Exploring Patterns of Stability and Change in Caregivers' Word Usage across Early Childhood

Hang Jiang[1], Michael C. Frank[2], Vivek Kulkarni[3], and Abdellah Fourtassi[4]

[1]Symbolic Systems Program, Stanford University, USA

[2]Department of Psychology, Stanford University, USA

[3]Department of Computer Science, Stanford University, USA

[4]Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France

Author Note

Abdellah Fourtassi

Laboratoire Informatique et Systèmes

Aix-Marseille University, France

Correspondence concerning this article should be addressed to Abdellah Fourtassi, LIS UMR 7020 CNRS / AMU / UTLN, Campus universitaire de Luminy. Bat. TPR2, 5ème étage, Bloc 1. 163 avenue de Luminy, 13288 MARSEILLE cedex 09. E-mail: abdellah.fourtassi@gmail.com

Abstract

The linguistic input children receive across early childhood plays a crucial role in shaping their knowledge about the world. To study this input, researchers have begun applying distributional semantic models to large corpora of child-directed speech, extracting various patterns of word use/co-occurrence. Previous work using these models has not measured how these patterns may change throughout development, however. In this work, we leverage NLP methods – originally developed to study historical language change – to compare caregivers' use of words when talking to younger vs. older children. Some words' usage changed more than others'; this variability could be predicted based on the word's properties at both the individual and category level. These findings suggest that caregivers' changing patterns of word use may play a role in scaffolding children's acquisition of conceptual structure in early development.

Keywords: Conceptual learning; meaning change; child-directed speech; language and cognition; word embeddings; distributional semantics

Exploring Patterns of Stability and Change in Caregivers' Word Usage across Early Childhood

Children learn from direct observation and interaction with the world around them (Piaget, 1954), but much of the knowledge they acquire comes also from the language they hear. Language provides children with information beyond what they can obtain through observation alone. A large body of work has established that language spoken to children can help them establish category boundaries, can teach them about causal structure, and can even provide the framework that guides their subsequent learning (Carey, 2009; Gelman, 2009; P. L. Harris, 2012; Xu, 2019).

Most previous work, however, treats language as a static source of information, rather than as a dynamic input source that reflects the child's own competence. Here we are interested in understanding whether and how information in Child-Directed Speech (CDS) changes as children develop. The study of changes in parents' talk is an important first step to understand how parents' presentation of knowledge is adapted — whether consciously or unconsciously — to children's developmental context, potentially optimizing the learning process (Vygotsky, 1978).

Previous research have shown that the overall quantity and complexity of CDS (both at the lexical- and sentence- level) change with development. This change correlates with the child's own linguistic development, suggesting a form of adaptation between parents and children (Newport et al., 1977; Huttenlocher et al., 2007; Huttenlocher et al., 2010; Kunert et al., 2011; Bergelson et al., 2019; Leung et al., 2021). The child's stage of motor development (e.g., crawling vs. walking) also elicits different kinds of caregiver speech. For example, Karasik et al. (2011) have shown that mothers of walkers provided more action directives to their children than mothers of crawlers.

Additional evidence of caregivers adapting their input to children's linguistic abilities come from recent work focusing on turn-by-turn analysis of caregiver-child conversations. These studies have shown that parents tend to align to children's production (more often than the other way around) at various linguistic levels (Spivey & Dale, 2006; Fernández & Grimm, 2014; Yurovsky et al., 2016; Misiek et al., 2020;

Fusaroli et al., 2021). This research points towards a rather "local" adaptation to the child production over a small time scale (e.g., a conversation). Though this kind of local adaptation is not the focus of the current study, it indicates that caregivers tend to use words and structures that are known to children. Local adaptation thus suggests – albeit in an indirect fashion – that caregivers' overall input properties undergo "global" changes as well, reflecting children's linguistic improvement over a developmental time scale.

In sum, a substantial body of work focuses on changes in caregivers' linguistic input, but this work has largely focused on either caregivers' choice of words or the complexity of their sentences. In contrast, our current work deals with the question of *semantic* word change which has received relatively less attention even though it is important to our understanding of both children's language acquisition and their conceptual development. In the remainder of our introduction, we introduce the computational approach that we take to represent word meaning and then discuss how we apply it to predicting meaning change over time in caregiver talk to children.

Distributional semantics and language learning

Here we focus on one aspect of semantics that emerges from the patterns of word use in language (Wittgenstein, 1953). More precisely, one can characterize the meaning of a word by the "company it keeps", that is, by the words it co-occurs with (Firth, 1935; Z. S. Harris, 1954). An important product of this characterization is what has come to be called the "distributional hypothesis." According to this hypothesis, words that share similar patterns of co-occurrence (or distribution) in speech tend to have similar meanings. For example, one can posit that the words "apple" and "banana" must refer to objects that share semantic properties because people tend to talk about both words in similar context using similar co-occurring words such as "eat", "kitchen", "dessert," and "tree."

The distributional hypothesis has been at the heart of models of semantic analysis of text in the field of Natural Language Processing. Distributional analysis has been implemented computationally in various ways, e.g., through counting occurrences in various contexts or, more recently, through optimizing the prediction of the context (see Baroni et al. (2014) for more details). Examples of influential distributional models include LSA (Dumais, 2004), HAL (Lund & Burgess, 1996), COALS (Rohde et al., 2006), BEAGLE (Jones et al., 2006), PMI (Recchia & Jones, 2009), topic models (Griffiths et al., 2007), GloVe (Pennington et al., 2014), and Word2vec (Mikolov, Sutskever, et al., 2013). Such models have been widely used in cognitive science (Günther et al., 2019) and have proven to be surprisingly good at predicting human similarity judgments (Landauer & Dumais, 1997; Mikolov, Sutskever, et al., 2013; Baroni et al., 2014; De Deyne et al., 2016; Mandera et al., 2017).

In our current work, we use Word2Vec, a prediction-based instantiation of the distributional hypothesis (Mikolov, Chen, et al., 2013). Word2Vec is a neural network model that maximizes the likelihood of predicting the linguistic context given a word (or predicting a word given the context). For each prediction made, the model derives an error signal obtained by comparing the predicted vs. observed context. The error signal is then backpropagated through the neural network, improving the ability of future predictions. The trained model outputs a high dimensional semantic space where words are represented as continuous vectors (or "embeddings"). Words that predict (or predicted by) similar linguistic context will end up having similar vectors.

Given the incremental and error-driven nature of its learning algorithm, word2vec might be thought of as slightly closer to human learning mechanisms than, for example, complex geometric transformations like singular value decomposition in Latent Semantic Analysis, (Landauer & Dumais, 1997). Further, word2vec also provides an excellent fit with human similarity judgments (Mikolov, Chen, et al., 2013; Baroni et al., 2014; Mandera et al., 2017). But overall, we choose word2vec in part because of its relatively strong performance and in part due to computational convenience – we do not expect that our results would be radically different using another model.

Distributional models have been applied fruitfully to language acquisition in a wide variety of studies. For example, distributional co-occurrence can facilitate the learning of several aspects of language from word forms to lexical-semantic organisation and word meanings (Andrews et al., 2009; Fourtassi et al., 2014; Frermann & Lapata, 2016; Huebner & Willits, 2018; Unger & Fisher, 2021). In addition, word co-occurrence distributions in the linguistic environment can influence learners' acquisition of new words as well as how these words are organized into a broader lexical-semantic network (Hills et al., 2010; Stella et al., 2017; Fourtassi, 2020; Unger et al., 2020).

