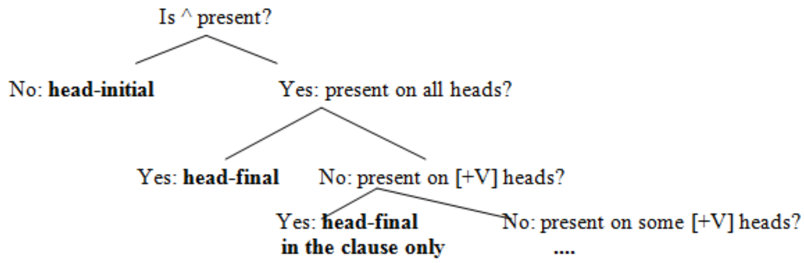
Timothy Bazalgette University of Cambridge

An algorithm for lexicocentric parameter acquisition.

Under the lexicocentric view of syntax (c.f. Baker 2008's "Borer-Chomsky conjecture"), parametric variation is viewed as simply involving differences in the features of lexical items, with the properties of FLN (e.g. Merge, Agree) being invariant. This is an attractively minimalist perspective, but as Roberts and Holmberg (2010) note, it predicts unconstrained microvariation of a multitude of independent parameters, making it difficult to explain e.g. diachronic stability of macroparameters, and implicational relationships between parameters, as well as placing a large burden on the acquirer. Following Roberts (2007), they propose this tension between descriptive and explanatory adequacy may be resolved using the notion of generalisation of the input – the learner systematically tries to quantify features over the largest possible class of heads, leading to emergent hierarchies of microparameters of the following form, which can also be viewed as learning pathways (see also Roberts 2012; in this particular hierarchy "^" is an abstract feature leading to head-finality):

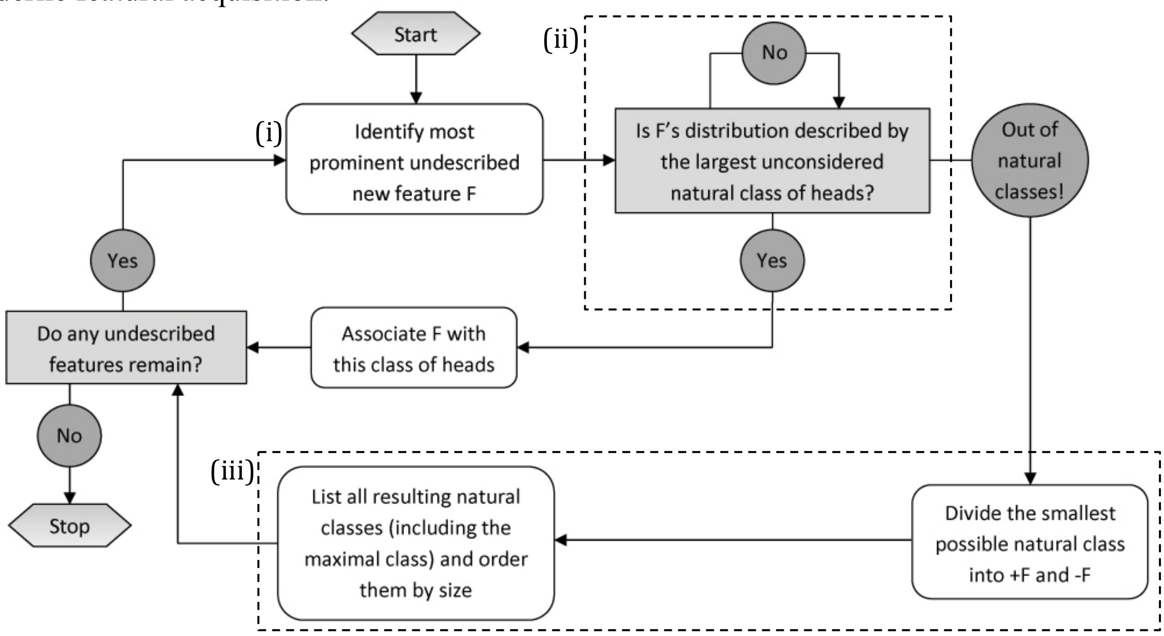
(1)



While this goes some way towards resolving the tension, Biberauer (2011) and Branigan (2012) both note that the top-down nature of the resulting hierarchies is susceptible to superset traps (c.f. Berwick 1985), and suggest a potential resolution: categories are not necessarily pre-given by UG, meaning that at different stages of development, different sets of categories are available to quantify over.

