Measuring regularity of plural forms

Rong Chen

A plural form is in two relationships: one is the relationship between singular and plural, while the other is that between plurals. The two relationships are either in harmony, as in English, or in conflict, as in Uchumataqu. Thus, regularity of plural forms can be measured in two dimensions.

In English, the suffix *-s* marked on nouns like *books* is taken as a plural marker for two reasons. One is that noun stems like *book* are usually singular forms. The other reason is that other plural nouns also take the suffix *-s*, e.g. *girls*, *pets* and *trees*. The majority of English nouns marked by *-s* equally have zero-coded singular counterparts. This implies that the relationship between plurals is in harmony with the relationship between singular and plural.

Distinguished from English, the two relationships are in conflict on Uchumataqu (Uru-Chipaya, Bolivia) pronouns. As seen in (1), except for the first person inclusive pronoun, all plural pronouns have the ending *-ka*. Nevertheless, the ending alone is hardly to be a plural suffix, if plural pronouns are compared with their singular forms. Otherwise segments like *-ču-* and *-na-* are not accounted for.

SINGULAR	**PLURAL**	
wir	wisnaka (EXCL)	1st
	učumi (INCL)	
am	amčuka	2nd
ni	inaka	3rd

(1) Uchumataqu personal nouns (Uru-Chipaya, Bolivia; Hanß 2008: 183)

There are two approaches to linguistic universals. One is the data-driven approach. With a little knowledge of interested linguistic phenomena, probably some intuition, typologists gather information from language descriptions. Subsequently, they generalize typological universals based on inductive reasoning. On the opposite, typologists can propose missing links between linguistic structures, based on previously accumulated knowledge or observations. Next, typologists gather cross-linguistic data and test hypothetical universals.

A hypothetical universal is proposed regarding the relationship between regularity and length difference as follows.

If regularity of plural forms increases, then their length differences with singular forms grow.

Guided by this hypothetical universal, regularity of plural forms is measured only by the relationship between singular and plural. A probability sample of 100 independent languages is made to test this hypothetical universal.

When measuring plural forms by the relationship between singular and plural, the morphophonemic connection between singular and plural plays a deterministic role. According to decomposability of plural forms, regularity of plural forms is measured at three levels: a) fully

decomposable, in other words, plural forms are directly developed from singulars, b) either decomposable or indecomposable, sharing the stem or arbitrary phonological materials respectively, c) absolutely indecomposable, showing no connection with singulars.

On the other hand, length of a linguistic expressions is measured by the number of segments. Length differences between singular and plural are results of the operation that subtract the length of singular forms from the length of plural forms.

Since the hypothetical universal propose a positive relationship between regularity and length, the statistic method of linear regression is adopted to measure the relationship. Because regularity as the predictor variable is categorical, the technique of dummy coding is used to transform values of the regularity variable into dichotomous variables. The issue of interest becomes a situation in which there are three predictors and one outcome. Given the cross-linguistic data, we can get a linear regression model which has significant p-values. Furthermore, the β values suggest simultaneous increase in both regularity and length difference. Therefore, the hypothetical universal is strongly supported by the data.