Final Project Report: Child Acquisition of Reduplication in Bangla

Shinjini Ghosh, Shuli Jones

Topic

There is existing work on reduplication rules, and classification of reduplication functionality in Bangla. Separately, there is a small amount of work on child learning of reduplication, although mostly focused on reduplication in English (where it is not a very productive paradigm). Some work also exists on child learning of other "paradigms" with different functionality and classification. We propose to bring together these areas of study and examine child learning/ classification abilities of reduplication in Bangla.

Because there is very little pre-existing work on the topic, our experimental design is an exploratory one. One particularly motivating research area is the theory that the codification of reduplication in some languages' adult grammars arose because of children's use of reduplication (which today occurs even in languages that do not have reduplication as a productive part of the adult grammar). Again preliminarily, we hope to explore this by looking at potential correlations between the kinds of reduplication that children master first and the ones that are more commonly in use by adults. (Of course, there are many possible explanations for such a correlation, so if it were to be found, followup research would be necessary.)

Experimental Overview and Background

Our experiment will focus on types of reduplication where the same sentence can have different meanings in more than one context. We want to see if children learn one of the possible meanings earlier than the other, or if they learn both at the same time. In addition to providing direct information about the trajectory of child acquisition of reduplication in Bangla, this information could also provide insight as to children's underlying expectations of what words in those syntactic positions are "supposed" to mean.

Types of reduplication that could be used for this experiment include:

- Adjectival: `choto choto ful` = "small small flowers" could ideally function as a plural, meaning many small flowers, or an intensifier, meaning a few tiny flowers; `sundor sundor mohila` =

"beautiful beautiful women" functions to mean many beautiful women rather than one very beautiful woman.

- Nominal: `kukure kukure` = "dog dog" means "among the dogs"; unlike the adjectival meaning, interpreting it as plural would be incorrect.
- Verbal: `jete jete` = "going going" means "while going" and functions as a tense modifier, instead of implying intensity/pace.
- Adverbial: `aste aste` = "slowly slowly" indicates intensity.

Experiment Design

We plan to run two sets of experimental trials. In one, children will be presented with two pictures, each of which is a correct depiction of one meaning of a sentence containing reduplication. The experimenter will ask, "can you show me the picture with...?" or "can you show me the picture where...?" and have the children point to a picture. In the second experiment, the experimenter will present both photos as before, but they will tell a story first and then end with the same question as in phase one. We expect that the story will bias the interpretation of the reduplication towards one of the two pictures.

A concrete example of a trial would be two images of flowers, one with just a few very small flowers and one with a greater number of relatively normal-sized flowers. (To make the second set of flowers still seem small, something like a large tree in the background might be used. We could also make the relative sizes of flowers in the two images different enough to make it more obvious.) Then, in the first experimental phase, the experimenter would ask "Can you show me the picture with the small small flowers?" In the second phase, they might instead tell a story about a girl who went on a walk and the things she saw, including flowers that were smaller than any flowers she had seen before! Then they would again ask "Can you show me the picture with the small small flowers?". In the first case, both answers are equally acceptable (with one being more common in colloquial usage than the other); in the second, the context makes it clear that the image with a few tiny flowers is the correct one.

Two-phase experiment. Phase 1:

"Show me the small small flowers."

Phase 1

Phase 2:

"I want to tell you a story about a girl who went out for a walk…
the plants she saw were smaller than she had ever seen before!
Show me the small small flowers."

Phase 2

The goal of the first phase is to see which meaning children pick when they have two equally good options. If children of a certain age always pick one meaning, we will know that they favour it. If they pick both at chance, they either don't know both, or know both, but are unaware that one is the more commonly used or more acceptable one in colloquial speech. It is also possible that children who

know both meanings would favour the more common one; we could also run adult trials of the same experiment to compare the children's performances against empirical adultlike behavior.

In the second phase, we assume that all adults will pick the meaning that the context is pointing to. If the child originally favoured the other meaning, we will know they only know that one if they continue to choose it; however, if they now switch to the contextually correct meaning, that indicates that they know both but prefer the other. If they originally picked at chance, then continuing to do so will show that they do not know either meaning; if they now pick the correct one, that will indicate that they know both.

Hypothesis

We believe that very young children will not know either meaning. As they grow, first they will learn the more common meaning, and then the less common one. So, we would expect to see chance picking in the youngest children, always favouring the more common meaning in the next-youngest group, and then picking the correct contextual meaning in the older groups. What is not known is if older children, as well as adults, will pick at chance when there is no context. We think it is more likely that both groups will be more likely to pick the more common meaning.

Discussion

We plan to address the questions raised in the presentation, particularly those on denoting relative "smallness" of flowers, designing sentences with adjectives that cannot be graded (an example is `chouko chouko khaat` = "rectangular rectangular bed") and alternative reasons for the order of reduplication learning.

