

Zero forms in morphology: Patterns and explanations

It has long been known that more frequent forms (i.e. words, constructions, grammatical markers) tend to be shorter than functionally equivalent but less frequent forms. The fact that one value of a grammatical feature can be left unexpressed or "zero-marked" in opposition to the other value(s) has also been related to frequency and economy in the typological literature. However, it is often assumed that zero forms pattern with shorter forms in opposition to longer forms, although this has not yet been tested quantitatively. The goal of this talk is to shed more light on the distribution of zero forms across languages and to discuss possible explanations for the patterns found.

Using data from the UniMorph database, the DoReCo collection and grammar mining, the results suggest that zero markers are not frequently used in morphological inflection paradigms of the world's languages, which may go against the common expectation in typology that zero forms are the default expression for frequent or "unmarked" values of morphological categories. The results show that no cells, neither in nominal nor in verbal paradigms, have a strong association with zero forms; in general, there is a strong crosslinguistic preference for overt exponents. We also find a very high degree of variation across languages and lexemes in the distribution of zero forms.

On the other hand, if we find zero forms, they tend to occur in certain morphosyntactic functions across languages which requires an explanation. It is plausible that frequency plays an indirect - rather than a direct - role in the development of zero forms in such functions. One possible process leading to the development of zero forms is the differential non-development of an overt exponent. In that case, an exponent for a different value of the same grammatical category (e.g. the plural) develops, while the zero form develops in relation to that exponent (e.g. in the singular), as it grammaticalizes. Frequency could play a role in this type of processes in that the functions for which exponents develop are likely to be less frequent, so that speakers start expressing those overtly for successful communication. The function that is eventually encoded by a zero form in opposition then is simply a function that was not sufficiently infrequent to motivate the development of an overt marker. Importantly, frequency does not play a direct role in such cases. Besides this process, zero forms can develop by a number of other diachronic processes. This shows that the distribution of zero forms cannot be accounted for by a single, efficiency-driven process but requires a more fine-grained explanation.