Title
The Semantic Spaces of Child-Directed Speech, Child Speech and Adult-directed Speech: a Manifold Perspective

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Authors
Sun, Hao
Pate, John

Publication Date
2017

Peer reviewed

The Semantic Spaces of Child-Directed Speech, Child Speech and Adult-directed Speech: a Manifold Perspective

Hao Sun (hsun7@buffalo.edu)
Department of Linguistics, University at Buffalo
Buffalo, NY 14260 USA

John Pate (johnpate@buffalo.edu)
Department of Linguistics, University at Buffalo
Buffalo, NY 14260 USA

Abstract

Child-directed speech (CDS) is a talking style adopted by caregivers when they talk to toddlers (Snow, 1995). We consider the role of distributional semantic features of CDS in language acquisition. We view semantic structure as a manifold on which words lie. We compare the semantic structure of verbs in CDS to the semantic structure of child speech (CS) and adult-directed speech (ADS) by measuring how easy it is to align the manifolds. We find that it is easier to align verbs in CS to CDS than to align CS to ADS, suggesting that the semantic structure of CDS is reflected in child productions. We also find, by measuring verbs vertex degrees in a semantic graph, that a mixed initialized set of verbs with high degrees and medium degrees has the best performance among all alignments, suggesting that both semantic generality and diversity may be important for developing semantic representations.

Keywords: child-directed speech; lexical development; manifold learning; distributional semantics; graph theory

Introduction

One of the biggest puzzles in cognitive science is how children learn language from language input, namely child-directed speech. Child-directed speech is characterized by simplified sentence structures, restricted vocabulary, exaggerated intonation, and hyperarticulation, and previous work has proposed that these features facilitate language acquisition (Golinkoff and Alioto, 1995; Snow, 1995; Thiessen, Hill, and Saffran, 2005). Here, we compare the semantic spaces of child speech, child-directed speech and adult-directed speech, spanned by verbs, using state-of-the-art computational tools.

The contributions of this paper are both theoretical and methodological. Theoretically, we explore various proposals about roles of verbs meanings in CDS, represented using a state-of-the-art distributional semantics approach. Distributional methods map each word to a point in high-dimensional space so that words with similar meanings are near each other. We view the semantic structure of the vocabulary as a high-dimensional surface in this space, called a *manifold*, and compare manifolds estimated from CDS to manifolds estimated from child speech (CS) and adult-directed speech (ADS). Young children often broaden the use of nouns and verbs and we model such differences in word meaning as a mismatch of data points in a semantic space.

Methodologically, we adapt a novel semi-supervised manifold alignment algorithm to compare semantic spaces (Ham et al, 2005), which maps two manifolds into a common subspace to measure the similarity of these manifolds. This algorithm takes as input a subset of initial points that must be aligned (i.e., pairs of points, one on each manifold, that correspond to the same verb), and produces an alignment for the rest of the verbs. We then measure the similarity of the manifolds in terms of the accuracy of the alignment: how often a verb is mapped to the same region of the common subspace.

We find that alignment between the CS and CDS is more accurate than the alignment between CS and ADS. Additionally, we obtain more accurate alignments when using verbs with many nearest neighbors (which have broader meanings) as the initial points than verbs with few near neighbors. Together, these results indicate that the semantic structure of CS reflects the semantic structure of CDS, and verbs with broad meanings may provide useful cues to children in acquiring the overall semantic structure of verbs. On the one hand, what children can learn from CDS deviates semantically from unfamiliar conversations in ADS, which suggests that further learning is required. On the other hand, caregivers might align their semantic spaces to children’s semantic spaces, which lies within the general framework of conversational alignment (Pickering & Garod, 2004).

Model Setting

We combine models from two different traditions into a general framework of semantic representation. To compare the semantic spaces of CS, CDS and ADS, we use a manifold-based algorithm. The similarities between semantic spaces are measured by how easy it is to map one semantic space to another. We represent the meaning of each verb by using the global vector model (Pennington, Socher & Manning, 2014) to embed words into a 50-dimensional space, which we call a semantic space. . Following the associationist tradition in psychology (Anderson, 1973), we represent the meaning structure of the verbal lexicon as a whole by considering how a collection of verbs is situated in this space, as expressed by a neighborhood graph (Steyvers & Tenenbaum, 2005).