Ideas for further study include:

- Examine not just which meaning children choose but also correctness (e.g. will they accept noun reduplication used to create a meaning that only adjective reduplication can license?)
- Investigate from a part-of-speech lens: if children can produce nouns but not adverbs correctly, how does that affect their command of noun reduplication vs. adverb reduplication? Do they assume adverb reduplication has the same functionality as noun reduplication, or do they not understand it at all?

Background Research

Summary of Ota & Skarabela (2017):

This paper investigates the hypothesis that very early infant word learning and segmentation is aided by word and syllable reduplication (e.g. 'mama' or 'choo-choo'). Other studies showed some advantage for infants in learning reduplicative words, but none explicitly tested whether they improve infants' ability to perform word segmentation. The study exposed nine-month-old native English learners to two different spoken passages, each containing a novel four-letter CVCV word chosen to be similar to English words phonetically. One word was reduplicative (the first CV syllable was the same as the second) and one had two different syllables. Next, they played those two words, as well as an additional two novel words the infants had not heard before. They found that the infants listened equally long to the two newest words and to the original non-reduplicative word, but listened longer or paid more attention to the original reduplicated word. That implies that the infants had only learned the reduplicated word despite being exposed to both equally. The authors' conclusion is that reduplication does indeed improve infants' ability to perform word segmentation.

Summary of Bandyopadhyay and Chakraborty (2010):

A rule-based approach has been used in two phases, the general expression level and the structural and semantic classifications. Notably, the authors found eight classes of reduplication at the 'sense' level, implying that there are multiple different functionalities of Bengali reduplication. Moreover, a rule-based system yielded on an average, over 90% precision, recall and f-score values, implying that it might also be possible for young children to apply similar segmentation using similar rules.

Summary of Senapati and Garain (2015):

The authors have studied the frequencies in which reduplicated words appear in Bengali. They have proposed an algorithm to identify the reduplicated words from a text corpus as well as a dictionary-based tuning technique to enhance the accuracy of identifying such words in the corpus. This implies the rule- based nature of Bangla reduplication.

Summary of (Anonymous):

A chapter on Bangla reduplication has provided 11 classifications of Bangla reduplications, based on various aspects. It would be interesting to look at if children (and even adults) do end up classifying all 11 separately eventually, of if they just make functional/morphological/phonetic classifications.

Summary of Culbertson et al. (2019):

Many languages with gender classification systems for nouns do not adhere perfectly to the rules of classification. As such, speakers rely on more than one type of clue in order to guess the gender of

novel nouns. Culbertson et al. look at phonological clues (e.g. class 1 nouns tend to end in -a) and semantic ones (e.g. animate objects tend to be class 2) and find that children tend to rely more on phonological clues for classification, whereas adults change their reliance between the two types depending on which seems more salient. However, they also found that both adults and children tend to rely more on whichever type of information was presented first. They put forward two theories, which can potentially both be true, for why children rely more on phonological clues; the first is that because children are exposed to phonological information even before they know word meanings, that is the information "presented first", which they over-rely on when classifying known words. The second is that, since children also over-rely with novel words, they have some kind of a priori bias for phonological clues. To explain this bias, they propose two potential explanations (not summarized here) based on other work.

It would be interesting to do a similar investigation, but focused on the different types of reduplication (and the different ways their meanings/functionality are signaled) rather than on different types of noun class.

Notes on Suzman (1997):

Looks at how children acquire Bantu noun classification and prefixing system; finds that children do initially start by acquiring a reduced subset of noun classes and prefixes. Looked at the type of errors children tend to make and found that they correctly segment the noun from its prefix (potential comparison for us: can children correctly segment partial reuplication?), but don't always add on the prefix in full or correctly.

Notes on Bar-Sever, Lee, Scontras, Pearl (2018)

The authors investigated how children learn adjective ordering by doing probabilistic modeling of data from the CHILDES database, for three separate hypotheses of lexical-class, subjectivity, and input frequencies. They discovered that both abstract representational hypotheses perform better as children age, implying the emergence of abstract knowledge about adjective ordering preferences. Between the two abstract-knowledge-based representations as well, they saw a preference for lexical-class-based knowledge rather than subjectivity-based. This offers insight into how experiments about child learning of preferences occur can be designed.

References

1. Ota & Skarabela, Reduplication facilitates early word segmentation

2. Bandyopadhyay & Chakraborty, Identification of Reduplication in Bengali Corpus and their Semantic Analysis: A Rule-Based Approach

3. Senapati & Garain, A Computational Approach for Corpus Based Analysis of Reduplicated Words in Bengali

4. Anonymous, Chapter on Bangla Reduplication

5. Culbertson, J., Jarvinen, H., Haggarty, F., & Smith, K. (2019). Children's sensitivity to phonological and semantic cues during noun class learning: Evidence for a phonological bias. Language 95(2), 268-293. doi:10.1353/lan.2019.0031.

6. Suzman, S., Acquisition of noun class systems in related Bantu languages

7. Bar-Sever G., et al, Little lexical learners: Quantitatively assessing the development of adjective ordering preferences