Year: 2020

Processing causatives in first language acquisition: A computational approach

You, Guanghao ; Daum, Moritz ; Stoll, Sabine

Originally published at:
You, Guanghao; Daum, Moritz; Stoll, Sabine (2020). Processing causatives in first language acquisition: A computational approach. In: Conference on Language Development / Proceedings of the 45th annual Boston University Conference on Language Development, Boston University, November 2020. Cascadilla Press, 818-828.

Processing Causatives in First Language Acquisition: A Computational Approach

Guanghao You, Moritz M. Daum, and Sabine Stoll

1. Introduction

One of the most challenging tasks for infant language learners is to extract meaning from the speech stream they hear. Linguistic contexts often stand as the main cues for children to infer the semantics of words. For instance, upon hearing the sentence "you broke the window", to understand the meaning of "broke", children have to first recognize the two participants "you" and "the window" in the utterance, and later establish the causal relation between them. The semantics embedded in linguistic contexts, often referred to as distributional semantics, plays an important role in children's semantic learning (Erickson & Thiessen, 2015; Theakston, Lieven, Pine, & Rowland, 2001; Tomasello, 2003). However, many studies have looked at only specific contextual patterns or selected semantic features to examine semantics in first language acquisition, therefore lacking comprehensiveness in their approach. In this study, we examined semantics with a computational approach that comprehensively accesses linguistic contexts to generate semantic representations. Using lexical causatives as a test case, we processed causative semantics in both English child-directed speech (CDS) and child speech (CS), thereby investigating the semantic development in parent-child interaction.

Children process rich information from their surrounding speech to later become a proficient speaker. It has been evidenced that CDS possesses

---

* Guanghao You, Department of Comparative Language Science, Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, guanghao.you@uzh.ch; Moritz M. Daum, Department of Psychology, Center for the Interdisciplinary Study of Language Evolution (ISLE), Jacobs Center for Productive Youth Development, University of Zurich; Sabine Stoll, Department of Comparative Language Science, Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich. Moritz M. Daum and Sabine Stoll share the last authorship. The results of this research are part of the project "The role of causality in early verb learning: language-specific factors vs. universal strategies" (Swiss NSF Grant agreement No. 100015_169712, Stoll, Daum), the project "Acquisition processes in maximally diverse languages: Min(ding) the ambient languages (ACQDIV)" that has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme FP7-2007-2013 (Grant agreement No. 615988; Stoll), and the NCCR Evolving Language, Swiss National Science Foundation Agreement No. 51NF40_180888 (Stoll, Daum).

 Proceedings of the 45th annual Boston University Conference on Language Development, ed. Danielle Dionne and Lee-Ann Vidal Covas, 818-828. Somerville, MA: Cascadilla Press.

idiosyncrasies specially tuned to ease children's learning tasks (Snow, 2017; Sokolov, 1993), and children indeed benefit from these "fine-tuned" properties (Kuhl, 2004; Soderstrom, 2007). This phenomenon of tuning has been researched at various linguistic levels, including syntax (Dale & Spivey, 2006) and speech complexity (Furrow, Nelson, & Benedict, 1979). Specific characteristics in CDS regarding semantic learning have been examined as well. Theakston et al. (2001), for instance, found that the acquisition of verbs is assisted with cross-situational contextual patterns in CDS. However, it remains unclear how such patterns develop over time and how children respond to these patterns in their own speech.

Previous studies have measured lexical and semantic learning in first language acquisition in various ways. For instance, Furrow et al. (1979) analyzed semantics by looking at indicators such as word classes (e.g. nouns, verbs) and utterance lengths, and linked their levels to semantic complexity. Likewise, Roy, Frank, and Roy (2009) employed the measures of speech complexity to indicate lexical learning, such as mean length of utterances for all utterances where a target word is situated. Others examined semantics by designing coding schemes with selected semantic features (Barnes, Gutfreund, Satterly, & Wells, 1983; Cross, 1979; Fenson et al., 1994; Gentner, 2006). Although semantic tuning could be implicated in these studies, they did not sufficiently attend to the subtleties of semantics in human language. While the verbs "raise" and "rise" can be discriminated with a scheme of transitivity, the discrimination between two transitive verbs, such as "hit" and "open", requires a more refined semantic taxonomy. The difficulty of categorization escalates when the vocabulary grows; worse, subjectivity of coding schemes becomes almost inevitable.