Estimating verb meanings from different datasets produces different semantic spaces, and we compare the spaces using a semisupervised manifold alignment algorithm (Ham et al., 2005). This algorithm maps verbal semantic graphs into a common semantic space and discovers the data point correspondences by finding pairs of points with the smallest Euclidean distances.

Lexical Semantic Representation

The past three decades saw efforts to model the mental representation of concepts (Launder & Dumais, 1997). The inspiration for recent computational work on lexical semantics dates back to Harris's (1954) hypothesis that synonymous words appear in similar contexts.

One of the most successful semantic representation models is proposed by Launder & Dumais (1997), known as Latent Semantic Analysis (LSA), which uses word-context co-occurrence matrices to produce a low-dimensional representation by singular value decomposition. The lexical semantic representation model used in this paper is based on a state-of-the-art algorithm, GloVe (Pennington, Socher & Manning, 2014), which is an extension of LSA. Instead of explicitly decomposing a word-context co-occurrence matrix, GloVe implicitly decomposes a word-context log-frequency matrix. GloVe uses a weighted regression objective function to reconstruct a log word-context count matrix log(X) with bias terms, as shown in Equation (1), where *w* and *b* are bias vectors, X is the co-occurrence matrix and *f* is a heuristic weighting function. The optimization problem is iteratively solved using AdaGrad (Duchi, Hazan & Singer, 2011).

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \widetilde{w}_j + b_i + \tilde{b}_j - \log X_{ij})^2 \quad (1)$$

Even though GloVe has better performance than traditional singular-value-decomposition-based LSA, careful analysis of the objective function suggests that GloVe is fundamentally probabilistic matrix factorizations (Levy & Goldberg, 2014).

Semantic Graphs

The manifold alignment algorithm we use approximates the underlying manifold by constructing a similarity graph G = (V, E), where the vertex set V is the set of verbs and the edge set E is a set of pairs of verbs that are near to each other. The weight of an edge is set to the cosine similarity between the verbs associated by the edge. The degree of a vertex is the sum of weights of all the edges linking to the vertex. In semantic networks, vertex degrees can be interpreted as contextual diversity. There are several ways to build such a similarity graph. Ozaki et al (2011) found that undirected mutual k nearest neighbor (*mkNN*) graphs give good performance for alignment of natural language data, so we use *mkNN* graphs. An *mkNN* graph has an edge (v_1, v_2) if either v_1 or v_2 is within the k nearest neighbors of the other. We set k to 15 for the first experiment. In the second experiment, we increase k to 20 to better investigate the degree effects. The unnormalized graph Laplacian (L) of graph W is defined in Equation (2). D is the degree matrix, a diagonal matrix with vertex degrees on the diagonal.

$$L = W - D \quad (2)$$

We use a symmetric graph Laplacian normalized by vertex degree (Shi & Malik, 2000), as

$$L_{sym} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{1/2} \quad (3)$$

Aligning Semantic Spaces

We compare the semantic spaces of CS, CDS and ADS using the semisupervised manifold alignment algorithm. A manifold is defined as a topological structure with every local point with a neighborhood similar to a Euclidean space. The goal of the manifold alignment algorithm is to pair up data points from two high-dimensional data sets. For example, the algorithm aims to match *give* in CS to *give* in CDS. A semisupervised algorithm, using both labeled and unlabeled data as input, combines the strength of supervised and unsupervised learning. The general goal of manifold alignment is to map two high-dimensional data sets to a common low-dimensional space simultaneously (Ham et al., 2005), which essentially is an extension of manifold-based nonlinear dimensionality reduction (Belkin & Niyogi, 2003). Manifold-based methods are based on the geometric assumption that data in high dimensional space lie in low-dimension manifolds.

Ham et al.'s algorithm defines a function $\boldsymbol{f}$ that maps the first manifold to a common space, and a function $\boldsymbol{g}$ that maps the second manifold to a common subspace. These functions strike a tradeoff between mapping labeled pairs to the same point in the common space, and respecting local structure on the original manifolds as expressed by the graph Laplacian L^x for the first space and L^y for the second space. As we have both labeled (l) and unlabeled (u) points, L^x and L^y are block matrices:

$$L^x = \begin{bmatrix} L_{lu}^x & L_{ul}^x \\ L_{lu}^x & L_{uu}^x \end{bmatrix} \quad (4)$$

