Bioinformatische Methoden in der historischen Linguistik

Historische Linguistik:
Die komparative Methode

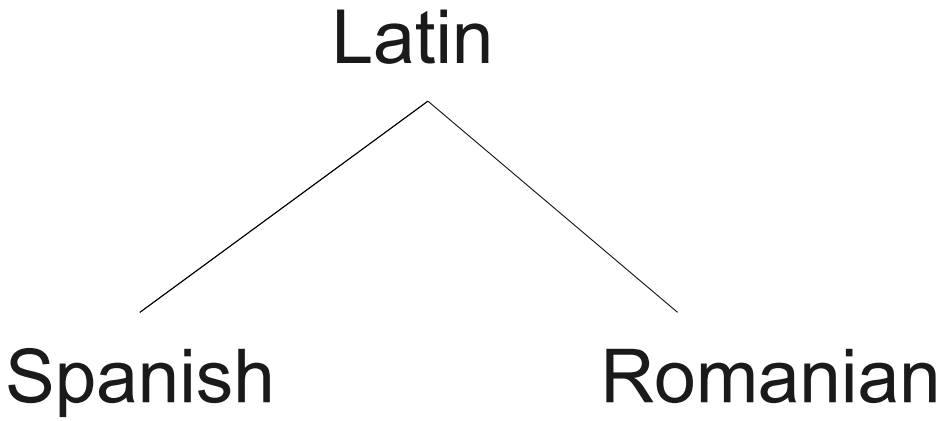
Gerhard Jäger
Forum Scientiarum
18. Januar 2013

Genetic language relationships

- Language communities sometimes split
- Parts undergo different changes
- Simplifying assumption: after a split, daughter languages change on independent trajectories
- In few cases, we have written records
  - lt. *[piskis]* → spn. *[peskado]* / → rom. *[peSte]*
  - lt. *[noks]* → spn. *[noCe]* / → rom. *[noapte]*
  - lt. *[pektus]* → spn. *[peCo]* / → rom. *[pyept]*

Genetic language relationships

- Tree model



Genetic language relationships

- In most cases, we do not have written records of earlier stages
- Regular **sound correspondences** provide evidence for genetic relationship though
  - Correspondences indicate common ancestor + different sound shifts
  - The more **cognates** two languages share and the fewer sound shifts separate them, the closer they are related

Example: Polynesian languages

- Taken from Crowley & Bowern (2010)

TABLE 5.1 Data from Four Polynesian Languages

	Tongan	Samoan	Rarotongan	Hawaiian	
1.	tapu	tapu	tapu	kapu	'forbidden'
2.	pito	pute	pito	piko	'navel'
3.	puhi	feula	puʔi	puhi	'blow'
4.	tafaʔaki	tafa	taʔa	kaha	'side'
5.	taʔe	tae	tae	kae	'feces'
6.	taŋata	taŋata	taŋata	kanaka	'man'
7.	tahi	tai	tai	kai	'sea'
8.	malohi	malosi	kaʔa	ʔaha	'strong'
9.	kalo	ʔalo	karo	ʔalo	'dodge'
10.	aka	aʔa	aka	aʔa	'root'
11.	ʔahu	au	au	au	'gall'
12.	ʔulu	ulu	uru	poʔo	'head'
13.	ʔufi	ufi	uʔi	uhi	'yam'
14.	afi	afi	aʔi	ahi	'fire'
15.	faa	faa	ʔaa	haa	'four'
16.	feke	feʔe	ʔeke	heʔe	'octopus'
17.	ika	iʔa	ika	iʔa	'fish'
18.	ihu	isu	putaŋio	ihu	'nose'
19.	hau	sau	ʔau	hau	'dew'
20.	tafuafi	siʔa	ʔika	hiʔa	'firemaking'
21.	hiku	siʔu	ʔiku	hiʔu	'tail'
22.	hake	aʔe	ake	aʔe	'up'
23.	huu	ulu	uru	komo	'enter'
24.	maŋa	maŋa	maŋa	mana	'branch'
25.	maʔu	mau	mau	mau	'constant'
26.	maa	mala	mara	mala	'fermented'
27.	naʔa	faʔaŋa	maninia	naa	'quieten'
28.	nofo	nofo	noʔo	noho	'sit'
29.	ŋalu	ŋalu	ŋaru	nalu	'wave'
30.	ŋutu	ŋutu	ŋutu	nuku	'mouth'
31.	vaka	vaʔa	vaka	waʔa	'canoe'
32.	vaʔe	vae	vae	wae	'leg'
33.	laho	laso	raʔo	laho	'scrotum'
34.	lohu	lou	rou	lou	'fruit-picking pole'
35.	oŋo	loŋo	roŋo	lono	'hear'
36.	ua	lua	rua	lua	'two'

