

The relation between vocabulary growth and syntactic complexity in child language development

lexicon-syntax

Eva Portelance

2019-02-20 14:27:02

Introductory note

I have chosen to use the opportunity of this class to complete the first steps of a larger research project. In the *Research question* section, I introduce the overarching research question I am interested in and I outline the specific substeps towards answering this question I hope to complete as part of this class.

Research question

Once new words are acquired, how do children learn to produce sentences? As children go through early language development they learn more and more new words and they produce more and more complex utterances. Though it is generally acknowledged that there is a strong correlation between vocabulary size and syntactic complexity in child language development, the nature of this relationship is still ill-defined. Within lexicalized frameworks, such as minimalist grammars or head-driven phrase structure grammars, we might expect that lexicon growth feeds syntactic complexity, but we also know that syntactic complexity can feed word learning from work on syntactic bootstrapping and errors in syntactic over generalizations. In order to define the relationship which exists between word learning and syntactic development I would like to design and compare computational models which explicitly state the interactions between lexicon and syntax that I am testing. I hope to use child directed speech data from corpora to train these models and child production speech for evaluating the quality of the output. In addition to potentially laying the foundations for speaking of the relationship between vocabulary growth and syntactic development in more formal terms, this project will inform our theoretical understanding of the relationship between the lexicon and syntax.

I plan to use the opportunity of this course to familiarize myself with the crosslinguistic repositories from which I will extract my data. I want to reproduce the finding of strong correlations between vocabulary size and complexity scores, as well as vocabulary size and syntactic category acquisition (Bates et al. 1994, Braginsky et al. (2015), M. Frank et al. (n.d.)). This is important given that the rest of my project assumes that the correlation between vocabulary growth and syntactic development (complexity scores and syntactic categories being different representational metrics of this) holds.

In addition to the plan outlined in the previous paragraph, if I have additional time this quarter, I would like to write a data processing pipeline for extracting example ‘modal’ child’s lexicon at various ages which will be used in the next part of this project: modelling a language learner.

Methods

I will work with existing datasets available through the Wordbank Project (<http://wordbank.stanford.edu>) which has developed an R API, wordbankr (<https://github.com/langcog/wordbankr>) for interfacing with the data. These datasets are collections of CDI forms (a parental report of their child’s language development) for a set of different languages. These forms are administered multiple times throughout the first 3 years of life of these children. They contain wordlists and syntactic complexity questions from which both the vocabulary size and the complexity scores are calculated.

Analysis and Report

The variables I am interested in for each observation (each CDI form) are the *vocabulary size*, *complexity score*, and the *raw counts for each syntactic category*. To plot and understand the relationship between *vocabulary size* and *complexity score*, I will use a mixed effect model since I cannot guarantee the independence of these variables. As for the relationship between *vocabulary size* and *raw counts for each syntactic category*, since these variables are clearly directly related, I will plot the proportion of words in each category as a function of vocabulary size.

If time permits, I will also work on the data processing pipeline for extracting lexicons of ‘modal’ children in different languages at different ages. In the report I plan to give an example of a modal child given an age (in months) and a language.

References

- Bates, Elizabeth, Virginia Marchman, Donna Thal, Larry Fenson, Philip Dale, J Steven Reznick, Judy Reilly, and Jeff Hartung. 1994. “Developmental and Stylistic Variation in the Composition of Early Vocabulary.” *Journal of Child Language* 21 (1). Cambridge University Press: 85–123.
- Braginsky, Mika, Daniel Yurovsky, Virginia A Marchman, and Michael C Frank. 2015. “Developmental Changes in the Relationship Between Grammar and the Lexicon.” In *CogSci*, 256–61.
- Frank, Michael, Mika Braginsky, Virginia Marchman, and Daniel Yurovsky. n.d. “Variability and Consistency in Early Language Learning: The Wordbank Project.”