The cost of the mapping is then:

$$\tilde{C}(\boldsymbol{f}, \boldsymbol{g}) = \frac{C(\boldsymbol{f}, \boldsymbol{g})}{\boldsymbol{f}^T\boldsymbol{f} + \boldsymbol{g}^T\boldsymbol{g}} \quad (5)$$

where μ expresses the tradeoff between mapping points exactly and preserving local structure on the original manifolds. The first term is the sum of distances between paired data points in the common space, and the second two terms represent faithfulness to the graph Laplacian. Ham et al. point out that Equation 4 is unsuitable for optimization, since it ignores simultaneous scaling of $\boldsymbol{f}$ and $\boldsymbol{g}$, and so instead minimize the Rayleigh quotient:

$$C(\boldsymbol{f}, \boldsymbol{g}) = \mu \sum_i |f_i - g_i|^2 + \boldsymbol{f}^T L^x \boldsymbol{f} + \boldsymbol{g}^T L^y \boldsymbol{g} \quad (6)$$

We set μ to positive infinity to impose a hard constraint for labeled pairs to be mapped directly on top of each other.

The analytic solution to the optimization is then given by the generalized graph Laplacian L^z in Equation 7.

$$L^z = \begin{bmatrix} L_{ll}^x + L_{ll}^y & L_{lu}^x & L_{lu}^y \\ L_{ul}^x & L_{uu}^x & 0 \\ L_{ul}^y & 0 & L_{uu}^y \end{bmatrix} \qquad (7)$$

The semisupervised manifold alignment algorithm adopted from Ham et al. 2005 is described in Algorithm 1.

Algorithm 1: Semisupervised Manifold Alignment Algorithm (Ham et al., 2005)

Input: data points from two data sets, with N initially aligned data point pairs
Output: a matching of data points

1. Construct similarity graphs G_1, G_2, for both data sets respectively, using m*kNN*
2. Compute the symmetric graph Laplacians of G_1 and G_2, L^x and L^y, using Equation (3)
3. Compute a graph Laplacian for a joint graph L^z using Equations (6) and (7)
4. Compute the eigenvectors of L^z and take eigenvectors corresponding to the smallest non-zero eigenvalues, the results of which are the vectors in a lower-dimensional space
5. Find the data points with smallest Euclidean distance weighted by the inverse of their respective eigenvalues

Experiment Setup

Corpora

The training set for CDS and CS is a combined data set from CHILDES (MacWhinney, 2000), which consists of all the data on American English-speaking monolingual 3 to 7 year-old children with typical language and cognitive development, excluding diary studies. To simplify data collection, only utterances annotated as child are considered child speech and only utterances annotated as mother and father are considered as child-directed speech. The CS and CDS corpora contain 5 million and 9 million word tokens, respectively. To prevent the CS from being similar to CDS purely due to priming effects, we divided the data into two halves so that the CDS and CS data were not drawn from the same contexts

Our ADS data is drawn from the spoken portion of the Corpus of Contemporary American English (COCA, Davies, 2008). Although this data may differ from more casual conversations, it provides a large amount of spontaneous speech in the form of unscripted conversations from 150 television and radio programs.

Materials

The target words used in this model are all verbs, which are understudied in the literature. We included the first 100 English verbs acquired by infants (Fenson et al., 1994), the most frequent 200 English verbs in adult language productions (Davies, 2008) and verbs that appear in three common constructions (Levin, 1993).

The classes of verbs are the ones that appear in 3 constructions: the ditransitive (*John gave Mary a book*), the locative (*The man loaded hay onto a truck*) and the conative (*The police shot at the criminal*). Since CHILDES suffers from data sparcity, verbs missing in either CS or CDS were excluded from analysis. We end up with 811 data points for CS, CDS and spoken COCA respectively.

Data Preprocessing

The adult-directed speech data from spoken COCA and the child speech and child-directed speech data from CHILDES data were preprocessed using regular expressions. Verbs in different inflectional forms were treated as separate verb types.

Model Training

Global Vector Training We used the implementation of GloVe from the Stanford NLP website to train 50-dimensional vectors for each of our three datasets (Pennington, Socher & Manning, 2014). We trained each set of vectors for 50 epochs with a context window size of 10, used a frequency cut-off of 2 for the CS and CDS datasets and a cut-off of 10 for the ADS dataset.

Similarity Graph Construction We construct *mkNN* graphs consistently throughout this paper. In the first simulations, we fix the number of mutual nearest neighbors to 15. In the second simulation, we test the effect of vertex degrees and we set the number of mutual nearest neighbors to 20 to increase the range of vertex degree.