Guidelines for reconstruction

- Only establish sound correspondences if you are reasonably sure the words are cognate
- Assume sound shifts that are plausible (are known to occur frequently)
- Assume as few sound changes as possible for reconstructing a proto-language
- The reconstructed proto-language should have a typologically plausible sound system

Polynesian example

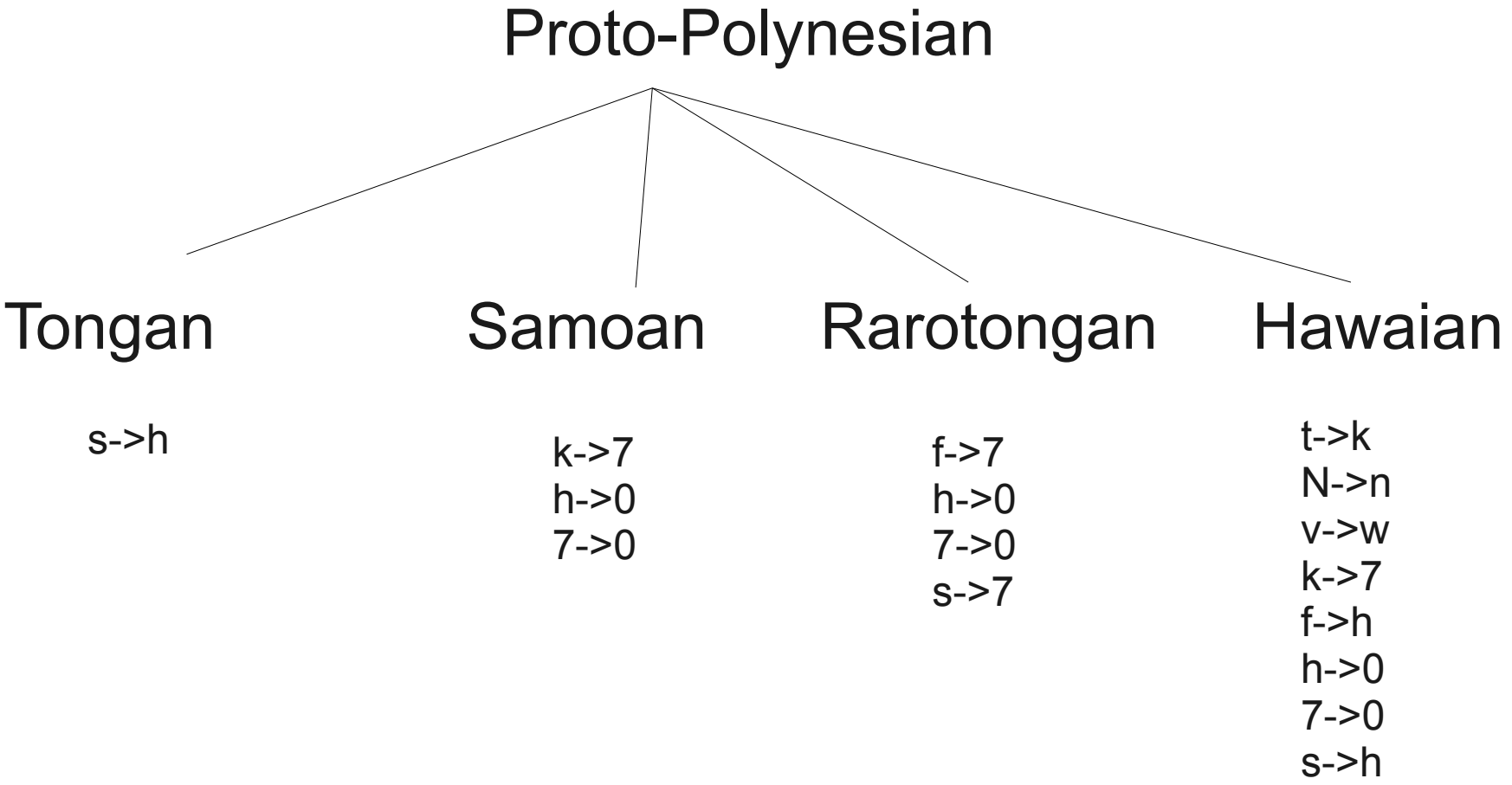
- Vowels in Proto-Polynesian are unchanged in daughter languages (otherwise we would stipulate unnecessary sound shift)
- Likewise, *p*, *m* and *n* are unchanged
- Majority rule:
  - pp. **t, *N, *v* → hw. *k, n, w*
- lenition is more likely than fortition
- also, Proto-Polynesian has *p* and *t*, so it should also have a *k*, hence:
  - pp. **k* → sm., hw. *7* (rather than *7 → tg./rg. *k*)

