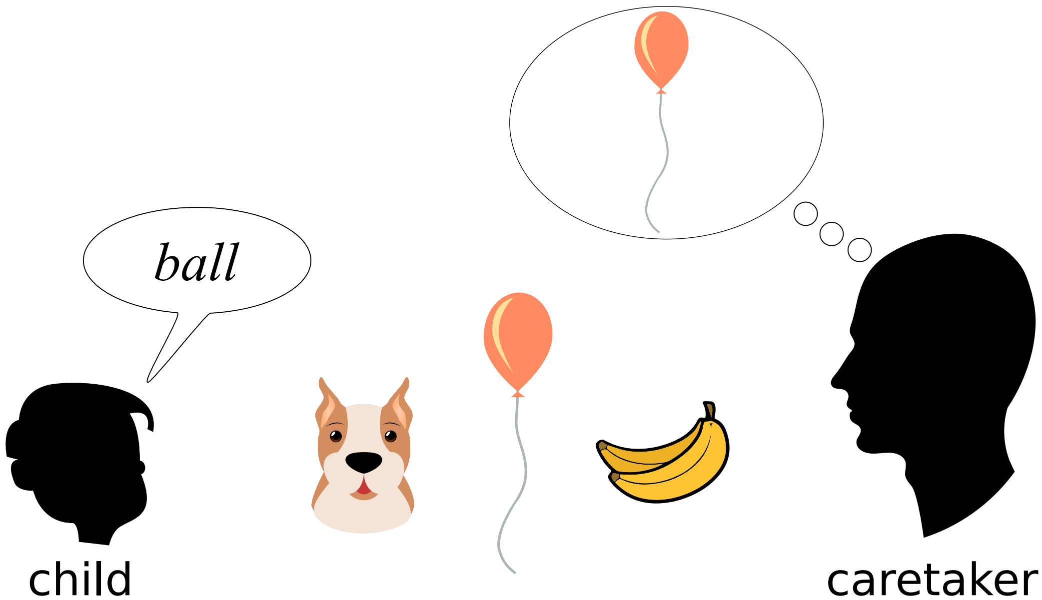

Figure 1.1: Illustration of overextension in child-caretaker communication.

ring to a squirrel); 2) *analogical* relation: overextension by linking objects with shared 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). An open question this work addresses is how to combine these types of relations to predict overextension strategies in early childhood. The framework I propose grounds word choices in the construction of a multimodal semantic space that underlies a probabilistic model of overextension.

Another puzzling aspect of overextension is its asymmetric patterns in children's linguistic production and comprehension. In addition to the vast literature documenting children's overextensions in production (e.g., Barrett, 1978; Clark, 1973; Rescorla, 1980), classic studies have shown that children's comprehension of words can also similarly overextend to other referents (Chapman & Thomson, 1980; Mervis & Canada, 1983; Thomson & Chapman, 1977). Although the extent of overextension in comprehension compared to production under different experimental or naturalistic conditions has been the subject of extensive debate (Chapman & Thomson, 1980; Fremgen & Fay, 1980; Mervis & Canada, 1983), two observations have been widely reproduced (Behrend, 1988;

Naigles & Gelman, 1995) and serve to ground most of this debate: 1) some overextension does occur in comprehension; and 2) children often overextend words in production even when they correctly comprehend the more appropriate adult word. This apparent asymmetry follows the common trend of comprehension leading production in linguistic development (Clark & Hecht, 1983), and psychologists have suggested that cognitive effort in word retrieval might help explain this disparity (Fremgen & Fay, 1980; Gershkoff-Stowe, 2001; Huttenlocher, 1974; Thomson & Chapman, 1977). My framework formally addresses this proposal by incorporating a measure of effort based on word frequency into a probabilistic model of overextension in production and comprehension.

A third question relates to what cognitive representations and processes support the convergence of children's language from early overextension patterns to conventional adult word meanings. Clark (1973) suggested a semantic narrowing hypothesis, which proposes that as children acquire semantic features restricting the meaning of a word, overextended referents are progressively excluded from the extension of that word; for example, *dog* could be overextended to a squirrel under the sole feature "has four legs", but the addition of feature "barks" would prevent this use of the word *dog*. However, an open question is how to integrate semantic narrowing with the observations that children rely on diverse semantic relations in their overextended word choices (Rescorla, 1980), or that the range of children's overextensions of a word in controlled experiments seems to defy an explanation based on a stable conjunction of features of a word (Thomson & Chapman, 1977). The model I propose encodes a developmental component that supports semantic narrowing over a multimodal space of semantic relations, and I show how this model can offer predictions ranging from early overextended speech to conventional adult word meanings.

Hence, in this work I investigate a formal computational theory of children's overextension, and offer computational answers to two fundamental topics of investigation in child development: what semantic representations and processes underlie children's overextension, and how can a unified model explain not only the occurrence of overextension in both production and comprehension, but also the asymmetries between the two? I show that this framework offers substantial predictability of children's overextended word choices from an extensive dataset collected from the developmental literature, and replicates patterns of production and comprehension observed in empirical studies of children's overextension. Furthermore, I show that this framework accounts for some of the developmental trends in the convergence from early overextended language to adult speech.

Another contribution of this work is an extensive dataset of noun pairs in overextension collected from a survey of developmental studies to evaluate my models. This dataset is described in more detail in Section 5.1 and is publicly available[1] for the research community.

This thesis is organized as follows. Chapter 2 surveys the relevant literature in developmental psychology on the problem of overextension, summarizing the main theories and open questions in the field. Chapter 3 reviews computational approaches to word learning, identifying how the present work relates to classic works in the literature, and it contrasts with previous research directions. Chapter 4 introduces the computational formulation of the proposed theory, specifying the probabilistic framework for children's overextension and the construction of the multimodal semantic space underlying that framework. Chapter 5 presents the data used in this study, including the novel dataset of children's overextension, as well as vocabulary and word frequency data. Chapter 6 presents the experimental results of model evaluations against the dataset of children's overextension and empirical data of production and comprehension from past experiments, and shows predictions of developmental trends in children's word choices. Chapter 7 discusses the findings of this work, summarizing the implications for research in developmental psychology and computational models of language acquisition, and sug-

[1] See Appendix D. I would especially like to thank Leslie Rescorla for providing a digitalized version of her dissertation work, from which I sourced precious data for this study.

gests directions for future research and extensions to historical language change and cross-linguistic variation.

Chapter 2

Psychological literature on overextension

Children's overextensions are intimately related to the notion of *complexive thinking*, which Vygotsky (1962) characterized as a crucial stage of early concept formation. In a 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 named this kind of structure *chain complex*, defined as a sequence of overextensions in which no single attribute characterizes the whole chain, but each link can be explained by some semantic relation. Vygotsky's account resonates with works in philosophy and cognitive linguistics which suggest that the complex structure of word meanings (e.g., Wittgenstein, 1953) emerge from a chaining process (Lakoff, 1987), where each referent is linked to the next in a chain-like formation.

Bowerman (1980) and Rescorla (1980) provided further evidence of complexive thinking in children's language with empirical data favouring the claim that most of children's overextensions form what Vygotsky called *associative complexes*, whereby a central prototype can be identified (e.g., a ball), and different semantic relations connect the central element to referential targets (e.g., apples, balloons, and beads). My framework draws from the notion of complexive thinking by relating a prototypical referent of a word to its overextended referents via multimodal semantic relations.

Psychologists have also investigated what cognitive mechanisms might underlie children's overextension in production and comprehension. Proposed theories can be classified into three main hypotheses: incomplete semantic system, pragmatic choice under limited vocabulary, and retrieval error.

The first hypothesis (Clark, 1973; Kay & Anglin, 1982; Mervis, 1987) posits that overextensions are caused by incomplete or loose semantic representations of word meanings. That is, children have incorrect definitions of words, and hence do not properly distinguish between regular and overextended word uses. While this theory helps explain the semantic leaps that children make in overextension, it is not sufficient to explain divergences between production and comprehension, because it predicts that once a word is correctly comprehended, it should be produced in conventional adult extension as well. However, this is not always the case, since correct comprehension coupled with overextended production of words has been widely documented in the literature (Fremgen & Fay, 1980; Rescorla, 1981; Thomson & Chapman, 1977).

A second proposal suggests that overextended word choices are pragmatic strategies to enable communication beyond vocabulary limitations (Bloom, 1973). In other words, children stretch word meanings to refer to objects whose adult words they do not know. This theory, while consistent with the observation that children often overextend known words to objects whose adult words they will acquire later (Rescorla, 1981), does not explain how children often overextend words to referents whose adult words they understand correctly in comprehension.

A third interpretation of overextension is the retrieval error hypothesis (Fremgen & Fay, 1980; Gershkoff-Stowe, 2001; Huttenlocher, 1974; Thomson & Chapman, 1977), which suggests that overextension in production can be the result not only of a missing vocabulary entry, but also of difficulty in retrieving the correct word at production time, causing a more accessible word to be retrieved and overextended instead. This hypothesis complements the pragmatic interpretation with an account of how words that are correctly comprehended can give way to overextension in production.

I suggest that a combination of these theories can best explain the diversity of phenomena in children's overextension. In particular, my computational framework incorporates multimodal semantic relations modulated by a developmental sensitivity parameter, thus expressing the semantic properties of overextension; a vocabulary that constrains possible word choices in the face of potentially unbounded communicative needs; and a cognitive effort-saving word prior that captures retrieval difficulty in a probabilistic setting, and enables the model to reproduce asymmetries between overextension in production and comprehension.

It is important to highlight that overextension is a general phenomenon that extends beyond just nouns (Clark, 1973). Psychologists have observed that children also systematically overextend a variety of other classes of words, for example: antonym pairs related to quantity (*less/more*; Donaldson & Balfour, 1968) and time (*before/after*; Clark, 1971); dimensional terms such as *big* for more specialized properties including *tall* and *high* (Clark, 1972); verbs such as *ask* and *tell* (Chomsky, 1969); kinship terms such as *brother* and *sister* (Piaget, 1928); spatial terms, with one general purpose term standing in for a variety of spatial relations (Clark, 1978), among others.

In this work, I focus on overextension of nouns, but some of the computational principles presented here, such as the probabilistic formulation integrating semantic relations and cognitive limitations, may still apply to overextension in some of these other domains. I leave it for future work to explore how to broadly define semantic relations applicable to other classes of words.

Chapter 3

Computational approaches to word learning

The computational approach to chilren's overextension presented here relates to the broader body of works in computational models of word learning and language acquisition. One important framework in this field is cross-situational word learning, which posits that children infer the meanings of words by exploiting the statistical co-occurrences of words in utterances and meanings in different situations (Fisher, Hall, Rakowitz, & Gleitman, 1994; Gleitman, 1990; Pinker, 1984; Siskind, 1996). Cross-situational word learning models have been developed under different methodological frameworks, such as symbolic (Siskind, 1996), connectionist (e.g., Davis, 2003; Li, Farkas, & MacWhinney, 2004; Li, Zhao, & Mac Whinney, 2007; Plunkett, Sinha, Møller, & Strandsby, 1992; Regier, 2005), associative probabilistic (e.g., Fazly, Alishahi, & Stevenson, 2010; Kachergis, Yu, & Shiffrin, 2017; Yu & Ballard, 2007), and Bayesian (e.g., Frank, Goodman, & Tenenbaum, 2009; Frank, Ichinco, & Saxe, 2009; Goodman, Tenenbaum, & Black, 2008). In contrast to this rich area of research, which focuses on modelling children's behaviour in learning conventional word meanings, my work models children's linguistic innovations under communicative and cognitive pressures.

