A first look at children's acquisition of 'zero': Interim summer report

Rose M. Schneider

Introduction

How do we learn about nothing? Our formal mathematics system relies partially on the premise that we are able to represent and perform logical operations with nothing, or 'zero.' This number takes on many special functions not only in symbolic mathematics, but also in language, where it can be used to indicate nonexistence, or an empty set. But how do children first learn about 'zero' and its functional roles? 'Zero' seems to present at least three distinct challenges that might hinder its acquisition in comparison to the other numerals:

1. The semantic meaning of 'zero' is literally the absence of something (Merritt & Brannon, 2013); thus, a child cannot rely on any mapping between a physical referent and a label, which inhibits word-learning (Gleitman, 1990; Medina, Snedeker, Trueswell, & Gleitman, 2011).
2. Reasoning about nonexistence might be conceptually difficult for children, requiring some potentially later-developing comprehension of negation (Nordmeyer & Frank, 2014; Reuter, Feiman, & Snedeker, 2017).
3. 'Zero' is not included in the canonical count routine, which is one of the main tools children leverage in acquiring the semantic meanings of numerals (Carey, 2004).

Additionally, 'zero' seems to be special kind of number not only for adults (Brysbaert, 1995), but also in history, with 'zero' added to the formal mathematics system much later than other numerals (Bialystock & Codd, 2000). Both the relatively late integration of 'zero' into the symbolic numerical system, and adults' difficulties with the integer indicate that reasoning about *nothing* is not intuitive. Given that 'zero' seems to occupy a special place in the numeric system both ontogenetically and historically, it seems likely that explorations of its acquisition might also reveal some insights into why this number is unique.

Here, I am defining a 'full acquisition' of zero as an integration of a) a mapping between both its symbolic forms and semantic meaning ('nothing', or the absence of a quantity) and b) an understanding of the ordinal relationship of 'zero' to other integers. Children progress towards a

full acquisition of 'zero' in distinct stages (Wellman & Miller, 1986). Starting at around three years of age, children first display a strict form-only-mapping from 'zero' to a symbolic digit 0. Children then acquire the semantic meaning of 'zero', mapping it to 'none' or 'nothing', but still do not show evidence of understanding its numeric function. Finally, at around six years of age, children fully acquire 'zero,' meaning that they have integrated the integer into the number line. Critically, 'zero' is acquired later than other positive numerals during this time period (Wellman & Miller, 1986).

There are many factors that could be contributing to the comparatively slow acquisition of 'zero.' I have grouped these broadly into *linguistic* and *conceptual* difficulties. The *conceptual* difficulties, as mentioned above, are that learning to reason about and acquire a label for *nothing* requires some abstract reasoning on the part of a child who (arguably) may still be developing this ability (Nordmeyer & Frank, 2014).

The *linguistic* difficulties of 'zero,' however, might also contribute to its prolonged acquisition. As mentioned above, it seems likely that children rarely (if ever) hear 'zero' in a count routine, a rich source of information about the meanings of numerals (Carey, 2004). Additionally, 'zero' seems to be used quite infrequently in relation to other number words, with a markedly different distributional pattern (Willits, Landy, & Jones, 2016). In fact, using a distributional pattern analysis with a large-scale corpus of number words in the Child Language Data Exchange System (CHILDES), Willits and colleagues found that 'zero' is treated differently in speech in comparison to words like 'two' or 'four.'

While the large-scale corpus analysis by Willits and colleagues provides an intriguing foundation for investigations of 'zero,' this summer project was designed to provide a more in-depth analysis of how 'zero' is used over development. The main question I set out to answer was whether its occurrences in both parental input and child output can shed light on some of the linguistic difficulties children face in its acquisition.

