Contents

Introduction
  Why is evolutionary linguistics interesting?
  Quantitative linguistics
  The Germanic languages
Data
Methods
  Maximum parsimony
  MCMC
Results & conclusions
Evolutionary linguistics

Charles Darwin offered languages as an illustrative example of evolution.

Languages show analogies to genetic features: mutation and inheritance.

The history of languages has a close correspondence to the history of humanity.

The origin of language ↔ the origin of modern humanity?
A brief history

Schleicher’s tree model

(Langauge families)

("Language stock")

(Ancestral language)
A brief history

Schleicher’s tree model of Indo-European

[Diagram showing the relationships between languages, starting from Proto-Indo-European (a) branching into German (Deutsch), Lithuanian (Lituauisch), Slavic (Slawisch), Celtic (Celtisch), Italic (Italisch), Albanian (Albanesisch), Greek (Griechisch), and Iranian (Iranisch).]
A brief history

Schmidt’s wave model
A brief history

Schmidt’s wave model
A brief history

Schmidt’s wave model
A brief history

Schmidt’s wave model
A brief history

Schmidt’s wave model - The Balkan Sprachbund

4 Indo-European language families (Greek, Romance, Albanian, Slavic) and the unrelated Turkish

Share many grammatical (and lexical) features not seen elsewhere
A brief history

Real linguistic evolution is driven by a combination of these processes

Analogous to genetic evolution: *inheritance* versus *lateral transfer* (in viruses)

Inheritance is dominant in sparsely populated regions, lateral transfer becomes important when there is much contact between unrelated languages

*(Strong influence of technology: writing, printing, internet...)*
Challenges facing evolutionary linguists

A (nearly) total absence of historical data!

Analysis must depend on observation of modern (i.e. written) languages, plus (more recently) modelling
Methodology-Swadesh lists

Very common for analyses to be based on lexical data: **Swadesh lists**

List of 100 common words thought to be particularly resistant to replacement by loanwords

- **German** (Meer, See)
- **English** (sea)
- **Italian** (mare)
- **Russian** (more)
- **Norwegian** (sjø)
- **Gaelic** (muir)
- **Dutch** (zee)
- **Greek** (thalassa)

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Quantitative linguistics

Swadesh lists allow for construction of a “genome” for languages

<table>
<thead>
<tr>
<th></th>
<th>Dutch</th>
<th>English</th>
<th>Gaelic</th>
<th>German</th>
<th>Italian</th>
<th>Norwegian</th>
<th>Russian</th>
<th>Greek</th>
</tr>
</thead>
<tbody>
<tr>
<td>M*r</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>S*</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

This is then used with similar machinery as used to compare amino acid or DNA sequences
What if non-lexical data are used?
Data set - the early Germanic languages

- Goths (C2nd - )
- Vikings (C8th)
- Angles (C5th)
- Saxons (C5th)
- Jutes (C5th)
- Franks (C4th)
- “Germans” (BC)
Data set - the early Germanic languages

Old English dialects
- Anglian
- West Saxon
- Kentish

Old Frisian

Old Saxon

Old High German

Gothic

Old Norse
Data set - the early Germanic languages

A classic data set...

Old English dialects
- Anglian
- Kentish
- West Saxon

Old Frisian
Old Norse
Gothic
Old High German
Old Saxon

Schleicher’s classification

Proto-Germanic

Gothic

German
Norse

Low German
“in a broader sense”

High German

Saxon
Frisian

Old Saxon

English
Dutch
Low German

Old English and the Continental Germanic Languages: A Survey of Morphological and Phonological Interrelations

Hans Frede Nielsen

Phylogeny of the early Germanic languages

Dario Papavassiliou
Sample entry:
“The [Indo-European genitive singular] ō-stem ending -ās is reflected in Gothic gibōs, ON skarar, OS geba
and OHG geba, but not in OE giefe and OFris. ieve,
where the original suffix has been analogically
replaced by the [dative singular] ending ([reflecting
Indo-European] -āī)...”
Data set - interpretation as binary genome

Sample entry:
“The [Indo-European genitive singular] ő-stem ending -ās is reflected in Gothic gibōs, ON skarar, OS geba and OHG geba, but not in OE giefe and OFris. ieve, where the original suffix has been analogically replaced by the [dative singular] ending ([reflecting Indo-European] -āi)...”