My computational formulation of a multimodal semantic space that underlies overextended word choices connects to previous work in visually-grounded word learning (Lazaridou, Chrupała, Fernández, & Baroni, 2016; Roy & Pentland, 2002; Yu, 2005). This research direction attempts to use visual features in the environment to model word learning as a process grounded in perception. My framework contrasts with these works by employing visual features to enable analogical use of language in overextension.

Although computational perspectives on children's linguistic innovations are still in their infancy, some computational works have begun exploring facets of this problem. Alishahi and Stevenson (2005, 2008) provided a probabilistic model of early argument structure acquisition that displayed a transient period of overgeneralized verb argument structure (e.g., *Mary fall toy*) followed by correct production. In another research direction, Beekhuizen, Fazly, and Stevenson (2014) and Beekhuizen and Stevenson (2016) studied the relationship between cross-linguistic variation in lexicalization and children's overextension of spatial prepositions an colour terms, and suggested that both word frequency and universal cognitive biases reflected in cross-linguistic data play a role children's overextension patterns. Contrasting with these works, my approach provides an explicit construction of the semantic relations that may underlie children's conceptual leaps in overextension, and shows how these relations can be combined into a predictive model of overextension. Another novel aspect of this work is the exploration of the relationship between production and comprehension, and a formal proposal of how cognitive effort can explain the asymmetry between the two. Finally, the flexible way in which my proposed models combine large-scale data sources into a parsimonious model enables a breadth of evaluations that goes beyond previous works; I evaluate my models against an extensive dataset of overextension over English nouns, shedding light into the predictive power of the models, the contribution of each component of the framework, and the limitations to be addressed in future work.

Chapter 4

Computational formulation of theory

I model overextension as a communicative process in which a child, in production, wishes to refer to a novel object given vocabulary and cognitive limitations (see Figure 1.1), and, in the opposite comprehension scenario, needs to infer the intended meaning of an utterance given potential referents in the environment. The framework has two main components: a generic probabilistic formulation of overextension in production and comprehension, and the specification of a multimodal semantic space that underlies this probabilistic process.

4.1 Probabilistic framework

4.1.1 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 or may not be in the child's vocabulary. Given a candidate word $w \in V$ for production (e.g., *ball*), I specify the following probabilistic model:

$$p_{\text{prod}}(w|c) = \frac{p(c|w)p(w)}{\sum_{w' \in V} p(c|w')p(w')} \tag{4.1}$$

In this formulation, the likelihood $p(c|w)$ is the appropriateness of categorizing concept c under word w, and is defined as a density function (specified in the next section) that depends on the semantic similarity between c and c_w (the concept corresponding to word w):

$$p(c|w) = f_{\text{sim}}(c|c_w) \tag{4.2}$$

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 (Beekhuizen & Stevenson, 2016), I define $p(w)$ as a frequency-based word prior:

$$p(w) = \frac{F(w)}{\sum_{w' \in V} F(w')} \tag{4.3}$$

where $F(w)$ is the total frequency of word w in a representative corpus children's linguistic environment.

4.1.2 Comprehension

In the case of comprehension, the child hears word w and must estimate the probability 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:

$$p_{\text{comp}}(c|w) = \frac{f_{\text{sim}}(c|c_w)}{\sum_{c' \in E} f_{\text{sim}}(c'|c_w)} \tag{4.4}$$

where E is the set of possible referents in the child's environment.

4.2 Multimodal semantic space

I define a multimodal semantic space that captures the three types of relational features in Rescorla (1980): categorical relation, visual analogy,[1] and predicate-based relation. I construct these relational features using a fusion of resources drawn from linguistics, deep learning networks, and psychological experiments, as illustrated in Figure 4.1.

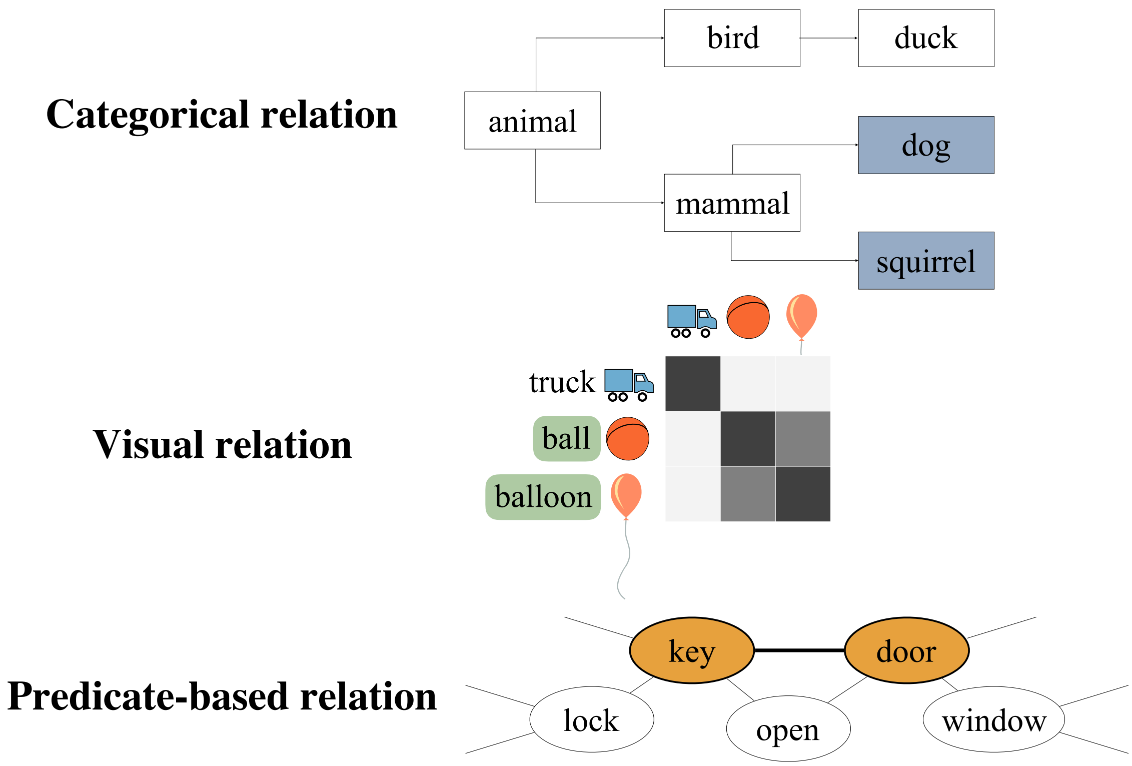

Figure 4.1: Types of semantic relations in multimodal space.

4.2.1 Categorical relation

I define the 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 hierarchy), the distance is:

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

$$d_c(c_1, c_2) = 1 - \frac{2N_{\text{LCS}}}{N_1 + N_2} \tag{4.5}$$

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, I took the taxonomy from WordNet (Miller, 1995) and annotated words by their corresponding *synsets* in the database. I used the *NLTK* package (Bird & Loper, 2004) to calculate similarities between referents for this feature.

4.2.2 Visual analogical relation

I define the visual analogical relation by cosine distance between vector representations of referents in visual embedding space. In particular, I 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, I 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, I extracted the corresponding visual feature vector from the first fully-connected layer after all convolutions: v_j^c. I then averaged the sampled k feature vectors to obtain an expected vector v^c for the visual vector representation of c.

4.2.3 Predicate-based relation

I define the predicate-based relation by leveraging the psychological measure of word association. I assume that two referents that frequently co-occur together should also be highly associable, e.g., in the case of key and door. Specifically, I 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 procedures 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. I 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. I used the implementation provided by the authors[2] to compute vector representations from the association probability matrix. I used cosine distance to compute predicate-based distances between pairs of referent vectors.

4.2.4 Multimodal space of relations

To complete the model formulation, I integrate the three types of semantic relations specified above into a density function based on concept similarity that measures the likelihood of concepts being associated in overextension in the probabilistic framework.

I take the Gaussian-Euclidean form of the generalized context model (GCM) or exemplar model of categorization (Nosofsky, 1986), 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

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

norm over the distance components in each psychological dimension:

$$d(c_1, c_2) = \left[d_c(c_1, c_2)^2 + d_v(c_1, c_2)^2 + d_p(c_1, c_2)^2\right]^{1/2} \tag{4.6}$$

In 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 Guassian kernel computes concept similarity as a decaying function of psychological distance:

$$\text{sim}(c_1, c_2) = \exp\left(-\frac{d(c_1, c_2)^2}{h}\right) \tag{4.7}$$

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. I empirically estimate the value of h from data in the 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:

$$f_{\text{sim}}(c|c_w) = \frac{\text{sim}(c, c_w)}{Z_h} \tag{4.8}$$

where Z_h depends only on h (concretely, $Z_h = \int \exp\left(-\frac{x^2}{h}\right) \mathrm{d}x$), and thus need not be explicitly computed in the models.

4.2.5 Orthogonality and assumptions

To ensure that the three types of relational features provide complementary information, I calculated their inter-correlation based on the 236 concept pairs that I used for my analyses. Although correlations were significant ($p < .001$), coefficients were low or moderate (Spearman's ρ; category & visual: 0.238; category & predicate: 0.445; visual

& predicate: 0.421), suggesting that each feature contributes to information encoded in the multimodal semantic space. I further explore the predictive contribution of each multimodal feature to overextension in Chapter 6.

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

Chapter 5

Data

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

5.1 Metadata of child overextension

I performed a meta survey of 12 representative studies from developmental psychology and collected a total of 323 overextension example word-referent pairs.[1] Each pair consists of an overextended word and the novel referent that word has been extended to. I kept word-referent pairs of nouns that overlapped with the available data from the three features I described, resulting in a total of 236 word-referent pairs. Table 5.1 shows examples from this meta dataset and their sources from the literature; see Appendix D for the complete dataset.

While the data I used for analysis may not constitute an unbiased sample of child overextension, two factors help to alleviate this concern. First, I 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 I

[1] I thank Yu B Xia for contributing to the preliminary collection of child overextension data.

Table 5.1: Examples of overextension data, one for each source included in this study.

Uttered word	→	Referent	Source
ball	→	onion	Thomson and Chapman (1977)
car	→	truck	Fremgen and Fay (1980)
apple	→	orange juice	Rescorla (1981)
ball	→	marble	Barrett (1978)
fly	→	toad	Clark (1973)
cow	→	horse	Gruendel (1977)
apple	→	egg	Rescorla (1976)
truck	→	bus	Rescorla (1980)

examined reduces the possibility of biasing the sample from any individual study.

5.2 Vocabulary from early childhood

To approximate children's vocabulary in early childhood, I 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). This database is based on the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2006) 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 (Clark, 1973) (that covers the range in Wordbank), I constructed a vocabulary V using all the nouns from Wordbank for which required semantic features could be obtained. The resulting vocabulary includes 317 out of the 322 nouns from the database (see Appendix E for a complete list).

5.3 Word frequencies in child-caretaker speech

To approximate the distribution of nouns in young children's environments, I collected a large set of child-caretaker speech transcripts from the CHILDES database (MacWhinney, 2014). Concretely, I collected all transcripts from studies performed in naturalistic child-caretaker situations for children between 1 and 2.5 years (the typical overextension period), resulting in 1713 transcripts with over 200K noun tokens in total.[2]

I 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*), I 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.

[2]Specifically, I collected transcripts from the studies of Bates, Bretherton, and Snyder (1988); Bernstein-Ratner (1985); Bloom (1973); Bloom, Hood, and Lightbown (1974); Braunwald (1971); Brent and Siskind (2001); Brown (1973); M. J. Demetras, Post, and Snow (1986); M. J.-A. Demetras (1986); Feldman and Menn (2003); Hayes and Ahrens (1988); Higginson (1985); Kuczaj II (1977); Leubecker-Warren and Bohannon III (1984); MacWhinney (2014); Masur and Gleason (1980); McCune (1995); McMillan (2004); Morisset, Barnard, and Booth (1995); Newman, Rowe, and Ratner (2016); Ninio, Snow, Pan, and Rollins (1994); Rollins (2003); Sachs (1983); Suppes (1974); Valian (1991).