Methods

Tools: One of the main components of this project has been learning the programming language Python to access and work with the CHILDES transcripts. While I was familiar with some of the basic functions of Python, this project required me to learn both the *nltk* (Natural Language Toolkit) package, and *regex* (Regular Expression). Most of the first month of this project was

spent learning how to use Python to extract the data from CHILDES. While the code for this project is still (very much) a work in progress, all scripts can be found in the Github repo for this project, located at https://github.com/rosemschneider/zero.

Corpora: All corpora come from the CHILDES, a repository of thousands of transcripts of child speech (http://childes.talkbank.org/). Based on the Willits et al. (2016) paper, I chose to search for instances of 'zero' over all typically-developing, English-speaking corpora in the CHILDES database (N = 55). Because 'zero' is so infrequent, however, only 19 corpora are included in this analysis.

Utterance selection: For this preliminary analysis, I began by extracting utterances that contain an exact match to 'zero.' This means that I am only including an analysis of the singular form of 'zero,' and did not include plural forms or other variations. Having read through these utterances, however, I feel that a second pass of these corpora should include both the plural form (which would likely capture more utterances about symbolic mathematics), and the letter 'o' in children's output, which seems to be frequently used in place of 'zero.' This latter point is discussed below in Future Directions.

Utterance coding: In total, I found 229 utterances containing 'zero' in these corpora. Each utterance was hand-coded by locating the utterance in the corpus transcript. I read the utterances before and after the target to determine the general context of the transcript; to what specifically 'zero' referred; the function of 'zero'; whether it referred to a physical object; and whether it was used correctly. Although I wrote the utterance extraction code to operate for either input or output, I also validated whether the utterance was parental input, or child output.

Because the coding was done by hand, it is necessarily subjective. Additionally, because very few transcripts are accompanied by a video recording of the interaction, some utterances were impossible to code.

For analyses, excluded were utterance without age data (N = 13), and for those in which a sibling (and not the target child) is speaking (N = 12). Finally, I filtered out any utterances that could not be coded, leaving a total of 180 utterances (N input = 105). All utterances and coding can be viewed on the Github repository.

Coding key: For each utterance, I subjectively determined what I felt the function of the 'zero' to be. With only a rough idea of what these utterances might contain, I began this coding process with a few categories (symbolic, nonexistence, and count). After coding all utterances, however, I narrowed these categories and added a few more, which required reclassifying some sentences. Thus, the final classifications of the utterances are partially informed by the data itself. Below are explanations of each code.

- Symbolic_digit: An occurrence of 'zero' referring to a physical object, such as a number in a book or in a puzzle.
- Symbolic_verbal: An occurrence of 'zero' not connected to a physical object, such as in a phone number, or in telling time.
- Count_number: An occurrence of 'zero' in either an ascending or descending count.
- Nonexistence_negation: An occurrence of 'zero' which indicates nonexistence, or in which 'zero' is used to indicate 'none' or 'nothing'
- Symbol: An occurrence of 'zero' where the word is mapped to a form that is not actually a number - the only instances of this are in a game of tic-tac-toe, where the mother and investigator call the 'O's 'zero.'

Age bins: Due to the sparsity of the data and the number of corpora over which I searched, I ended up binning utterances into 6-month age intervals for ease of analysis. While this binning has been helpful in visualizing the data, future work should explore other binnings of this data, as it might reveal other developmentally interesting information.

Results

The qualitative analysis of utterances containing 'zero,' revealed, as expected, that this number does not occur in a standard count routine. There are only a handful of instances of 'zero' coded as a 'count_number,' but these are exclusively instances where a child is counting down from a number (often in the context of playing with a spaceship). Thus, this corpus analysis provides support for the argument that 'zero' is difficult to learn because it is not used like other numbers, especially 'one' through 'ten.'

Also as predicted, ‘zero’ is a low-frequency word. I was unable to quantify its frequency in comparison to a word like ‘one’ through ‘ten:’ These words are so high-frequency that I lacked sufficient computing resources to extract their utterances. ‘Zero’ occurs roughly as frequently as words like ‘hundred’ or ‘thousand.’

