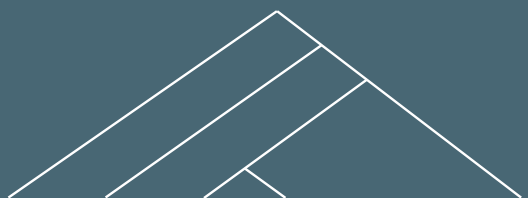
Theoretical (In)Compatibilities between the Comparative Method and Cladistics

Erik Elgh, Department of Linguistics and Philology, Uppsala University

DfGS 2021



Phylogenetic analyses

- Computer assisted
- ~20 years old in linguistics (e.g. Gray & Atkinson 2003, Holden 2002, Rexová et al. 2003)
- Adapted from biology
- Utilize data generated by the comparative method

Phylogenetic analyses

Contents

- The Comparative Method and Issues
- Computer assisted phylogeny from biology
- Data from the comparative method in CAP
- Possible issues ”spilling over” from CM

The Comparative Method

- outline and issues

The Comparative Method

English	Sanskrit	Greek	Latin	Gothic
foot	pad-	pod-	ped-	fotus
father	pitár-	patér	pater	fadar
nephew	nápat	-	nepos	-

Examples modified from Campbell 2013, p. 136.

The Comparative Method

English	Sanskrit	Greek	Latin	Gothic
foot	pad-	pod-	ped-	fotus
father	pitár-	patér	pater	fadar
nephew	nápat	-	nepōs	-

Examples from Campbell 2013, p. 136.

The Comparative Method

English	Sanskrit	Greek	Latin	Gothic
foot	pad-	pod-	ped-	fotus

Sound correspondence set:

f	p	p	p	f

The Comparative Method

English	**Sanskrit**	**Greek**	**Latin**	**Gothic**
f	p	p	p	f

/p/ is reconstructed as the ”proto-sound”.

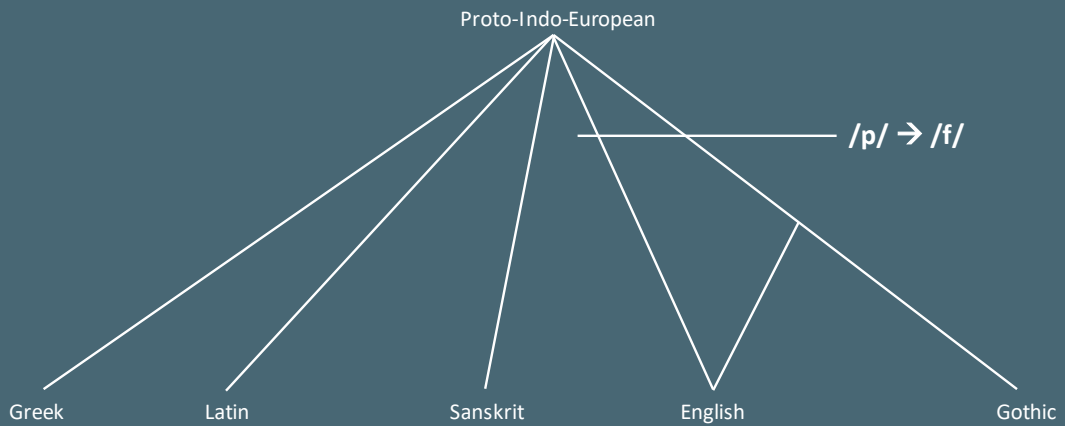
The Comparative Method

English	Sanskrit	Greek	Latin	Gothic
f	p	p	p	f

/p/ is reconstructed as the ”proto-sound”.

English and Gothic thus share an innovation, or apomorphy, and can be grouped together in a clade.

The Comparative Method



The Comparative Method

1. Combined homology hypotheses - which words and sounds are homologous?

2. Ancestral state reconstruction for the MRCA

3. Estimate phylogeny by shared apomorphies

4. Improve reconstruction and phylogeny assessment

Ancestral State Reconstruction

- The reconstruction depend on the tree

In fact:

- ”[…] the tree being used is the true tree […]” – Omland 1999

- Reconstructions *without* a tree topology must lead to assumptions of said topology

Ancestral State Reconstruction

English	Sanskrit	Greek	Latin	Gothic
f	p	p	p	f

/p/ is reconstructed as the ”proto-sound”.