The Current Study

Though distributional semantic models have enjoyed substantial uptake in the study of CDS, to our knowledge no previous work has explored whether and how patterns of co-occurrence distribution in caregiver talk change over the course of children's development. Yet there are some intuitive reasons to suspect that, as children develop, caregivers change the way they talk.

First, caregivers might change how they talk simply because they want children to understand. They might not use combinations of words that are too complex for younger children. Second, the contexts of interaction for older children are different than for younger children and so words might be used differently. While for a baby, food might be associated with *high chair* and *bib*, an older child might have a stronger association with *restaurant*. Finally, the books or media that caregivers select to read and watch might contain different contexts and content for younger and older children.

Given that contextual distributions of words likely change developmentally, it is important to understand how these changes lead to differences in words' meaning that children learn across development. The current work aims at filling this gap in the literature.

More precisely, we ask the following questions. Do caregivers use words with different linguistic contexts when talking to younger and older children? If caregivers do use words in different (or more diverse) contexts across development, how does this difference in use give rise to differences in the distributional meaning of words? Does this change in meaning affect all words or do some words' meaning change more than others? How can we explain the patterns of stability and change, and how could these patterns support conceptual development? In the remainder of this introduction, we foreshadow the research methods we will leverage to answer these questions.

In the majority of existing work studying CDS with distributional semantics, models are typically trained on a single corpus, extracting "average" word embeddings by aggregating across all available ages to make the best use of the extant data (Hills et al., 2010; Huebner & Willits, 2018; Fourtassi, 2020). Further, if word vectors are induced from different corpora (e.g., representing caregiver linguistic input at different ages), up until recently there have been only limited methods for comparing these distinct sets of embeddings.

Here we make use of innovations in distributional models — originally introduced to study historical change in word use (Hamilton et al., 2016b; Rudolph & Blei, 2017) — to compare word embeddings based on language directed to older and younger children. In particular, we construct different embeddings at different points in development and compare the geometry of these spaces to understand whether caregivers use words in different linguistic contexts (i.e., with different meanings) when they are talking to younger vs. older children.

The paper is organized as follows. First, we introduce our dataset and explain the modeling procedure that allows us to compare word embeddings based on language directed to younger and older children. Then, we present a set of analyses exploring the patterns of stability and change in word meanings, controlling for various confounding sources of change. We study the extent to which these patterns can be predicted using both word- and category-level properties. Finally, we discuss if some of these patterns could be used by children to scaffold the acquisition of some difficult word meanings, especially words that belong to the category of number, color, and time.

Methodology

Data

Corpus. We constructed a corpus by aggregating over all English-language CDS transcripts, including 513 children from 39 corpora in the US and 149 children

Epoch	Number of tokens	Number of utterances	Age range (months)
1	1,967,578	496,013	3-24
2	1,864,348	425,615	25-28
3	1,983,094	447,645	29-32
4	1,924,849	429,827	33-36
5	1,976,926	404,729	37-50
6	1,958,298	397,511	51-144

Table 1

Statistics of the corpus per epoch. In our study, we focus on comparing epoch 1 and 6.

from 10 corpora in the UK from CHILDES (MacWhinney, 2014) keeping only utterances by adults. The words in the corpus were lemmatized.[1]

For our developmental analysis, we divided the corpus into a set of time epochs. To avoid biasing our developmental analyses due to possible data sparsity, we chose our epochs to be equally spaced with respect to the number of tokens produced by caregivers. Out of a corpus of about 12 million tokens in total, we constructed six epochs, each with around 2 million tokens. Preliminary analyses showed that finer-grained epochs (e.g., 1 million tokens or less) led to noisier representations. The statistics of these epochs are in Table 1.[2]

Target Words. Since analytic resources were available for these words, we primarily limited our analysis to the vocabulary contained in the MacArthur-Bates Communicative Development Inventory (CDI), a commonly-used instrument for measuring children's early vocabulary (Fenson et al., 1994). This vocabulary provides a consensus list of relevant early words as well as a categorization into syntactic/semantic categories that are hypothesized to be relevant to children's language use (e.g., "action words" or "places").[3] Aside from the categories present in CDI, we also included words

[1] Data were downloaded on June 19th, 2020.

[2] Epoch 1 covers a very large age range but a relatively small portion of the data (10%) occurs before 12 months. The vast majority of the data are from 12–24 months, leading us to characterize this epoch informally as "toddler language."

[3] CDI categories are not hypothesized to be true mental categories – merely a guide for grouping trends in a way that might make them more interpretable than the use of broader categories derived

from CHILDES related to the categories of emotions, colors, and numbers. The final list included 824 words: 665 words from CDI, 124 emotion words (Ridgeway et al., 1985), 7 color words, and 28 number words.

Procedure

Following the method outlined by Hamilton et al. (2016b), we trained Word2Vec models on each epoch of CHILDES corpus, aligned the vector spaces, and computed a measure of change between vectors from different epochs. We elaborate on each of these steps below.[4]

Constructing Semantic Spaces. We used the Skip-gram with Negative Sampling (SGNS) version of Word2vec (Mikolov, Sutskever, et al., 2013) to learn word embeddings. We used the Gensim implementation.[5] Following Hamilton et al. (2016b), we trained the models for 50 iterations on the utterances from each period individually and used the following hyperparameters: 100 dimensions, window size 5, and negative sample size 5.

Aligning the Semantic Spaces. The stochastic nature of the word2vec training process produces, for each epoch, a semantic space that has a unique coordinate system. To be able to study how the same words behave across different epochs (i.e., across different semantic spaces), we need to align these spaces so that they all have a unified coordinate system. A standard way to do such an alignment is called the orthogonal Procrustes method (Hamilton et al., 2016a, 2016b). Using this method, we aligned all vector spaces to the vector space obtained with the first epoch.

Measuring Meaning Change. We next computed a measure that allowed us to compare word embeddings across epochs. In Hamilton et al. (2016a), the authors derived both a “global” measure that quantified global shift in a word’s embedding across epochs and a “local” measure that focused on local changes across epochs of a

from adult syntax.

[4] All code necessary to reproduce our analyses can be found here: https://osf.io/98jfh/?view_only=cd1f2e0f4a774ec0b7b300a7094be0ec

[5] https://radimrehurek.com/gensim/models/word2vec.html

word's nearest semantic neighbors. They found the global measure to be a good predictor of regular processes of linguistic drift, e.g., the fact that the word "must" shifted from an obligation usage ("you must do X") to an epistemic one ("X must be such and such"). The local measure, in contrast, was a better predictor of change due to historical and cultural developments, e.g., the change in the meaning/use of the words "virus" or the word "gay." In our work, we aim to study change in caregivers' speech that reflect changes in the child's developmental context, a phenomenon that is more similar to historical/cultural change than to linguistic drift. Thus, we adopted the local measure (see Hamilton et al. (2016a) for more technical details) and make use of this measure in all subsequent analyses of change.

Controls

Our main goal was to measure changes in caregivers' word usage across development. However, observed changes between the input to younger and older children might not only be due to genuine developmental changes. They could also be caused by several other factors such as sampling noise, the slightly different contexts in which the corpora were recorded, the gender of the child, and/or the social role of the caregiver. Below, we explain the comparisons that allow us to quantify – and hence control for – these additional sources of variation.