Manifold Alignment The parameters that we need to specify in the manifold alignment module include the initial labeled alignments and the dimensionality of the manifold. In addition to the number of labeled data, the identity of the labeled data can also influence the quality of alignment. The dimensionality of the manifold controls the abstraction of semantic information contained in the word vectors. The lower the dimension, the more abstract the representation.

Evaluation

Because the alignment algorithm pairs up labeled data points exactly, we only evaluate alignments on unlabeled data. We use a random alignment averaged over 5 times as the baseline condition. Ideally, corresponding data points from two data sets should be mutual nearest neighbor in the lower dimensional space. We relax the evaluation requirements by giving every alignment a k-nearest neighborhood evaluation radius. If one data point is one of the k-nearest neighbors of the corresponding point, we take it as a hit. When the evaluation neighborhood radius equals 1, the measures quantify the exact alignment.

Simulation 1: Mapping CS to CDS and COCA

In this section, we demonstrate that CS-CDS alignment is a less demanding task than CS-COCA alignment even when potential priming effects from linguistic and non-linguistic contexts are removed. We also predict that with the increase of labeled data, the alignment accuracy also increases.

Method

We performed verb semantic graph alignments of CS to CDS and to ADS for alignment spaces of dimensionality from 5 to 30. The unlabeled precisions are evaluated by the window-size at 1 and at 20, as demonstrated in the contour heat maps in Figure 1. The colors of different areas in the contours indicate different levels of unlabeled accuracy and the data points with the same unlabeled accuracy are connected by the isolines in the maps.

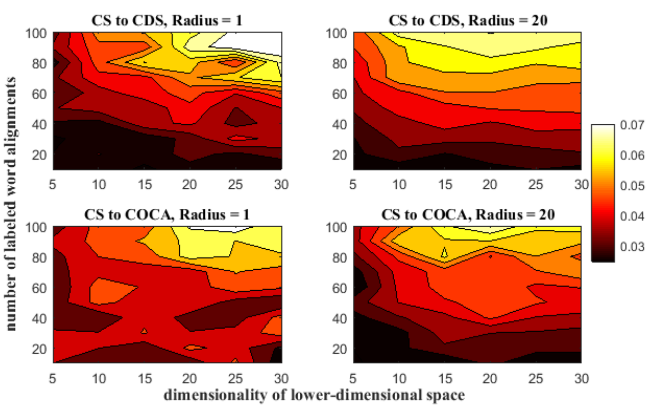

Figure 1 Accuracies of mapping CS to CDS and COCA

Results

The general trend is that the highest unlabeled precisions are found in the upper right corners of the contour maps whereas the lowest unlabeled precisions tend to lie close to the x-axis. The dimensionality of the embedding space can be interpreted as the granularity of children's representations.

The result of the alignments is demonstrated graphically in Figures 1 and 2. In the alignments from CS to CDS and CS to COCA, the CS-COCA alignment achieves only 50% to 60% of the unlabeled precision of the CS-CDS alignment. The unlabeled precision of the CS-CDS alignment is consistently higher than the unlabeled precision of the CS-COCA alignment across all conditions. Both alignments have much larger unlabeled accuracy than the random baseline.

The CS data are aligned to both the spoken COCA and CDS corpora. The CS-CDS alignment precision wins over the CS-COCA precision across all conditions. In other words, child speech is much easier to map to child-directed speech than to spoken COCA. This easier alignment can be interpreted as similarity in semantic spaces across corpora.

Since the CS and the CDS word vectors are trained on speech data from different experiments, the relative similarity between CS and CDS lexical semantics, this similarity does not reflect mere priming effects. There are two possible interpretations for this result. First, the result can be viewed as an imitation effect in which children mirror child-directed speech semantically. Second, adult caregivers might adapt their mental representations to children's when they talk to children, which sits well with the conversational alignment theory (Pickering & Garrod, 2004). The big semantic gap between initial language input and adult-to-adult conversations on TV shows or radios suggests that learning from CDS alone is not sufficient for real world language processing. Adapting to TV or radio conversations constitute one part of further learning, which supports a continuous theory of language development.

Simulation 2: Semantic generality

In Simulation 2, we use a fixed list of labeled data to investigate the effect of initialization in alignment, instead of random initialization. The motivation is that language scientists argue for the importance of a few important "path-breaking" word exemplars in language learning (Ninio, 1999; Goldberg, Casenhiser & Sethuraman, 2004). Some words attract more vertices than others, which is known as preferential attachment in network growth (Steyvers & Tenenbaum, 2005). We evaluate the proposal that semantically general verbs are better starting points for language learning than semantically specific verbs, by measuring the vertex degrees.