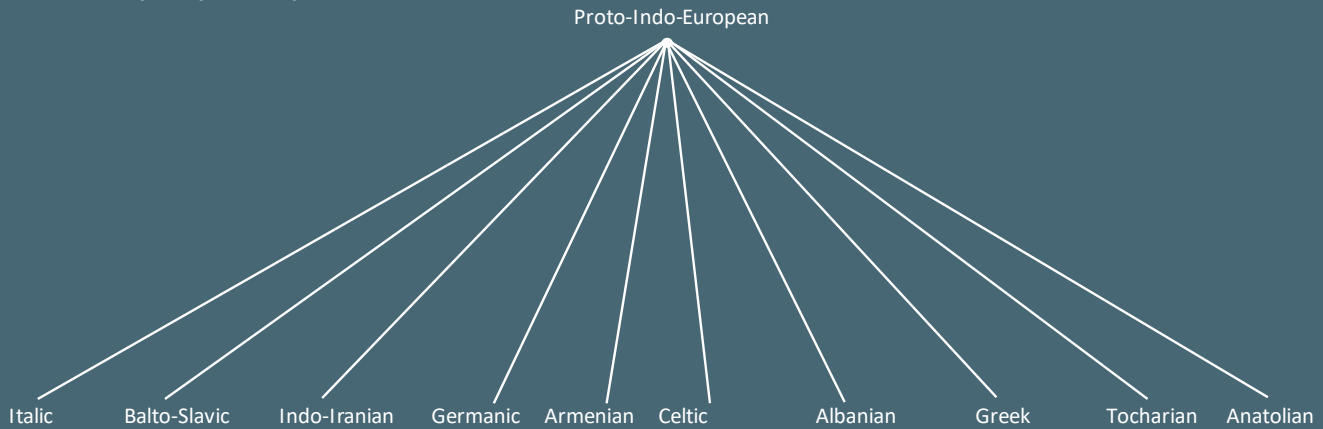
Ancestral State Reconstruction

English	Sanskrit	Greek	Latin	Gothic
f	p	p	p	f

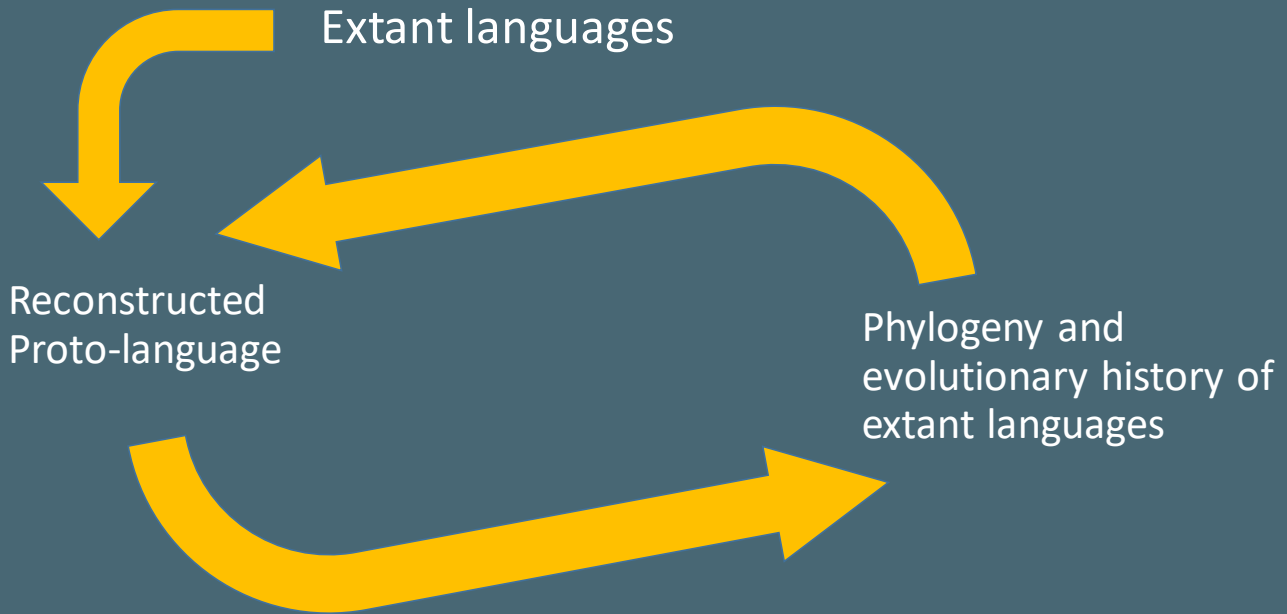
/p/ is reconstructed as the ”proto-sound”.

