On the scope of the form-frequency correspondence principle

Martin Haspelmath
MPI-SHH Jena & Leipzig University
haspelmath@shh.mpg.de

In earlier work (Haspelmath 2008; Haspelmath et al. 2014), I made a strong claim:

(1) The form-frequency correspondence principle in grammar
When two minimally different grammatical patterns (i.e. patterns that form an opposition) occur with significantly different frequencies, the less frequent pattern tends to be overtly coded (or coded with more coding material), while the more frequent pattern tends to be zero-coded (or coded with less coding material).

This is illustrated by a wide range of coding asymmetries: singular/plural, nominative/accusative, affirmative/negative, cardinal/ordinal, present/future, active/passive, same-subject/different-subject, and many others (as first identified and catalogued by Greenberg 1966).

In addition, I also claimed that the reverse also holds: all systematic form asymmetries correspond to frequency asymmetries. However, since cases of ellipsis also qualify as form asymmetries and they are due to contextual predictability rather than to frequency, the more general explanatory statement is that form asymmetries are due to predictability, whether contextual or frequency-based.

In this presentation, am going to discuss a few general issues that arise for this research programme, using concrete examples: (A) What kinds of factors can override the preference for short coding, leading to symmetric coding? (B) Can we treat non-occurrence of a pattern as a special case of coding asymmetry (e.g. the non-existence of certain relativization patterns)? (C) Given that the frequency asymmetries are claimed for meanings, rather than specific forms, is there a way to measure frequency of meanings independently of frequency of forms? (D) Can we find a general factor underlying frequency asymmetries (maybe some kind of “cognitive asymmetry”), which might also explain the form asymmetries? Is there thus a possible alternative to the proposed causal chain (frequency > predictability > shortness of coding)?