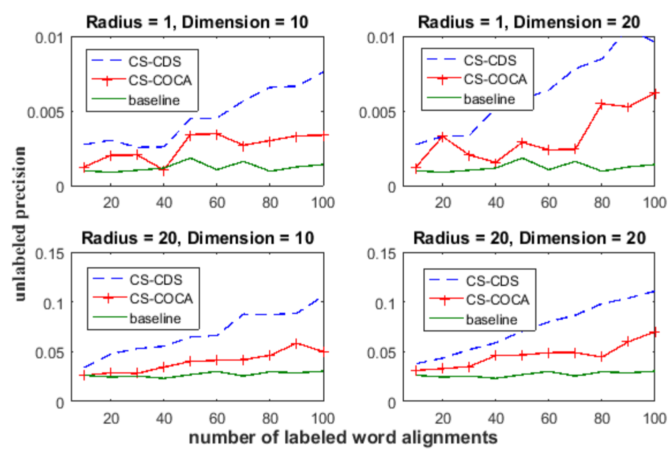

Figure 2 Unlabeled accuracies of CS-CDS and CS-COCA alignments with a random alignment as the baseline

The degree of a vertex measures the association between a vertex and its neighboring vertices. The prediction is that vertices with large degree are better labeled data than vertices with small degree. Cognitively, the verbs with high degree are semantically general verbs whereas the verbs with low degree are the ones with less general meanings.

Method

Verbs are ranked based on their vertex degree in a semantic network. As shown in Table 1, what we use as labeled data is 100 verbs with the largest degrees, 100 with the smallest degrees, and medium-degree verbs with degree rank of 201 to 300. We also mixed half of high degree verbs with half of medium degree verbs in the mixed condition. The baseline condition is averaged over 5 random initializations. We set the number of mutual nearest neighbors, the evaluation radius and the dimensionality all to 20.

Results

The alignment precisions shown in Figure 3 show a clear advantage of high-degree and medium degree conditions over the low degree condition, but both high-degree and low-degree have below random performances. We can also see an advantage of medium degree initialization, which is parallel to the basic level categorization theories. When we use a mixed set of high-degree and medium-degree verbs, we get the best results on all the conditions, which suggests that a diverse-degree initialization facilitates semantic space alignment.

Table 1: Verbs with the largest, medium and smallest vertex degrees in ADS

largest	medium	smallest
get	giving	tickles
go	tearing	points
want	taken	shooting
put	poured	design
think	tipping	tapping

General Discussion

In Simulation 1, we demonstrate that CS has semantic properties very similar to CDS in comparison to ADS. This result supports a usage-based approach to language acquisition: children imitate their caregivers. The results can be interpreted in multiple perspectives. First, the result suggests that child speech is built upon restricted linguistic contexts. One of the biggest characteristics of human memory is context-dependency. Early language experience is built upon restricted contexts and usages requires further learning to achieve the adult form. Second, child-directed speech is used in young children's living environments. Children seem to use words highly consistent with their caregivers. Third, talking to children in child-directed speech is a double-edged sword. On the one hand, children might have an easier time initializing their language capacities at an early language development stage because their hypothesis space is restricted by child-directed speech. On the other hand, the mismatch between child-directed speech and adult-directed speech requires children to shift their semantic representations at later development stages.

In Simulation 2, we show empirically that semantically moderately general verbs are better starting points for language development. Our simulations show mixed results for the "path-breaking" argument that semantically generic verbs are important for language learning (Ninio, 1999). Our results suggest that both semantic generality and semantic diversity play a role here. Although semantically general verbs help in general, verbs that are semantically too general may not be that helpful.