The raw distributional information surrounding words can serve as a less biased yet informative source for disentangling word meanings. As Riordan and Jones (2011) suggested, distributional models trained on CDS can capture the feature representations of human lexical knowledge. For instance, to distinguish the semantics between "raise" and "rise", instead of explicitly clarifying their transitivity, the number of arguments appearing in the contexts can already indicate their difference, as in "I raise the chair, so the chair rises". Further, for "hit" and "open", the typical words that co-occur with these verbs can differ. The object for "hit" can show a higher level of animacy, whereas "open" points to an explicit "opening" event of its object, as in "I open the window, so the window opens". The distinctive semantics of each word manifests itself after recurring cross-situationally, and the contextual patterns become crucial if the semantics is statistically inferred. In fact, it has been shown that distributional learning, namely learning from distributional information, is a fundamental mechanism for children in language understanding (Erickson & Thiessen, 2015; Naigles, 1996; Tomasello, 2003; Yuan, Fisher, & Snedeker, 2012). Children at the age of as young as several months are able to statistically infer the patterns from their surrounding speech (Aslin & Newport, 2014; Saffran, Aslin, & Newport, 1996). Particularly, in terms of verb learning, abundant evidence has been given regarding children's ability to recognize co-occurring words in an utterance, which could assist them with understanding verb actions (Fisher, Gertner, Scott, & Yuan, 2010; Fisher, Hall, Rakowitz, & Gleitman, 1994; Mintz, 2003; Moran et al., 2018). These studies, however, focused on limited patterns and did not involve the semantic subtleties that can be inferred for individual words from their respective rich contexts. Such comprehensive examination of distributional information has been realized by recent development of computational algorithms (Devlin, Chang, Lee, & Toutanova, 2019; Mikolov, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014). Besides, these algorithms provide further evidence for the reliability of distributional information in semantic inference. For example, the word embeddings algorithm is able to generate efficient semantic representations by simply relying on contextual words (Mikolov et al., 2013). Distributional information can thus be a reliable interface for accessing the subtleties of word meanings.

In this study, we focused on a specific semantic category, namely lexical causatives that express causal meaning, for which the processing of word meanings is particularly bound with distributional information. As exemplified above, words such as "raise" and "open" pertain to this category, where the verb describes a causal scene and the caused event is prominent ("the chair rises" for "raise" and "the window opens" for "open") (Comrie & Polinsky, 1993; Haspelmath, 1993; Shibatani, 2002). Lexical causatives lack formal marking that explicitly states the causative semantics. Therefore, the processing of lexical causatives chiefly relies on co-occurring words with their semantic nuances. Also, You et al. (You, Bickel, Daum, & Stoll, 2020) show that causatives can be differentiated from non-causatives with the word embeddings algorithm in CDS. In addition, to examine how causative semantics has been generalized via distributional learning, we employed complex network to connect similar words based on their semantics generated from the computational algorithm, namely word embeddings in our study. Complex network with distributional semantics has been shown to capture the connections between words and can therefore facilitate the investigation of semantics as a whole (Chen, 2020; Utsumi, 2015).

With a novel computational approach that combines distributional semantics with complex network, we attempted to comprehensively examine causative semantics in both CDS and CS, and thereby investigated how semantics is related between these two speech genres. We aimed to show the developmental trajectory of semantics in first language acquisition and answer whether adaptation occurs between CDS and CS at the semantic level.

2. Methods

2.1. Data

Our data comes from the Manchester corpus (Theakston et al., 2001), a longitudinal corpus consisting of both CS and CDS for 12 children in their interaction with the caregivers. Table 1 shows the age span of the recordings and the size of the speech in the corpus. The age span of each child is approximately 1 year, mostly in their third year. Particularly, we accumulated the speech by month, that is, the accumulated data set for each age of month included the data up to this time point, so that each data set recorded the complete traces of development.