Chapter 6

Results

I evaluate the proposed computational framework in three aspects: 1) model accuracy in predicting children's word choices in overextension; 2) model reproduction of the production-comprehension asymmetry; and 3) illustration of developmental trends predicted by the model.

6.1 Predicting word choices in overextension

I evaluated the production model in Section 4.1 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. I assessed the model by finding the maximum *a posteriori* probability (MAP) of all the overextension pairs under the single sensitivity parameter h, which I optimized to the MAP objective function via standard stochastic gradient descent:

$$\max_h \prod_i p_{\text{prod}}(w_i|c_i; h, V) = \max_h \prod_i \frac{p(c_i|w_i; h)p(w_i)}{\sum_{w \in V} p(c_i|w; h)p(w)} \tag{6.1}$$

To assess the contribution of the three relational features, I 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. I 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.

I evaluated all models under two metrics: Bayesian Information Criterion and performance curves.

The Bayesian Information Criterion (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). The score is defined as $\text{BIC} = \log(n)k - 2\log(\hat{L})$, 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).

As a second assessment, I produced performance curves that measure model accuracy at different numbers of allowed model predictions m. Concretely, for each level of m, I measured the predictive accuracy of the model from its choice of top m words, or 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.

Table 6.1 summarizes the BIC scores of the family of production models. I 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 I considered. 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. Third, models with featural integration performed better than those with isolated features (i.e., all features $<$ feature pairs $<$ single features

Table 6.1: Bayesian Information Criterion (BIC) scores for production models with respect to overextension dataset ($N = 236$).

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

in BIC score), suggesting that children rely on multiple kinds of semantic relations in overextensional word choices.

Figure 6.1 further confirms these findings in performance curves that show the 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 6.1 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. In many of the cases of incorrect prediction, the model still predicts words that are closely related to the target referent, and since the overextension dataset cannot be taken as an exhaustive account of child speech, this performance measure should be seen as a lower bound on the ability of the model to reconstruct children's overextension strategies. Appendix B shows sample model outputs for both correct and incorrect predictions.

Figure 6.2 shows the estimated contribution of each semantic relation in the multimodal space toward characterizing the overextension dataset, obtained from a logistic regression analysis that achieved 84% accuracy in distinguishing true overextension word

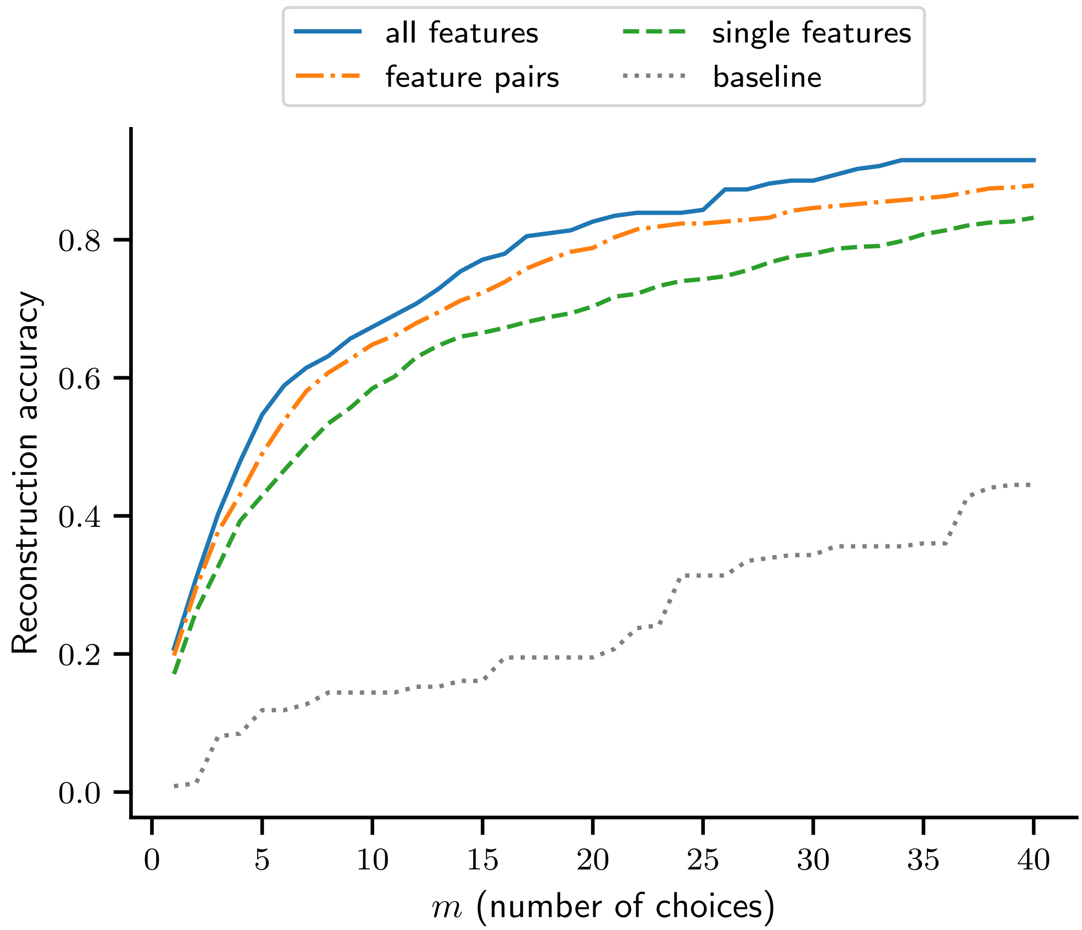

Figure 6.1: Performance curves for production models showing cross-validated model accuracies in reconstructing overextended word choices. Aggregated results (single features and feature pairs) show mean accuracy over individual models; see Appendix B for a fine-grained comparison of all models.

pairs from control pairs (see the complete analysis in Appendix A), 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.

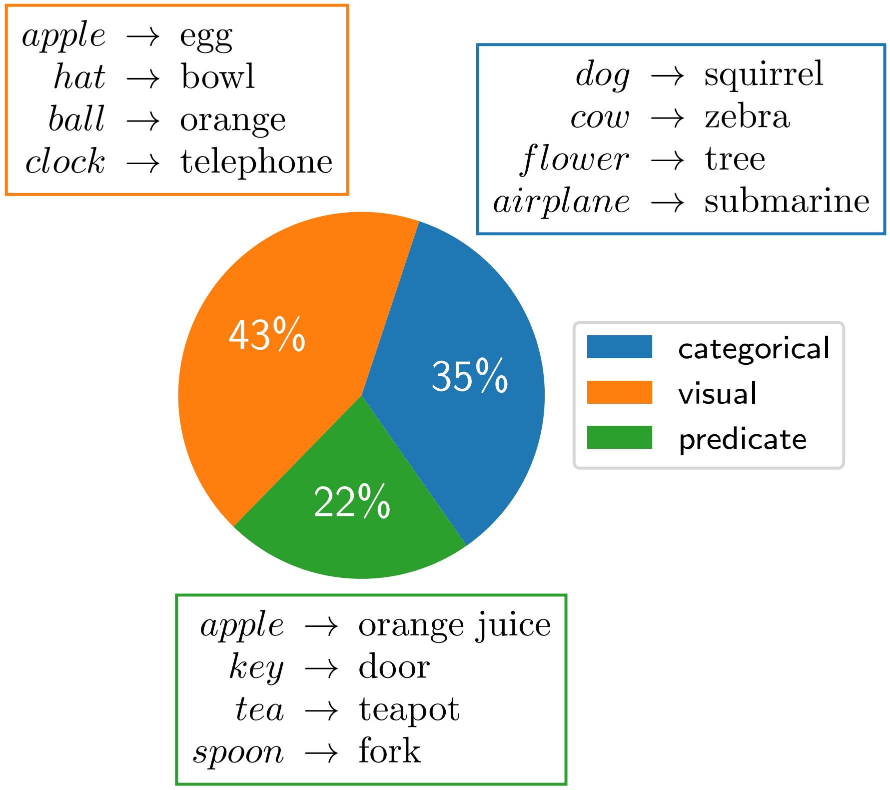

Figure 6.2: Percentage shares and examples explained by the three types of features from the overextension dataset.

6.2 Predicting production-comprehension asymmetry

To assess the ability of my model to capture the asymmetry between overextension in production and comprehension, I performed a computational replication of the experimental study conducted by McDonough (2002). That study tested 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

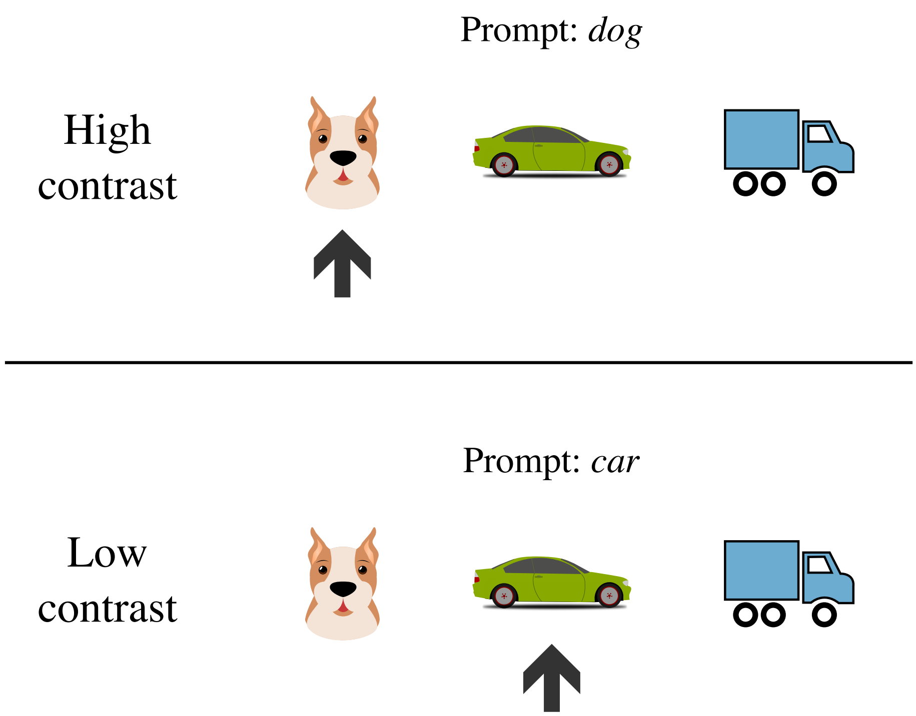

Figure 6.3: Two conditions in comprehension experiment devised by McDonough (2002).

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 had 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 6.3). Table 6.2 shows the stimuli triplets and conditions.

I replicated these experiments in my computational models as follows. For the production experiment, I presented the production model in Equation 4.1 (with fixed parameter value from the first experiment) with each stimulus c, and measured the probability of correct (target word) production, $p_{\text{prod}}(w|c)$, versus all other words in the child vocabulary V. For the comprehension experiment, I presented the comprehension model in Equation 4.4 (with same parameter setting) with each target word w, and computed the probability of the target referent versus the two distractors in the triplet; i.e., if the triplet

Table 6.2: Experimental stimuli from McDonough (2002). 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

of stimuli were (c_1, c_2, c_3) with c_2 as the target, then the probability of correct selection would be $p_{\text{comp}}(c_2|w)$ with referential environment $E = \{c_1, c_2, c_3\}$ in Equation 4.4.

