

Learning that numbers are the same, while learning that they are different

Jon A. Willits (jwillits@indiana.edu), Michael N. Jones (mnjones@indiana.edu),
and David Landy (dlandy@indiana.edu)

Department of Psychological and Brain Sciences, Indiana University
1101 E. 10th St. Bloomington, IN 47408 USA

Abstract

It has been suggested that the way that number words are used may play an important role in the development of number concepts. However, little is currently known about the overall ways in which number words are used in child-directed speech. To address this, we performed an analysis of how number words are used in the CHILDES database. We looked at four statistics: 1) lexical frequency, 2) contextual diversity, 3) word co-occurrence, and 4) distributional similarity, to see if these distributional statistics suggest why some aspects of number acquisition are easy and others are hard, and if these statistics are informative about specific debates in number acquisition. We found that there are many important differences in how small and large number words are used (such as differences in frequency, co-occurrence patterns, and distributional similarity), differences that may play a role in shaping hypotheses about children's acquisition of number concepts.

Keywords: number representation, language acquisition, concept acquisition, statistical learning, corpus analyses

Introduction

Numbers are an odd category of things. Although often treated as objects themselves, numbers can be observed only as relational features—features of sets of objects. Their structure is as regular as one could hope to find in a category, and they are extremely important in modern life. But numbers are very challenging for children and students to understand and use, and many simple-to-phrase questions about the counting numbers remain unanswered even by professional mathematicians. In short, numbers are among the most frequent and regular, and at the same time most abstract and difficult, of the concepts in a young child's environment. Here, we focus on the role of the statistics of the child's linguistic environment to interrogate two key challenges for the learner of numbers: how do we learn that number words form a category, that they are all the same kind of thing? But in contrast, how do children also learn how each number word is different from the others?

One important mechanism by which children can learn about semantic categories is by learning about the distributional patterns of the words that refer to them. Much research has demonstrated that children can learn a lot about a word's semantics, based on the word's pattern of co-occurrence. For example, toddlers can learn whether a verb's meaning is transitive or intransitive by counting the number of nouns that occur in the same utterance as the verb (Gleitman, 1990; Yuan, Fisher, & Snedeker, 2012). Toddlers can also learn about the meaning of a novel noun if it occurs in distributions that are informative about the category to which it belongs (Lany & Saffran, 2011).

If children can infer aspects about semantic content from words lexical distributions – words' frequencies, the words with which they co-occur, and the words to which

they are distributionally similar, what can they learn about number concepts based on the ways that number words are used in child-directed speech, and what might this tell us about the development of children's number concepts and number word usage?

The Conceptual Structure of Number

The cognitive structure of numbers is a matter of intense current debate. At this point, a general consensus has emerged that the processing of numbers under about 4 is qualitatively different from that of numbers roughly 4 and over (Feigenson, Dehaene, & Spelke, 2004). Numbers over 4, on this general account, are processed by a general magnitude system, which discriminates relatively numerous sets from relatively sparse sets, but does not have distinct ways of capturing precise numbers. This system is generally thought to be 'log-scaled' in the sense that the reliability of discrimination between two sets is a function of the ratio between their numerosities, regardless of the absolute numerosity. At the same time, numbers under 4 are processed by a very different mechanism that individuates particular objects, but may not be all that involved in estimating magnitudes. This mechanism has been tied to subitization (Trick & Pylyshyn, 1994; Revkin et al, 2008), and to object tracking (e.g., Spelke & Kinzler, 2007). On some accounts, linguistic experience is used to coordinate these two systems, bootstrapping a new "positive whole number" system (Carey, 2009), which combines representation of exactness with large magnitudes, comprising concepts like "exactly 7."

The general dual-systems account outlined above has been applied to several behavioral paradigms beyond subitization. First, when asked to give an experimenter a certain number of objects (the *give-N* task), children tend to show discrete changes in capacity, from being able to give one object only successfully, to being able to give one or two, to being able to give up to three, before passing into a final "four or more" stage. A common interpretation is that the first three stages are associated with the small-number processing system, while the final stage reflects invocation of the full count-list and magnitude system into the task.

