

The Biological Lumbering of Linguistic Trees

A Comparison of Traditional and “New School” Approaches for Phylogenetic Reconstruction in Historical Linguistics

Jens Fleischhauer, Johann-Mattis List
Heinrich Heine University Düsseldorf

2010/07/03

Structure of the Talk

Introduction

Proof of Genetic Relationship

- Lexical Similarities

- Sound Correspondences

Methods for Phylogenetic Reconstruction

- The Comparative Method

- Lexicostatistics

- The Method of the “New School”

Comparison of the Methods

- Similarities

- Differences

Conclusion

“Old” and “New School” Approaches

- ▶ Historical processes are inferred from synchronic structures.
- ▶ This is a general inference pattern, not unique to linguistics but also used in e.g. biology and geology (cf. Christy 1983).
- ▶ Intra- and interdisciplinarily different methods for such an inference exist.
- ▶ Marris (2008) claims that the methods of historical linguistics are “Old School” while the biological methods (applied to language data) are “New School”.

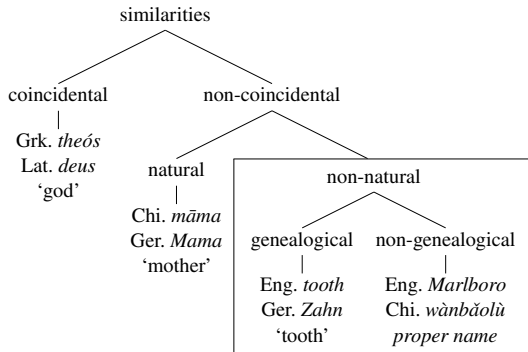
“Old” and “New School” Approaches

- ▶ Historical processes are inferred from synchronic structures.
- ▶ This is a general inference pattern, not unique to linguistics but also used in e.g. biology and geology (cf. Christy 1983).
- ▶ Intra- and interdisciplinarily different methods for such an inference exist.
- ▶ Marris (2008) claims that the methods of historical linguistics are “Old School” while the biological methods (applied to language data) are “New School”.
- ▶ **Leading question of the talk:** What are the differences between “Old” and “New School”?

Proof of Genetic Relationship

- ▶ Lexical Similarities
- ▶ Sound Correspondences

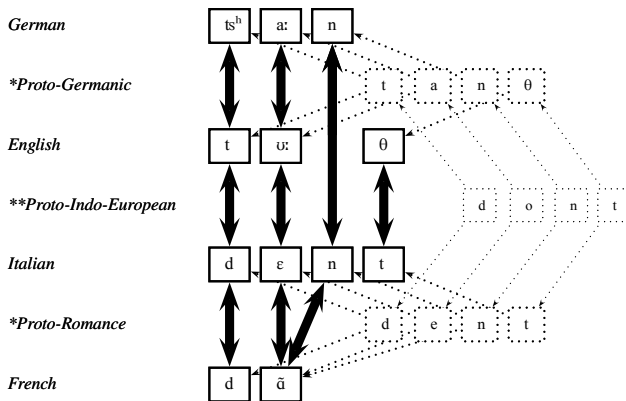
Lexical Similarities



Sound Correspondences

- ▶ In genetically related languages we can find a certain amount of semantically similar words which are structurally similar with sounds occurring regularly in the same position of the words. (cf. Trask 2000: 336)
- ▶ As opposed to other kinds of language change, sound change happens to be quite *regular* affecting *large parts of the lexicon* of a language. (cf. Trask 1996: 52-101)

Sound Correspondences



Methods for Phylogenetic Reconstruction

- ▶ The Comparative Method
- ▶ Lexicostatistics
- ▶ The Method of the “New School”

The Comparative Method

Key assumptions of the comparative method:

- ▶ Innovations in languages which are not reflected in other genetically related languages indicate that the respective languages have evolved separately.
- ▶ Shared innovations allow to reconstruct a phylogenetic tree which depicts the process of how an ancestor language split into several descendents.

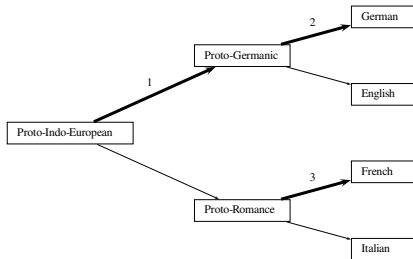
The Comparative Method

Working procedure:

1. **Proof of Genetic Relationship:** Carry out a comparative analysis of languages previously assumed to be genetically related. Search the languages for possible cognates, thereby identifying regular sound correspondences and carrying out putative reconstructions.
2. **Shared Innovations:** Search the languages for shared innovations.
3. **Phylogenetic Reconstruction:** Reconstruct a language tree which explains the identified shared innovations of the different subgroups in a most parsimonious way.

The Comparative Method

Phylogenetic Reconstruction:



No.	Innovation	Languages	Example
1	First Germanic Consonant Shift	Germanic Languages	PIE *p > PGM *f
2	Second Germanic Consonant Shift	High German	PGM *p > GER pf
3	Vowel Syncopation	Western Romance	LAT cinērēm > FRE cendre

Lexicostatistics

Key assumptions of lexicostatistics:

- ▶ The lexicon of every human language contains words which are relatively resistant to borrowing and relatively stable over time due to the meaning they express: these words constitute the *basic vocabulary* of languages.
- ▶ The process of replacement of words belonging to the realm of *basic vocabulary* is reflected in the amount of shared cognates in genetically related languages.
- ▶ Shared cognates in the *basic vocabulary* of genetically related languages reflect their degree of genetic closeness and allow to reconstruct their phylogeny.