First, I show the results of the comprehension task. Figure 6.4 shows the probability of correct and incorrect referent selections in low and high contrast conditions, for early and late nouns, from empirical results by McDonough (2002) and as predicted by the model. I observe that the model predictions reflect the qualitative trends from experimental data: while correct comprehension was the dominant response in every condition, 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 experimental results, it is worth highlighting that these predictions followed from training the model on production data only, and thus the qualitative match shows that the model is able to predict patterns of overextension in comprehension without seeing any such data before.

Figure 6.5 shows the probability of correct word production in response to early versus late item stimuli, from empirical results by McDonough (2002), predictions by the production model with frequency-based word prior, and predictions by a production model with a uniform word prior. I observe that the production model with frequency-based prior was able to replicate the experimental finding that correct labels were produced for the early items ($n = 16$, $M = 0.68$) more often than for late items ($n = 14$, $M = 0.30$), for which incorrect production was the dominant response, and the difference between the two groups was significant (Welch's $t(23) = 6.08$, $p < 0.001$). On the other hand, the model with uniform prior was unable to capture this difference ($M(\text{early}) = 0.56$, $M(\text{late}) = 0.55$, $t(27) = 0.36$, $p = 0.73$), evidencing the importance of the prior in capturing the patterns of children's word production.

Comparing the results in the two tasks, I make two observations. First, the semantic space and probabilistic formulation enable the model to make predictions in both production and comprehension, even without re-tuning its sensitivity parameter from the first task. This result suggests that the framework captures relevant features of young children's linguistic abilities. Second, the frequency-based word prior was essential in enabling the model to capture the asymmetry between the production and comprehension: while correct comprehension was dominant in every condition, with no difference

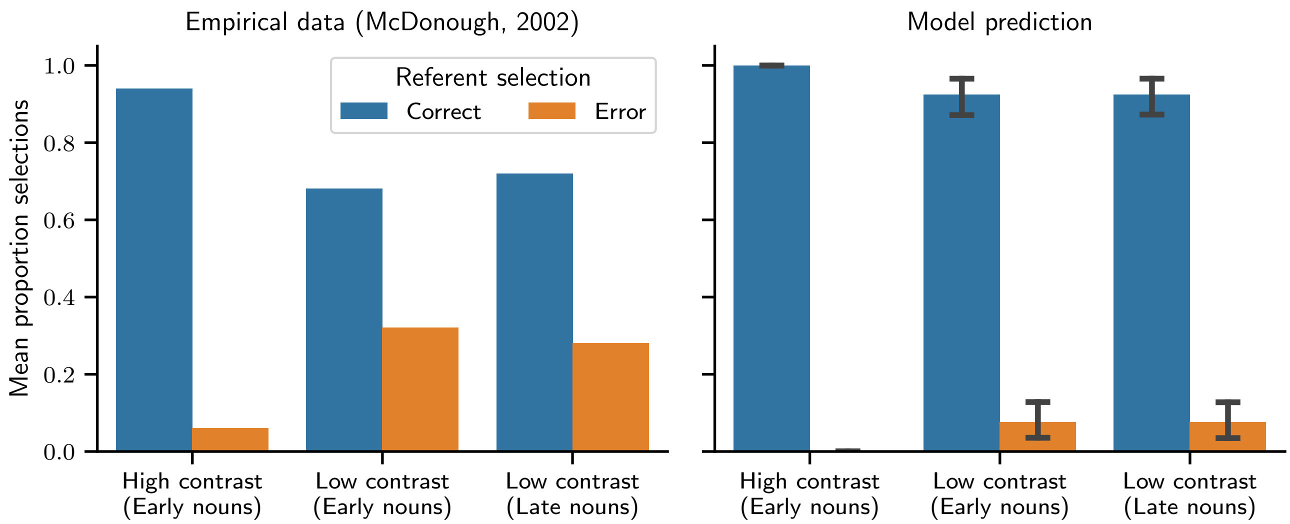

Figure 6.4: Results of comprehension experiments from McDonough (2002) and from model reproduction. Error bars represent bootstrap 95% confidence intervals.

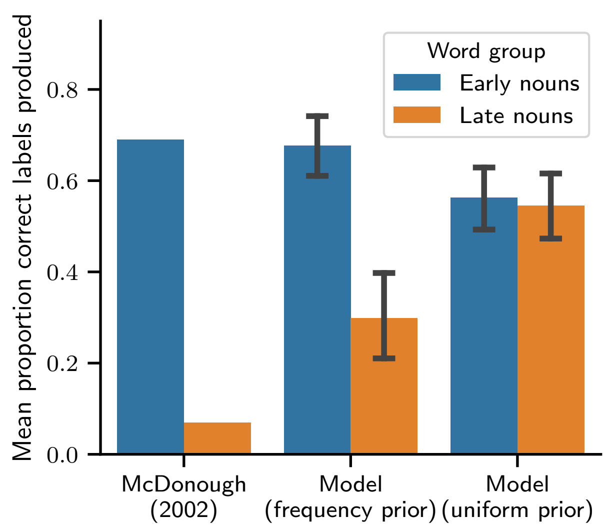

Figure 6.5: Results of production experiments from McDonough (2002) and from model reproduction. Error bars represent bootstrap 95% confidence intervals.

in performance between early and late items, performance in production was higher for early than for late items, for which incorrect production was the dominant response. Thus the model is able to capture an intriguing phenomenon in developmental psychology: that children often overextend words even when they seem to correctly understand the adult words in comprehension. As proposed by the retrieval error hypothesis, the model suggests that taking cognitive effort into account in production is the crucial step in explaining this asymmetry.

6.3 Predicting developmental trends in word learning

Children typically overextend words until they are about 2.5 years old; as their productive speech converges to adult patterns, categories undergo a restructuring process in which the semantic space is "split up" as children start producing the correct words for concepts that they previously named in overextension (Clark, 1973).

In this experiment, I explored the developmental trends captured by the model as a result of varying its sensitivity parameter. Concretely, given that, as children develop, they become more sensitive to the semantic appropriateness of labels to objects (i.e., they cease overextending words), I hypothesized that I could reproduce characteristics of the development of children's productive speech in the model by shrinking the kernel width parameter h. In Equations 4.1 and 4.7, shrinking the parameter h has the effect of making the similarity-based likelihood $p(c|w)$ steeper with respect to distance in the multimodal space; hence, given two candidate words w_1 and w_2 to refer to a concept c, with respective probabilities $p_{\text{prod}}(w_1|c) \propto p(c|w_1)p(w_1)$ and $p_{\text{prod}}(w_2|c) \propto p(c|w_2)p(w_2)$, the ratio between their frequency priors $p(w_1)/p(w_2)$ becomes less relevant than the ratio between their semantic appropriateness likelihoods $p(c|w_1)/p(c|w_2)$ for the model prediction $p_{\text{prod}}(w|c)$ as h shrinks. Thus, the parameter h modulates the trade-off between semantic appropriateness and cognitive effort in the model.

Figure 6.6a shows the predicted rate of overextension of the production model as its sensitivity parameter changes in a simulation of development. The rate of overextension is defined as the proportion of words in the child vocabulary applied in overextension when the model is exposed to all concepts in the vocabulary and in the overextension

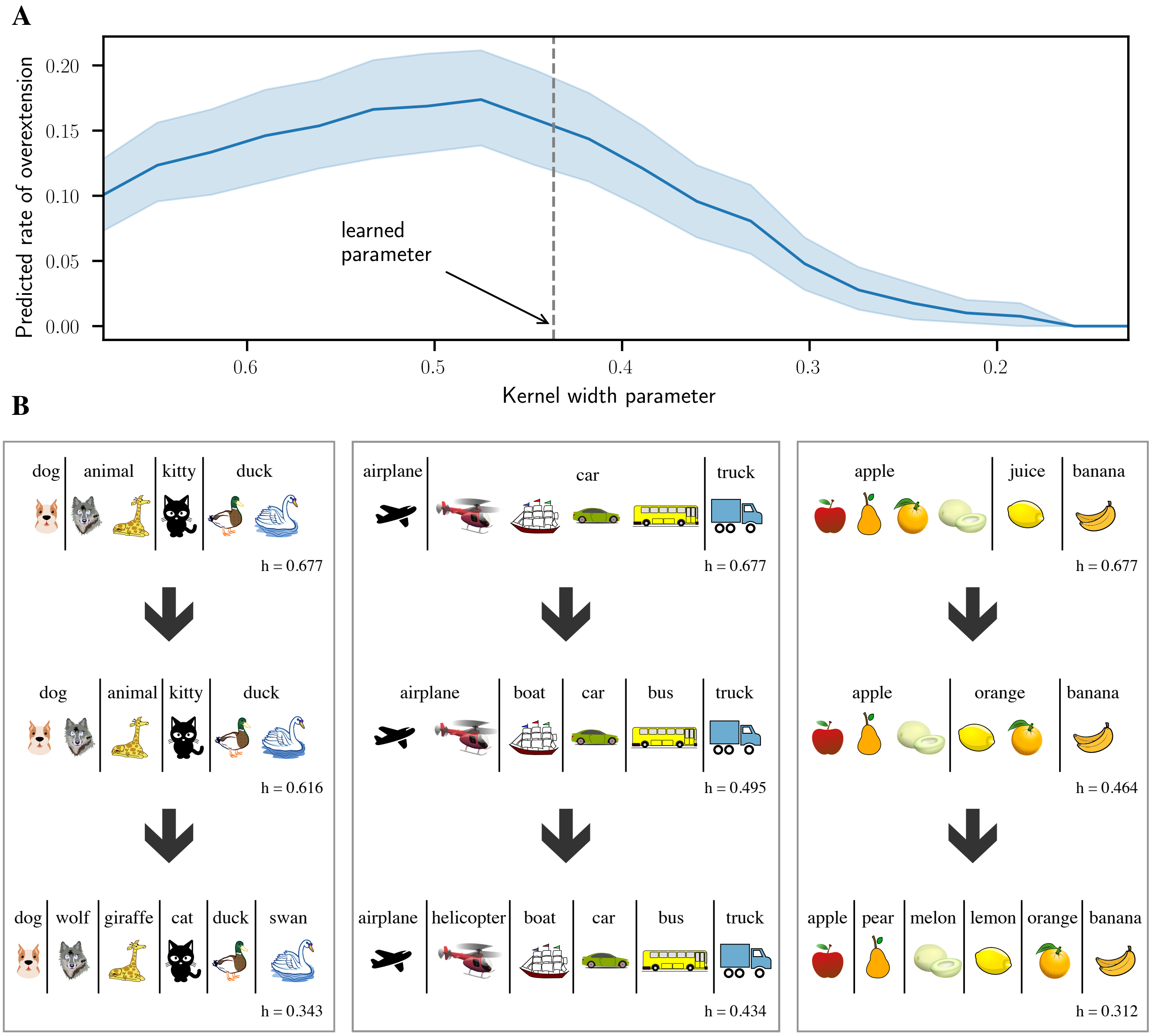

Figure 6.6: Results of developmental simulations. (A) Predicted rate of overextended words in the child vocabulary over development, from early overextension patterns to convergence to adult speech. Shaded region represents bootstrap 95% confidence interval. (B) Sample model predictions of children's object naming in three domains over development. The top row shows the earliest predictions, the middle row shows an intermediate stage, and the bottom row shows convergence to adult words. Each stage indicates the kernel width (h) value for which the model first produces that pattern.

dataset. I observe that the rate of overextension reaches a peak close to the parameter learned from the overextension data, consistent with the fact that that dataset represents the wide variety of overextension pairs reported in the literature. To the left of the peak (earlier in developmental time), there is a slight decrease in overextension rate, which can be explained as the model becoming less sensitive to semantic appropriateness, which causes cognitive effort to dominate production, i.e., earlier words are applied more broadly in overextension over later words. To the right of the peak (later in developmental time), the opposite story unfolds: as sensitivity to semantic appropriateness increases and surpasses the pressure of cognitive effort, overextension gives way to new words and becomes increasingly rare, until it completely ceases and the model productions converge to adult speech.