Sampling Noise. Previous work using similar computational methods such as Kulkarni et al. (2015) and Dubossarsky et al. (2017) demonstrated that part of the observed changes in embeddings across epochs can be due to sampling variation. That is, even when using two similar corpora (e.g., from the same epoch), vectors for the same words can be slightly different, leading to the spurious impression of changes in meaning. Thus, following Dubossarsky et al. (2017), we systematically compare changes across developmental epochs to changes across size-equivalent "epochs" from a randomly shuffled version of the corpus where utterances in each "epoch" were randomly sampled across all real epochs. This method removes all temporal structure from the data – in each shuffled epoch there is a mixture of utterances from different developmental epochs.

Context of Adult-child Interaction. As we mentioned earlier, we used all English data in CHILDES. These corpora have not all been collected in the same interaction context. Even though most are recordings during home visits, the specific activities being performed and the set of participants that were present likely differed considerably. These differences could lead to seeming developmental changes in word usage that are in fact driven by random differences in corpus dynamics rather than true developmental change (which can of course lead to real changes in context based on the activities that older vs. younger children can engage in). To control for this potential source of variation, we compared changes that result from using different corpora (within the same epoch) to changes across developmental epochs, thus quantifying the effect of corpus variation.

Child's Gender. The CHILDES aggregated corpus we use in this study is roughly equally balanced for female and male children (47.1% versus 52.9%). To evaluate how much change is due to the child's gender, we compared changes across male and female children within the same epoch to changes across developmental epochs.

Adult's Identity. Finally, though the majority of adult speakers in CHILDES are the children's mothers, we can also find other speakers such as the father, sibling, or investigator (i.e., the researcher recording the interaction). The distribution is shown in Appendix A. We evaluate changes due to adult speakers to changes across developmental epochs. Since the overwhelming majority of speakers are children's mothers, we restricted the analysis to the binary distinction of Mother vs. non-Mother.

Analyses

We separate our analyses into three parts: The first focuses on describing patterns of stability and change, the second focuses on testing the predictors of change, and the third on how this change can scaffold children's conceptual development.

We start the first set of analyses with a qualitative analysis of some of the words that underwent the most and the least change across the first and the last epochs. Then we present a more quantitative analysis of change for all words across syntactic and semantic categories, showing that this change occurred beyond sampling noise. Next, we test if developmental/diachronic change occurred over and above a few synchronic (i.e., occurring within the same epoch) sources of change such as the adult-child interaction context, the child's gender, and the adult's identity. Finally, though we did most of the analysis by contrasting the first and last epochs, we also test whether change actually occurred continuously across all six epochs.

In the second set of analyses, we test and compare predictors of change including both word-level predictors (word frequency and polysemy) and category-level predictors (the density and centrality of the category to which the word belongs).

The analyses we conducted are exploratory in nature. We did not have specific a priori predictions about the size or direction of the effects we measured.

Qualitative Analysis

Table 2 shows a selection of words that changed the most and least across the first and last epochs, using our measure of change. Take for example the word "quack." This is one of the words that changed the most. Investigation of a few contexts of occurrence across the first and last epochs suggests that this word is initially used predominantly by adults to imitate a duck sound (e.g., "duck duck duck"). Later, adults start using the word as a verb with diverse meanings (e.g. "quack the duck", "quack the march behind her").

Similarly, some verbs that changed the most (e.g. "shoot", "mash", "skip", "suck", "pee") as well as some animals (e.g., "lamb" and "turkey") were used with a single meaning at the early stages of development and they later gained polysemous traits. To illustrate, the word "turkey" can uttered in a food-related context with words such as "meat" and "dinner" early in development. It then gets enriched as caregivers start using it also to talk about "Turkey" in an animal-related context and, thus, becomes also associated with other animal words like "deer" and "cow" and not just with food items.

The words that changed the least related to time ("o'clock") and number (e.g. "twelve", "seven") and also included pronouns ("you") and function words ("because", "if", "can"), suggesting that these words were used in more stable linguistic contexts over time.

top 25	everyday, steady, quack, found, base, choice, teddy, shoot, pat, mash, worm, everywhere
greatest change	angry, chimney, each_other, lamb, skip, magic, patty, bee, alligator, suck, clown, low, pee
top 25	eight, ten, twenty, eleven, pardon, because, five, nine, four, you, think, thirteen
least change	it, can, if, not, seventeen, quarter, three, seven, twelve, do, nineteen, o'clock

Table 2

A selection of top words with the greatest and least change across the first and last epochs in the dataset.

Developmental Change beyond Sampling Noise

Figure 1 shows average changes across syntactic categories. As indicated above, change is computed by comparing the first and last epochs. There was more overall change in the real corpus than there was in the control, time-shuffled corpus.

Further, we found a difference between words belonging to different syntactic categories: function words changed less than content words. While this finding was observed in both the real and shuffled corpora, it was substantially larger in the real corpus, suggesting that at least part of this effect can be attributed to a genuine change in word usage across time. Statistical modeling confirmed this observation: A linear model predicting word change as a function of syntactic category and corpus condition showed a significant interaction between these two predictors. The estimates are shown in Table 3. The simple effect of syntactic category within the real corpus alone was $\beta = 0.80$ ($SE = 0.12$, $p < 0.001$).

Figure 2 shows average changes in word embedding by functional/semantic (CDI) category, providing a finer-grained view than syntactic category. For example, within the syntactic category of function words, quantifiers underwent more change than connecting words. Within the category of nouns, animals and people changed

Table 3

Estimates of a linear model predicting change by syntactic class and condition (real vs. shuffled corpus). The model compares the categories with largest and smallest values of change (i.e., nouns vs. function words)

	Estimate	*Standard Error*	*Significance Level*
(Intercept)	0.433	0.084	***
Syntactic_category	1.271	0.100	***
Condition	−0.842	0.092	***
Syntactic_category:Condition	1.047	0.109	***
$R^2 = 0.39$			

Note: *p<0.05; **p<0.01; ***p<0.001

more than food and clothes. These patterns of change differed from those observed in the time-shuffled condition, though some of the same trends were present to a lesser degree. Thus, at least part of the variability between semantic categories can be attributed to genuine differences in the way the words in these categories were used differently at different developmental epochs. Indeed, a linear model that predicted change as a function of semantic category and corpus condition showed a significant interaction term. The estimates are shown in Table 4. The simple effect of semantic category within the real corpus alone was $\beta = -1.70$ ($SE = 0.16$, $p < 0.001$).

Comparing Diachronic and Synchronic Change

We next address the possibility that the results obtained above could be due not to genuine diachronic (i.e., developmental) changes but rather to synchronous (i.e., non-developmental) factors whose distribution may still differ across developmental epochs, causing the appearance of a developmental change. For example, maybe corpora with data from younger children in CHILDES happen to be from male children, while female children are better represented in older corpora. Although this particular account is unlikely, differences caused by confounding synchronous factors are a real

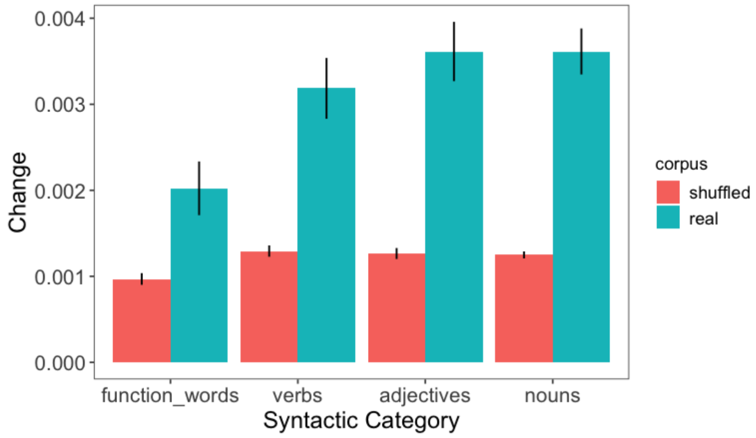

Figure 1. Word changes averaged by syntactic category (real vs. time-shuffled corpus). Error bars indicate 95% confidence intervals across words in each category.

