

## Lexical splits and “complete typologies”

A key notion in understanding language is ‘possible word’. While some words (lexemes) are internally homogeneous and externally consistent, others have splits in their internal structure (morphological paradigm) and inconsistencies in their external behaviour (syntactic requirements). I first analyse the most straightforward lexemes, in order to establish a point in the theoretical space from which we can calibrate the real examples we find. We can then schematize the interesting phenomena which deviate from this idealization, including suppletion, deponency, syncretism and defectiveness. These phenomena have been centre stage for morphologists over the last decade. I now shift the perspective from the phenomena to the different resulting segments into which lexemes can be ‘split’, irrespective of the phenomenon inducing the split. The key point is the dividing line between the two (or more) segments of the lexeme’s paradigm. I set out a typology of possible splits, along four dimensions:

- (i) splits based on the composition/feature signature of the paradigm versus those based solely on morphological form. Thus a Russian verb has two segments: one with a feature signature requiring person and number and one requiring number and gender. This type of split is to be contrasted with one where the feature signature is the same but the morphological form differs (as when one segment has a stem mutation and the other does not);
- (ii) motivated (following a boundary motivated from outside the paradigm, such as singular-plural) versus purely morphology-internal (‘morphomic’);
- (iii) regular, extending across the lexicon, versus irregular (lexically specified);
- (iv) externally relevant versus irrelevant: we expect splits to be internal to the lexeme, but some have external relevance (they require different syntactic behaviours).

I identify instances of these four dimensions separately: they are orthogonal, and therefore not dependent on each other. Their interaction gives a substantial typology, which can be insightfully represented as a Boolean lattice, with 16 possibilities. Drawing on a range of languages, including Archi, Georgian, Kayardild, Krongo and Sanskrit, I demonstrate that the typology is surprisingly complete. All the 16 possibilities specified by the typology are in fact attested.

From the perspective of classical typology, this could be seen as a disappointing outcome: there is no unattested cell whose absence we should justify and attempt to explain. From a canonical perspective, the typology offered a set of possibilities (some of which appeared highly unlikely), and this set indicated the directions in which to look. In a sense, the typology provided the research programme rather than being the result. The fact that a “complete typology” was established is both surprising and significant. Furthermore, since the typology allows for the unexpected patterns of behaviour to overlap in particular lexemes, it helps us to recognize some remarkable examples. Such instances show that the notion ‘possible word’ is more challenging than many typologists have realized.