I also analyzed how individual category members were labelled throughout this developmental simulation. Inspired by the longitudinal study of Rescorla (1976), I took concepts from the domains of animals, fruits, and vehicles, and categorized them according to the words produced by the model to name those concepts at each point in simulated developmental time. Figure 6.6b shows the resulting categories. I observe several commonalities with observations by developmental accounts such as those by Rescorla (1981) and Clark (1973): 1) Initial period in which words such as *car* and *apple* are broadly overextended, serving as central elements for their respective domains; 2) Progressive narrowing of categories, with emerging words "subdividing" the space by correctly labelling referents, and possibly being themselves overextended to other closely related concepts (e.g. *orange* overextended to lemons after being acquired); and 3) Eventual convergence to adult naming patterns, with words finally being applied in correct extension.

This experiment also shows some limitations of my model in the developmental setting. First, the model does not have a mechanism for rejecting the production of a label altogether, which children routinely do by not naming every object in a scene. Adding this pragmatic component to my formulation might more closely replicate children's strategies. Second, the model also predicts the usage of superordinate terms such as *animal*, which is not common in this period of development (Rescorla, 1981). F. Xu and Tenenbaum (2007) showed that older children (ages 3 to 4) are sensitive to the taxonomic properties of words and can generalize novel subordinate, basic level, and superordinate terms to new referents. Hence, a joint account of these phenomena should explain how children transition from favouring the overextension of basic level nouns to the acquisition and appropriate application of superordinate terms.

Chapter 7

Discussion

In this thesis, I presented a formal computational account of children's overextension. I formulated the problem of overextension in production and comprehension in a probabilistic setting, and showed that a combination of multimodal semantic relations (namely categorical, visual analogical, and predicate-based relations), integrated in a minimally-parameterized categorization model, explains substantial variation in children's overextended word choices in a large dataset of word-referent pairs from the developmental literature. Furthermore, I showed how cognitive effort in word retrieval, encoded in my model as word frequency, can account for the asymmetries between overextension in production and comprehension, thus successfully reconstructing both processes in a unified computational framework. Finally, I showed how the model can account for some of the developmental trends of child language, from broad overextension to adult naming patterns, by the simple manipulation of a parameter that trades off sensitivity to semantic appropriateness and cognitive effort.

I view this work as a contribution to several research efforts. In developmental psychology, it aids in resolving the debate among the three main theories about the mechanisms behind overextension – incomplete semantic system, pragmatic choice under limited vocabulary, and retrieval error – by showing how a single formal model can unify the main elements of each theory into a cohesive framework. Namely, my model incorporates the looseness of children's semantic systems via a probabilistic categorization process that encodes multimodal semantic relations and is modulated by a varying sensitivity parameter; it expresses pragmatic choices motivated by a limited vocabulary by overextending in-vocabulary words to out-of-vocabulary referents under communicative need; and it incorporates cognitive effort, encoded as word frequency, into the model of word production, so that even words that are correctly understood can probabilistically give way to more easily retrievable words at production time.

The present work also adds to a body of computational works modelling broader linguistic phenomena in children. While previous works have shown substantial progress in modelling the conventional learning component of developmental phenomena, e.g., by learning word meanings from word-referent co-occurrence statistics in the environment (Fazly et al., 2010; Frank, Goodman, & Tenenbaum, 2009; Yu & Ballard, 2007), inferring the taxonomic structure of words from limited exposure to data (F. Xu & Tenenbaum, 2007), or integrating syntax and semantics to bootstrap more advanced stages of language learning (Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017; Niyogi, 2002), relatively little attention has been given to another crucial characteristic that distinguishes children from adults: their cognitive limitations, the mistakes that arise from them, and how children creatively innovate around their limitations to achieve their communicative goals. Among steps in this direction are the work of Alishahi and Stevenson (2005, 2008), which showed evidence of overgeneralization in a model of verb argument structure acquisition, and Beekhuizen et al. (2014) and Beekhuizen and Stevenson (2016), which investigated the extent to which cognitive biases evidenced by cross-linguistic data can recapitulate the overextension of spatial prepositions and colour terms. However, underexplored directions regard the specific semantic relations that may underlie overextension in general domains, and the task of predicting the diversity of overextension strategies that children employ to name a wide range of objects under communicative need. Hence, my work provides a step in this direction by showing how children's linguistic innovations help them communicate about the world despite their cognitive limitations, and my formal computational framework grounds this discussion in a concrete predictive setting.

I have tested my models of children's overextended word choices against a large dataset of word-referent pairs that I collected from the developmental literature, and by making this aggregated data source available to the research community, I hope to facilitate future investigations of young children's linguistic creativity.

Some limitations of the present work also provide clues for future work. Although my incorporation of cognitive effort into the model reconstructs a developing pattern of word retrievals as emerging from a fixed vocabulary, children's vocabularies certainly evolve in comprehension over development. Thus, a full developmental account should incorporate a mechanism that allows for vocabulary growth over time. One way to model this evolution would be to integrate word learning and overextension strategies into a unified model. Thus, future work should investigate how a single model can reproduce both word learning and overextension phenomena from sequences of naturalistic stimuli.

I conclude by suggesting two possible areas for extensions of the present work: historical patterns of word sense extension, and cross-linguistic phenomena of colexification and cross-linguistic influence.

First, as argued in the introduction, Vygotsky's (1962) account of children's early concept formation resonates with the view of historical meaning change as a chaining process whereby referents of a word grow over time in a chain-like structure (Lakoff, 1987). Indeed, recent work has shown that chaining predicts word sense extension in the history of English (Ramiro, Srinivasan, Malt, & Xu, 2018) and other languages (Y. Xu, Regier, & Malt, 2016). Furthermore, some of the same cognitive mechanisms may underlie the chaining processes in both instances of word sense extension (e.g., the extension of *mouse* from rodent to computer device may be evidence of visual analogy). Thus, future work could explore to what extent linguistic innovation in early childhood and over the span of generations of speakers are subject to similar cognitive pressures and mechanisms.

Second, cross-linguistic diversity offers clues about cognitive biases in human categorization in two ways. Colexification patterns may indicate that concepts that are jointly labelled by the same word in distinct language families are more closely connected in semantic space (François, 2008), thus raising the question of whether the same cognitive mechanisms that drive concepts together in languages also induce co-labelling via overextension in children. Indeed, this seems to be the case at least for the domains of spatial prepositions (Beekhuizen et al., 2014) and colour terms (Beekhuizen & Stevenson, 2016). Future work could explore the ways in which cross-linguistic variation and overextension agree, and where they depart from each other in broader domains. A second direction of exploration could investigate to what extent adult language learners' or bilinguals' naming patterns under cross-linguistic influence relate to children's overextended word choices. In contrast with children, adult speakers of a second language are influenced by their first language (Jarvis & Pavlenko, 2008), and have an adult conceptual system guiding and constraining their word choices, so that some of the flexible, multimodal innovations displayed by children (e.g., *ball* overextended to an apple) may not be displayed by those speakers. On the other hand, it is possible that some of the mechanisms that emerge from the pressures of communicative need and cognitive limitations in children, e.g., favouring a more easily-retrievable word in trade-off with semantic appropriateness, also play a role in the communicative strategies of speakers of a second language.

In conclusion, I have provided a formal computational theory of children's overextension that integrates major theories in developmental psychology, elucidates in computational terms how children can make use of multimodal semantic relations to achieve communication while constrained by cognitive limitations, and explains overextension patterns over a large dataset of examples collected from the developmental literature. I believe that this work paves the way toward a formal characterization of children's linguistic innovations, and provides clues about the cognitive principles that underlie word sense extension as a core ability of humans as constant language creators and innovators.

References

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

Alishahi, A., & Stevenson, S. (2005). A probabilistic model of early argument structure acquisition. In *CogSci 27.*

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

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

Bates, E., Bretherton, I., & Snyder, L. S. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* Cambridge, MA: Cambridge University Press.

Beekhuizen, B., Fazly, A., & Stevenson, S. (2014). Learning meaning without primitives: Typology predicts developmental patterns. In *CogSci 36.*

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

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

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

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.

Bowerman, M. (1980). The structure and origin of semantic categories in the language learning child. In *Symbol as sense.*

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

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

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

Chapman, R. S., & Thomson, J. (1980). 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), 575–578.

Chomsky, C. (1969). *The acquisition of syntax in children from 5 to 10.* MIT Press.

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

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

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.

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

Davis, M. H. (2003). Connectionist modelling of lexical segmentation and vocabulary acquisition. In *Connectionist models of development: Developmental processes in* *real and artificial neural networks.*

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.

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

Demetras, M. J.-A. (1986). *Working parents' conversational responses to their two-year-old sons* (Unpublished doctoral dissertation). The University of Arizona.

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.*

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

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

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

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., & Bates, E. (2006). *Macarthur-bates communicative development inventories: User's guide and technical manual, 2nd edition.* Baltimore, MD: Brookes.

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

François, A. (2008). Semantic maps and the typology of colexification. In *From polysemy to semantic change: Towards a typology of lexical semantic associations.*

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.

Frank, M. C., Ichinco, D., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. In *CogSci 31.*

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.

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

Goodman, N., Tenenbaum, J. B., & Black, M. J. (2008). A bayesian framework for cross-situational word-learning. In *Advances in neural information processing systems* (pp. 457–464).

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

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

Higginson, R. P. (1985). *Fixing: Assimilation in language acquisition* (Unpublished doctoral dissertation). Washington State University.

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

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition.* Routledge.

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.

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

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.*

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

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

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

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

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

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

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

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

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

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

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

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

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

Newman, R. S., Rowe, M. L., & Ratner, N. B. (2016). 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), 1158–1173.

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

Niyogi, S. (2002). Bayesian learning at the syntax-semantics interface. In *CogSci 24.*

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

Piaget, J. (1928). *Judgment and reasoning in the child.* London: Kegan Paul, Trench, Trubner.

Pinker, S. (1984). *Language learnability and language acquisition.* Cambridge, MA: Harvard University Press.

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

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.

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

Rescorla, L. A. (1976). *Concept formation in word learning* (Unpublished doctoral dissertation). Yale University.

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

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

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

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

Sachs, J. (1983). 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.*

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), 103–114.

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.

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

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, F., & Tenenbaum, J. B. (2007). Word learning as bayesian inference. *Psychological Review*, *114*(2), 245–272.

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.

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

Appendix A

Evidence for multimodality in overextension

To examine directly how the multimodal semantic space I constructed accounts for variation in the overextension data, I performed a logistic regression analysis. In particular, I considered two sets of data: the *attested set* of overextension word-referent pairs, and a *control set* that shuffles the word-referent mappings from the attested set. I then performed a binary classification task via logistic regression to assess whether the attested set could be distinguished from the control set, given the same three relational features that I used for my previous analyses. Concretely, for each word-referent pair, the logistic regression factors were the z-scores of categorical, visual analogical, and predicate-based distances, normalized over the entire dataset, and the response was a binary indicator for the attested/control set. Finally, I also labelled each word-referent pair in the attested set according to the top-scoring feature in the logistic regression model as a way to approximate which multimodal feature best explains each instance of overextension, and thus assess the degree of multimodality in the overextension data.