The Linguistic Structure of Number

Number words are an extremely common part of our language. For example, in the CHILDES database (MacWhinney, 2000) that we examined below, the word *one* is slightly more common than the word *mom*, while *two* is slightly more common than *dad*. The words *one* and *two* taken together comprise almost exactly 1% of all the words in CHILDES. And while these two number words are certainly the most common, the remaining number words [0, 3-20, 30, 40, 50, 60, 70, 80, 90], *hundred*, *thousand*, and

million constitute about an additional 0.5% of the total words in the corpus, a truly large number of tokens for such a small number of types (31).

The structure of language has been implicated in particular aspects of number learning. For instance, phonological irregularity has been argued to slow the learning of teen words in English (Fuson & Kwon, 1991), and numbers between 20 and 40 in languages where such words are partially irregular or inverted (Brysbart, Fias, & Noël, 1998). Processing of numbers in large ranges (one thousand to one billion) may also be affected by how these words are used (Landy, Silbert, & Goldin, 2013; Landy, Charlesworth, & Ottmar, 2014).

Although number-elaboration accounts rely on language to connect number experiences, the specific ways that numbers are structured in language are less thoroughly explored than number-related behaviors. In particular, the information available in the morphemic structure of the speech stream—though useful in distinguishing categories in language—has to our knowledge not been systematically explored in the case of number words. Although the literature often refers to a “count list” (Carey, 2009), it is unclear how much of a role this list plays in the number-related linguistic experience of young children, though compelling recent work suggests that counting and label-present sets with children have a positive impact on number learning (Gunderson & Levine, 2011). More generally, we do not yet know how the striking behavioral distinctions between numbers under and over 4 are mirrored in the structure of child-directed language, nor how the statistics of language differentiate numbers over and under 10.

Starting from the basis that number words are very common, that they are unevenly distributed and structured, and that they are implicated in number concept learning, we ask what information about numbers exists in the statistical structure of the speech stream, that could potentially help children understand numbers’ complex category structure. In particular, we see the central challenge as one of learning that number words form a category, but at the same time learning how number words differ from one another. To the degree that number words are used similarly, identifying them as a category should be facilitated. At the same time, the easier it is to learn that items belong to the same category, the harder it should be – at least for a time – to distinguish members of the category from one another (Rogers & McClelland, 2004).

Analysis and Data Set

To address some of these questions and provide an introductory analysis of what children can learn about number words based on their distributional patterns in child-directed speech, we analyzed these distributional patterns in the CHILDES corpus, a collection of transcripts of caregivers and family members interacting with children in a wide range of situations and contexts (MacWhinney, 2000). For our analyses, we built a composite corpus containing all documents involving typically developing children from American English speaking households up to 72 months of age. This resulted in a corpus containing 4,561 documents, 32,580 word types, and 7,016,488 million word

tokens. Compared to estimates of the number of words children typically hear, the size of this corpus represents approximately 5-10% of the input a typical child hears by age six, thus providing a very good model of the available input to a child.

We performed four analyses of how number words are used in the CHILDES corpus. First, what is the frequency of number words, how do frequencies between different numbers compare to each other, and how do these frequencies change across corpora to children of different ages? Second, what is the contextual diversity of number usage? Do some numbers occur in broader or narrower ranges of contexts? And how does this change in speech to children of different ages? Third, what are the other words with which number words co-occur, and is there a difference in co-occurrence patterns or different kinds of numbers? Finally, how do number words vary in terms of their distributional similarity to other number words? Which numbers can most easily be categorized as a number word based on their distributions? Which numbers are most prototypical? Answers to questions like these can begin to help us understand how the usage of number words contributes to the development of number representations.

Number Frequency and Contextual Diversity

One of the most basic factors that may affect the learnability of number words is how frequently they are used, how frequency varies for different number words, and how frequency varies in speech to children of different ages. For an empirical answer to these questions, we took the American English CHILDES corpus and subdivided it into 10 sub-corpora, designed to reflect children’s cumulative experience at evenly divided points. Thus, we created corpus M12 composed of all corpora with target children of up to 12 months of age, M18 composed of all corpora with target children of up to 18 months of age, and so on, up to M72, composed of all corpora for children up to 72 months of age. We then calculated the frequency of a large set of number words, and converted this frequency count into a “parts per million” count (i.e. divided the frequency count by the corpus size, and then multiplied that proportion by 1,000,000) designed to control for differences in sizes of the specific sub-corpora. The results of this analysis are shown in Figure 1.

