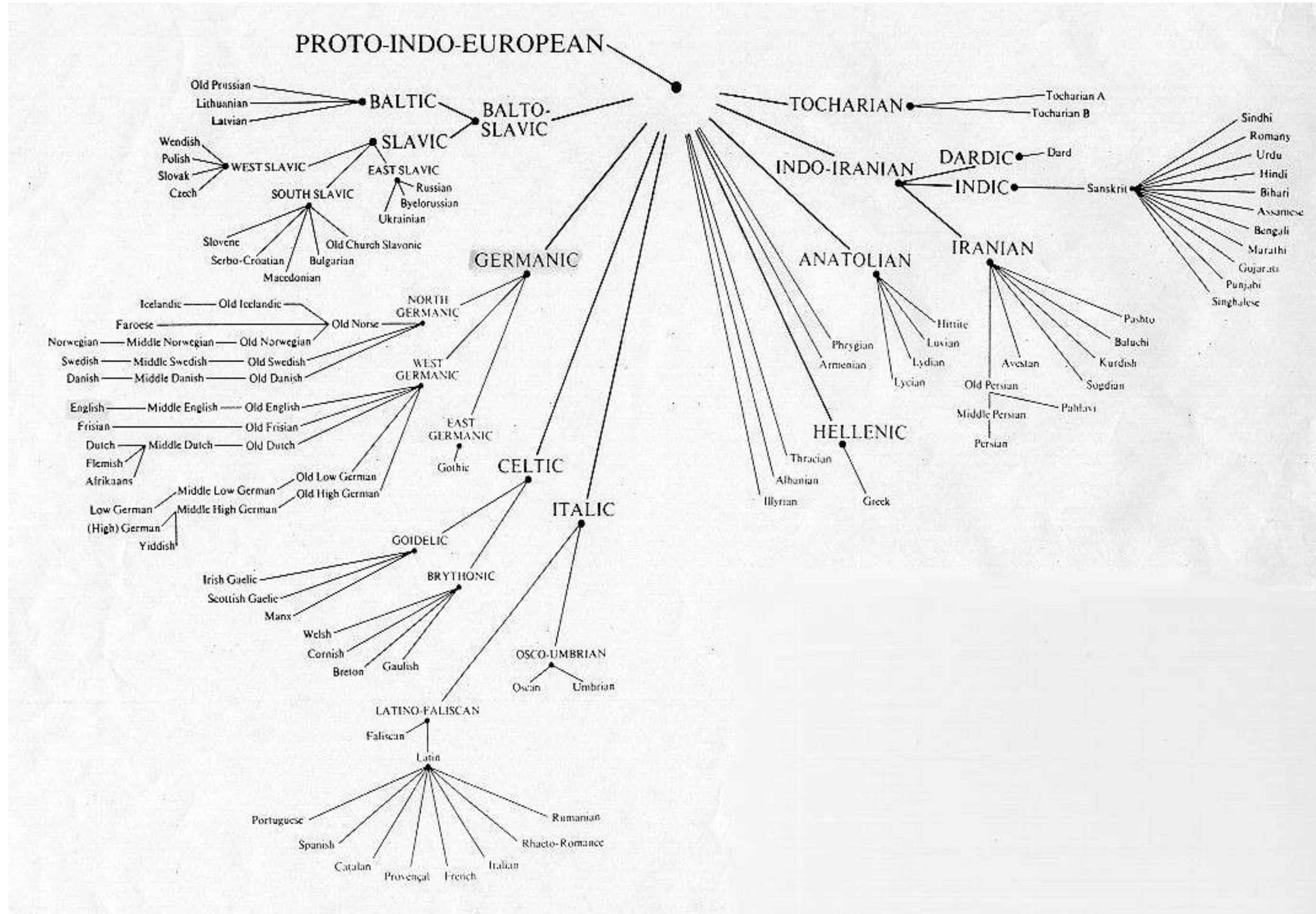


# Developing Algorithms for Computational Comparative Diachronic Historical Linguistics

Over time, languages change by regular, systematic processes. It is possible, by looking at the state of a language now and in the past, to deduce the exact changes that occurred, and the order in which they occurred. These changes also split languages, therefore it is also possible to, by using modern languages as input, induce the probable structure of their parent language.

I have developed algorithms to form a web of languages, similar to that to the side, given a set of modern languages. The languages would be the leaf nodes on the tree.

They function recursively. Given a group of languages, one algorithm forms a proto-language, a sort of average from which all of the others are probably derived. The algorithm then splits the group into different groups of languages more closely related to each other than the proto-language. Each of these groups is treated in the same fashion until each group contains a single language.



I have successfully run my web-formation algorithms on various words in eight languages: English, Dutch, Danish, German, Portuguese, Romanian, Scots Gaelic, and Polish. They somewhat successfully separated the languages on individual word basis, and if all of the results were combined and analyzed, a coherent and correct picture of their relations to each other could be created.

This process can be applied to any abstract structure, not merely languages. This makes it useful outside of my project, but if any use is going to be made of it in practical historical linguistics, it must be applied to languages.

The ideal way to deal with language historically is purely phonetically, only through sound. As traditional alphabets are woefully ill-conceived, it is necessary to create an organized system to store phonetic data.

I developed a 5-dimensional array structure, generally following the lines of the International Phonetic Alphabet (IPA).

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

Phonetic data can also be used in lower-level attempts to compare languages to each other, but direct phonetic comparison and building from the bottom-up has proved far less effective than building language relations from the top down, which requires far less intense phonetic computation.

Daniel Wright  
2005  
TJHSST  
Computer Systems Lab