Table S1 shows the BIC scores and cross-validated accuracies of the full multimodal model, partial models consisting of feature pairs, and partial models consisting of single features. The full multimodal model best distinguishes attested overextension from control word pairs in BIC score and predictive accuracy. Furthermore, all three features of the full multimodal model reached significance in the logistic regression ($p < .001$). These results suggest that a combination of semantic relations provides significant predictability of concept pairs that might undergo overextension.

Table S1: BIC scores and cross-validated predictive accuracies of logistic regression models ($N = 472$). Standard errors for accuracies are shown in parentheses.

Model	**BIC score**	**Accuracy**
categorical (cat.)	495	0.778(19)
visual (vis.)	464	0.767(19)
predicate (pred.)	469	0.763(20)
vis. + pred.	408	0.807(18)
cat. + vis.	393	0.816(18)
cat. + pred.	422	0.805(18)
all features	**374**	**0.839(17)**

Appendix B

Detailed model results for predicting word choices

Predictive performance of individual partial models. Section 6.1 shows the average performance curves of models containing single features, feature pairs, and all three multimodal features. Figure S1 shows the performance curves of all individual models, thus elucidating the relative performance of each combination.

Sample predictions of production model. Table S2 shows the top 5 words predicted by the production model for a random sample of the overextension dataset. Recalling that the top-5 model accuracy is approximately 55%, I observe that in the cases of incorrect model prediction, many of the predicted words are still closely related to the target concept, suggesting that even predictions that do not match the recorded child production are often sensible predictions. Since the overextension dataset is not an exhaustive record of child speech, this performance measure should be seen as a lower bound on the ability of the model to reconstruct children's overextension strategies.

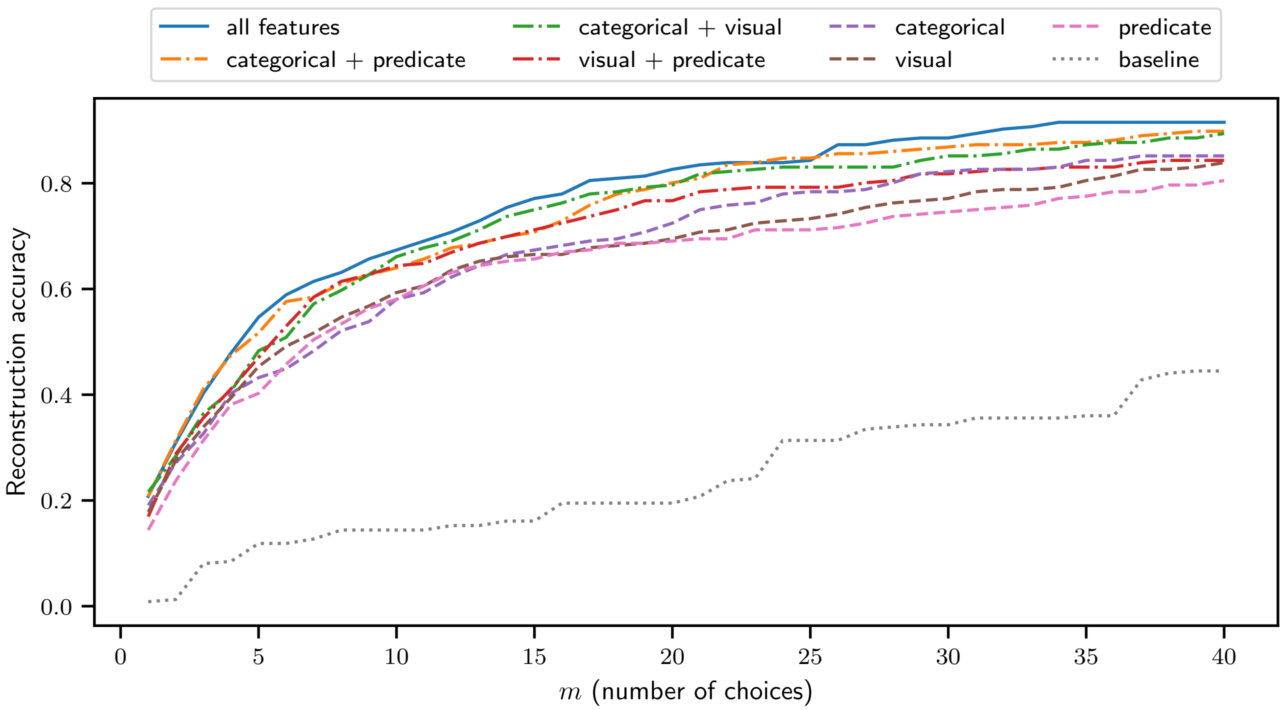

Figure S1: Performance curves for all production models showing cross-validated model accuracies in reconstructing overextended word choices.

Table S2: Top 5 model-predicted word choices for a random sample of the overextension dataset, stratified by correctness of model predictions. The upper panel shows examples for which the model predicts the true child production, and the lower panel shows examples for which model predictions do not include the child word. Each row displays the child production and intended referent from the overextension dataset, and the top 5 words predicted by the model, all denoted by their corresponding WordNet *synsets*.

Production	**Referent**	**1**	**2**	**3**	**4**	**5**
apple.n.01	banana.n.02	**apple.n.01**	orange.n.01	fruit.n.01	pea.n.02	juice.n.01
horse.n.01	deer.n.01	elk.n.01	animal.n.01	sheep.n.01	cow.n.01	**horse.n.01**
truck.n.01	train.n.01	car.n.01	bus.n.01	**truck.n.01**	bicycle.n.01	toy.n.03
sock.n.01	stocking.n.01	shoe.n.01	**sock.n.01**	hat.n.01	shirt.n.01	chair.n.01
dad.n.01	man.n.01	son.n.01	girl.n.01	baby.n.01	**dad.n.01**	animal.n.01
bicycle.n.01	motorcycle.n.01	car.n.01	**bicycle.n.01**	truck.n.01	wheel.n.01	train.n.01
catsup.n.01	mayonnaise.n.01	cheese.n.01	egg.n.02	peanut_butter.n.01	**catsup.n.01**	juice.n.01
ball.n.01	marble.n.02	**ball.n.01**	toy.n.03	chair.n.01	table.n.02	box.n.01
cheese.n.01	butter.n.01	**cheese.n.01**	milk.n.01	food.n.01	egg.n.02	toast.n.01
shoe.n.01	slipper.n.01	**shoe.n.01**	sock.n.01	blanket.n.01	hat.n.01	boot.n.01
kitten.n.01	horse.n.01	domestic_ass.n.01	pony.n.01	cow.n.01	animal.n.01	hog.n.03
cow.n.01	fish.n.01	tuna.n.03	animal.n.01	duck.n.01	child.n.01	baby.n.01
cat.n.01	lamb.n.01	sheep.n.01	animal.n.01	kitten.n.01	baby.n.01	dog.n.01
ball.n.01	peach.n.03	apple.n.01	orange.n.01	fruit.n.01	plum.n.02	grape.n.01
horse.n.01	jaguar.n.01	tiger.n.02	animal.n.01	lion.n.01	bear.n.01	cat.n.01
catsup.n.01	dressing.n.01	food.n.01	juice.n.01	cheese.n.01	pizza.n.01	pickle.n.01
bubble.n.01	marble.n.02	ball.n.01	toy.n.03	chair.n.01	table.n.02	box.n.01
horse.n.01	dog.n.01	puppy.n.01	animal.n.01	cat.n.01	kitten.n.01	kitty.n.04
cat.n.01	hog.n.03	cow.n.01	baby.n.01	animal.n.01	dog.n.01	bear.n.01
banana.n.02	tomato.n.01	apple.n.01	potato.n.01	juice.n.01	carrot.n.03	cheese.n.01

Appendix C

Validation of linguistic data from McDonough (2002)

The computational replication of McDonough (2002) in Section 6.2 follows the original experiment in dividing the stimuli into early and late items (in age of acquisition). Figure S2 shows that the relative frequency data collected from child-directed speech reflects this division, i.e., on average, early nouns are more frequent than late nouns. This "sanity check" verifies that my data is in sufficient agreement with McDonough (2002) for a meaningful computational reproduction.

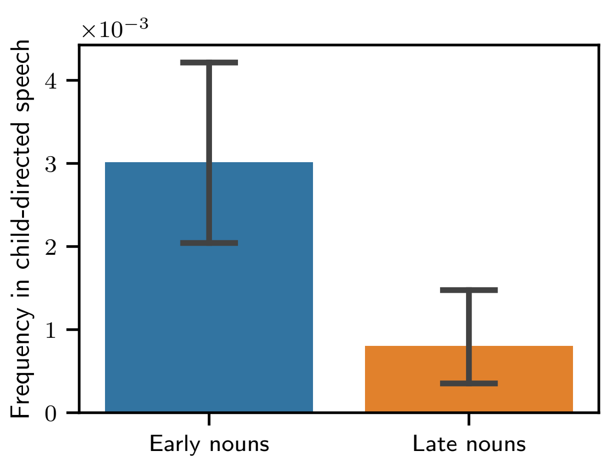

Figure S2: Relative frequencies of early and late nouns from McDonough (2002) in child-directed speech. Error bars represent bootstrap 95% confidence intervals.

Appendix D

Dataset of children's overextension

Table S3 shows all examples of overextension collected from the literature for this study, along with their sources. Each example includes a word and its intended referent, along with WordNet *synsets* coded for each.

Table S3: Dataset of children's overextension. Each row shows the word overextended in production and its intended referent, along with WordNet *synsets* for each.

Production	**Synset**	**Referent**	**Synset**	**Source**
airplane	airplane.n.01	boat	boat.n.01	Rescorla (1976)
airplane	airplane.n.01	helicopter	helicopter.n.01	Rescorla (1976)
airplane	airplane.n.01	rocket	rocket.n.01	Rescorla (1976)
airplane	airplane.n.01	submarine	submarine.n.01	Rescorla (1976)
airplane	airplane.n.01	train	train.n.01	Rescorla (1976)
apple	apple.n.01	ball	ball.n.01	Thomson and Chapman (1977)
apple	apple.n.01	balloon	balloon.n.02	Rescorla (1976)
apple	apple.n.01	banana	banana.n.02	Rescorla (1981)
apple	apple.n.01	cherry	cherry.n.03	Thomson and Chapman (1977)
apple	apple.n.01	egg	egg.n.02	Rescorla (1976)
apple	apple.n.01	fruit	fruit.n.01	Rescorla (1976)
apple	apple.n.01	grape	grape.n.01	Rescorla (1981)
apple	apple.n.01	lemon	lemon.n.01	Rescorla (1981)
apple	apple.n.01	melon	melon.n.01	Rescorla (1981)
apple	apple.n.01	onion	onion.n.01	Thomson and Chapman (1977)
apple	apple.n.01	orange	orange.n.01	Thomson and Chapman (1977)
apple	apple.n.01	orange juice	orange_juice.n.01	Rescorla (1981)
apple	apple.n.01	peach	peach.n.03	Thomson and Chapman (1977)
apple	apple.n.01	pear	pear.n.01	Thomson and Chapman (1977)
apple	apple.n.01	plum	plum.n.02	Rescorla (1981)
apple	apple.n.01	pumpkin	pumpkin.n.02	Thomson and Chapman (1977)
apple	apple.n.01	strawberry	strawberry.n.01	Thomson and Chapman (1977)
apple	apple.n.01	teapot	teapot.n.01	Thomson and Chapman (1977)