Table 4

Estimates of a linear model predicting change by semantic class and condition (real vs. shuffled corpus). The model compares the categories with largest and smallest values of change (i.e., animals vs. number)

	Estimate	*Standard Error*	*Significance Level*
(Intercept)	2.504	0.081	***
Semantic_category	−2.955	0.134	***
Condition	−2.453	0.088	***
Semantic_category:Condition	2.205	0.145	***
$R^2 = 0.76$			
Note:	*p<0.05; **p<0.01; ***p<0.001		

possibility given that the aggregated CHILDES corpus we use is largely cross-sectional and quite heterogeneous.

To address this issue, we test synchronic changes due to the adult-child interaction context (i.e., the corpus), the child's gender, and the adult's identity. In each of these cases, we compared change within epochs to change across epochs as follows. First, we split epochs into two synchronous parts by 1) the original corpus

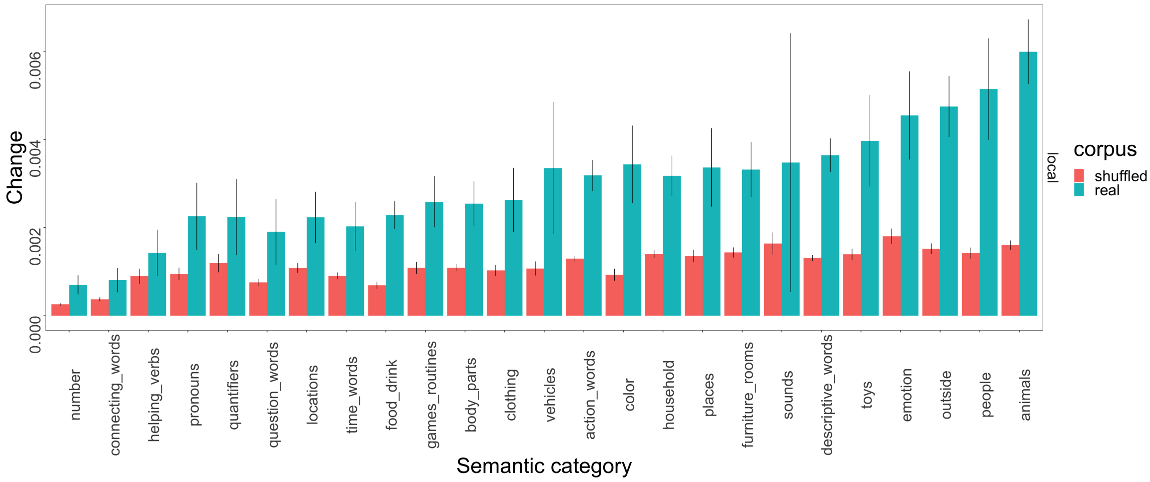

Figure 2. Word changes averaged by semantic category (real vs. time-shuffled corpus). The error bars indicate 95% confidence intervals.

(averaged across five different random splits), 2) the child's gender (female vs. male), and 3) speaker's identity (mother vs. non-mother). Second, in each case, we computed synchronous change by comparing embeddings between the two synchronous parts of a given epoch (averaging across epochs). Finally, we computed diachronic change by comparing one part from the first epoch to another part from the last epochs (averaging across all combinations).

Figure 3 shows the results comparing synchronic to diachronic changes. In all cases, diachronic change was larger in magnitude than synchronic change. While Figure 3 only shows average scores, we found the same pattern at a finer-grained level across syntactic and semantic categories (results not shown). These results demonstrate that observed developmental changes cannot be explained away by synchronous factors projected in time and that – at least part of – this development is due to changes in the way adults use words when talking to younger vs. older children.

Comparing Diachronic Changes across Epochs

While all analyses above were done by contrasting the first and last epochs, we also investigated the nature of change across intermediate epochs. In Figure 4, we show average changes obtained by taking epoch 1 as a reference and contrasting it with epoch 2 to 6 (for both the real and time-shuffled corpus). We observed a rather continuous and

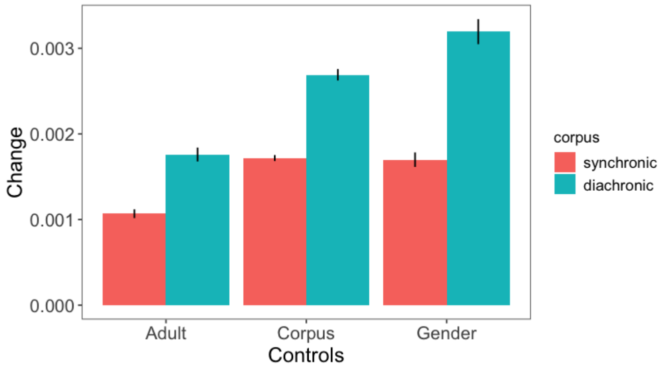

Figure 3. Word average changes (diachronic vs. synchronic corpora). Error bars indicate 95% confidence intervals.

monotonic pattern of change across developmental time. Here again, we also observed more overall change in the real corpus than in the control, time-shuffled corpora.

Predictors of Change

Previous work using similar methods has focused on studying the properties of words that predict their change. For example, Hamilton et al. (2016b) proposed two laws of change whereby words that undergo most change tend to be low-frequency and high-polysemy (but see Dubossarsky et al. (2017)). Following this work, we test the extent to which these same laws – that have been hypothesized to operate over different historical periods – can also explain the change that occurs in caregiver talk to children across development. In addition, we explore whether the way words relate to one another between and across categories also predicts their level of change across development.

Word-level Predictors. We used the natural logarithm of word frequency and polysemy as our primary word-level predictors. Frequency was computed as simple token frequency across the entire corpus. As a measure of polysemy, we counted the number of different senses each word had in WordNet (http://wordnet.princeton.edu), a resource devoted to cataloging word senses. Figure 5 shows word change as a function of frequency (left) and as a function of polysemy (right). For better comparison, the values for predictors and their corresponding values of change were centered and scaled.

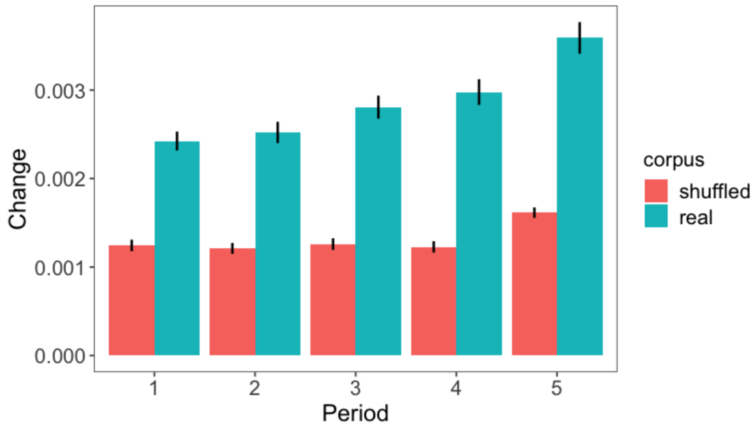

Figure 4. Word changes across 6 epochs (real vs. time-shuffled corpus). Each timestep t corresponds to the change between epoch $t+1$ and 1. The error bars indicate 95% confidence intervals.

