

George Starostin

(*Russian State University for the Humanities, Moscow / Higher School of Economics, Moscow / Santa Fe Institute, New Mexico*).

## Chinese basic lexicon from a diachronic perspective: implications for lexicostatistics and glottochronology

Languages with lengthy attested written histories are crucial for a better understanding of the tendencies and peculiarities of the historical evolution of basic lexicon, one of the most diachronically important layers of any language. In this respect, the study of Chinese is particularly crucial not just because it is one of the very few languages in the world whose written history goes back more than three thousand years, but also because it represents the only example of a (still living) non-Indo-European language with such a deep history. Consequently, detailed studies of the rates of lexical change across written Chinese monuments, the distributions of the changes across various sub-layers of the basic lexicon, and possible reasons for these changes, may all yield significant insights into the more general picture of lexical evolution as a whole.

Unfortunately, studying lexical change across different chronological layers of Chinese is also a complex challenge for the scholar, beset with a host of philological, semantic, and pragmatic problems that are inevitable for languages with closed text corpora as a whole, and exacerbated for Chinese texts, with their script ambiguities and mixtures of colloquial and literary language registers, in particular. This means that any such study will be flawed by definition, and requires a very high level of familiarity with different parts of the corpus for optimal results. Nevertheless, even today we are in a better position to analyze gradual diachronic changes in Chinese basic lexicon than we were a decade ago, thanks to the presence of digitalized corpora that may be relied upon instead of (or in addition to) more traditional dictionaries.

In my talk, I will describe the procedure and demonstrate the results of the creation of new Swadesh-type 110-item wordlists for four distinct chronological layers of Chinese: Early Zhou Chinese (as represented by the texts of the earliest literary monuments, the *Shàngshū* and the *Shījīng*), Classic Old Chinese (primarily represented by texts of the Lù school of Confucianism — *The Analects* and *Mencius*), Middle Old Chinese (represented by one of the most notoriously colloquial-oriented texts of the Táng era, the *Records of Linji*), and Modern Literary Chinese (data from modern so-called "dialects" of Chinese will be occasionally considered, but does not constitute a primary object of this study). I will discuss several technical and theoretical problems that are encountered during the assembly process of the wordlists, address the issue of marking cognations between potentially or "partially" related items, and present some new findings on rates of change between these four stages, which can also be viewed in comparative perspective alongside data from Indo-European languages with attested histories.