<table>
<thead>
<tr>
<th></th>
<th>Reflects IE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gen</td>
</tr>
<tr>
<td>OE Anglian</td>
<td>0</td>
</tr>
<tr>
<td>OE Kentish</td>
<td>0</td>
</tr>
<tr>
<td>OE W Saxon</td>
<td>0</td>
</tr>
<tr>
<td>O Frisian</td>
<td>0</td>
</tr>
<tr>
<td>O Saxon</td>
<td>1</td>
</tr>
<tr>
<td>O H German</td>
<td>1</td>
</tr>
<tr>
<td>O Norse</td>
<td>1</td>
</tr>
<tr>
<td>Gothic</td>
<td>1</td>
</tr>
</tbody>
</table>

Phylogeny of the early Germanic languages

Dario Papavassiliou

Wednesday, 3 September 2014
Data set - interpretation as binary genome

Missing data marked with ?
Omitted data (duplicate entries, “insignificant/late”, too subtle) marked with - and disregarded

Results in a ‘genome’ of 531 characters for each language

Can be filtered into sub-genomes for different linguistic categories (nouns, verbs, numerals..., vowels, consonants) and (in principle) weighted
A very basic indication of the completeness of the data.
A very basic indication of the completeness of the data

Gothic under-represented (due to a lack of Gothic sources)

Old English dialects over-represented (due to subject of book)
Since book focuses on relationships between languages it does not discuss traits seen in only one language.

Traits seen in all, or none, of the species are uninformative.

Flat distribution $\rightarrow$ timescale of evolution is long.
**Statistics - distance matrix**

Form distance matrix by counting differences in genome

Some relationships immediately apparent
Minimal spanning tree

A very crude quantification of distances between languages

Construct a full graph with edge weights defined as distance

Delete edges with large weight to give minimal spanning tree
Methods

Maximum parsimony

Minimises number of changes over tree to obtain observed genomes

Implemented using the *Fitch algorithm*

Repeated for each character in genome, then for each possible tree topology
Maximum parsimony

Unless ancestral state is a leaf, the tree is *unrooted*
Maximum parsimony

Unless ancestral state is a leaf, the tree is unrooted
Maximum parsimony

Unless ancestral state is a leaf, the tree is *unrooted*

Gothic chosen as outgroup due to distance from other languages
Methods

Maximum parsimony

Gives a sensible tree topology, but unrooted tree → cannot resolve EG/WG/NG split!

Gives only information on topology, not chronology

Dario Papavassiliou
Phylogeny of the early Germanic languages
Wednesday, 3 September 2014
Methods

Markov chain Monte Carlo - Dollo model

Evolution modelled as a collection of Poisson processes:
- Trait **born** with rate $\lambda$  
- Trait **dies** with rate $\mu$  
- Lineage **splits** with rate $\theta$  

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Markov chain Monte Carlo - Dollo model

Evolution modelled as a collection of Poisson processes:
- Trait **born** with rate $\lambda$
- Trait **dies** with rate $\mu$
- Lineage **splits** with rate $\theta$

**Catastrophe** occurs with rate $\rho$: each trait dies with $P(\kappa)$,
$\text{Poisson}(\kappa\lambda/\mu)$ new traits born

Equivalent to an edge lengthening

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Methods

Markov chain Monte Carlo - Implementation

Implemented using the TraitLab package*

MCMC scheme example moves

Change tree topology

*Geoff Nicholls, Oxford

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Methods

Markov chain Monte Carlo - Implementation

Implemented using the TraitLab package*

MCMC scheme example moves

*Geoff Nicholls, Oxford

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Markov chain Monte Carlo - Implementation

Implemented using the *TraitLab* package*

MCMC scheme example moves

Vary model parameters

Vary locations of catastrophes

*Geoff Nicholls, Oxford*
Markov chain Monte Carlo - Implementation

1,000,000 steps performed

First 100,000 discarded (equilibration)

Remaining sampled every 100 steps

Samples averaged to give a consensus tree
Given a set of $N$ trees, a consensus tree representing an ‘average’ topology is constructed:

Root node ➔ Most common split ➔ ... 

$\times \%$
Results

Obtain same tree (topologically) as from parsimony

Chronological resolution groups NG with WG

Very good consensus between samples
Results

Morphology

Phonology - vowels

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
We obtain the following phylogeny...
Results

Conclusions

Compared (mostly) favourably to Schleicher’s classification

as well as quantitative (lexical) analyses by others

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Conclusions

Compares (mostly) favourably to Schleicher’s classification as well as quantitative (lexical) analyses by others.

Dario Papavassiliou
Phylogeny of the early Germanic languages
Wednesday, 3 September 2014
Conclusions

*Phonetic, particularly vocalic, data emphasise later contact...*

Proto-Germanic

- Gothic
- Old Norse
- Anglian
- Old English
- West Saxon
- Kentish
- Old Frisian

Old German: Continental

Old High German

Old English

North Sea Germanic

Old Norse

Old Frisian

Old Saxon
Conclusions

*Phonetic, particularly vocalic, data emphasise later contact...*

...criterion to determine breakdown of phylogenetic model?

Dario Papavassiliou

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Thanks

Keith Briggs
UWE & BT Research

Dario Spanò
Warwick

Geoff Nicholls
Oxford

Dario Papavassiliou
Phylogeny of the early Germanic languages

Wednesday, 3 September 2014
Thanks

Keith Briggs  
UWE & BT Research

Dario Spanò  
Warwick

Geoff Nicholls  
Oxford

Phylogeny of the early Germanic languages

Wednesday, 3 September 2014