In Appendix C, we show these patterns within each syntactic category.

Overall, less frequent words changed more (and more in the real data than the time-shuffled data). In contrast, polysemy (as operationalized by the Wordnet sense count) was not related to meaning change in either corpus. We provide statistical analyses quantifying these observations below.

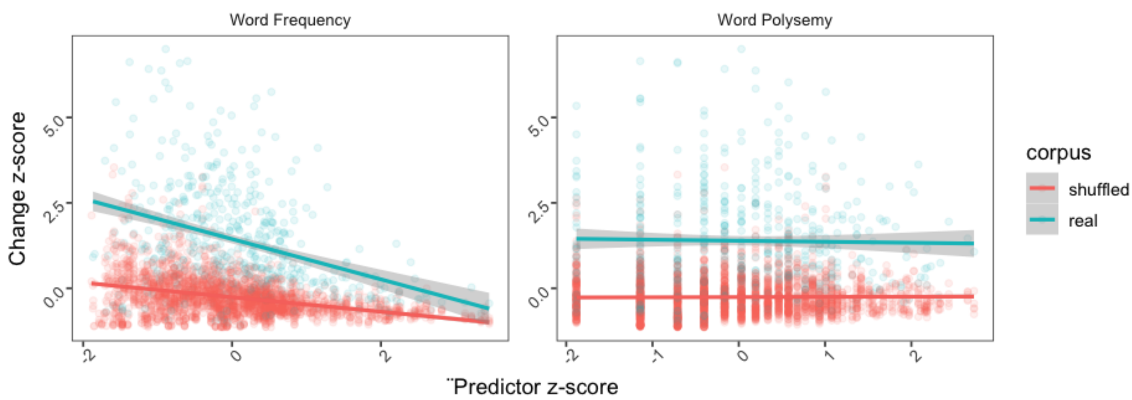

Figure 5. Relations between meaning change and log word frequency on the one hand and log word polysemy on the other hand. Each point represents a word, with lines indicating linear model fits. For better comparison, the values for predictors and their corresponding values of change were centered and scaled.

Category-level Predictors. We next explored the extent to which the way words are related to one another within and across categories can explain change variance above and beyond word-level properties. In this analysis, we again adopted the MacArthur-Bates CDI categories used above as a proxy for psychologically meaningful semantic categories. Following Callon et al. (1991), we characterize categories by their density and centrality. **Density** quantifies the strength of the relationships that tie together the words making up the category. In our context, it measures the extent to which words in this category have a similar meaning, i.e., a similar pattern of use. It is calculated simply as the mean similarity of pairs of words belonging to the category (similarity in the Word2Vec vector space). **Centrality**, on the other hand, quantifies the category's strength of connections with other categories. The greater this quantity, the more central the category is to the meaning/use of other categories. The centrality of a given category is computed as the mean similarity of pairs made of words belonging to one category and words belonging to other categories.

In order to test density and centrality as *predictors* of change, we computed their values using word vectors from the first epoch. The reason is that we wanted to measure how the initial values of density and centrality in the first epoch predict later word changes in the last epoch.

Figure 6 shows a diagram characterizing each category in terms of its density and centrality. We observe a correlation between these two measures ($r = 0.58$, $p < 0.01$): Denser words tend to be more central and vice-versa. Figure 7 shows the words' rate of change as a function of the density and centrality. The less dense and central a category was, the more it tended to change in meaning with development.[6]

Comparing Predictors of Change. Next, we test how word-level and category-level predictors interact in predicting meaning change over developmental

[6] The reason values of centrality (Figure 7, right) are slightly higher in the "shuffled" condition compared to the "real" one is the fact the shuffled condition contains utterances from different epochs, whereas the real condition contains utterances from the first epoch only. It follows that words in the "real" appear in a relatively limited context, whereas in the "shuffled," these same words appear with more other words, many belonging to different categories, leading to overall higher values of centrality.

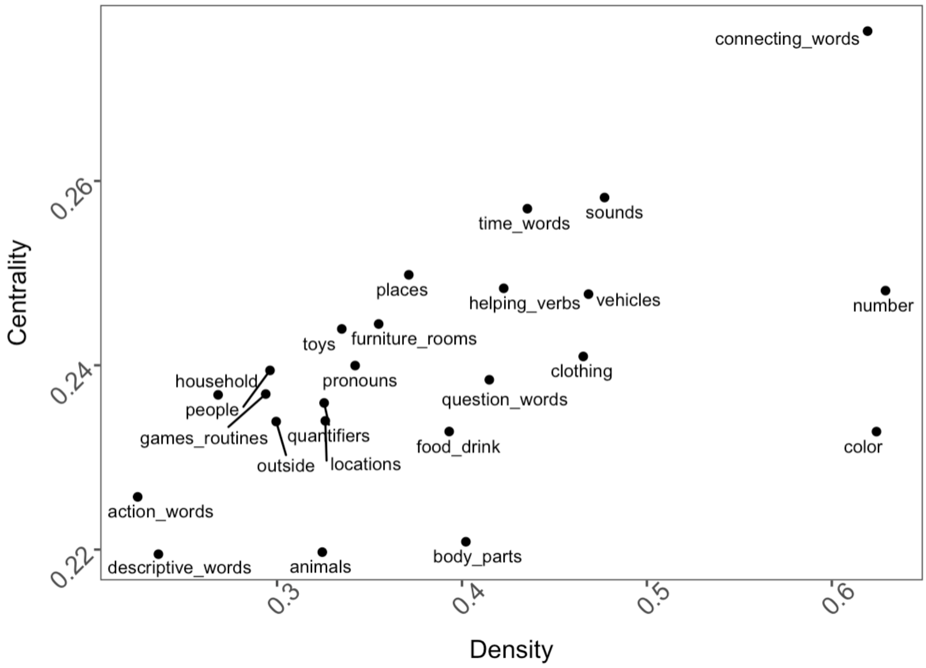

Figure 6. Density vs. centrality computed for each of the semantic categories for our vocabulary.