Table 1. Age span and data size for each child in the Manchester corpus

			CDS		CS	
Child	Sex	Age span	Utterances	Words	Utterances	Words
1	F	1;10.7 - 2;9.10	37,554	152,940	20,701	48,831
2	M	1;11.12 - 2;10.28	36,155	210,191	17,577	47,919
3	F	2;0.7 - 2;11.15	27,988	111,304	24,399	58,489
4	M	1;8.22 - 2;8.15	23,026	96,594	25,567	68,914
5	M	1;10.25 - 2;10.16	37,971	145,981	21,914	46,949
6	F	1;11.27 - 2;11.12	29,272	11,8162	17,277	42,751
7	F	1;11.1 - 2;10.11	33,428	131,666	18,755	44,285
8	M	1;11.15 - 2;10.24	21,201	92,123	13,763	30,478
9	F	1;11.9 - 2;10.18	20,900	89,577	16,660	41,329
10	M	2;0.25 - 3;0.10	33,143	141,491	17,771	32,897
11	F	1;11.15 - 2;11.21	36,786	151,148	20,824	40,760
12	M	1;10.6 - 2;9.20	25,334	129,393	17,179	49,565

2.2. Word embeddings modeling

We employed the skip-gram word embeddings to generate semantic representations of words for both CDS and CS for each child at each age in month. The algorithm scans all the words in a corpus iteratively in the same manner. By using a neural network, the algorithm predicts the contextual words within a window from the current word, and the true contexts in the training texts are used to adjust the matrices in the network to ultimately render optimal predictions. In our training, we limited the window size to 3, namely 3 words on each side, since the mean length of utterances is generally low in both speech genres (4.33 in CDS and 2.38 in CS; see more details in Table 1). Figure 1 shows more details of the neural network. Each data set was iterated 100 times during training, and the dimension of the generated word vectors was set to 200 to capture sufficient semantic nuances (Lai, Liu, He, & Zhao, 2016).

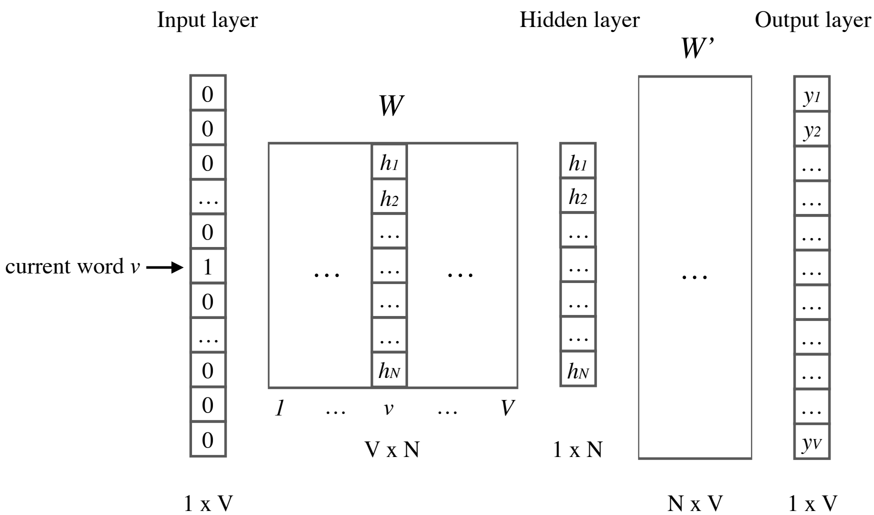

Figure 1. The network of skip-gram word embeddings. W denotes weight matrix that contains the information of vectors, and W' stands for the weight matrix that computes the relations between words and the contexts. The output layer is later transformed into a vector of probabilities that determine the predicted contexts.

2.3. Causative semantics in complex network

For each word embeddings model, we constructed an unweighted and undirected graph to represent the causative semantics by following three major steps. First, we started our search of pivotal causative semantics within a set of prototypical causatives that commonly exist across languages (Haspelmath, 1993).

- Causatives (23): *begin, boil, break, burn, change, close, destroy, dry, fill, finish, freeze, gather, kill, lose, melt, open, raise, roll, sink, spread, stop, teach, turn*

These words marked the central semantics in the network, and were placed as the vertices. Second, we extended our search to the *n* most similar words of each prototypical causative in the word embeddings model, and set these words as the vertices in the network as well. Here, *n* was determined by a fixed ratio 0.01 to the vocabulary size, so as to control for the vocabulary differences between data sets. Meanwhile, we linked these similar pairs of words as the edges in the network. As the last step, we removed the leaf vertices, namely the ones with degree 1 in the network, as they did not bridge any words in the network and hence were of little relevance regarding the generalization of causative semantics. This constructed graph was then intended to represent the causative network with the most relevant semantics.

