How to learn recursive rules:
Productivity of prenominal adjective stacking in English and German

Lydia Grohe[1], Petra Schulz[1], Charles Yang[2]
[1]Goethe-University Frankfurt, [2]University of Pennsylvania
grohe@lingua.uni-frankfurt.de, P.Schulz@em.uni-frankfurt.de, charles.yang@ling.upenn.edu

Although the ability for recursion may be innately available [1], languages differ in the depth and structures of recursion in specific syntactic domains [2], which children must acquire based on experience. For example, prenominal adjectives can be recursively stacked (e.g., *the big red shiny car*) in many languages including English and German, a property, which children master around age 2 [3]. Which learning mechanism enables such early acquisition?

We propose that recursion is an instance of productivity, according to which a rule applies to a category irrespective of lexical identities. In the case of English determiners [4,5], productivity is defined as the structural substitutability of *a* and *the* in combination with nouns. We propose that adjective stacking can be similarly viewed as productivity of adjective placement irrespective of their structural position: for a noun phrase A_1A_2N, if an adjective can appear in position A_1, it can also appear in position A_2 (with the specific positions being determined by possibly universal adjective ordering constraints [6]).

This conception of recursion as productivity enables us to apply learning models such as the Tolerance/Sufficiency Principle [TSP;7]: a rule defined over N lexical items productively generalizes iff $e \leq N/\ln N$ where e is the cardinality of the subset of items *not* attested under the rule. A crucial property of the TSP is that N pertains to the child learner's vocabulary, which is about a few hundred words at age 2 [8,9]. Thus, the evidence for rule productivity must come from a small set of early words, which can be approximated by examining the distributional properties of the most frequent types (here: adjectives) in child-directed input.

For English, we focus on the 49 adjectives in the 550 words known to typical 3-year-olds [10]. We use a part-of-speech tagger to extract "A_1A_2N" sequences from a 5.5-million-word child-directed English corpus. All 49 adjectives appear in either A_1 or A_2 position, of which only 3 fail to appear in both, trivially clearing the TSP threshold ($49/\ln 49=13$). A_1 and A_2 are fully substitutable: adjective stacking is productive and recursive. For German, we analyze five child-directed corpora (CHILDES, 3.5-million words). We focus on the 40 most frequent adjectives and extract all "A_1A_2N" sequences). 38 of the 40 adjectives appear in either A_1 or A_2 position, of which only 7 fail to appear in both, clearing the TSP threshold ($38/\ln 38=10$). We conclude that the productivity of English and German adjective stacking can be rapidly acquired on a distributional basis from Level 1 data.

Our approach lends itself to other NP-structures including recursive PP embedding [11]. This way of distributional learning predicts that a rule is either infinitely recursive or must stop at level one. The latter can be detected if an insufficient number of lexical items are structurally substitutable, as in German possessives (**Marias Nachbars Buch,*'Maria's neighbour's book'), hitherto an unresolved problem for the theory of recursion and its acquisition [2], but cf. [13]. We leave open whether placement preferences [6] are also acquired by probabilistic learning models [12] or whether they are hard-wired.

500 words

References

1. Berwick & Chomsky (2016). *Why only us.* MIT Press, Cambridge MA.
2. Roeper (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics*, *5*(1-2), 57-86.
3. Bar-Sever, Lee, Scontras & Pearl (2018). Little lexical learners: Quantitatively assessing the development of adjective ordering preferences. In *Proceedings of the 42nd annual Boston university conference on language development*, 58-71.
4. Valian, Solt & Stewart (2009). Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of child language*, *36*(4), 743-778.
5. Yang (2013). Ontogeny and phylogeny of language. *PNAS, 110(16), 6324-6327.*
6. Cinque (2010). *The syntax of adjectives: A comparative study.* MIT press, Cambridge MA.
7. Yang (2016). *The price of linguistic productivity.* MIT Press, Cambridge MA.
8. Hart & Risley (1995). *Meaningful differences in the everyday experience of young American children*. Brookes, Baltimore MD.
9. Szagun, Steinbrink, Franik & Stumper (2006). Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory. *First Language* 26(3), 259–280
10. Carlson, Sonderegger & Bane (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159-180.
11. Pérez-Leroux, Peterson, Castilla-Earls, Béjar, Massam & Roberge (2018). The acquisition of recursive modification in NPs. *Language*, *94*(2), 332-359.
12. Yang (2002). *Knowledge and learning in natural language.* OUP, New York, NY.
13. Li, Grohe, Schulz & Yang (in press). The distributional learning of recursive structures. In: *Proceedings of the 45th annual Boston University Conference on Language Development.*