Lexicostatistics

Working procedure:

1. **Swadesh-List Compilation:** Compile a list of basic vocabulary items (a Swadesh list).
2. **Swadesh-List Translation:** Translate the items into the languages that shall be investigated.
3. **Cognate Judgments:** Search the language entries for cognates.
4. **Cognate Percentages:** Compute percentages of shared cognates for every language pair.
5. **Subgrouping:** Construct a graphical representation out of the information on percentages of shared cognates (this is usually, but not necessarily, a genealogical tree).

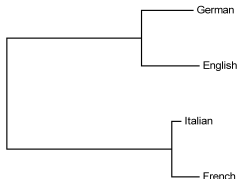
Lexicostatistics

Cognate Judgments:

Basic Concept	German	ID	English	ID	Italian	ID	French	ID
HAND	Hand	1	hand	1	mano	2	main	2
BLOOD	Blut	3	blood	3	sangue	4	sang	4
HEAD	Kopf	5	head	6	testa	7	tête	7
TOOTH	Zahn	8	tooth	8	dente	8	dent	8
TO SLEEP	schlafen	9	sleep	9	dormir	10	dormir	10
TO SAY	sagen	11	say	11	dire	12	dire	12
...

Lexicostatistics

Phylogenetic Reconstruction:



10

	German	English	Italian	French
German	100	82	40	38
English	82	100	40	36
Italian	40	40	100	94
French	38	36	94	100

The Method of the “New School”

Key assumptions of the “New School” approach:

- ▶ Borrowing is rare in basic vocabulary.
- ▶ There are significant similarities between linguistic and biological evolution which allow the same methods to be applied for phylogenetic reconstruction.
- ▶ The distribution of cognate sets over a sample of languages can be used to model linguistic evolution as a process of gain and loss within a phylogenetic tree.

The Method of the “New School”

Working procedure:

1. **Swadesh-List Compilation:** Compile a list of basic vocabulary items (a Swadesh list).
2. **Swadesh-List Translation:** Translate the items into the languages that shall be investigated.
3. **Cognate Judgments:** Search the language entries for cognates.
4. **Binarization of Cognate Information:** Convert the data into a binary matrix reflecting for each cognate set its presence (1) or absence (0) in the respective language.
5. **Subgrouping:** Use phylogenetic software to construct a phylogenetic tree which explains the distribution of cognate-sets best.

The Method of the “New School”

Binarisation of Lexicostatistical Wordlists:

Basic Concept	German	ID	English	ID	Italian	ID	French	ID
HAND	Hand	1	hand	1	mano	2	main	2
BLOOD	Blut	3	blood	3	sangue	4	sang	4
HEAD	Kopf	5	head	6	testa	7	tête	7
TOOTH	Zahn	8	tooth	8	dente	8	dent	8
TO SLEEP	schlafen	9	sleep	9	dormir	10	dormir	10
TO SAY	sagen	11	say	11	dire	12	dire	12
...

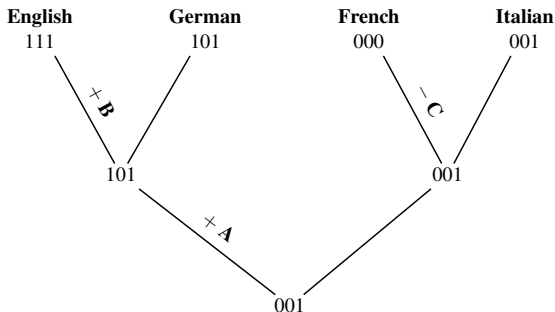
The Method of the “New School”

Binarisation of Lexicostatistical Wordlists:

ID	Proto-Form	Basic Concept	German	English	Italian	French
1	PGM *xanda-	HAND	1	1	0	0
2	LAT mānus	HAND	0	0	1	1
3	PGM *bloḍa-	BLOOD	1	1	0	0
4	LAT sanguis	BLOOD	0	0	1	1
5	PGM *kuppa-	HEAD	1	0	0	0
6	PGM *xawbda-	HEAD	0	1	0	0
7	LAT tēsta	HEAD	0	0	1	1
8	PIE *h ₃ dont-	TOOTH	1	1	1	1
9	PGM *slēpan-	TO SLEEP	1	1	0	0
10	LAT dormīre	TO SLEEP	0	0	1	1
11	PGM *sagjan-	TO SAY	1	1	0	0
12	LAT dicere	TO SAY	0	0	1	1
...

The Method of the “New School”

Phylogenetic reconstruction:



Char.	English	German	French	Italian
A	1	1	0	0
B	1	0	0	0
C	1	1	0	1

Comparison of the Methods

- ▶ Similarities
- ▶ Differences

Similarities

- ▶ inference of language phylogenies from synchronic data
- ▶ reconstruction of evolutionary trees depicting the historical processes
- ▶ inference of language splits

Differences

The comparative method vs. lexicostatistics and the “New School” approach:

- ▶ qualitative vs. quantitative data
- ▶ innovations vs. character distributions

Lexicostatistics vs. the “New School” approach:

- ▶ replacement vs. gain/loss

Different Theories – Same Problems

- ▶ Conflicting data (as a result of undetected borrowing events or linguistic convergence) are problematic for all methods.
- ▶ The evolutionary process is supposed to be tree-like and reticulate evolution is rejected.

The Need for a New “New School” Approach

- ▶ Contradicting Marris (2008), the “New School” approach is not superior to the “Old School” approaches.
- ▶ A real “New School” account should allow to solve the well known problems and abandon the *a priori* assumption of tree-likeness.

The End

Thanks for Your Attention!