In addition to the networks built for the Manchester corpus, we set up a baseline network of adult-directed speech (ADS) to indicate adults' normal level of semantic generalization, so as to render further comparison between speech genres. We conducted the same steps for a subset of the spoken corpus (326,359 utterances) in the British National Corpus (*British National Corpus, version 3 (BNC XML Edition)*, 2007) to obtain this baseline.

To assess the generalization of causative semantics in each network, we calculated the average degree of vertices (average node degree) in the network as the metric. This represents the connectedness of a network and can therefore indicate how well causative semantics is connected.

2.4. Analyses

We conducted two main analyses. In the first analysis, we ran local regression to observe the dynamics of causative semantics in both CDS and CS in comparison with the level of ADS. In the second analysis, in order to discover how CDS changes in relation to CS, we built a hierarchical Bayesian regression model for the difference between CDS and CS on age, with children as the random factor.

3. Results

3.1. Analysis 1

As shown in Figure 2, there is generally an uptrend for the development of causative semantics in both CDS and CS.

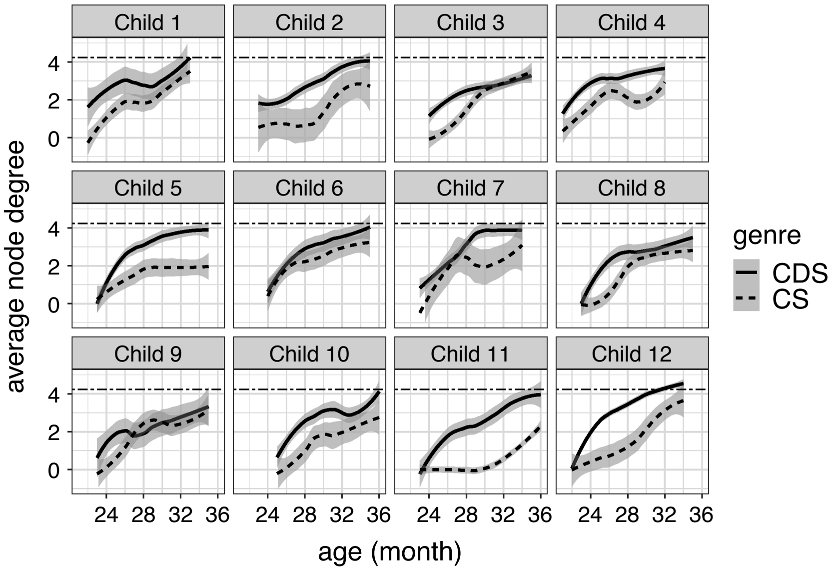

Figure 2. Local regression (span = 0.75) for the average node degree of causative network of both CDS and CS for all 12 children in the Manchester corpus. The dash-dotted line is the performance for ADS in the British National Corpus).

The average node degree in CDS, unsurprisingly, mostly exceeds that in CS. Nonetheless, by the last recording at the age of around 3 years, all 12 children have developed a causative network with a moderate average node degree, ranging from 2.00 to 3.54. Figure 3 shows an example of the development of causative network in CS for Child 1. It is clearly displayed that connections between causative-related words quickly grow, and prototypical causatives are substantially bridged in the last recordings.

As compared to the level in ADS, the semantic connectedness in CDS starts from a much lower level at an early age, and slowly increases over time before eventually approaching the level of ADS in some cases. In addition, the corresponding level in CS has not yet overcome the deficit with ADS.

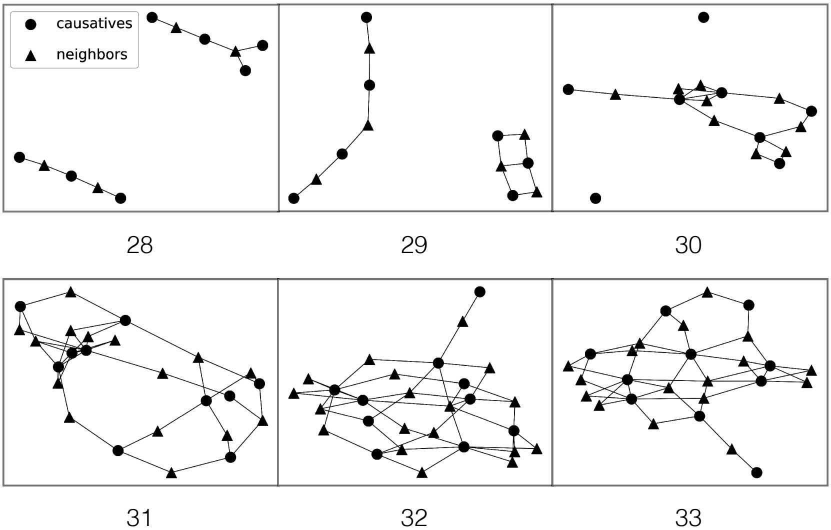

Figure 3. Visualization of the development of the causative network in CS: Child 1 from 28 to 33 months. The circles (causatives) denote the prototypical causatives, while the triangles (neighbors) stand for the extended similar words from the prototypes.