Polynesian example

- majority rule:
  - pp. **f* → rg. *7*, hw. *h*
- not enough data to reconstruct the *l* and *r*
- majority rule:
  - pp. **h, *7* → sm., rg., hw. *0*
- change *s* → *h* is known to be more common than *h* → *s*, hence (against majority rule):
  - pp. **s* → tg./hw. *h*, rg. *7*

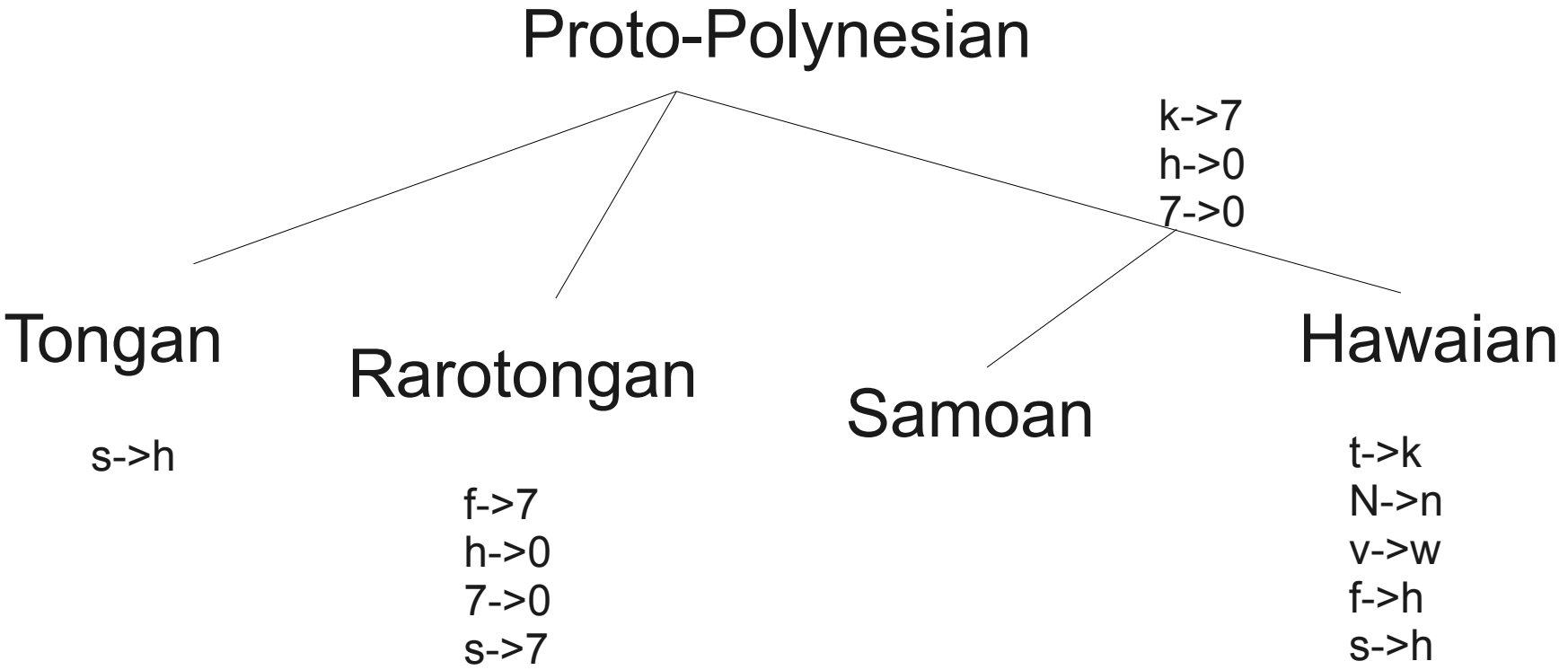
Polynesian example

- constructing a tree

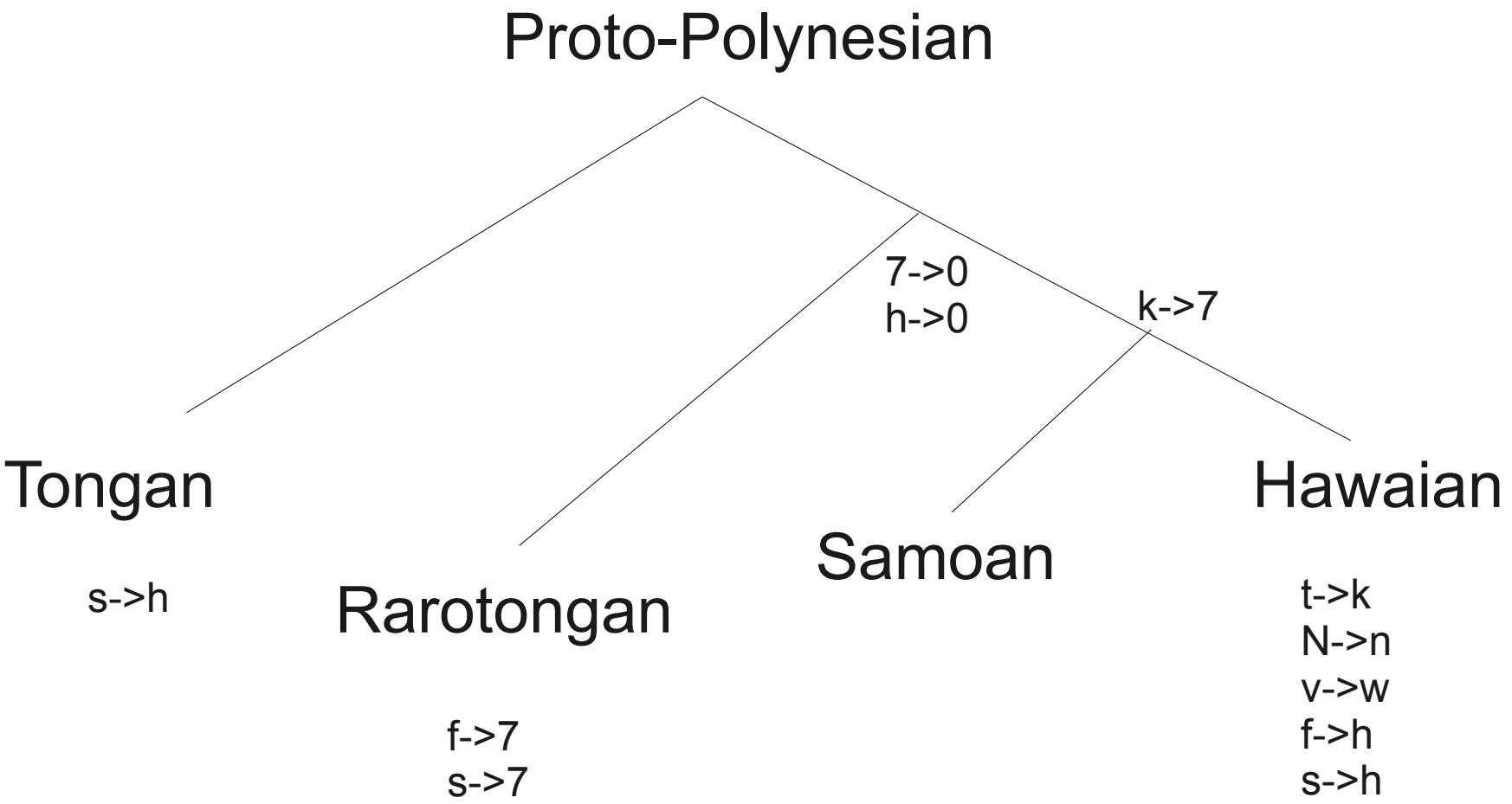


Polynesian example

- constructing a tree



Polynesian example



Polynesian example

- reconstruction seems reasonable because
  - only one shift is assumed twice (s->7), and this type is known to occur frequently
  - reconstruction assumes (pull-) chain shifts
    - Rarotongan and Proto-Samoan/Hawaian restore the lost 7
    - Hawaiian additionally restores the lost k and h
- this procedure started from a reconstructed proto-language; usually tree construction and reconstructon of ancestral forms go hand in hand

Heuristics for identifying language families

- shared cognates, as shown by establishing sound correspondences
- shared grammatical features
- similarities in different parts of the language system
