Language variation and change

Gerhard Jäger

SfS, October 28, 2016

Examples of language change

1. Modern English (*The New English Bible*, 1961):
   Shortly afterwards the bystanders came up and said to Peter, 'Surely you are another of them; your accent gives you away!'
2. Early Modern English (*The King James Bible*, 1611):
   And after a while came vnto him they that stood by, and saide to Peter, Surely thou also art one of them, for thy speech bewrayeth thee.
3. Middle English (*The Wycliff Bible*, fourteenth century):
   And a litil aftir, thei that stooden camen, and seiden to Petir, treuli thou art of hem; for thi speche makith thee knowun.
4. Old English (*The West-Saxon Gospels*, c. 1050):
   þa æfter lytlum fyrste genēalǣton þa ðe þær stodon, cwædon to petre. Soðlice þu eart of hym, þyn spræc þe gesweotolað.
   [Literally: then after little first approached they that there stood, said to Peter. Truly thou art of them, thy speech thee makes clear.]

Examples of language change

Modern German:
Vater unser, der du bist im Himmel, geheiligt werde dein Name.

Middle High German:
Got vater unser, dâ du bist in dem himelrîche gewaltic alles des dir ist, geheiliget sô werde dîn nam

Old High German:
Fater unser thû thâr bist in himile, si giheilagôt thîn namo

Gothic:
Atta unsar þu in himinam, weihnai namo þein

further examples

The comparative method

(from Ross and Durie 1996)

- dominant paradigm in historical linguistics
- developed during the 19th century
- originally applied mostly to Indo-European, but applicable to all language families
- central axiom:

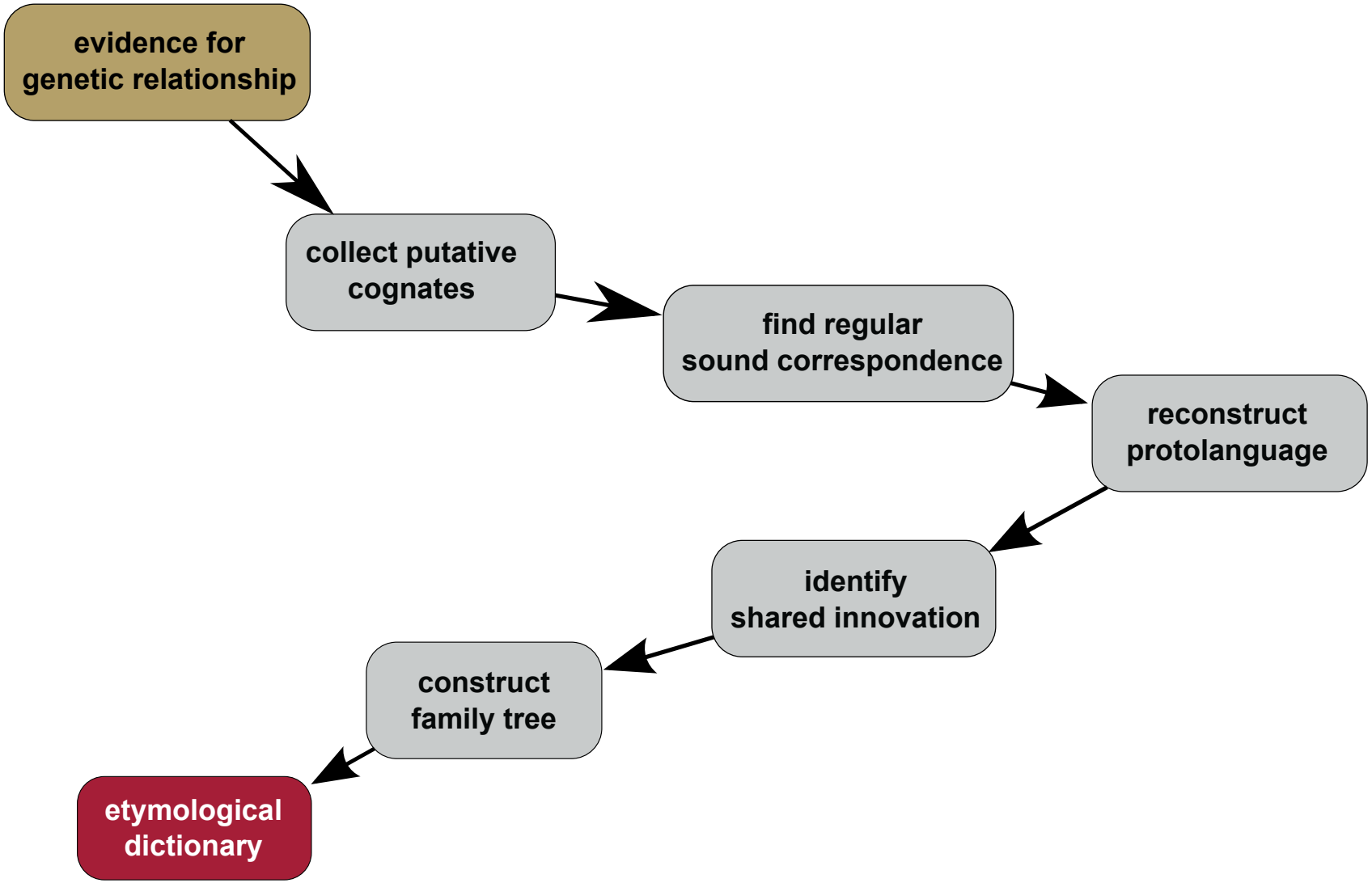
  Neogrammarian Hypothesis Sound laws apply without exception.