3.2. Analysis 2

We regressed the difference of average node degree between CDS and CS on age. The model shows that age is not a significant predictor for the gap between CDS and CS [95% quantile-based interval of the coefficient: (-0.058, 0.046)]. Predicted values of the model also suggest negligible effect by age (see Figure 4).

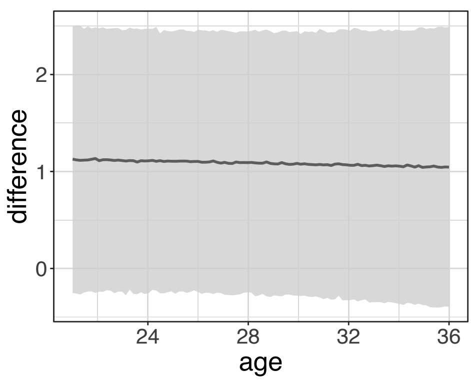

Figure 4. Posterior predicted values with 95% quantile-based interval for the predictor "age" in the Bayesian regression model

4. Discussion

The results of both analyses suggest an important role of CDS in children's semantic development in their own speech. First of all, CDS does not stay invariant over time, but rather exhibits dynamics, with a general uptrend in terms of causative semantic generalization. This suggests an active role of adults in their language interaction with children. Secondly, while the same uptrend is detected in CS, the gap between CDS and CS does not show a significant trend of either enlarging or shrinking, indicating an adaptive manner between the two genres. That is, adults might monitor the development of children's semantic understanding and thus increase the semantic complexity in their speech; meanwhile, children benefit from the tuning of CDS and the reduced complexity in CDS compared to ADS, thus sustaining their pace of development and prompting adults to continuously increase semantic complexity.

It should also be noted how our approach is able to exploit distributional semantics in an unbiased way and showcases the complex semantic network with intricate semantic representations. On the one hand, visualization as in Figure 3 can facilitate the qualitative diagnosis of children's semantic development. On the other hand, from a quantitative point of view, graph theory offers a great number of options for analyzing the characteristics in a network, such as distance and centrality measures. This allows for investigations into semantics both holistically and with a narrower scope.

Lastly, while a general uptrend is observed, there is indeed stage-like behavior along the trajectory, especially some periods of stagnation in the middle of the development (e.g. Child 1 and Child 10 in Figure 2). It is therefore potentially beneficial to examine these dynamics in segments within each speech genre. This can also help identify the break points of substantial changes, so as to shed light on the directionality of the adaption of semantic development between CDS and CS.

Our general conclusion is that CDS and CS are closely related to each other in their distributional semantics, and this coupling effect could potentially facilitate children's semantic learning at an early age. The examination of semantics in our study is less biased with the help of the computational approach, where connections between words are established based on their distributional information. We thus avoid the subjectivity of pattern and feature selection in investigating semantics. Further studies should capitalize more on the semantic network generated from distributional information, and therewith discover the intricacy and the directionality of the adaptation between CDS and CS in first language acquisition.

References

Aslin, Richard N, & Newport, Elissa L. (2014). Distributional language learning: Mechanisms and models of category formation. *Language Learning*, *64*(s2), 86–105.

Barnes, Sally, Gutfreund, Mary, Satterly, David, & Wells, Gordon. (1983). Characteristics of adult speech which predict children's language development. *Journal of child language*, *10*(1), 65–84.

British national corpus, version 3 (bnc xml edition). (2007). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Retrieved from `http://www.natcorp.ox.ac.uk/`

Chen, Alvin Cheng-Hsien. (2020). Words, constructions and corpora: Network representations of constructional semantics for mandarin space particles. *Corpus Linguistics and Linguistic Theory*, *1*(ahead-of-print).