Figure 1 demonstrates a number of important properties about the frequencies of number words. As noted above, the frequency of the word *one* is an order of magnitude above the other number words. This reflects the importance of the concept of *one*, but also the fact that one is used in a number of ways that aren’t explicitly numerical. However, even these non-numerical uses of one are still informative about numerical content (e.g. “*another one*”, “*the one I love*”, “*one of the toys*”, “*it is a good one*”). Beyond one, the frequency of the words are roughly in cardinal order. Of the “small” numbers, 2 is much more frequent than 3, which is much more frequent than 4 and 5, which are much more frequent than everything else. The rank orderings are all very sensible (especially when one considers usages like dates, times, money, and other measurement contexts).

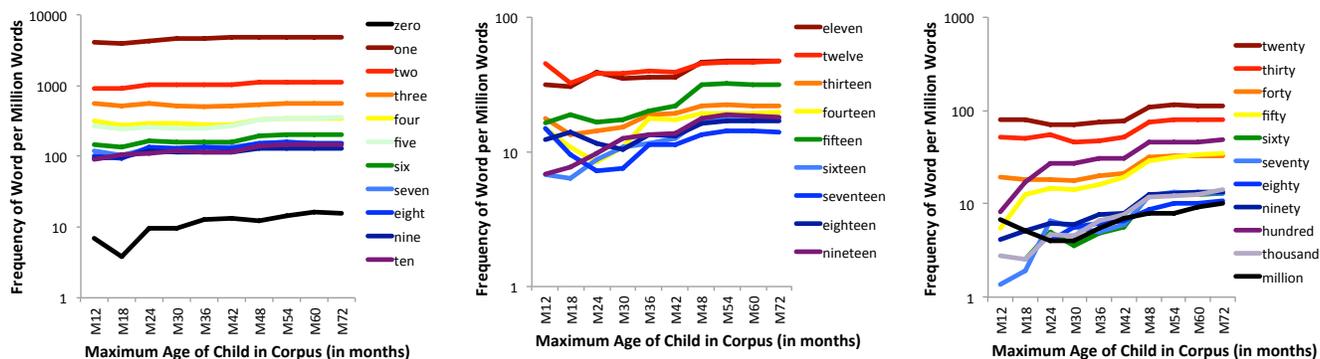


Figure 1. Cumulative frequency per million words of various number words plotted on a log10 scale. The x-axis represents different sub-corpora for children of different ages. M12 is in all corpora for all children up to 12 months of age, M72 is for all corpora for all children up to 72 months of age). Note that across the three sub-figures, the scale of the y-axis changes, reflecting the large differences between big and small numbers.

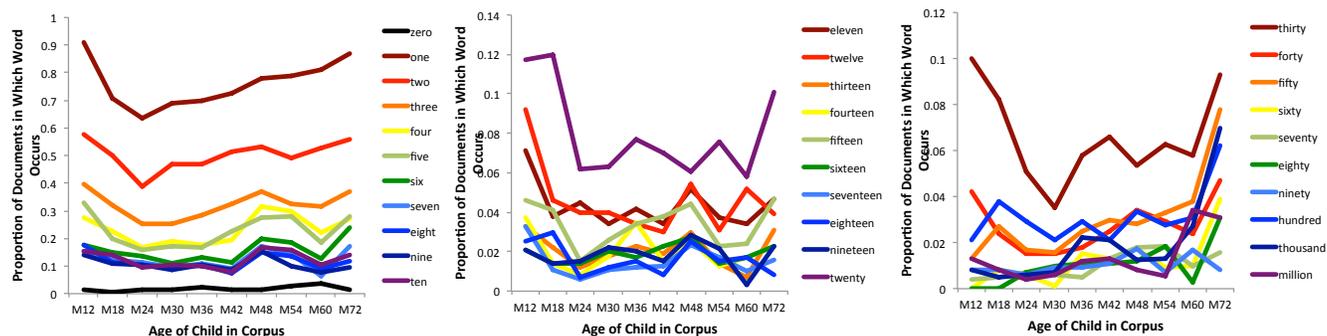


Figure 2. The proportion of documents in which the word occurs, for various number words across development. Note that across the three sub-figures, the scale of the y-axis changes, reflecting the large differences between big and small numbers.