Figure 1 shows the total number of utterances in each age bin. This distribution generally reflects the age distribution of the CHILDES corpora, with the oldest ages underrepresented.

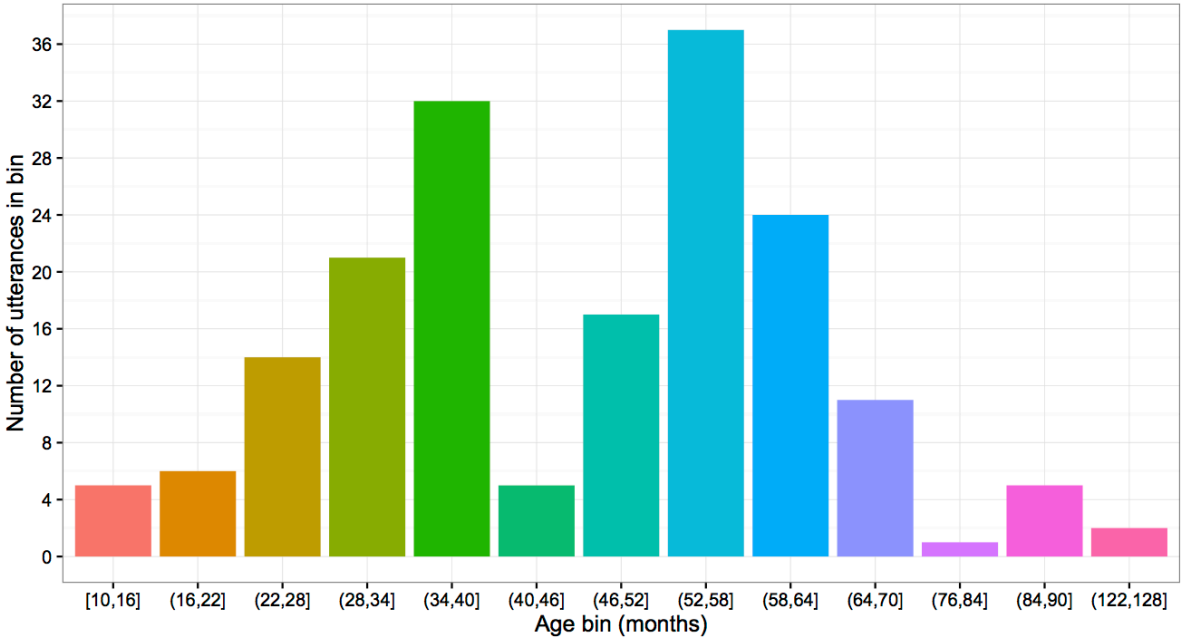

Figure 1: Number of utterances per age bin contained in corpus analysis

Figure 2 breaks down these utterances by input and output. As expected, ‘zero’ is most frequent in parental input until approximately three years of age. From three years onward, the frequency of utterances containing ‘zero’ is split roughly evenly between input and output. It is worth noting that this is the period during which children are acquiring the meaning of ‘zero’ (Wellman & Miller, 1986).