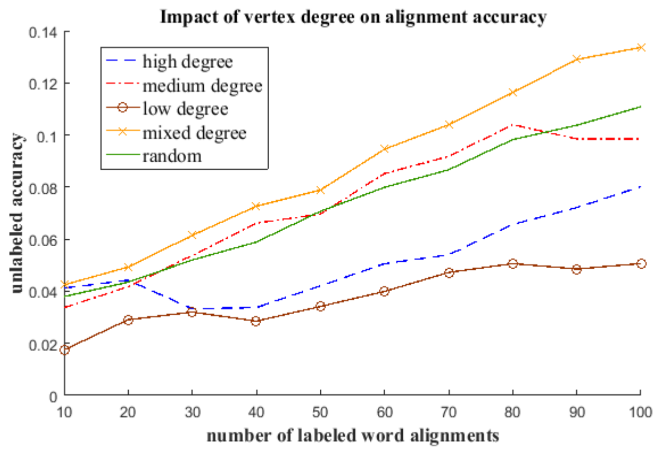

Figure 3 Unlabeled accuracies of alignments with high-degree, medium-degree, low degree, mixed-degree and random initializations

Speaker Normalization by Manifold Alignment

In speech recognition and perception, speaker normalization is the task of automatically adjusting to acoustic differences between different speakers. Our work is inspired by Plummer et al. (2010), who proposed manifold alignment as an account for how young children learn to handle phonetic variability in vowel production during language acquisition.

Aside from working with semantic, rather than acoustic, representations, our work differs from theirs in two respects. First, they used synthesized data as input, while we used naturalistic corpus data. Second, since two token pronunciations of vowels will never be the same, they imposed only a soft alignment constraint that labeled pairs be aligned, while we imposed a hard constraint.

Crosslinguistic Alignment of Polysemous Words

Youn et al. (2016) investigated semantic universals by constructing networks of corresponding polysemous nouns from 81 languages sampled from different language families. Using an approach reminiscent of thesaurus-based synonym induction, they established semantic correspondences between nouns using bilingual dictionaries. The target polysemous words were selected from the Swadesh 200 basic vocabulary list. The procedure described in this paper is automatic and takes into consideration the matching of semantic spaces in one language, whereas Youn and colleagues manually establishes semantic correspondences for a few basic words in bilingual data.

Conclusions

The contribution of this paper is a novel integrated framework that compares semantic spaces of children and their caregivers based on naturalistic language productions. We combined methods from three traditions, distributed semantic representations, graph theory, and manifold alignment, into one framework for approaching the semantic structure of the lexicon. We used naturalistic language productions from CHILDES to compare the semantic spaces spanned by verbs and demonstrated that (i) that CDS is more similar to ADS than CS in terms the semantic spaces spanned by verbs and that (ii) verbs with relatively large and diverse degrees are especially useful for aligning semantic structures.

While the general computational framework proposed in this paper does not provide an account of how children might exploit this manifold-based and graph-theoretic information, it does suggest that useful information about the structure of the adult lexicon is available to children. Even though our framework is on the computational level, using Marr's terminology (1982), it is very likely that semantic manifold alignment plays a role in children's semantic development. Additionally, this framework may be of use to other fields that are interested in the semantic structure of different lexicons. For example, this approach may be useful for performing semantic comparisons between languages or across time over the course of language change, and understanding the semantic organization of bilingual lexicons.

References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Psychology press.

Belkin, M., Niyogi, P., & Sindhwani, V. (2006). Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, *7*(Nov), 2399-2434.

Davies, Mark. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Diaz, F., & Metzler, D. (2007, January). Pseudo-Aligned Multilingual Corpora. In *IJCAI* (pp. 2727-2732).

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*(Jul), 2121-2159.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*(1), 71-99.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... & Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i-185.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive linguistics*, *15*(3), 289-316.

Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of child language*, *17*(01), 171-183.

Golinkoff, R. M., & Alioto, A. (1995). Infant-directed speech facilitates lexical learning in adults hearing Chinese: Implications for language acquisition. *Journal of Child Language*, *22*(3), 703-726.

Ham, J., Lee, D. D., & Saul, L. K. (2005, January). Semisupervised alignment of manifolds. In *AISTATS* (pp. 120-127).

Harris, Z. S. (1954). Distributional structure. *Word, 10*(2-3), 146-162.

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

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems* (pp. 2177-2185).

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. The MIT Press.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of child language*, *26*(03), 619-653.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global Vectors for Word Representation. In *EMNLP* (Vol. 14, pp. 1532-1543).

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(02), 169-190.

Plummer, A. R., Beckman, M. E., Belkin, M., Fosler-Lussier, E., & Munson, B. (2010). Learning speaker normalization using semisupervised manifold alignment. In *INTERSPEECH* (pp. 2918-2921).

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, *7*(1), 53-71.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, *22*(8), 888-905.

Snow, C. (1995). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. *The handbook of child language*, ed. by Paul Fletcher and Brian MacWhinney. Oxford: Blackwell.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41-78.

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., ... & Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, *113*(7), 1766-1771.