

Towards the computational simulation of sound change in Indo-European

Most applications of computational linguistics in the diachronic research on language concentrate on creating programs that find cognates, align the phonetic and semantic similarities between words in the lexicon, establish sound correspondences between languages and reconstruct the proto-forms (cf. Hewson 1974, Kondrak 2002, Steiner-Stadler-Cysouw 2011). The other prevalent approach is that of the grouping of different languages in terms of their ancestry by investigating shared innovations, i.e. cladistics (cf. Ringe-Warnow-Taylor 2002, Chang-Cathcart-Hall-Garrett 2015). Along these two trends there has also been the problem of 'historical derivation' which received some attention in the past (cf. Kondrak 2002: 12-15). The aim of such approach was to use the computer program to process the input data through a set of regular expressions modeled on the sound changes which are usually assumed to have occurred in the development of the languages being analyzed. The computer program applied the sound changes in their relative chronology to the input database of the language and generated the next phase of its development which was then compared with the actual attested data (e.g. Latin as input data, sound changes to Spanish as rules and the generated output compared to the actual attested material of Modern Spanish). This approach was pursued basically in order to describe the sound changes within their relative chronology, for didactic purposes and for testing the hypothesis of the regularity of sound change.

The previous approaches to the problem of the 'historical derivation' (i.e. the computational simulation of sound changes) were numerous: Smith 1969 for Russian from Proto-Indo-European, Burton-Hunter 1976 for Old French from Latin, Eastlack 1977 for Ibero-Romance from Latin, Hartman 2003 for Spanish from Latin, Kondrak 2002: 141-143 for Polish from Proto-Slavic but only took into account the basic forms of words leaving aside the whole paradigms.

The purpose of this talk is to describe and evaluate the programs created so far, discuss the possible modifications which can be made for further applications and present the results of the application of several Indo-European sound changes to the dataset of forms, including the inflectional paradigms. Additionally, the application of historical derivation procedures to the database of the reconstructed Proto-Indo-European language will be discussed including the possible modeling of morphological changes as well as the phonological ones. Such an approach could be helpful in formulating of the hypotheses concerning the typology, relative chronology and tendencies of sound change and also of analogical leveling based on fairly complete empirical data. The results could confirm the existing theories on sound change and analogical remodeling to a certain degree and could form the basis for the future electronic historical grammars and etymological dictionaries. The method itself could be applied to the other language families and thus it could contribute to the research on the universal tendencies of sound change.

Burton-Hunter, Sarah K. 1976. "Romance etymology: a computerized model". *Computers and the Humanities*, 10:217-220, 1976.

Chang, Will - Cathcart, Chundra - Hall, David – Garrett, Andrew. 2015. "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis", *Language* 91 (2015) 194-244

Eastlack, Charles L. 1977. "Iberochange: a program to simulate systematic sound change in Ibero-Romance". *Computers and the Humanities*, 11:81-88, 1977.

Hartman, Lee. 2003. "Phono (Version 4.0): Software for Modeling Regular Historical Sound Change". In: Leonel Ruiz Miyares, Celia E. Álvarez Moreno, and María Rosa Álvarez Silva (eds.), *Actas: VIII Simposio Internacional de Comunicación Social: Santiago de Cuba, 20-24 de Enero del 2003*, I, 606-609.

Hewson, John. 1974. "Comparative reconstruction on the computer". In: *Proceedings of the 1st*

International Conference on Historical Linguistics, p. 191-197.

Kondrak, Greg. 2002. *Algorithms for Language Reconstruction*. Unpublished Ph.D. Thesis.
University of Toronto.

Ringe, Don - Warnow, Tandy - Taylor, Ann. 2002. "Indo-European and computational cladistics".
Transactions of the Philological Society 100:1, pp. 59-129.

Smith, Raoul N. 1969. "A computer simulation of phonological change". *ITL: Tijdschrift voor
Toegepaste Linguïstiek*, 1(5):82-91, 1969.

Steiner, Lydia – Stadler, Peter - Cysouw, Michael. 2011. "A Pipeline for Computational Historical Linguistics".
Language Dynamics and Change 1(1). 89-127.

Dariusz R. Piwowarczyk
Assistant Professor of Classics
Jagiellonian University Kraków
Poland

Fulbright Fellow at the Department of Linguistics
University of Pennsylvania, Philadelphia