apple	apple.n.01	tomato	tomato.n.01	Thomson and Chapman (1977)
baby	baby.n.01	child	child.n.01	Barrett (1978)
baby	baby.n.01	statue	statue.n.01	Clark (1973)
ball	ball.n.01	apple	apple.n.01	Clark (1973)
ball	ball.n.01	balloon	balloon.n.02	Barrett (1978)
ball	ball.n.01	button	button.n.01	Rescorla (1976)
ball	ball.n.01	circle	circle.n.01	Thomson and Chapman (1977)
ball	ball.n.01	dome	dome.n.04	Barrett (1978)
ball	ball.n.01	egg	egg.n.02	Rescorla (1976)
ball	ball.n.01	grapefruit	grapefruit.n.02	Rescorla (1976)
ball	ball.n.01	helmet	helmet.n.01	Thomson and Chapman (1977)
ball	ball.n.01	marble	marble.n.02	Barrett (1978)
ball	ball.n.01	onion	onion.n.01	Thomson and Chapman (1977)
ball	ball.n.01	orange	orange.n.01	Rescorla (1976)
ball	ball.n.01	oval	ellipse.n.01	Thomson and Chapman (1977)
ball	ball.n.01	peach	peach.n.03	Rescorla (1976)
ball	ball.n.01	radish	radish.n.01	Clark (1973)
ball	ball.n.01	squash	squash.n.02	Thomson and Chapman (1977)
ball	ball.n.01	toy	plaything.n.01	Clark (1973)
banana	banana.n.02	apple	apple.n.01	Thomson and Chapman (1977)
banana	banana.n.02	cherry	cherry.n.03	Thomson and Chapman (1977)
banana	banana.n.02	fruit	fruit.n.01	Rescorla (1976)
banana	banana.n.02	lemon	lemon.n.01	Thomson and Chapman (1977)
banana	banana.n.02	orange	orange.n.01	Thomson and Chapman (1977)
banana	banana.n.02	peach	peach.n.03	Thomson and Chapman (1977)
banana	banana.n.02	pear	pear.n.01	Rescorla (1981)
banana	banana.n.02	raisin	raisin.n.01	Rescorla (1981)
banana	banana.n.02	strawberry	strawberry.n.01	Thomson and Chapman (1977)
banana	banana.n.02	tomato	tomato.n.01	Thomson and Chapman (1977)
bath	bathtub.n.01	water	water.n.06	Rescorla (1976)
bear	bear.n.01	rabbit	rabbit.n.01	Thomson and Chapman (1977)
bear	bear.n.01	seal	seal.n.09	Thomson and Chapman (1977)
bee	bee.n.01	bug	bug.n.01	Rescorla (1976)
beer	beer.n.01	soy sauce	soy_sauce.n.01	Thomson and Chapman (1977)
bike	bicycle.n.01	motorcycle	motorcycle.n.01	Rescorla (1976)
bike	bicycle.n.01	scooter	scooter.n.02	Rescorla (1976)
bike	bicycle.n.01	tricycle	tricycle.n.01	Rescorla (1976)
bike	bicycle.n.01	wheelchair	wheelchair.n.01	Rescorla (1976)
bird	bird.n.01	cat	cat.n.01	Clark (1973)
bird	bird.n.01	cow	cow.n.01	Clark (1973)
bird	bird.n.01	dog	dog.n.01	Clark (1973)
bird	bird.n.01	duck	duck.n.01	Fremgen and Fay (1980)
boat	boat.n.01	airship	airship.n.01	Barrett (1978)
boat	boat.n.01	bulldozer	bulldozer.n.01	Rescorla (1976)
boat	boat.n.01	raft	raft.n.01	Rescorla (1976)
book	book.n.01	magazine	magazine.n.02	Rescorla (1976)
bottle	bottle.n.01	cup	cup.n.01	Rescorla (1976)
box	box.n.01	drawer	drawer.n.01	Clark (1973)
box	box.n.01	TV	television_receiver.n.01	Rescorla (1976)
boy	son.n.01	child	child.n.01	Barrett (1978)
bubble	bubble.n.01	egg	egg.n.02	Rescorla (1976)
bubble	bubble.n.01	marble	marble.n.02	Rescorla (1976)
bunny	bunny.n.02	squirrel	squirrel.n.01	Gruendel (1977)
bus	bus.n.01	cable car	cable_car.n.01	Rescorla (1976)

bus	bus.n.01	firetruck	fire_engine.n.01	Rescorla (1976)
bus	bus.n.01	train	train.n.01	Rescorla (1981)
bus	bus.n.01	truck	truck.n.01	Rescorla (1981)
butter	butter.n.01	peanut butter	peanut_butter.n.01	Rescorla (1976)
cake	cake.n.03	candy	candy.n.01	Barrett (1978)
candy	candy.n.01	cherry	cherry.n.03	Barrett (1978)
car	car.n.01	airplane	airplane.n.01	Rescorla (1981)
car	car.n.01	bike	bicycle.n.01	Rescorla (1981)
car	car.n.01	bus	bus.n.01	Rescorla (1981)
car	car.n.01	cassette	cassette.n.01	Rescorla (1976)
car	car.n.01	stroller	baby_buggy.n.01	Rescorla (1981)
car	car.n.01	train	train.n.01	Rescorla (1981)
car	car.n.01	truck	truck.n.01	Rescorla (1981)
car	car.n.01	vehicle	vehicle.n.01	Rescorla (1976)
cat	cat.n.01	bear	bear.n.01	Rescorla (1981)
cat	cat.n.01	camel	camel.n.01	Rescorla (1981)
cat	cat.n.01	chicken	chicken.n.02	Rescorla (1981)
cat	cat.n.01	coyote	coyote.n.01	Rescorla (1981)
cat	cat.n.01	dog	dog.n.01	Rescorla (1976)
cat	cat.n.01	giraffe	giraffe.n.01	Rescorla (1981)
cat	cat.n.01	horse	horse.n.01	Rescorla (1981)
cat	cat.n.01	lamb	lamb.n.01	Rescorla (1981)
cat	cat.n.01	lion	lion.n.01	Rescorla (1976)
cat	cat.n.01	pig	hog.n.03	Rescorla (1976)
cat	cat.n.01	rabbit	rabbit.n.01	Rescorla (1981)
cheese	cheese.n.01	butter	butter.n.01	Rescorla (1976)
clock	clock.n.01	bracelet	bracelet.n.02	Rescorla (1976)
clock	clock.n.01	meter	meter.n.02	Rescorla (1976)
clock	clock.n.01	radio	radio_receiver.n.01	Rescorla (1976)
clock	clock.n.01	telephone	telephone.n.01	Rescorla (1976)
clock	clock.n.01	timer	timer.n.01	Rescorla (1976)
clock	clock.n.01	watch	watch.n.01	Rescorla (1976)
comb	comb.n.01	centipede	centipede.n.01	Rescorla (1976)
cow	cow.n.01	ant	ant.n.01	Thomson and Chapman (1977)
cow	cow.n.01	bear	bear.n.01	Thomson and Chapman (1977)
cow	cow.n.01	buffalo	american_bison.n.01	Thomson and Chapman (1977)
cow	cow.n.01	camel	camel.n.01	Thomson and Chapman (1977)
cow	cow.n.01	elephant	elephant.n.01	Thomson and Chapman (1977)
cow	cow.n.01	fish	fish.n.01	Thomson and Chapman (1977)
cow	cow.n.01	gorilla	gorilla.n.01	Thomson and Chapman (1977)
cow	cow.n.01	hippopotamus	hippopotamus.n.01	Thomson and Chapman (1977)
cow	cow.n.01	horse	horse.n.01	Gruendel (1977)
cow	cow.n.01	kangaroo	kangaroo.n.01	Thomson and Chapman (1977)
cow	cow.n.01	leopard	leopard.n.02	Thomson and Chapman (1977)
cow	cow.n.01	lion	lion.n.01	Thomson and Chapman (1977)
cow	cow.n.01	polar bear	ice_bear.n.01	Thomson and Chapman (1977)
cow	cow.n.01	reindeer	caribou.n.01	Thomson and Chapman (1977)
cow	cow.n.01	seal	seal.n.09	Thomson and Chapman (1977)
cow	cow.n.01	zebra	zebra.n.01	Thomson and Chapman (1977)
dada	dad.n.01	man	man.n.01	Rescorla (1976)
dada	dad.n.01	mom	ma.n.01	Rescorla (1976)
deer	deer.n.01	horse	horse.n.01	Thomson and Chapman (1977)
dog	dog.n.01	bear	bear.n.01	Thomson and Chapman (1977)
dog	dog.n.01	cat	cat.n.01	Rescorla (1976)

dog	dog.n.01	doe	doe.n.02	Thomson and Chapman (1977)
dog	dog.n.01	fish	fish.n.01	Thomson and Chapman (1977)
dog	dog.n.01	fox	fox.n.01	Thomson and Chapman (1977)
dog	dog.n.01	frog	frog.n.01	Rescorla (1981)
dog	dog.n.01	giraffe	giraffe.n.01	Rescorla (1981)
dog	dog.n.01	hippopotamus	hippopotamus.n.01	Thomson and Chapman (1977)
dog	dog.n.01	horse	horse.n.01	Fremgen and Fay (1980)
dog	dog.n.01	lamb	lamb.n.01	Rescorla (1981)
dog	dog.n.01	lion	lion.n.01	Fremgen and Fay (1980)
dog	dog.n.01	rabbit	rabbit.n.01	Gruendel (1977)
dog	dog.n.01	raccoon	raccoon.n.02	Thomson and Chapman (1977)
dog	dog.n.01	rhinoceros	rhinoceros.n.01	Thomson and Chapman (1977)
dog	dog.n.01	squirrel	squirrel.n.01	Rescorla (1981)
dog	dog.n.01	turtle	turtle.n.02	Rescorla (1981)
dog	dog.n.01	wolf	wolf.n.01	Thomson and Chapman (1977)
door	door.n.01	shutter	shutter.n.02	Clark (1973)
duck	duck.n.01	bird	bird.n.01	Fremgen and Fay (1980)
duck	duck.n.01	chicken	chicken.n.02	Fremgen and Fay (1980)
duck	duck.n.01	goose	goose.n.01	Rescorla (1981)
duck	duck.n.01	pigeon	pigeon.n.01	Rescorla (1981)
duck	duck.n.01	platypus	platypus.n.01	Thomson and Chapman (1977)
duck	duck.n.01	swan	swan.n.01	Rescorla (1981)
fish	fish.n.01	bee	bee.n.01	Thomson and Chapman (1977)
fish	fish.n.01	butterfly	butterfly.n.01	Thomson and Chapman (1977)
fish	fish.n.01	crab	crab.n.01	Thomson and Chapman (1977)
fish	fish.n.01	firefly	firefly.n.01	Thomson and Chapman (1977)
fish	fish.n.01	mosquito	mosquito.n.01	Thomson and Chapman (1977)
fish	fish.n.01	moth	moth.n.01	Thomson and Chapman (1977)
fish	fish.n.01	seal	seal.n.09	Rescorla (1976)
fish	fish.n.01	spider	spider.n.01	Thomson and Chapman (1977)
flower	flower.n.01	tree	tree.n.01	Rescorla (1976)
fly	fly.n.01	insect	insect.n.01	Clark (1973)
fly	fly.n.01	frog	frog.n.01	Clark (1973)
fruit	fruit.n.01	applesauce	applesauce.n.01	Rescorla (1981)
fruit	fruit.n.01	pear	pear.n.01	Rescorla (1981)
grandma	grandma.n.01	grandpa	grandfather.n.01	Rescorla (1976)
hat	hat.n.01	bowl	bowl.n.01	Rescorla (1976)
hat	hat.n.01	bucket	bucket.n.01	Rescorla (1976)
hat	hat.n.01	crown	crown.n.04	Rescorla (1976)
hat	hat.n.01	mitten	mitten.n.01	Rescorla (1976)
hat	hat.n.01	scarf	scarf.n.01	Rescorla (1976)
helicopter	helicopter.n.01	rocket	rocket.n.01	Rescorla (1976)
horse	horse.n.01	camel	camel.n.01	Rescorla (1976)
horse	horse.n.01	cow	cow.n.01	Rescorla (1976)
horse	horse.n.01	deer	deer.n.01	Thomson and Chapman (1977)
horse	horse.n.01	donkey	domestic_ass.n.01	Rescorla (1976)
horse	horse.n.01	fox	fox.n.01	Thomson and Chapman (1977)
horse	horse.n.01	giraffe	giraffe.n.01	Rescorla (1976)
horse	horse.n.01	goat	goat.n.01	Rescorla (1976)
horse	horse.n.01	hyena	hyena.n.01	Thomson and Chapman (1977)
horse	horse.n.01	kangaroo	kangaroo.n.01	Thomson and Chapman (1977)
horse	horse.n.01	llama	llama.n.01	Thomson and Chapman (1977)
horse	horse.n.01	reindeer	caribou.n.01	Thomson and Chapman (1977)
horse	horse.n.01	dog	dog.n.01	Thomson and Chapman (1977)