Possible effects on results?

Giant polytomy:



Possible effects on results?



Computer assisted phylogeny

- practices in biology

Computer assisted phylogeny - Biology

1. Vertebrate scapula
2. Variation in the Scapula → Character states
3. Infer phylogeny
4. Ancestral state of scapula in the MRCA and other internal nodes

Computer assisted phylogeny - Biology

1. Homology hypotheses a) - which characters are homologous?
2. Homology hypotheses b) – which character states are homologous?
3. Estimate phylogeny by tree search and optimality criteria
4. Ancestral state reconstruction

Computer assisted phylogeny - Biology

No assumption of tree topology in the coding (ideally)

Phylogeny & Ancestral States - Comparison

Comparative Method

1. Combined homology hypotheses - which words and sounds are homologous?
2. Combined homology hypotheses - which words and sounds are homologous?
3. Ancestral state reconstruction for the MRCA
4. Estimate phylogeny by shared apomorphies

Biology

1. Homology hypotheses a) - which characters are homologous?
2. Homology hypotheses b) – which character states are homologous?
3. Estimate phylogeny by tree search and optimality criteria
4. Ancestral state reconstruction

Phylogeny & Ancestral States - Comparison

Cognate data (homologous words) come without assumptions of topology:

English	**Sanskrit**	**Greek**	**Latin**	**Gothic**
foot	pad-	pod-	ped-	fōtus
f	p	p	p	f

→ /p/ is reconstructed

Phylogeny & Ancestral States - Comparison

Comparative Method

1. Combined homology hypotheses - which words and sounds are homologous?
2. Combined homology hypotheses - which words and sounds are homologous?
3. Ancestral state reconstruction for the MRCA
4. Estimate phylogeny by shared apomorphies

Biology

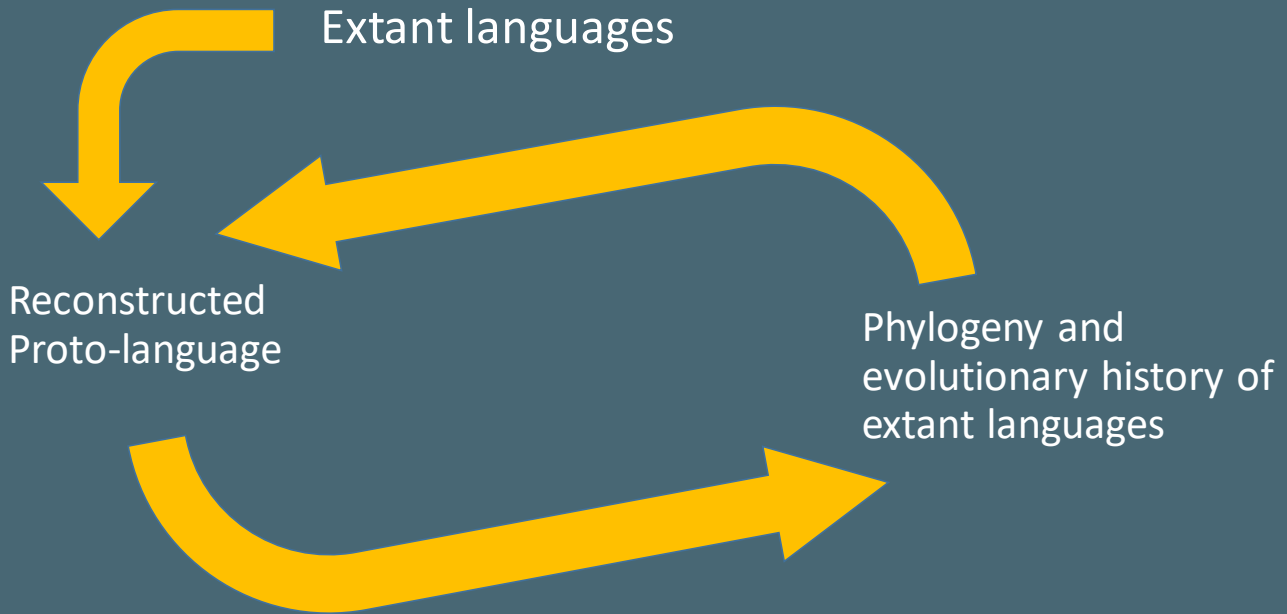
1. Homology hypotheses a) - which characters are homologous?
2. Homology hypotheses b) – which character states are homologous?
3. Estimate phylogeny by tree search and optimality criteria
4. Ancestral state reconstruction