One notable and perhaps surprising finding from these analyses was how little the frequency ratios of the number words changed across the different age samples in the CHILDES corpus. All numbers become slightly more frequent over time (though for some numbers like *one* this difference is very small). Words for larger numbers do make up a larger proportion of the number words as the children in CHILDES increase in age, although this difference is not as large as it appears given the log-scaled differences between big and small numbers. For example, the numbers one through ten make up 95.0% of all number word usage in the sub-corpora from kids age 12 months and below, whereas they make up 90.7% of the usage for kids age 48 months and above.

We also investigated the contextual diversity (CD) of number word usage, operationally defined as the proportion of different CHILDES documents in which the number word occurred. *High* CD has been found to be a strong predictor (above and beyond word frequency) of lexical access in adults (Adelman & Brown, 2008) and of age of acquisition in children (Hills et al., 2013). However, lower CD (i.e. usage consistency) has been linked to faster learning of words in a number of behavioral experiments (Akhtar & Tomasello, 1997). Thus, the exact role of contextual diversity on word learning is obviously complex and not yet well understood.

We found that contextual diversity of number words in CHILDES was highly correlated with their frequencies (as is always true), with words like *one* and *two* much more diversely used than larger numbers. However, it is also

notable that CD of number words showed differences across the corpora from children of different ages, with trajectories that are intriguing with regard to possible developmental trends, and the role that usage diversity plays on word learning. For almost all numbers, the contextual diversity is relatively higher in speech to children 12 months and younger, drops to a much lower rate in corpora for children in the middle range of the CHILDES sample, and then rose again in corpora for older children. The data is consistent with the idea that when children are first introduced to a word, caregivers use that word in a very high number of contexts, then they selectively use that word in a much narrower range of contexts (facilitating the acquisition of the word’s specific meaning and core properties), before reverting to a more base-rate level of usage of the word, in terms of diversity of contexts (Tomasello, 2003).

Number Word Co-occurrence

The second set of statistics we investigated were word co-occurrence patterns of the number words, to see if they suggested anything about the learnability of number concepts. To do this, we calculated the co-occurrence of number words with other words within a 12-word window. The most frequently co-occurring words with various number words are shown in Table 1.

When looking at co-occurrence of number words with all other words in the corpus, interesting patterns emerge which are suggestive of what children can learn about numbers based on these co-occurrence patterns. First, one of the most apparent facts is how similar all the numbers are in

terms of the words with which they frequently co-occur. This is a big part of why (as we show in the next section) it would be easy to learn that number words belong to the same category. Next, one can clearly see how grammatical number and conceptual number are interrelated, with the number *one* co-occurring with the plural form “s” much less often than other numbers (of which “s” would be the most frequent collocation for many). Finally, it is apparent that many of the frequent collocations of numbers are other numbers (and numbers that are nearby in order), evidence of the extent to which counting and other ordering is highly frequent in child directed speech.

Table 1. Highest co-occurring words for various number words within a 12-word window in CHILDES. The morpheme “_S” is demonstrating how often the number word had the plural “s” in its 12-word window, when “s” is treated as a distinct morpheme in the corpus (i.e. broken off of the word to which it was attached).

one	two	three	four
you	_S	one	_S
one	you	_S	three
is	one	two	one
the	two	you	you
that	is	three	two
I	I	is	is
this	the	four	four
_S	that	the	give
eight	eighteen	eighty	hundred
is	twenty	eighty	a
you	is	is	and
eight	_S	you	you
seven	eighteen	and	_S
nine	you	five	is
I	I	I	I
six	seventeen	eight	the
the	the	the	to

The relationship involving the co-occurrence of number words with other number words is shown for all pairs of numbers in Figure 3. In this Figure you can see many interesting facts about number word co-occurrence. First, the pattern of number words being more likely to co-occur with other number words of similar size is shown by the brightness along the diagonal. Especially interesting is how the breadth of this “counting: window grows as numbers get larger. *Three* is very likely to co-occur with *two*. *Four* is as well, though less so. *Six* and *seven* are barely more likely to co-occur with *two* than we would expect them to based on their baserate frequencies. However, while a simple model based on logarithmic scaling might assume that number words would continue to be more broadly associated as they increased in magnitude, we find instead that particular number ranges seem to form co-occurrence clusters. For numbers in the teens, for example, all the teens are highly likely to co-occur with each other, even 13 and 19. Many other idiosyncratic (but meaningful) highly probable collocations can be seen, including those that encode relations due to time (such as numbers like *six* and *twelve* with *thirty*) and dates (such as *nineteen* and *ninety*).

