Phylogenetics of the Germanic languages

Keith Briggs   `Keith.Briggs@bt.com`

BT Innovate & Design, Adastral Park, Martlesham IP5 3RE

2012 January 16 3:40 P.M.

The aim is to fit trees to the set of early Germanic languages (Old German, Dutch, Frisian, Flemish, English, Norse etc.), and to study robustness of these trees with respect to choice of distance metric, feature subset, etc.

A novel feature of the work would be to use data on comparison of languages which is non-lexical (i.e. grammatical features are taken into account, not just vocabulary). All previous work in this area has been lexical only. The required data is all in the book Hans Nielsen, Old English and the continental Germanic languages (2nd ed., Innsbruck 1985).

Previous work: the best work is by Geoff Nicholls in Oxford, who has agreed to act as external advisor. The Warwick co-supervisor will be Dario Spano. Papers by Geoff Nicholls are available here: http://www.stats.ox.ac.uk/~nicholls/.

There might be difficulties:

1. decoding the highly technical material in Nielsen's book will be hard. Also finding a way to encode it for input to phylogenetic algorithms.

2. the easier algorithms work only on a distance matrix (http://evolution.genetics.washington.edu/phylip/software.html). However, these are now thought to be not good enough for application to language evolution and Nichols and others are now using Markov chain Monte-Carlo methods and other methods. It is probably not possible to get into these more difficult methods in a mini-project.

The techniques learnt would be applicable in the future to mathematical biology.