

Productivity and Recursion in English Compounding

Charles Yang (UPenn)

charles.yang@ling.upenn.edu

Problem English noun-noun compounding comes in at least two varieties (Lieberman & Sproat 1992): simple compounds (e.g., *engine oil*) and argument-predicate compounds (e.g., *engine driver*) in which the first noun and the second, deverbal, noun form a clear syntactic and semantic relation. Both are productive and recursive, and children acquire them very early (Clark et al. 1985, Gordon 1985): naturalistic examples from CHILDES are given in (1). Clearly, the productivity of these structures has to be learned from the distributional evidence. As is familiar, compounding is not productive across the board even in English: for example, adjective-noun compounds (e.g., *sweetheart*, *blackmail*, typically written as a single typographical word) are not productive and generally have non-compositional meanings (Adam 1978).

Proposal We propose that recursion be understood as *substitutability* of two positions in a formal structure (Li et al. 2020). In the case of compounds such as (2), the structure is recursive if a noun that appears in N_1 can also appear in N_2 (and/or vice versa). This view of recursion allows for the acquisition of infinite recursion on the basis of level-one data. It also enables the application of the Tolerance Principle (TP; Yang 2016) which has proven robust in accounting for children’s acquisition of productivity.

The TP asserts that a process defined over a set of N items in the input is productive iff the cardinality of the subset not attested under the process is no more than $N/\ln N$. Crucially, the TP allows for a proportionally higher level of exception when N , i.e., the child’s vocabulary, is small. Thus, if a sufficiently number of nouns in the learner’s vocabulary are substitutably attested in N_1 and N_2 , the child concludes that compounding is productive.

Results We extracted noun-noun sequences in a 13-million-word annotated corpus of child-directed English from CHILDES. Table 1 and 2 summarize the distribution of the most frequent nouns—and thus the most likely items in children’s vocabulary—as they appear in compounds. The productivity is impossible to miss under the TP (or indeed any reasonable learning model).

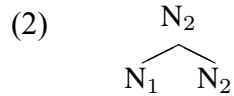
By contrast, adjective-noun compounds cannot be productive. Based on the largest collection of English compounds (Gagné et al. 2019), we found only 65 unique adjective-noun compounds in the input corpus, which involve only 44 unique adjectives (Table 3). It is impossible for adjectives to reach the threshold for productive compounding. Similar conclusions can be reached for verb-noun (e.g., *turncoat*, *pickpocket*) compounds.

Conclusion Compounding can be learned distributionally by an independently motivated principle of generalization. Furthermore, recursion can be reduced to the simple notion of structural substitutability. In the talk, we will also discuss how the headedness and stress pattern of compounds are similarly acquired, and how the semantic interpretative regularity of compounds is at least partly independent from their formal productivity.

(1) Innovative and recursive noun-noun compounds in child English (CHILDES contains numerous examples)

a. cat food machine, dragon water bug (dragonfly), mommy daddy room (parents’s bedroom), water fountain toy

b. picture taker (camera), diaper changer (referring to parent), kid driver (school bus driver that drives children), house seller (real estate agent), raisin keeper (raisin container), television tape recorder (VCR)



| Top N | Needs | In N ₁ | In N ₂ |
|-------|-------|-------------------|-------------------|
| 50 | 38 | 49 | 49 |
| 100 | 79 | 99 | 99 |
| 150 | 121 | 149 | 149 |
| 200 | 163 | 199 | 196 |

Table 1. Almost all nouns in early vocabulary are attested in both N₁ and N₂.

| Top -er | Needs | In N ₂ |
|---------|-------|-------------------|
| 50 | 38 | 50 |
| 100 | 79 | 100 |
| 150 | 121 | 150 |
| 200 | 163 | 169 |

Table 2. Almost all -er deverbals appear in compounds.

| Top adj. | Needs | In compound |
|----------|-------|-------------|
| 50 | 38 | 6 |
| 100 | 79 | 16 |
| 150 | 121 | 22 |
| 200 | 163 | 27 |

Table 3. Very few adjectives appear in compounds.

References

- Adams, V., 1978. *An introduction to modern English word-formation*. Routledge.
- Clark, E.V., Gelman, S.A. and Lane, N.M., 1985. Compound nouns and category structure in young children. *Child development*, 84-94.
- Gagné, C.L., Spalding, T.L. and Schmidtke, D., 2019. LADEC: the large database of English compounds. *Behavior research methods*, 51(5), 2152-2179.
- Gordon, P., 1985. Level-ordering in lexical development. *Cognition*, 21(2), 73-93.
- Li, D., Grohe, L., Schulz, P. and Yang, C. 2021. The distributional learning of recursive structures. *Proceedings of BUCLD 45*. 471-485.
- Lieberman, M. and Sproat, R., 1992. The stress and structure of modified noun phrases in English. *Lexical matters*, 131-181.
- Yang, C., 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.