Problem or not?

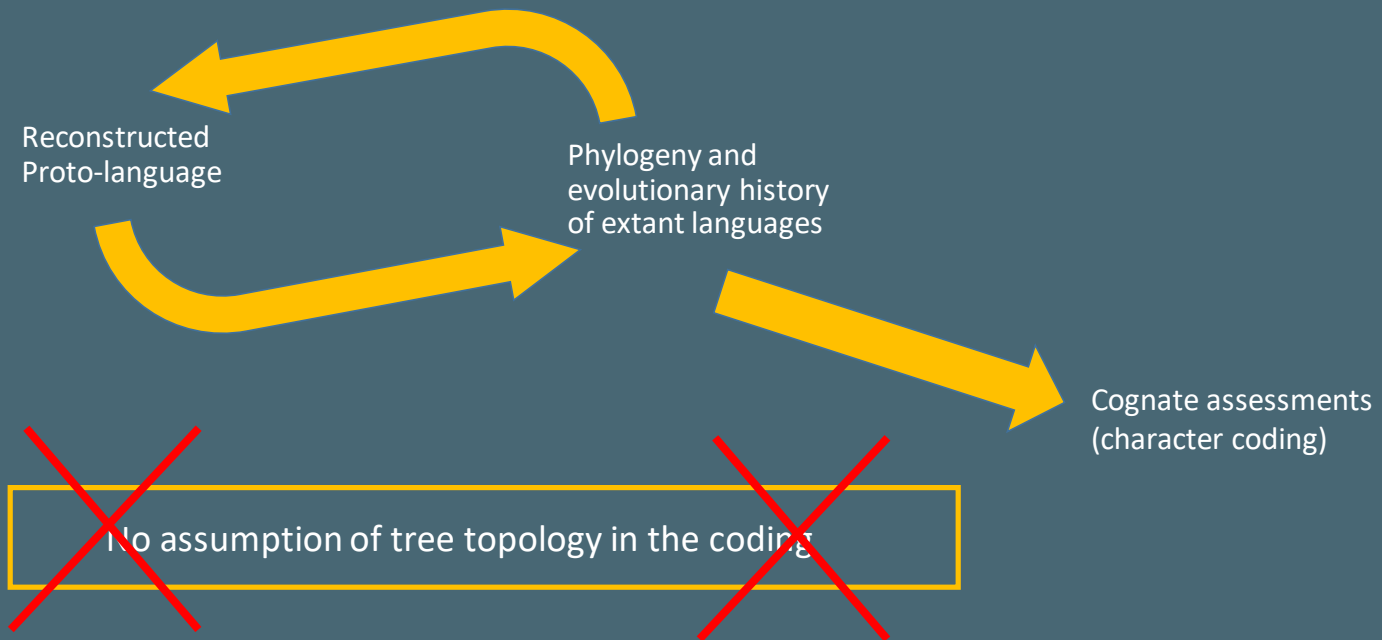
Phylogeny & Ancestral States - Circularity

1. Combined homology hypotheses - which words and sounds are homologous?
2. Ancestral state reconstruction for the MRCA
3. Estimate phylogeny by apomorphies
4. Improve reconstruction and phylogeny assessment

Phylogeny & Ancestral States - Circularity



Phylogeny & Ancestral States - Circularity



Phylogeny & Ancestral States - Circularity

Cognate data (homologous words) come without assumptions of topology

- This is true only if solely based on joint hypothesis of cognates and sound correspondences
- What are the effects if that isn’t the case?