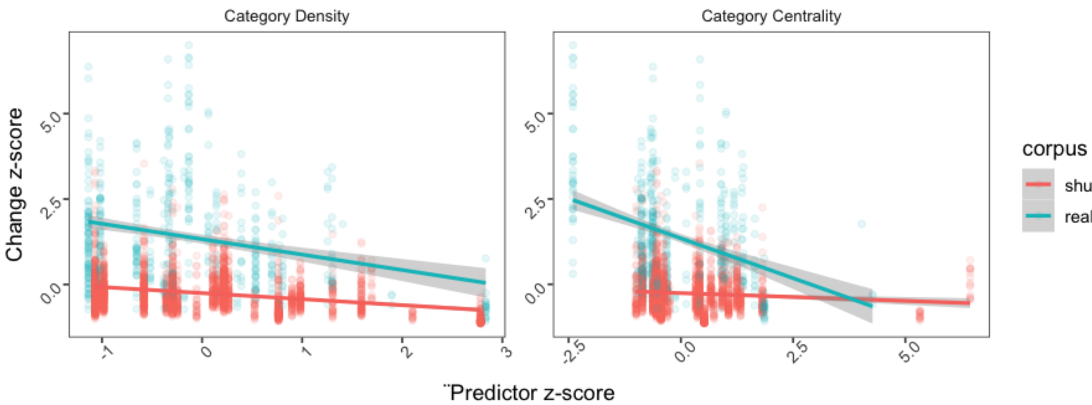

Figure 7. Change as predictors by the category-level property of density and centrality. Each point represents a word, with lines indicating linear model fits. For better comparison, the values for predictors and their corresponding values of change were centered and scaled.

time. We fit a mixed-effects linear regression model predicting the rate of change of a given word (contrasting the fist and last epochs) by its `frequency`, `polysemy`, and `density`. We did not include `centrality` as a factor due to its high correlation with density. We included `condition` as a binary predictor denoting whether data come from the real or time-shuffled corpora. We also added interaction terms between each predictor and `condition` to control for sampling noise.

The mixed-effects model was specified as follows.[7]

`change` $\sim$ `(freq + density + polysemy) * condition + (freq | lexical_class)`.

Regression coefficients are shown in Table 5. For `frequency`, we observe an interaction with `condition`, i.e., confirming the above observation that frequency relates to change above and beyond sampling noise (see Figure 5, left). The simple effect of frequency within the real corpus alone was $\beta = -0.64$ ($SE = 0.14$, $p < 0.01$). For `polysemy`, as expected from Figure 5 (right), we did not find an effect on change. As for `density`, we found an interaction with `condition`, both confirming the above qualitative observation (Figure 7, left) and indicating that `density` is associated with change above and beyond `frequency`. The simple effect of density within the real corpus alone was $\beta = -0.75$ ($SE = 0.08$, $p < 0.001$). Since all predictors were centered and scaled, we can interpret coefficient magnitudes as indicating that the effects of frequency and density were quite similar in magnitude.

Scaffolding conceptual change?

Previous work has shown that for some word categories that are notoriously hard for children to acquire (e.g., words for number, color, and time), learning the class of lexical alternatives (even devoid of referential meanings) is a first step towards the acquisition of the adult-like meaning structure (Carey, 2009; Wagner et al., 2016). It follows that words in such classes would actually benefit from being used in a very similar way (i.e., high density) and also in a way that is quite distinct from words from other classes (i.e., low centrality). Figure 6 shows that this is indeed the case for number and color words.

We refine the analysis by defining a measure of "purity", which we believe

[7] The random effects structure of the model was motivated by our observation that frequency relates to change within lexical categories, as shown in Appendix C.

Table 5

Estimates of the fixed effects in a mixed-effects model predicting change by frequency, category density, polysemy, and condition.

	Estimate	*Standard Error*	*Significance Level*
(Intercept)	1.344	0.077	***
Freq	−0.683	0.055	***
Density	−0.655	0.036	***
Polysemy	−0.034	0.035	
Condition	−1.610	0.034	***
Freq:Condition	0.393	0.036	***
Density:Condition	0.354	0.038	***
Polysemy:Condition	0.065	0.038	
$R^2 = 0.55$			

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

operationalizes better the idea of a class as defined in previous work (Carey, 2009; Wagner et al., 2016). We define purity as follows. For a given category (e.g., numbers) of size N, we iterate over words belonging to the category and we compute, for each word, its N nearest word neighbors, i.e., the most similar to that instance in the entire vocabulary. Finally, we compute what proportion of these N neighbors are themselves members of the same class.

Figure 8 shows that number words and – to a lesser extent – color words (but not time words) both have a higher than average purity. This result supports the hypothesis that the pattern of use of these categories in the input might facilitate learning them as a distinct class of alternatives.

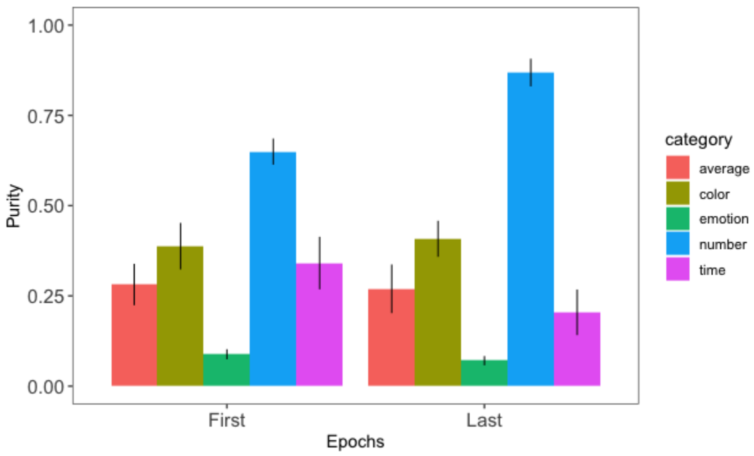

Figure 8. The measure of purity for the categories of number, color, emotion, and time. A value of 1 means that each word in the category has all its N-nearest neighbors as members of the same category (where N is the size of the category).

Discussion

The language children hear around them has been hypothesized to play a major role in supporting children's linguistic and cognitive development (Carey, 2009; Gelman, 2009; P. L. Harris, 2012; Xu, 2019). In particular, the pattern of word use (in terms of co-occurrence with other words) in the linguistic input to children is a rich source of semantic knowledge (Andrews et al., 2009; Hills et al., 2010; Frermann & Lapata, 2016; Stella et al., 2017; Huebner & Willits, 2018; Fourtassi, 2020; Unger et al., 2020; Unger & Fisher, 2021). The current work asked whether and how parents *change* their pattern of word use as children develop. Answering this question is a crucial first step towards a broader understanding of how parents tune their speech to the children's developmental context, scaffolding the acquisition of sophisticated concepts (Vygotsky, 1978).

We measured change in parents' pattern of word use by comparing word embeddings derived across different periods of developmental time. Overall, we found that parents did use words differently when talking to younger vs. older children beyond sampling noise. Further, change across time was larger in magnitude compared to changes in word use due to the context of interaction (at least as captured by the diversity of CHILDES corpora), adult speakers, or the child's gender.

Not all words changed similarly. At a broad level, function words tended to be more stable than content words. Finer-grained semantic categories also showed different levels of variability (e.g., clothing more stable, animals changing more). We speculate part of this variability is due to the fact that semantic categories vary in terms of their dependence on a specific context of use. Some categories are clearly context-dependent (e.g., clothing, food/drinks, and games): Children hear words from these categories used in a specific routine (e.g., it is typical during a meal to talk about food/drink-related words). Thus, the linguistic context around these words is relatively limited and does not change much across early development, leading to rather stable meanings.

Other categories of words are less tied to specific contexts/routines (e.g., animals, people, outside, and emotion). Children hear words from these categories in a diversity of contexts. For example, the word "dog" can be mentioned during meal time, when playing outdoors, or even while driving a car. Perhaps the meaning of these words changes more because older children become able to hear them in a more diverse set of linguistic contexts than younger children.

We captured this intuition through the measure of category "density," characterizing how words – that make up a given category – are similar to each other in their context of use. Words from rather context-dependent categories (e.g., food) tend to share very similar pattern of use, leading to high category density. In contrast, words from rather context-independent categories (e.g., people) may be used in a diversity of situations, leading to low density. Words belonging to high-density categories had a more stable meaning across development, whereas words belonging to low-density categories changed more. The effect of category density on word change was strong, even controlling for frequency.