horse	horse.n.01	panther	jaguar.n.01	Thomson and Chapman (1977)
juice	juice.n.01	drink	drink.n.01	Rescorla (1976)
juice	juice.n.01	milk	milk.n.01	Rescorla (1976)
juice	juice.n.01	orange	orange.n.01	Rescorla (1976)
juice	juice.n.01	water	water.n.06	Rescorla (1976)
ketchup	catsup.n.01	dressing	dressing.n.01	Thomson and Chapman (1977)
ketchup	catsup.n.01	mayonnaise	mayonnaise.n.01	Thomson and Chapman (1977)
ketchup	catsup.n.01	pancake syrup	syrup.n.01	Thomson and Chapman (1977)
key	key.n.01	door	door.n.01	Rescorla (1976)
key	key.n.01	hook	hook.n.02	Rescorla (1976)
kitty	kitten.n.01	beaver	beaver.n.07	Thomson and Chapman (1977)
kitty	kitten.n.01	fox	fox.n.01	Thomson and Chapman (1977)
kitty	kitten.n.01	gorilla	gorilla.n.01	Thomson and Chapman (1977)
kitty	kitten.n.01	horse	horse.n.01	Thomson and Chapman (1977)
kitty	kitten.n.01	lamb	lamb.n.01	Thomson and Chapman (1977)
kitty	kitten.n.01	lion	lion.n.01	Fremgen and Fay (1980)
kitty	kitten.n.01	skunk	skunk.n.04	Thomson and Chapman (1977)
kitty	kitten.n.01	tiger	tiger.n.02	Thomson and Chapman (1977)
kitty-cat	kitty.n.04	fox	fox.n.01	Thomson and Chapman (1977)
kitty-cat	kitty.n.04	raccoon	raccoon.n.02	Thomson and Chapman (1977)
kitty-cat	kitty.n.04	skunk	skunk.n.04	Thomson and Chapman (1977)
mom	ma.n.01	dada	dad.n.01	Rescorla (1976)
mom	ma.n.01	woman	woman.n.01	Rescorla (1976)
man	man.n.01	boy	son.n.01	Rescorla (1976)
octopus	octopus.n.02	porcupine	porcupine.n.01	Thomson and Chapman (1977)
onion	onion.n.01	fruit	fruit.n.01	Rescorla (1976)
onion	onion.n.01	potato	potato.n.01	Rescorla (1976)
peach	peach.n.03	plum	plum.n.02	Rescorla (1976)
pen	pen.n.01	pencil	pencil.n.01	Rescorla (1976)
plum	plum.n.02	applesauce	applesauce.n.01	Rescorla (1976)
plum	plum.n.02	peach	peach.n.03	Rescorla (1981)
salt	salt.n.02	coffee	coffee.n.01	Thomson and Chapman (1977)
shoe	shoe.n.01	boot	boot.n.01	Rescorla (1976)
shoe	shoe.n.01	slipper	slipper.n.01	Rescorla (1976)
sock	sock.n.01	stockings	stocking.n.01	Rescorla (1976)
sock	sock.n.01	tights	tights.n.01	Rescorla (1976)
spoon	spoon.n.01	fork	fork.n.01	Rescorla (1976)
squirrel	squirrel.n.01	chipmunk	chipmunk.n.01	Thomson and Chapman (1977)
squirrel	squirrel.n.01	polar bear	ice_bear.n.01	Thomson and Chapman (1977)
squirrel	squirrel.n.01	skunk	skunk.n.04	Thomson and Chapman (1977)
tea	tea.n.01	coffee	coffee.n.01	Rescorla (1976)
tea	tea.n.01	cup	cup.n.01	Rescorla (1976)
tea	tea.n.01	teapot	teapot.n.01	Rescorla (1976)
tiger	tiger.n.02	lion	lion.n.01	Thomson and Chapman (1977)
truck	truck.n.01	bulldozer	bulldozer.n.01	Rescorla (1976)
truck	truck.n.01	bus	bus.n.01	Rescorla (1980)
truck	truck.n.01	train	train.n.01	Rescorla (1981)
water	water.n.06	hose	hose.n.03	Rescorla (1976)
wheel	wheel.n.01	ring	ring.n.08	Barrett (1978)
wheel	wheel.n.01	wagon	wagon.n.01	Barrett (1978)
wheel	wheel.n.01	wheelbarrow	barrow.n.03	Barrett (1978)

Appendix E

Vocabulary from early childhood

Table S4 shows the approximate vocabulary from early childhood extracted from Wordbank and used in my analyses, with each word manually coded as a WordNet *synset* to enable its representation in the semantic space.

Table S4: Approximate vocabulary from early childhood. Each cell shows the WordNet *synset* corresponding to one word in the vocabulary.

Synset			
airplane.n.01	alligator.n.02	animal.n.01	ant.n.01
apple.n.01	applesauce.n.01	aunt.n.01	baby.n.01
baby_buggy.n.01	bag.n.04	ball.n.01	balloon.n.01
banana.n.02	basement.n.01	basket.n.01	bat.n.01
bath.n.01	bathroom.n.01	bathtub.n.01	beach.n.01
bean.n.01	bear.n.01	bed.n.01	bedroom.n.01
bee.n.01	beer.n.01	belt.n.02	bench.n.01
beverage.n.01	bicycle.n.01	bird.n.01	bite.n.04
black.n.01	blanket.n.01	block.n.03	blue.n.01
boat.n.01	book.n.01	boot.n.01	bottle.n.01
bowl.n.01	box.n.01	breakfast.n.01	broom.n.01
brush.n.02	bubble.n.01	bucket.n.01	bug.n.01
bulge.n.01	bunny.n.02	bus.n.01	business_district.n.01
butter.n.01	butterfly.n.01	button.n.01	cake.n.03
camera.n.01	camping.n.01	can.n.01	candy.n.01
car.n.01	carrot.n.03	cat.n.01	catsup.n.01
chair.n.01	chamberpot.n.01	cheese.n.01	chewing_gum.n.01
chicken.n.02	child.n.01	chip.n.04	chocolate.n.03
church.n.02	clock.n.01	cloud.n.02	clown.n.02
coat.n.01	coca_cola.n.01	cock.n.04	coffee.n.01
comb.n.01	corn.n.01	cow.n.01	cracker.n.01
crayon.n.01	crib.n.01	cup.n.01	cupboard.n.01
dad.n.01	dance.n.02	deer.n.01	diaper.n.01

dinner.n.01	dish.n.01	doctor.n.01	dog.n.01
doll.n.01	domestic_ass.n.01	door.n.01	doughnut.n.02
drawer.n.01	dress.n.01	dryer.n.01	duck.n.01
dwelling.n.01	egg.n.02	elephant.n.01	elk.n.01
face.n.01	fire_engine.n.01	fireman.n.04	fish.n.01
flag.n.01	flower.n.01	fly.n.01	food.n.01
foot.n.01	fork.n.01	french_fries.n.01	friend.n.01
frog.n.01	fruit.n.01	game.n.09	garage.n.01
garbage.n.03	garden.n.01	gelatin.n.02	giraffe.n.01
girl.n.01	glass.n.02	glove.n.02	goose.n.01
grandfather.n.01	grandma.n.01	grape.n.01	grass.n.01
green.n.01	gym_shoe.n.01	hair.n.06	hamburger.n.01
hammer.n.02	hand.n.01	hat.n.01	head.n.01
helicopter.n.01	hen.n.01	hog.n.03	horse.n.01
hose.n.03	house.n.01	ice.n.01	ice_cream.n.01
ice_lolly.n.01	jacket.n.01	jar.n.01	jean.n.01
jello.n.01	juice.n.01	key.n.01	kitchen.n.01
kitten.n.01	kitty.n.04	knife.n.01	lady.n.01
lamb.n.01	lamp.n.01	lawn_mower.n.01	light.n.02
lion.n.01	lip.n.02	living_room.n.01	lollipop.n.02
lunch.n.01	ma.n.01	man.n.01	melon.n.01
menagerie.n.02	milk.n.01	mitten.n.01	monkey.n.01
mother.n.01	motorcycle.n.01	mouse.n.01	mouth.n.01
movie.n.01	muffin.n.01	nail.n.02	napkin.n.01
necklace.n.01	nurse.n.01	nut.n.01	octopus.n.02
onion.n.01	orange.n.01	oven.n.01	owl.n.01
paint.n.01	pancake.n.01	paper.n.01	party.n.02
patty.n.01	pea.n.02	peach.n.03	peanut_butter.n.01
pen.n.01	pencil.n.01	penguin.n.01	people.n.01
person.n.01	pickle.n.01	picnic.n.03	pillow.n.01
pizza.n.01	plant.n.02	plate.n.04	playground.n.02
plum.n.02	pony.n.01	pop.n.02	popcorn.n.01
porch.n.01	potato.n.01	pretzel.n.01	pudding.n.01
pumpkin.n.01	puppy.n.01	puzzle.n.02	radio.n.01
raisin.n.01	rear.n.05	refrigerator.n.01	rock.n.01
rocking_chair.n.01	roof.n.01	room.n.01	salt.n.02
sandwich.n.01	sauce.n.01	scarf.n.01	school.n.02
scissors.n.01	sheep.n.01	shirt.n.01	shoe.n.01
shop.n.01	short_pants.n.01	shoulder.n.01	shovel.n.01
shower.n.01	sidewalk.n.01	sink.n.01	sister.n.01
skate.n.01	sky.n.01	sled.n.01	slide.n.04
slide_fastener.n.01	slipper.n.01	snow.n.01	sock.n.01
sofa.n.01	son.n.01	soup.n.01	spaghetti.n.01
spectacles.n.01	spoon.n.01	sprinkler.n.01	squirrel.n.01
stairs.n.01	star.n.03	stick.n.01	stove.n.01
strawberry.n.01	street.n.01	sun.n.01	swab.n.02
sweater.n.01	swing.n.02	table.n.02	tape.n.04
tea.n.01	teacher.n.01	telephone.n.01	television.n.01
tiger.n.02	tights.n.01	toast.n.01	tooth.n.02
toothbrush.n.01	towel.n.01	toy.n.03	tractor.n.01
train.n.01	tray.n.01	tree.n.01	tricycle.n.01
trouser.n.01	truck.n.01	tuna.n.03	turkey.n.01
turtle.n.02	underpants.n.01	vacuum.n.04	vanilla.n.01
vitamin.n.01	walker.n.04	wash.n.01	washer.n.03

watch.n.01	water.n.06	wheel.n.01	white.n.02
window.n.01	wolf.n.01	yellow.n.01	yogurt.n.01
zebra.n.01			