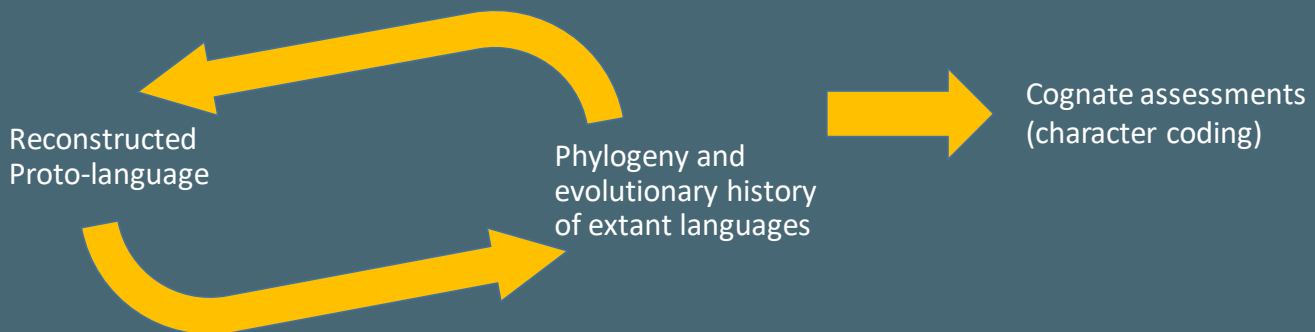
Possible effects on results?

- Wrong tree
  - Topology congruent with assumptions
  - Unclear what this means

Possible effects on results?

English	Sanskrit	Greek	Latin	Gothic
f	p	p	p	f

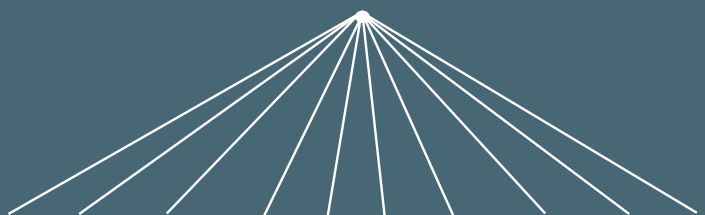
/p/ is reconstructed as the ”proto-sound”.



Possible effects on results?

- Wrong tree
  - Topology congruent with assumptions?
  - Unclear what this means

- Polytomy?



Conclusions

- Comparative method crucial for homology assessments
- When used for reconstruction it comes with assumptions on tree topology
- Reconstructions depend on the tree → don't use them to estimate the tree
- Cognate data (homologous words) can be used if they are independent of the reconstructed proto-language
- Next step: Do this in practice

Conclusions

- Next step: Do this in practice:

    1. Use cognate assessments done without reconstructions and circular ”improvement”

    2. Run phylogenetic analysis

    3. Reconstruct ancestral states

Conclusions

- Why?

    1. Circularity moved higher up in the hierarchy of assumptions

    2. Hypotheses that are more testable

    3. (hopefully) Better tree and thus better reconstructions

Thanks to:

The audience

Philipp Rönchen, Harald Hammarström for discussions and input

Virginia Panara for help with layout

Members of Systematic Biology research programme, Department of Organismal Biology, Uppsala University for listening to and commenting on an earlier version of this presentation

References

- Campbell, L. 2013. Historical Linguistics An Introduction. 3rd e. Edinburgh University Press.
- Gray, R., Atkinson, Q. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature 426:435–439.
- Holden, C. J. 2002. Bantu language trees reflect the spread of farming across sub-Saharan Africa: a maximum-parsimony analysis. Proc. R. Soc. Lond. B. 269:793–799.
- Omland, K. E. 1999. The Assumptions and Challenges of Ancestral State Reconstructions. Systematic Biology 48(3):604-611.
- Rexová, K., Frynta, D. and Zrzavý, J. 2003. Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. Cladistics 19:120-127.