The concept of category density can also help us explain the stability in the closed class of function words relative to content words. Though function words are context-independent, semantically speaking, they differ from content words (e.g., "people" or "animals") that are also context-independent by the fact that function words co-occur with almost all content words, making them very similar to each other in that particular sense. We hypothesize that this fact is what leads function words to have high category density and less change over development.

Previous work has suggested that low-density words that occur in a diverse set of contexts are more likely to be learned earlier by children (Hills et al., 2010; Stella et al., 2017; Fourtassi et al., 2020). This fact could be explained by statistical learning strategies such as cross-situational learning (Smith & Yu, 2008): Hearing the same word in a diversity of visual situations allows the learner to narrow down the set of possible word-referent mappings (at least for concrete nouns).

What the current study adds to our understanding about this phenomenon is the suggestion that, though the first meanings for these words may be acquired early, these meanings are not fixed: They continue to be enriched and nuanced as the context of use continues to increase/change across development. To test this hypothesis, future experimental work should explore if observed meaning change in the input leads to similar change in the children's meaning representations (Ameel et al., 2008).

In contrast to low-density words, high-density words tend to occur in specific semantic contexts, making their word-referent disambiguation harder rather than easier (in a cross-situational learning scenario). Nevertheless, their high density could make them easier to learn in a different way.

In particular, we know that for some word categories that have been shown to be challenging to learn for children (especially words for numbers, color, and time), it helps children to first learn a class of lexical alternative, i.e., a class of words that are closely related in terms of context of occurrence/use (Carey, 2009; Wagner et al., 2016). For example, it helps to have first learned that the words {"red", "blue", "green,"...} "go together" before children learn how to map them to referents in the world. We found that some of these concepts (especially number and colors) do form high density (and even high purity) categories, supporting their acquisition by children.

Conclusions

In sum, this work quantifies developmental changes in the way caregivers use words across early childhood. We documented patterns of stability and change across various syntactic and semantic categories and explained these patterns based on the concept of category density that characterizes the extent to which words occur in specific vs. broad contexts. We suggest that low-density, highly changing words could influence children's meaning learning as these words are used in a richer set of contexts over time. As for high-density, stable words, they also could influence children's meaning learning as they facilitate the acquisition of classes of lexical alternatives, especially for number and color words. These predictions invite more experimental work to compare the change in the input to change in children's mental representation across development.

One limitation of the current work is the focus on English-speaking children's input. We attempted to generalize this work cross-linguistically but found ourselves limited by the sparsity of CHILDES corpora beyond English – fitting word embedding models typically takes millions of words of text and these millions of words are simply unavailable at the present time. Further, future work should allow us to go beyond the analysis of global trends towards a deeper understanding of fine-grained word change in theoretically important cases (e.g., spatial vocabulary, animate/inanimate words). However, again this work will likely require larger datasets.

Another limitation is our choice to use vector-space representations and word2vec in particular. While these models are flexible and powerful proxies for human semantic judgments, they clearly incorporate only a small part of human meaning. Besides, although our study focuses on understanding how parents explicitly tune their speech to child developmental abilities, these word2vec vectors would also encode information beyond caregivers' explicit linguistic tuning across developmental stages, such as the environment and activity changes mentioned in CDS as children grow older. As natural language processing research progresses, we hope that new models of meaning are used to confirm and extend our observations.

Finally, the current study remains largely exploratory. That said, the effects we reported in our analyses were very large and unlikely to be due to overfitting to the current dataset. Future work should however seek to generalize to other datasets and other languages.

In conclusion, we hope our work here will contribute to more concrete, synergistic interaction between corpus-based studies of CDS and experimental work on early semantic development and change.

References

Ameel, E., Malt, B., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language*, *58*(2), 262–285.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors, In *Proceedings of the 52nd annual meeting of the association for computational linguistics*.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715.

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205.

Carey, S. (2009). *The origin of concepts*. Oxford University Press.

De Deyne, S., Perfors, A., & Navarro, D. J. (2016). Predicting human similarity judgments with distributional models: The value of word associations., In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*.

Dubossarsky, H., Weinshall, D., & Grossman, E. (2017). Outta control: Laws of semantic change and inherent biases in word representation models, In *Proceedings of the 2017 conference on empirical methods in natural language processing*.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, *38*(1), 188–230.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i–185.

Fernández, R., & Grimm, R. (2014). Quantifying categorical and conceptual convergence in child-adult dialogue, In *Proceedings of the annual meeting of the cognitive science society*.

Firth, J. R. (1935). The technique of semantics. *Transactions of the Philological Society*, *34*(1), 36–73.

Fourtassi, A. (2020). Word co-occurrence in child-directed speech predicts children's free word associations, In *Proceedings of the workshop on cognitive modeling and computational linguistics*.

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847.

Fourtassi, A., Dunbar, E., & Dupoux, E. (2014). Self-consistency as an inductive bias in early language acquisition, In *Proceedings of the annual meeting of the cognitive science society*.

Frermann, L., & Lapata, M. (2016). Incremental bayesian category learning from natural language. *Cognitive Science*, *40*(6), 1333–1381.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2021). Caregiver linguistic alignment to autistic and typically developing children.

Gelman, S. A. (2009). Learning from others: Children's construction of concepts. *Annual review of psychology*, *60*, 115–140.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*(2), 211.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016a). Cultural shift or linguistic drift? comparing two computational measures of semantic change, In *Proceedings of the* *conference on empirical methods in natural language processing*. NIH Public Access.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016b). Diachronic word embeddings reveal statistical laws of semantic change. *Association for Computational Linguistics*.

Harris, P. L. (2012). *Trusting what you're told: How children learn from others*. Harvard University Press.

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

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133.

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, *43*(5), 1062.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365.

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development*, *82*(4), 1199–1209.

Kulkarni, V., Al-Rfou, R., Perozzi, B., & Skiena, S. (2015). Statistically significant detection of linguistic change, In *Proceedings of the 24th international conference on world wide web*.

Kunert, R., Fernández, R., & Zuidema, W. (2011). Adaptation in child directed speech: Evidence from corpora. *Proc. SemDial*, 112–119.

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.

Leung, A., Tunkel, A., & Yurovsky, D. (2021). Parents fine-tune their speech to children's vocabulary knowledge. *Psychological Science*, *32*(7), 975–984.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208.

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

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *International Conference on Learning Representations, Workshop Track Proceedings*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality, In *Advances in neural information processing systems*.

Misiek, T., Favre, B., & Fourtassi, A. (2020). Development of multi-level linguistic alignment in child-adult conversations, In *Proceedings of the workshop on cognitive modeling and computational linguistics*.

Newport, E., Gleitman, H., & Gleitman, L. (1977). Mother, id rather do it myself: Some effects and non-effects of maternal speech style. *Cognitive theory*, *2*.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation, In *Proceedings of the 2014 conference on empirical methods in natural language processing.*

Piaget, J. (1954). *The construction of reality in the child.* New York, NY, US: Basic Books.

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, *41*(3), 647–656.

Ridgeway, D., Waters, E., & Kuczaj, S. A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology*, *21*(5), 901.

Rohde, D. L., Gonnerman, L. M., & Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*, *8*(627-633), 116.

Rudolph, M., & Blei, D. (2017). Dynamic bernoulli embeddings for language evolution. *arXiv preprint arXiv:1703.08052*.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*.

Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, *60*, 100949.

Unger, L., Savic, O., & Sloutsky, V. M. (2020). Statistical regularities shape semantic organization throughout development. *Cognition*, *198*, 104190.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, *23*(3), 34–41.

Wagner, K., Tillman, K., & Barner, D. (2016). Inferring number, time, and color concepts from core knowledge and linguistic structure. *Core Knowledge and Conceptual Change*, *105*.

Wittgenstein, L. (1953). Philosophical investigations, trans. *GEM Anscombe*, *261*, 49.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological review*, *126*(6), 841.

Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level., In *Cogsci*.

Appendix A

Adult's identity

The distribution adults' identities the CHILDES aggregated corpus we use in the current study (Figure A1).

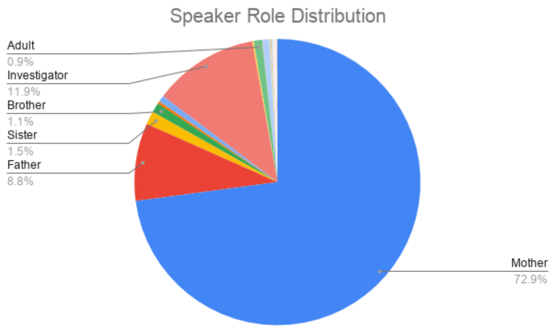

Figure A1. Distribution of adults' utterances as function of identity (i.e., their relationship to the target child)

.

Appendix B

Context of adult-child interaction

The list of corpora we used in the current study and some summary statistics.

family	min age (months)	max age (months)	utterance number	token number
McMillan	28	28	106	456
Bliss	27	73	1,017	4,462
Nelson	19	32	1,628	10,992
Hicks	61	132	5,281	31,601
Warren	18	74	5,853	26,711
Cornell	18	60	6,005	26,603
Feldman	14	27	6,670	27,585
Bohannon	36	36	6,757	27,724
PetersonMcCabe	48	113	7,219	30,374
Demetras1	24	47	8,294	35,528
Bernstein	13	25	8,469	31,176
VanKleeck	37	48	8,756	38,158
VanHouten	28	43	9,306	36,507
Higginson	11	35	9,672	40,051
Howe	19	24	9,675	37,699
Bates	20	28	10,082	37,637
Forrester	12	60	10,458	47,708
McCune	12	24	10,927	34,847
Demetras2	25	33	11,119	47,616
Sachs	14	57	12,227	54,294
Tommerdahl	29	45	13,594	64,169
Tardif	17	23	14,302	52,252
Rollins	3	20	19,131	56,742
Morisset	30	39	19,399	69,780
Snow	29	45	21,235	100,916
Clark	26	38	24,283	134,655
Kuczaj	28	60	25,853	123,348
Peters	15	25	25,914	106,761
Fletcher	36	87	26,383	128,279
Soderstrom	5	12	27,057	98,333
Valian	21	32	27,831	124,907
Post	19	32	29,139	121,875
MacWhinney	16	92	33,412	162,071
Suppes	23	39	35,793	185,012
Bloom70	20	37	36,351	158,958
Gleason	25	62	38,813	175,588
Wells	17	60	46,331	160,321
Braunwald	15	84	47,984	171,909
NewEngland	13	33	48,126	144,417
Weist	25	60	54,120	309,022
Belfast	24	54	81,350	424,376
Brown	18	62	87,485	367,242
Lara	21	39	102,077	363,551
Hall	54	57	125,095	605,391
HSLLD	42	144	201,200	957,014
Providence	11	48	286,770	1,356,684
MPI-EVA-Manchester	24	37	328,309	1,227,692
Manchester	20	36	373,936	1,454,060
Thomas	24	59	376,734	1,991,541

Table B1

Corpora from 49 families in CHILDES English dataset.

Appendix C

Frequency and polysemy and Change within syntactic categories

We provide two fine-grained plots to show how frequency and polysemy influence word change within a given syntactic category.

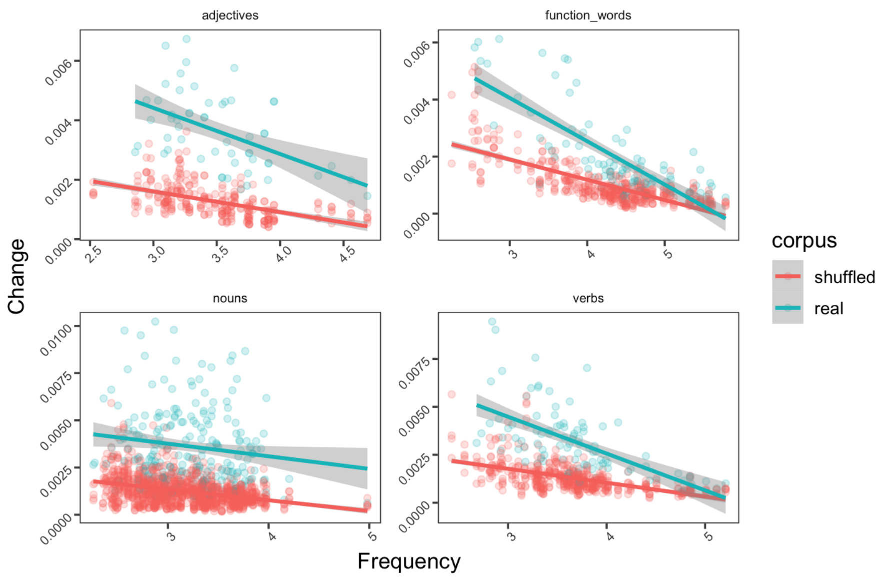

Figure C1. Frequency against change per syntactic category.

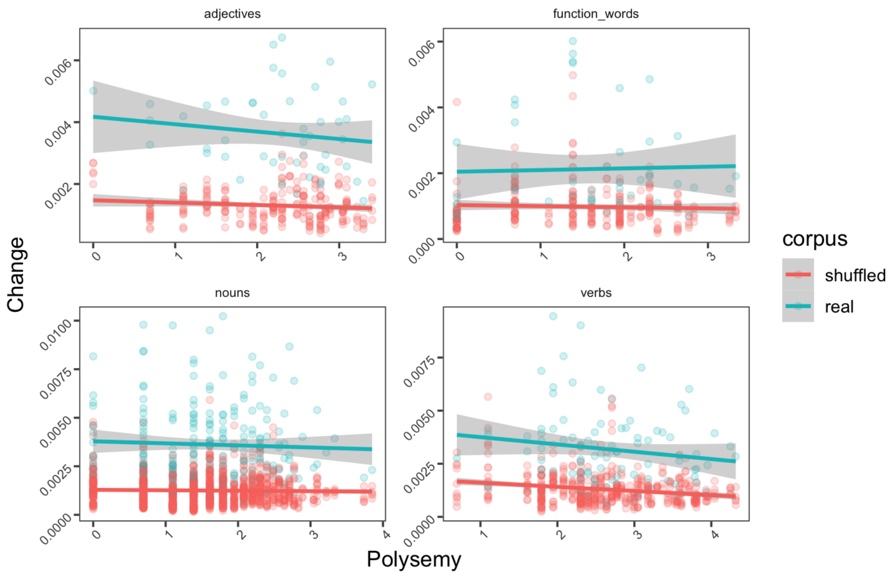

Figure C2. Polysemy against change per syntactic category.