Valery Solovyev

Is Grammochronology Possible?

Up to now, only lexical, but not grammatical information has been used for determining the age of language groups. This can generally be explained by two reasons. The first one is the lack of grammatical descriptions of various languages in a systematic and standartized fashion which would allow for statistical methods to be applied. The second reason is the widespread opinion that grammatical features change very irregularly across different languages.

The first of these problems can be considered solved with the appearance of large typological databases like WALS and *Jazyki Mira* [1]. As for the second, the opinion that grammars change irregularly is mostly a subjective statement and requires verification.

In this report the results of several preliminary studies based on the *Jazyki Mira* database are presented. The most important goal at the current stage is to develop a methodology for research in this direction.

We use the standard Hamming distance as a measure of differences between languages, i. e. the number of features whose values are different for the compared languages. If the speed with which grammatical features change is significantly different for various languages, then the distances between them will vary substantially even for languges with the same divergence date.

The average distances between languages among 9 different genetic groups of approximately the same age of 2-3 thousand years are as follows: Indo-Iranian — 242, Italic — 195, Celtic — 215, Germanic — 226, Balto-Slavic — 234, Finno-Ugric — 234, Turkic — 193, Mongolian — 158, Tunguso-Manchurian — 177. The data show that while there is a difference between the distances, it is not as large as to exceed 50%.

The speed with which the lexicon of different languages changes is also not entirely regular. Calculations based on Starostin's adjusted formula [2] show that 4 to 6 words from the 100-item Swadesh list change in different languages during a period of 1000 years. This means that the variation here is also limited to no more than 50%.

Therefore, it would seem that grammatical changes occur at a similar rate in different languages. One could probably receive a better result by counting not all of the grammatical features contained in a database, but only a subset of the most stable ones. Preliminary data on the stability of features for WALS and *Jazyki Mira* can be found in [3,4].

An important part of Swadesh's approach was the notion that the rate of temporal change is static. Sergey Starostin has later adjusted this proposition [2] by introducing a formula with a nonlinear dependence between the number of lexical changes and time. The dependence of the speed of grammatical change is probably more complex and determining it is a task to be accomplished in the future.

The main difference between lexicon and grammar is that while the number of words is virtually limitless, the number of grammatical features is relatively low: the *Jazyki Mira* database contains 3821 such features. Therfore, a language evolving over time in the limit space of grammatical features would inevitably return to its earlier state — these are so-called *back mutations*, which are almost impossible in lexical evolution [5]. A possible mathematical model for this process would perhaps be the movement of points in non-Euclidean space (a hyperboloid in Lobachevsky's geometry). Grammochronology, if established, could be useful in determining the age of language families and macrofamilies on greater time depths than glottochronology.

Preliminary data shows that grammatical change can be considered suitable for determining the times of divergence for languages, although a lot of work still needs to be done for an adequate mathematical model to be created.

References

1. Polyakov V., Solovyev V. Computer Models and Methods in Typology and Comparative Linguistics. Kazan: Kazanskiy Gosudarstvennyy Universitet. 2006.

2. Starostin S. Comparative-historical linguistics and Lexicostatistics. In: Historical linguistics and lexicostatistics. Melbourne, 1999. pp. 3-50.

3. Wichmann S., Holman E.M. Assessing temporal stability for linguistic typological features. http://email.eva.mpg.de/˜wichmann/WichmannHolmanIniSubmit.pdf

4. Polyakov V., Solovyev V., Wichmann S., Belyaev O. Using WALS and Jazyki Mira. (submitted).

5. Nakhley L., Ringe D., Warnow T. Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. Language, v.81, pp. 382-420.