As a final analysis of word-occurrence patterns for numbers, and how important number words were to the structure of child-directed speech, we performed a principal

component analysis (PCA) of the co-occurrence patterns between the 8,000 most frequent words in the corpus. We then looked at the extent to which number words loaded on the 30 highest components, shown in Figure 4.

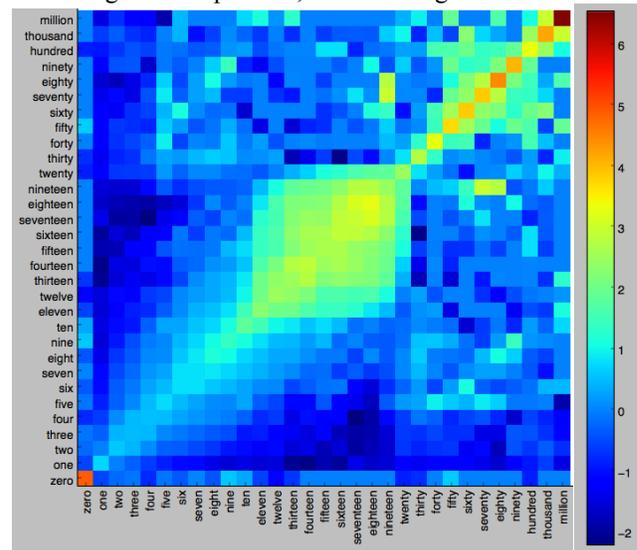


Figure 3. A heat map showing the pointwise mutual information (PMI) scores of number words with other number words at 72 months, demonstrating the degree to which they co-occur with each other more or less one would expect due to their base rate frequencies. A value of zero means the two co-occur exactly as often as one would expect, with each value greater or less than one meaning they co-occur 10 (or 100 for a score of +/- 2, or 1000 for a score of +/-3) times more or less than one would expect them to, given their word frequencies.

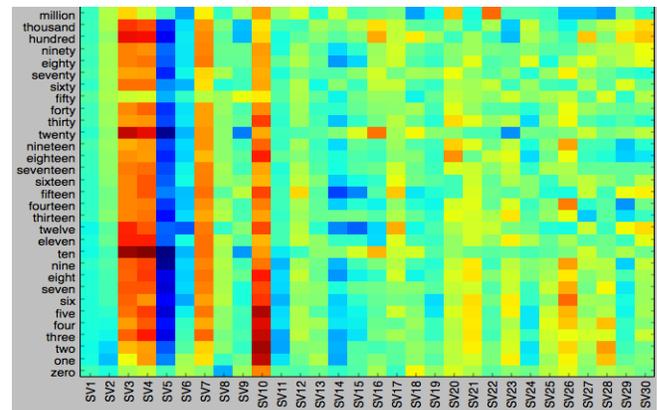


Figure 4. The singular value loadings for various number words across the first 30 singular value dimensions.

Figure 4 makes it clear just how important number words are to the structure of child-directed speech. Excepting principal components 1 and 2 (which tend not to carry structural information in these kinds of analyses), at least some number words load extremely highly on most of the primary principal components). Though beyond the scope of this short paper, follow-up analyses of these principal components show that many of the components are interpretable as representing contextually different ways in which numbers are used, such as counting, money, time, dates, and measurement. Some of these relationships can be identified in Figure 4 by noting which numbers do, and do not, load on particular principle components.

Number Distributional Similarity

In our final analysis, we attempted to assess the extent to which distributional statistics can be used to learn that numbers belong to the same category, as well as to tell that and tell how number words are different from one another. To do this, we derived these similarity scores for each of the 8,000 most frequent words in the corpus, by computing the correlation of each word pair's 30-element principal component vectors (such as those shown in Figure 4). Thus, the extent to which each pair of words was similar depended on the extent to which the two words shared the same co-occurrence patterns in language (Riordan & Jones, 2011).