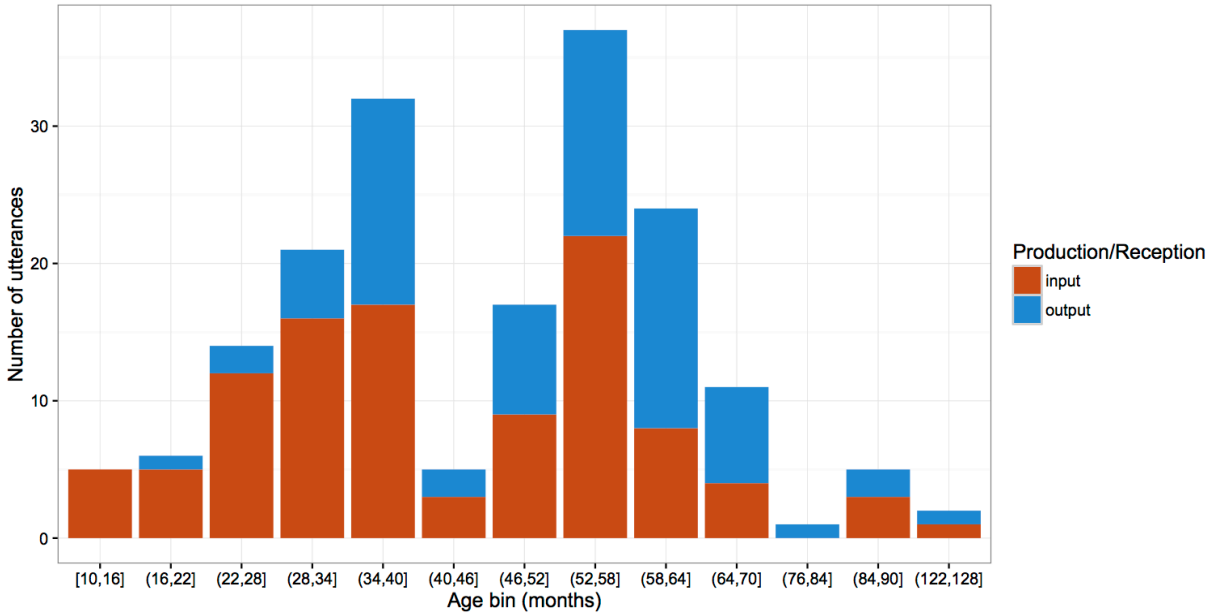

Figure 2: Number of utterances in production and reception for each age bin

Next, I turned my attention to the function of 'zero' in these utterances. Excluded from Figure 3 are frequencies for 'zero' as a count, symbol, and other/undefined due to low frequency. While the nonexistence/negation function of 'zero' is also relatively low frequency, its connection to the conceptual challenges posed by 'zero' makes its inclusion relevant here.

Figure 3: Histogram of 'zero' types contained in production and comprehension over development

Interestingly, ‘zero’ is used most overwhelmingly to refer to a physically present symbolic digit. The second most frequent function of ‘zero’ is to refer to the digit in the absence of a symbolic form, most frequently as a phone number. Perhaps reflecting the conceptual difficulty of this number, ‘zero’ is very infrequently used to indicate nonexistence (e.g., discussing ‘zero gravity’ or ‘zero dollars’). Of the 13 instances of ‘zero’ used to indicate nonexistence, only 5 occurred in output, with 4 correct usages (e.g., ‘zero dollars’), and one incidence of a partially correct use (a child tries to indicate a very small quantity).

Finally, Figure 4 shows the breakdown of actual referents of ‘zero’ in both input and output. Reflecting the high occurrence of ‘zero’ referring to a symbolic digit, the most frequent referent is a toy. Children do not begin hearing about ‘zero’ in the context of mathematics until fairly late in development, and even then the frequency is low. Aggregating over a larger corpus of older children, and perhaps expanding the search to include ‘zeros’ may yield more mathematically-focused utterances, however.

.

