Compounding as a universal, and the trouble with ‘word-formation’
Steve Pepper, University of Oslo

Compounding, the combination of lexemes into larger words (as in windmill < ‘wind’ + ‘mill’), is the most frequent method of word formation in many of the world’s languages (Booij 2007). It is easy to learn for children, occurs widely in pidgin languages, and appears to be near-universal (Bauer 2009). Dressler (2006: 23) claims that “compounds are present in all languages of the world (as far as described in grammars)” and (Gagné & Spalding 2006: 145) concur: “Compounding is a common word formation process in all languages.” But how common is compounding, and what alternative strategies are available to languages that disprefer compounding?

This paper addresses those questions for the subset of nominal compounds that consist of two nouns [N+N], which Guevara & Scalise (2009) found to be by far the most frequent type cross-linguistically. It is based on an onomasiological study of noun-noun compounds and their functional equivalents. The function of such constructions is to name a (complex) concept via a strategy of nominal modification (Croft forthc.). This functionally-defined comparative concept (Haspelmath 2018) is called a ‘binominal lexeme’ (or ‘binominal’ for short).

The study is based on a set of 100 meanings that were carefully selected in order to maximize the yield of noun-noun compounds. Translation equivalents were collected from 105 languages representing 70 genera and then subjected to morphological analysis. Those that consisted primarily of two nouns, or a noun and a nominalising affix, were considered to be binominals.

On the basis of the data collected, a total of eight morphosyntactic strategies were identified. They are exemplified by the words in (1), all of which denote the meaning NOSTRIL. (1a) **cmp** represents the strategy of noun-noun compounding; (1b)-(1g) represent the seven alternative strategies. (Note that the labels are mnemonic, not definitional.)

```
(1)  (a) cmp        (b) prp        (c) adj        (d) gen
    Saramaccan     Tagalog      Slovak       Amharic
    nisid-báaku    butas ng ilong nos-ná dierka ye-‘āfinich’a k’edada
    nose-hole      hole LK nose  nose-ADJZ hole  GEN-nose hole

    (d) con         (e) dbl       (f) der        (g) els
    Anindilyakwa   Takia        Central Yupik   Murui Huitoto
    eme-dhvrra emindha  ḥdu-n awa-n  pacig-uaq  defo
    INAL-hole nose   nose-3SG mouth-3SG gills-UAQ¹  nose.CLF(cavity)
```

In this talk I present the typology and characterise each of the binominal types in terms of the form of the nominal constituents and the form, locus and number of markers (cf. Croft 2003 §2.1). I then discuss the frequencies of each type, their areal distribution, and the degree to which they compete within individual languages. I will show that noun-noun compounding is not a universal and that, based on the present sample, other binominal word-formation strategies are preferred in about half of the world’s languages. This prompts questions regarding the status of those other strategies in the context of word-formation as traditionally understood; it exposes a problem inherent in the division of grammar into syntax and morphology, and the need for

---

¹ UAQ: A postbase (affix) in Central Yupik that signifies an “imitation N, thing similar to or reminiscent of N” (Jacobson 2013: 890). Other abbreviations are from the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel 2015) and/or Croft (2003).
field linguists to ensure that binominals do not fall through the cracks (between the two ‘compartments’ of grammar) when documenting strategies for enriching the nominal lexicon.

Stop press: Further analysis has revealed the need for a ninth morphosyntactic strategy, to be announced in Pavia!

References