To assess how useful these similarity scores would be for guessing that number words belong to the same category, we performed a discrimination analysis of using each of the 31 number words from the above analyses, and a

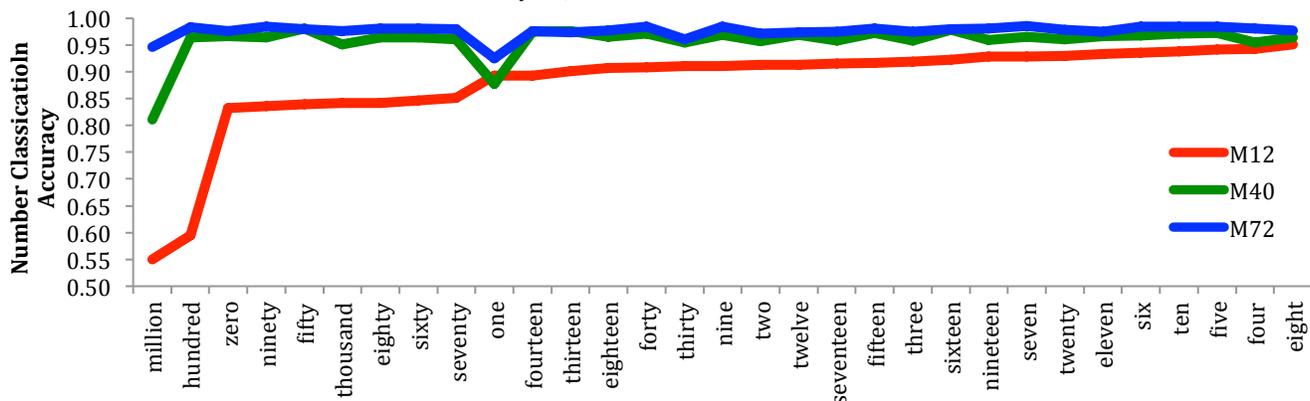


Figure 5. Classification balanced accuracy for various number words using distributional similarity to predict same vs. different category membership, for various number words at various stages of amount of input (up to 12 months, up to 40 months, up to 72 months).

This attempt to classify number words as same and different from other words based on their distributional statistics resulted in a number of interesting findings about the statistical structure of number words. The first is how remarkably successful children could be at classifying number words based on their distributional statistics. This was true even in the “youngest” corpus, with only the child-directed speech to children 12 months and below, with a mean accuracy across all number words of 88.0% (SE = 1.6%). Using all speech to children up to 40 months, accuracy rose to 95.8% (SE = 0.6%). Using all speech to children up to 72 months, accuracy peaked at 97.6% (SE = 0.2%). Number words are used remarkably consistently (in fact, number words can be classified better than words from any other category, Willits & Jones, in review), and learning they belong to the same category based on their usage statistics would be a very easy task. Also interesting, however, were the specific words that were more difficult, especially for the “youngest” corpus. As expected, *one* was difficult, but not the most difficult. Lower frequency (and less consistently used) words like the decade words and *million* were the words that were most difficult to classify.

To assess what children could learn about the internal structure of the number category based on language statistics, we performed a hierarchical cluster analysis of words based on their co-occurrence patterns. This cluster

set of approximately 600 other words from a collection of 30 other semantic categories. We took each pair of these words (both the number-number pairs as well as the number-“not-number” pairs), and used their similarity score to guess whether that pair should be judged as belonging to the same category or to different categories. We used a constant similarity threshold across all comparisons ($r = 0.30$), and we used the accuracy of these judgment to compute balanced accuracy scores (i.e. we evenly averaged the score on all trials where the correct answer was “yes” with the score on all trials where the correct answer was “no”, to account for the fact that there were many more “no” trials). We performed this discrimination analysis on three different sub-corpora to get a sense of how the results changed across speech to children with of different ages. The results of this analysis are shown in Figure 6.

analysis is shown in Figure 6.

As can be seen in Figure 6, the sub-category clusters that emerge are incredibly sensible and informative about what children might be learning about number from number words. The most distinctively used words are *one*, *two*, and *three*, which form their own subcluster. Other small numbers group together, as do the teens, decades, and large numbers. One notable exception is 19, which groups with the decade numbers, driven by the fact that most of the CHILDES data was collected in the 1980s and 1990s.

Discussion

We find that the use of numbers in child-directed speech provides evidence of many of the same distinctions noted in children's number-related behavior. In the frequency analyses, the power-law nature of number words' frequency distribution and diversity of usage (with *one*, *two*, *three*, and *four* each being used successively and dramatically less than the next) could be important factor in children's difficulty learning to properly use words for larger numbers.