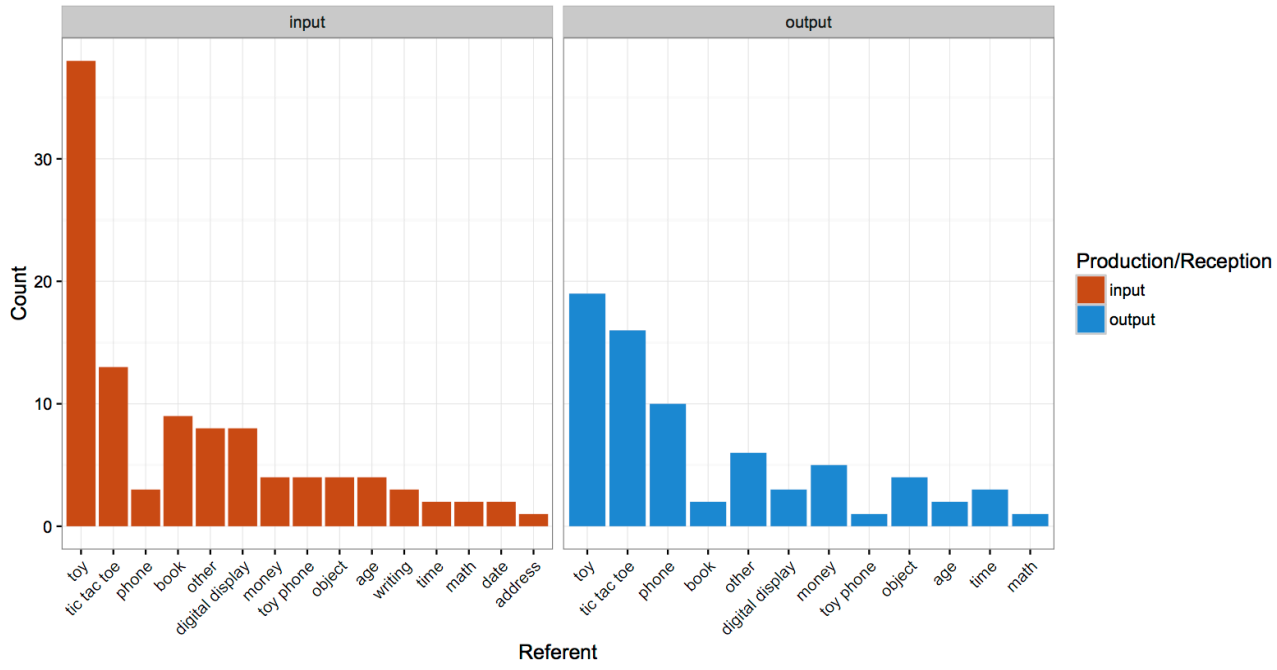

Figure 4: Actual referent of ‘zero’ in utterance.

Discussion

One of the most interesting insights to come out of this corpus analysis is the extremely low occurrence of 'zero' being used to indicate the absence of a quantity, especially in consideration with the high frequency of 'zero' in conjunction with a symbolic digit. I find this compelling for two reasons: The first is that the data from parent input and child output provide evidence for the progressive acquisition of 'zero' observed by Wellman and Miller (1986), wherein children acquire a mapping from word to symbolic form in the absence of deeper understanding of semantic meaning. While further empirical work on this progressive acquisition of 'zero' needs to be done, this corpus analysis does provide some preliminary support for Wellman and Miller's findings.

The second reason this is interesting is due to the lack of other data for determining the semantics of 'zero.' As mentioned above, 'zero' refers to the absence of a quantity, making it impossible for children to use a physical referent to determine its meaning. This is not the case with other number words. Additionally, other number words often enjoy the benefit of appearing in the count routine; 'zero' does not appear in a canonical count routine. Thus, 'zero' lacks two important data sources that children use in acquiring the meaning of number words.

Critically, however, the low incidence of 'zero' meaning 'none' or 'nothing' in input may also be an impediment to acquiring its semantic meaning, and may act as a bottleneck which inhibits children's ability to later integrate it into their number line.

Taken together, the results of this corpus analysis provide some support and explanation for Wellman & Miller's (1986) observed developmental trajectory of 'zero' acquisition. While the frequency of 'zero' is low in comparison to other number words, when 'zero' does appear in input or output it almost exclusively refers to a symbolic digit. With the absence of 'zero' in a count routine and its infrequent use to indicate nonexistence/negation, children have very little data over which to aggregate in determining 'zero's exact semantics. Taken together, these linguistic factors could perhaps be contributing to the delayed 'zero'-'none' mapping and slow integration into a child's formal number system.

