

“Deep irregularity”: when the expected lexical splits follow the wrong pattern

Greville G. Corbett and Matthew Baerman
Surrey Morphology Group, University of Surrey

We give a broad characterization of irregularity and lexical splits. Against this background we highlight a special type of inflectional irregularity which is concealed under an apparent but spurious regularity.

A useful starting point is the notion of the maximally regular (canonical) lexical item. Internally, the idealized lexical item would be fully consistent: for inflection this means that it would have distinct realizations for all combinations of the relevant morphosyntactic features. And when compared across the lexicon it would be just like other items of its class. Of course, we actually find interesting instances of lexical items which are not consistent internally: they show *splits*. Equally some items stand out from their class in external comparison: they are *irregular*, in the traditional sense. These two notions are significantly related, but are independent of each other.

Consider Russian *čitat'* ‘read’. Its present tense is *čitaju* (1SG), *čítaeš'* (2SG), *čítaet* (3SG), *čítaem* (1PL), *čítaete* (2PL), *čitajut* (3PL), marking person and number. The past tense, however, has the forms *čital* (M.SG), *čitala* (F.SG), *čitalo* (N.SG) and *čitali* (PL), showing gender and number but not person. The lexeme is therefore internally inconsistent; it is split, by the morphosyntactic features to which it is sensitive (person and number vs gender and number). (If we were to specify each form redundantly for all three features, we would produce two large patterns of syncretism, again splitting the lexeme.) The split according to feature sensitivity lines up with a semantic split (tense), and largely with a morphological split, since there are distinct (though in general regularly related) stems for present versus past (plus infinitive). The three splits coincide; in principle they could be distinct.

When we compare *čitat'* ‘read’ externally (across the lexicon), however, we find that every other Russian verb has a similar split in terms of feature sensitivity; thus *čitat'* is fully regular. Contrast this with the verb *exat'* ‘go (by transport)’, present *edu* (1SG), *edeš'* (2SG), *edet* (3SG) ... , with past tense *exal* (M.SG), *exala* (F.SG), *exalo* (N.SG), *exali* (PL). This verb has a semi-suppletive relation between the stems. This is highly irregular, marking out the verb *exat'* from most others, if we compare externally. Importantly, this irregularity splits the verb, but along a familiar line. We can therefore consider irregularity in terms of (i) the phenomena involved (irregular inflection, suppletion, and so on); (ii) the number and type of lexical items covered, or (iii) as a way of investigating how lexemes can split.

We now focus on the typology of these splits. We have established a **first** dimension for the typology: splits may be regular or irregular. A **second** dimension to the typology is that splits may be ‘inward’ or ‘outward’. The splits described so far relate to the lexeme’s internal structure – they have no outward effect. Contrast this with verbs in Georgian where, as part of a complex system, verbs in the aorist may govern different cases from their requirements in other tense/aspect combinations. This split is outwardly relevant (it has syntactic effects), but it is not necessarily irregular. Contrast this with Bosnian / Croatian / Serbian (BCS) where a minority of nouns split singular versus plural in showing irregular stem alternation *and* in requiring different

gender on agreement targets. Such nouns are few. Thus ‘outward’ splits, like ‘inward’ splits, may be regular (Georgian verbs) and irregular (BCS nouns).

A **third** dimension concerns the relation between the parts of the lexeme which are split. Our examples so far have been *motivated* in morphosemantic or in morphosyntactic terms: they have involved a complete tense, or a complete number. But many splits produce no such pattern: they are morphology-internal or *morphomic*. These may involve morphological form, as with the familiar 1PL and 2PL present stem alternation in various Romance languages (Maiden 2004). There is no morphosemantic or morphosyntactic motivation for these forms to be split off. Consider then the French verb ‘go’, in the present: *vais, vas, va, allons, allez, vont*. The forms show suppletion, a relation of maximal irregularity. In another respect, however, the verb is regular: the split runs across a substantial portion of the verb inventory. Thus the distinction motivated versus morphomic also cross-cuts regular versus irregular. Morphomic splits may also involve the distribution of morphosyntactic features: in Kayardild the case inventory is split between ‘normal’ cases and ‘verbalizing’ cases, which mark TAMP features (Evans 1995); this split in the paradigm is not motivated, but it extends right across the lexicon.

Given this typology, we should consider the remarkable case of the Russian verb *xotet’* ‘want’, looking now just at the present tense, where it has these forms:

	SINGULAR	PLURAL
1	xočú	xotím
2	xóčeš’	xotíte
3	xóčet	xotját

This verb shows a split, demonstrated by three separate phenomena. (i) its singular inflections are from the first inflectional class and those of the plural from the second. (ii) there is a consonant alternation in the singular only, while alternations split Russian present tense verbs in

three possible ways (namely 1SG vs the rest, 1SG and 3PL vs the rest, alternation right through the present). And (iii) the stress (´) patterns of singular and plural are not consistent. Each of these three morphological indicators splits singular versus plural. We have a simple, apparently unremarkable singular-plural split. Yet this is deeply irregular. Splits within the present in Russian verbs are always morphomic. By seeming to follow a morphosyntactic motivation, and not invoking a morphomic pattern, *xotet’* ‘want’ goes against all the form classes of Russian verbs. This key example shows a deep type of irregularity, and one which is often overlooked.

Thus lexemes may be split, that is, they may show internal inconsistency. Such a split may or may not be reflected in a partition of the lexical inventory: that is, the split may mark an irregularity or it may not. Splits may have outward relevance, or not; again this distinction cross-classifies with (ir)regularity. Finally, the split within the lexeme may be motivated or it may be morphomic, and this distinction too is orthogonal to (ir)regularity. By splitting along a motivated distinction, an item may in fact be deeply irregular, as in the case of Russian *xotet’* ‘want’.

Corbett, Greville G. 2007. Canonical typology, suppletion and possible words. *Language* 83.8-42.

Evans, Nicholas D. 1995. *A Grammar of Kayardild: with historical-comparative notes on Tangkic*. Berlin: Mouton de Gruyter.

Maiden, Martin. 2004. When lexemes become allomorphs—On the genesis of suppletion. *Folia Linguistica* 38.227–56.

Stump, Gregory T. 2006. Heteroclis and paradigm linkage. *Language* 82.279-322.