

Phylogenetische Methoden in der historischen Linguistik

Die IELex-Datenbank
Maximum Parsimony

Gerhard Jäger
Forum Scientiarum
2. Dezember 2014

Kognate

- Kognate Wörter haben sich aus derselben Wortform entwickelt
- z.B. *It. filius* → *frz. fils* / *it. figlio*
- Faustregel: je näher zwei Sprachen verwandt sind, um so mehr Kognatenpaare gibt es

IELex

- *Indo-European Lexical Cognacy Database*
- Öffentlich zugänglich unter <http://ielex.mpi.nl/>
- Basiert auf der älteren Kruskal-Dyen-Datenbank
- Geleitet von Michael Dunn am Max-Planck-Institut für Psycholinguistik Nijmegen

IELex

- Swadesh-Listen mit 207 Konzepten für 157 indoeuropäische Sprachen
- z.T. Orthographie und/oder phonetische Transkription (uneinheitlich)
- Jeder Eintrag wird einer **Kognatenklasse** zugeordnet

Kognatenklassen

- Konzept 'mountain':
 - **Klasse A:** armenisch *sar*, serbisch/russisch/polnisch *gora*, cz, slowak., ukr. *hora* ...
 - **Klasse B:** dt. *Berg*, fries. *berch*, dän. *bjerg*, ...
 - **Klasse C:** alban. *mal*
 - **Klasse D:** armen. *ler*
 - **Klasse E:** panjabi *par*, nepali *parbat*, marathi *parwat* ...
 - **Klasse F:** engl. *mountain*, fr. *montagne*, it. *monte*, breton. *menez*, gäl. *monadh...*
 - ...
- Elemente der selben Kognatenklasse sind untereinander kognat;
Elemente verschiedener Klassen sind nicht kognat
- Zuordnung basiert auf Expertenurteilen (nicht automatisierbar)

Kognaten als phylogenetische Charaktere

- Durch Bedeutungswandel ändern sich Kombination Sprache/Konzept/Kognatenklasse
- z.B. Konzept 'bone':
 - ahd. *Bein* (Kognatenklasse B) → nhd. *Knochen* (Kognatenklasse G)
- Vergleichbar zu einer biologischen Mutation

Kognaten als phylogenetische Charaktere

- Kognatenklassen können wie biologische *Charaktere* behandelt werden
- Allerdings: Häufig Synonyme Einträge
 - z.B. 'guts' → dt. *Eingeweide, Gedärm*
 - Keine eindeutige Zuordnung von Charakterzuständen zu Sprachen
- Alternative: **Binarisierung**

Binarisierte Charaktere

- Jede Kognatenklasse (für jedes Konzept) ist ein Charakter
- Zwei mögliche Zustände: 0/1
 - 0: Sprache L verwendet kein Element der Kognatenklasse für das Konzept
 - 1: Sprache L verwendet ein Element der Kognatenklasse für das Konzept
- Wechsel wie ahd. *ubil* → *nhd. schlecht* entspricht also zwei Mutationen

Praktische Fragen

- Gutes Programm für Maximum-Parsimony-Analyse: **paup***
- Verfügbar von
http://people.sc.fsu.edu/~dswofford/paup_test/

Praktische Fragen

- Starte paup* von dem Verzeichnis aus, in dem die Nexus-Datei liegt

```
> execute Ielex_binarizedFull.nex
```

```
> Hsearch
```

```
Do you want to increase 'Maxtrees'? (Y/n): Y
```

```
Enter new value for 'Maxtrees' (100): 10000
```

```
Action if limit is hit:
```

```
(1) Prompt for new value
```

```
(2) Automatically increase by 100 (= AUTOINC)
```

```
(3) Leave unchanged, and don't prompt: 2
```

```
> SaveTrees file='ielexFull_MP.tree' format=Newick brlens=yes
```

```
> q
```

Das Newick-Format

- Siehe http://en.wikipedia.org/wiki/Newick_format
- Darstellung mit **Dendroscope** (<http://ab.inf.uni-tuebingen.de/software/dendroscope/>) oder **SplitsTree** (<http://www.splitstree.org/>)

Konsens-Methoden

- Häufig ergibt eine phylogenetische Inferenzmethode mehr als einen Kandidaten-Baum (manchmal über 1000)
- Konsens-Bäume und -Netzwerke sind kompakte Repräsentationen einer Menge von Bäumen