- shared suppletive forms are very strong evidence, such as
  - grm. *gut – besser – am besten* vs. engl. *good – better – best*
  - engl. *I – me* vs. fr. *je - moi*

Heuristics for identifying language families

- Red herrings:
  - grammatical properties that are typologically common
    - *ergative case system, vowel harmony, SVO or SOV word order, tone, …*
  - onomatopoiea, sound symbolism, nursery forms, eg. *mama* for 'mother'
  - chance similarities (especially for short words such as pronouns, the chance of false positives is non-negligible)
  - effects of language contact

Trask (2001) presented an interesting example in which a proposed genetic relationship between Basque and Etruscan fails because the evidence on both sides is spurious. Both Basque and Etruscan, at least to the present, have no known relatives. In this example, a Spanish scholar announced a “breakthrough” showing these two languages to be related, uncritically reported in leading newspapers, including Le Monde in Paris and The Times in London. The single pair of words reported which he supposed demonstrated the relationship was Basque dulla ‘scythe’ and Etruscan dula ‘scythe,’ which he regarded as “practically identical,” and therefore strong evidence for joining these two languages. As Trask points out, the alleged Etruscan word dula does not exist. No word of this form is found in the Etruscan corpus, regardless of meaning, and moreover, such a word would be impossible: Etruscan had no /d/; the Etruscan alphabet, taken from Greek, eliminated the letter “D” – they could not even write a word such as dula. Worse, Basque has no word dulla either. In Trask’s words, “these scholarly breakthroughs are so much easier to achieve, of course, if you’re allowed to invent your own data. Real data can be so tiresomely unhelpful.” As Poser (1992:224) observes, spurious forms “are of no comparative value, no matter what methodology one may favor.”

Campbell/Poser (2008-01-07). Language Classification (pp. 209-210). Cambridge University Press. Kindle Edition.

Subgrouping

- collect data from languages known to be related
- reconstruct the proto-language
- identify sound changes
- establish a relative chronology
- group together languages with shared innovations
- shared unusual changes are strong evidence, because common process may occur independently in different branches