In the co-occurrence analyses, we discovered many ways in which the co-occurrence patterns varied, that could lead to differences in learnability. Lower numbered words were relatively less likely to co-occur with one another, suggesting that these words are used in a number of diverse ways other than strictly counting and comparatively quantifying, which may lead to some of the difficulty

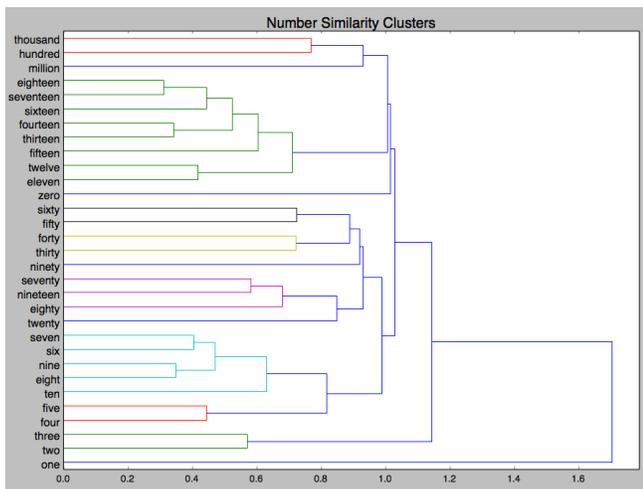


Figure 6. Hierarchical clustering of number words based on their co-occurrence patterns in CHILDES.

children have at extending their ability to quantify small numbers to larger quantities.

Likewise, the similarity analyses demonstrate a number of facts about number word usage with implications for number concept acquisition. In particular, the two broadest divisions in the hierarchical cluster analysis were between *one* and other numbers, and between *two* and *three* and other numbers. *One* in particular seems to be used quite differently from other numbers, which is not terribly surprising given its less obviously cardinal roles in language, e.g., as a specification of an agent, or in idioms like “one time”. The typical adult number categories (4-10, the teens, the decades, and the larger numbers) form distinct clusters as well, while zero forms its own subcategory.

Previous behavioral evidence has indicated special trouble understanding the computational role of the number words *million* (Landy et al., 2013), *zero* (Evans, 1983), and *hundred* and *thousand* (Rips, 2013; Siegler & Opfer, 2003). Statistical analyses confirm that these are treated differently than other number words: they are more difficult for a statistical algorithm to categorize as numbers based on usage (especially in early months), and form distinct sub-clusters in the hierarchical clustering analysis.

Just as striking is the nature of the clusters. The teen words tend to strongly co-occur, to form a tight cluster, and to contain strong mutual information. One plausible interpretation is that, compared to other numbers, these numbers appear primarily together in count lists, and appear relatively rarely in other contexts.

References

Adelman, J. S., & Brown, G. D. A. (2008). Modeling lexical decision: The form of frequency and diversity effects. *Psychological Review*, *115*, 214–229.

Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology*, *33*, 952-965.

Brybaert, M., Fias, W., & Noel, M. P. (1998). The Whorfian hypothesis and numerical cognition: is twenty-four processed in the same way as four-and-twenty? *Cognition*, *66*, 51-77.

Carey, S. (2009). *The Origin of Concepts*. Oxford U. Press.

Evans, D. W. (1983). *Understanding zero and infinity in the early school years*. Doctoral Dissertation.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307-314.

Fuson, K. C., & Kwon, Y. (1991). Chinese-based regular and European irregular systems of number words: The disadvantages for English-speaking children. *Language in mathematical education: Research and Practice*, 211-226.

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

Gunderson, E. A., & Levine, S. C. (2011). Some types of parent number talk count more than others: Relations between parents' input and children's cardinal number knowledge. *Developmental Science*, *14*, 1021-1032.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language. *Journal of Memory and Language*, *63*, 259-273.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive science*, *37*(5), 775-799.

Landy, D., Charlesworth, A., & Ottmar, E. (2014). Cutting In Line: Discontinuities in the Use of Large Numbers by Adults. In Proceedings of the 36th Annual Conference of the Cognitive Science Society.

Lany, J., & Saffran, J. R. (2010). From statistics to meaning infants' acquisition of lexical categories. *Psychological Science*, *21*, 284-291.

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

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*, 607-614.

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

Rips, L. J. (2013). How many is a zillion? Sources of number distortion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1257.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychological science*, *14*(3), 237-250.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89-96.

Tomasello, M. (2003). *Constructing a language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*, 80-102.

Yuan, S., Fisher, C., & Snedeker, J. (2012). Counting the nouns: Simple structural cues to verb meaning. *Child Development*, *83*, 1382-1399.