Comrie, Bernard, & Polinsky, Maria. (1993). *Causatives and transitivity* (Vol. 23). John Benjamins Publishing.

Cross, Toni G. (1979). Mothers' speech adjustments and child language learning: Some methodological considerations. *Language Sciences*, *1*(1), 3 - 25.

Dale, Rick, & Spivey, Michael J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, *56*(3), 391–430.

Devlin, Jacob, Chang, Ming-Wei, Lee, Kenton, & Toutanova, Kristina. (2019). *Bert: Pre-training of deep bidirectional transformers for language understanding.*

Erickson, Lucy C, & Thiessen, Erik D. (2015). Statistical learning of language: theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108.

Fenson, Larry, Dale, Philip S, Reznick, J Steven, Bates, Elizabeth, Thal, Donna J, Pethick, Stephen J, … Stiles, Joan (1994). Variability in early communicative development. *Monographs of the society for research in child development*, i–185.

Fisher, Cynthia, Gertner, Yael, Scott, Rose M, & Yuan, Sylvia. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(2), 143–149.

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

Furrow, David, Nelson, Katherine, & Benedict, Helen. (1979). Mothers' speech to children and syntactic development: Some simple relationships. *Journal of Child Language*, *6*(03), 423–442.

Gentner, Dedre. (2006). Why verbs are hard to learn. In Kathryn Hirsh-Pasek & Roberta M. Golinkoff (Eds.), *Action meets words: how children learn verbs* (p. 544-564). Oxford: Oxford University Press.

Haspelmath, Martin. (1993). More on the typology of inchoative/causative verb alternations. In Bernard Comrie & Maria Polinsky (Eds.), *Causatives and transitivity* (Vol. 23, pp. 87–121). John Benjamins.

Kuhl, Patricia K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843.

Lai, Siwei, Liu, Kang, He, Shizhu, & Zhao, Jun. (2016). How to generate a good word embedding. *IEEE Intelligent Systems*, *31*(6), 5–14.

Mikolov, Tomas, Chen, Kai, Corrado, Greg, & Dean, Jeffrey. (2013). Efficient estimation of word representations in vector space. In *Proceedings of iclr workshops track.*

Mintz, Toben H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117.

Moran, Steven, Blasi, Damián E, Schikowski, Robert, Küntay, Aylin C, Pfeiler, Barbara, Allen, Shanley, & Stoll, Sabine. (2018). A universal cue for grammatical categories in the input to children: Frequent frames. *Cognition*, *175*, 131–140.

Naigles, Letitia R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, *58*(2), 221–251.

Pennington, Jeffrey, Socher, Richard, & Manning, Christopher D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1532–1543).

Riordan, Brian, & Jones, Michael N. (2011). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, *3*(2), 303–345.

Roy, Brandon C., Frank, Michael C., & Roy, Deb. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society, CogSci 2009*.

Saffran, Jenny R., Aslin, Richard N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926 - 1928.

Shibatani, Masayoshi. (2002). *The grammar of causation and interpersonal manipulation* (Vol. 48). John Benjamins Publishing.

Snow, Catherine E. (2017). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. *The handbook of child language*, 179–193.

Soderstrom, Melanie. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

Sokolov, Jeffrey L. (1993). A local contingency analysis of the fine-tuning hypothesis. *Developmental psychology*, *29*(6), 1008.

Theakston, Anna L., Lieven, Elena V. M., Pine, Julian M., & Rowland, Caroline F. (2001). The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language*, *28*, 127-152.

Tomasello, Michael. (2003). *Constructing a language: a usage-based theory of language acquisition*. Harvard, MA: Harvard University Press.

Utsumi, Akira. (2015). A complex network approach to distributional semantic models. *PloS one*, *10*(8), e0136277.

You, Guanghao, Bickel, Balthasar, Daum, Moritz M, & Stoll, Sabine. (2020, Aug). *Child-directed speech is statistically optimized for meaning extraction.* PsyArXiv. Retrieved from `psyarxiv.com/zmktd` doi: 10.31234/osf.io/zmktd

Yuan, Sylvia, Fisher, Cynthia, & Snedeker, Jesse. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*(4), 1382–1399.

Proceedings of the 45th annual Boston University Conference on Language Development

edited by Danielle Dionne
and Lee-Ann Vidal Covas

Cascadilla Press Somerville, MA 2021

Copyright information

ISSN 1080-692X
ISBN 978-1-57473-067-8 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com