Future Directions

While this detailed qualitative corpus analysis has yielded some intriguing insights into why children's acquisition of 'zero' lags behind other small integers, a similar analysis needs to be performed with some other small numbers. Specifically, the hypothesis that children do not

receive enough data to determine the semantics of 'zero' needs to be tested. Although there have been other corpus analyses of number in child-directed speech, these have focused mainly on syntactic construction (Bloom & Wynn, 1997), or frequency and co-occurrence patterns (Willits, Landy, & Jones, 2016). A qualitative analysis of the content and context of these corpora still remains to be done. Thus, future work will include a comparison of the acquisition of 'zero' to that of other small numbers, such as 'two' or 'three.' Due to the frequency of these words in CHILDES, this comparison will most likely be done over a much smaller subset of corpora.

As mentioned previously, one limitation of this corpus analysis is that it included only exact matches to 'zero.' Evaluation of these utterances revealed, however, that (especially young) children used the letter 'o' as a stand-in for 'zero,' perhaps because the initial phoneme of 'zero' is difficult to produce. Additionally, 'o' or 'oh' is also used interchangeably with 'zero' in adult output as well (e.g., "It's five-oh-three"). Future work should include utterances containing 'o' or 'oh;' it is likely that these utterances will reveal evidence of earlier output, and may also shed some light on whether children display uncertainty between 'oh' and 'zero' across development. Such ambiguity between forms might also contribute to the delayed acquisition of 'zero.'

Finally, this corpus analysis lays the groundwork for empirical work on the acquisition of 'zero.' Previous work has indicated that while four-year-olds are unable to order '0' in relation to other numerical symbols, they do show distance effects in making judgements with nonsymbolic arrays (Merritt & Brannon, 2013). This corpus analysis suggests that children may lack the appropriate linguistic input with which to construct the semantic meaning of 'zero,' thus constraining their ability to integrate the integer into the number line. A key test of this hypothesis could involve training the 'zero'-'none' relationship (thus appropriately linking form to meaning), and measuring the effect on children's ordinal judgments involving '0.'

Conclusion

Number is an incredibly difficult cognitive ability to acquire, requiring children to aggregate over both perceptual and linguistic input. When one or both of those sources of data are disrupted or otherwise under informative, however, number acquisition becomes much more difficult. In the case of 'zero,' children face the double challenge of trying to match a label to the *absence* of a referent, and of sparse linguistic input.

This qualitative corpus analysis explored that linguistic input in an effort to determine why children have difficulty acquiring 'zero.' The results suggest that, in addition to its low frequency, 'zero' occurs in nonstandard number contexts, and rarely occurs in a semantically supportive context. While this theory needs to be empirically tested, this work has underscored the importance of exploring not just the frequency or syntactic use of number words, but also their roles in the context of real speech.

References

Bialystok, E., & Codd, J. (2000). Representing quantity beyond whole numbers: Some, none, and part. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *54*(2), 117.

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child language*, *24*(3), 511-533.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of experimental psychology: General*, *124*(4), 434.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59-68.

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

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, *108*(22), 9014-9019.

Merritt, D. J., & Brannon, E. M. (2013). Nothing to it: Precursors to a zero concept in preschoolers. *Behavioural processes*, *93*, 91-97.

Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, *77*, 25-39.

Reuter, T., Feiman, R., & Snedeker, J. (2017). Getting to No: Pragmatic and Semantic Factors in Two-and Three-Year-Olds' Understanding of Negation. *Child Development*.

Wellman, H. M., & Miller, K. F. (1986). Thinking about nothing: Development of concepts of zero. *British Journal of Developmental Psychology*, *4*(1), 31-42.

Willits, J. A., Jones, M. N., & Landy, D. (2016). Learning that numbers are the same, while learning that they are different. Proceedings of the 38th Annual Meeting of the Cognitive Sciences Society.