The below algorithm provides a computational model that interprets both of these insights from a radically minimalist perspective, providing a general system that could underlie featural acquisition:

(2)



In (i), a specific feature attested in the data is chosen to be described. Process (ii) then attempts to assign this feature to a natural class that is already in the system, working through them from largest to smallest – it is this step that generates the hierarchies of the kind seen in (1). If, however, no natural classes match the distribution of the feature in question, then process (iii) creates new natural classes to describe the featural distribution seen, which can in turn be made reference to in process (ii) in subsequent loops of the algorithm.

This has the overall effect of taking a multiset of items, each of which has a number of associated properties (the characterisation of the linguistic input), and from this constructing a categorial system that provides a structured representation of these properties, with non-category-defining properties being represented as features associated with natural classes.

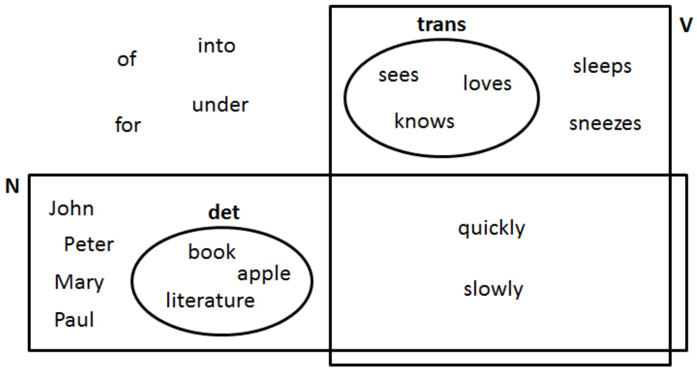
Thus when additionally given a sequential order of features to acquire (presumably resulting from a combination of cognitive biases and overtness in the data; see e.g. Gentner 1982, and Harley and Ritter 2000 for evidence of sequential categorial and featural acquisition) the algorithm deterministically results in a structure, which can be demonstrated using a toy fragment of English:

Items: John, Mary, Peter, Paul, book, apple, literature, sees, loves, knows, sneezes, sleeps, slowly, quickly, of, under, for, into

Features (in order of prominence): N, V, case-assigner, “free”-adjoiner, pied-piping possible, transitive, can take a determiner, proper noun

Supposing (for expositional simplicity, rather than theoretical validity) the distinctive featural category system of Chomsky (1981), the algorithm gives the following output:

(3)



[-N] are case-assigners

[-N,-V] allow pied-piping

[+N,+V] are “freely” adjoining

[+N,-V,-det] are proper nouns

Though much of this example is overly simplistic, the toy grammar nonetheless demonstrates some key properties of the algorithm:

- Where possible, it assigns features to existing natural classes, which reduces the complexity of the system, and may be why many syntactic features (e.g. uCase, vPhi, EPP etc.) seem to be privative rather than distinctive. Such privative features do not create new categorial distinctions, and so hierarchies of the type in (1) are predicted not to interact with one another.
- Features assigned by process (iii) do make categorial distinctions, and so here the order of prominence may affect the output, potentially underlying e.g. microvariation in lexical meanings, with cognitive biases perhaps preventing such variation in syntactic acquisition.
- When categories do need to be distinguished, sub-categories are preferred over top-level ones, which considerably reduces computational load, and also may explain e.g. syntactic sub-categories and potentially feature-geometric structures.
- Distinction of new high-level categories is especially dispreferred late in the acquisition pathway, as this creates even more complexity, which may underlie linguistic tendencies to regularise.

More generally, the advantages associated with the hierarchies in (1) also apply here, meaning that the algorithm can be seen to underlie a wide range of known properties of syntactic variation and change. Furthermore, it is equally applicable to phonological data (e.g. Ito and Mester’s 1994 phonotactic analysis of Japanese can be readily adapted to this approach), yielding further insights, and even to non-linguistic forms of categorisation, and so is a plausible candidate for a naturally-selected cognitive process that is nonetheless a crucial part of FLB – in the sense of Chomsky (2005), a third factor.