

A typology of lexical splits

A key notion in understanding language is ‘possible word’. While some words (lexemes) are internally homogeneous and externally consistent, we find others with splits in their internal structure (morphology) and inconsistencies in their external behaviour (syntactic requirements). I thus apply Booij’s notion of lexical splits to inflectional morphology. I begin by exploring the characteristics of the most straightforward lexemes, adopting the approach of Canonical Typology. In this approach, we push our definitions to the logical limit, in order to establish a point in the theoretical space from which we can calibrate the real examples we find. By defining canonical inflection, we can schematize the interesting phenomena which deviate from this idealization, including suppletion, syncretism, deponency and defectiveness. We can then look at the different ways in which lexemes are ‘split’ by these phenomena.

Consider the French verb *aller* ‘go’, which is split by suppletion. Some of its forms are based on the stem *all-* (as in *allons*), some on *v-* (as in *vont*) and some on *ir-* (as in *irons*). These splits demonstrate that a lexeme’s forms need not have any phonology in common. In this respect, the split is as radical as it could be; in certain other respects there are more remarkable examples. I therefore set out a typology of possible splits, along four dimensions:

- 1) form versus composition/structure of the paradigm: in the French example the split concerns forms only and does not affect the structure of the paradigm; contrast this with the deeper split in the Slovak verb, where different segments of the paradigm are sensitive to different featural requirements (gender is marked in the past: *niesol* ‘he carried’, *niesla* ‘she carried’ but not in the present: *nesie* ‘he/she carries’).
- 2) motivated versus morphology-internal (morphomic): the Slovak split follows a boundary which is motivated from outside the paradigm (it follows tense), while the French split is purely morphology-internal.
- 3) regular versus irregular: splits may be fully regular, extending across the lexicon (all Russian verbs share a featural split), or they may be lexically specified, as we find in Archi (Daghestanian), where particular cells of individual personal pronouns must be specified as taking agreement (while the remaining cells do not).
- 4) externally relevant versus irrelevant: we would expect such splits to be internal to the lexeme, as with English *go~went*, but some have external relevance, in that they lead to different syntactic requirements. Instances include the different alignments found with certain TAM forms in Georgian, and lexemes whose splits bring with them different gender values, as in Czech, Scots Gaelic and the Tromsø dialect of Norwegian.

The typology specifies these four dimensions independently. They are orthogonal to each other, so that the unexpected patterns of behaviour may co-occur in particular lexemes, giving rise to some remarkable examples, especially where periphrasis is involved. These examples show that the notion ‘possible word’ is more challenging than many linguists have realized. And, given the unlikelihood of finding some of the theoretical combinations, the typology proves remarkably complete.