- **Workflow**

1. Determine on the strength of diagnostic evidence that a set of languages are genetically related, that is, that they constitute a 'family';
2. Collect putative cognate sets for the family (both morphological paradigms and lexical items).
3. Work out the sound correspondences from the cognate sets, putting 'irregular' cognate sets on one side;
4. Reconstruct the protolanguage of the family as follows:
   a. Reconstruct the protophonology from the sound correspondences worked out in (3), using conventional wisdom regarding the directions of sound changes.
   b. Reconstruct protomorphemes (both morphological paradigms and lexical items) from the cognate sets collected in (2), using the protophonology reconstructed in (4a).

5. Establish innovations (phonological, lexical, semantic, morphological, morphosyntactic) shared by groups of languages within the family relative to the reconstructed protolanguage.
6. Tabulate the innovations established in (5) to arrive at an internal classification of the family, a 'family tree'.
7. Construct an etymological dictionary, tracing borrowings, semantic change, and so forth, for the lexicon of the family (or of one language of the family).

Workflow


evidence for genetic relationship
collect putative cognates
find regular sound correspondence
reconstruct protolanguage
identify shared innovation
construct family tree
etymological dictionary

Diagnostic evidence for genetic relatedness

- sometimes self-evident (e.g. Slavic)
- similarities in morphological paradigms (example from Clackson 2007, 124)

	PIE	Sanskrit	Greek	Latin	Gothic	Lith.	O.C.S.
1.	**h₁és-**mi**	*ásmi*	*eimí*	*sum*	*im*	*esmì*	*jesmĭ*
2.	**h₁és-**si** (or **h₁ési*)	*ási*	*eî*	*ess, es*	*is*	*esì*	*jesĭ*
3.	**h₁és-**ti**	*asti*	*estí*	*est*	*ist*	*ẽsti*	*jestŭ*
4.	**h₁s-**mé**	*smás*	*esmén*	*sumus*	*sijum*	*esme*	*jesmŭ*
5.	**h₁s-**té**	*sthá*	*éste*	*estis*	*sijuþ*	*este*	*jeste*
6.	**h₁s-**énti**	*sánti*	*eisí*	*sunt*	*sind*		*sǫtŭ*

- overwhelming lexical similarities ... $\Rightarrow$

Diagnostic evidence for genetic relatedness

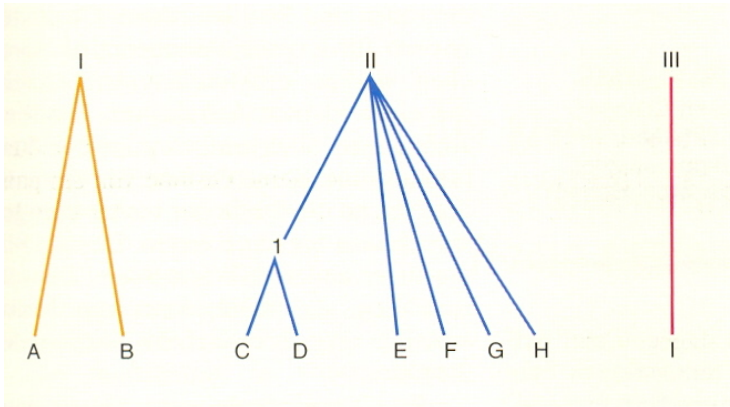
Eine Klassifikationsübung nach der vergleichenden Methode à la Merritt Ruhlen:

Sprache	zwei	drei	ich	du	wer?	nicht	Mutter	Vater	Zahn	Herz	Fuß	Maus	er trägt
A	ʔiθn-	θalāθ-	-ni	-ka	man	lā	ʔumm-	abū	sinn	ɩubb	rijl-	fār	yaḥmil-
B	ʃn-	šaloš	-ni	-ka	mi	lo	ʔem	aβ	šen	leβ	regel	ʕaḵbɔr	nośeh
C	duvấ	tráyas	mấm	tuvám	kás	ná	mātár	pitár-	dant-	hṛd-	pád	muṣ-	bhárati
D	duva	θrāyō	mạm	tuvəm	čiš	naē-	mātar-	pitar-	dantan-	zərəd	paiðya		baraiti
E	duo	treîs	eme	sú	tís	ou(k)	māter	pater	odṓn	kardiā	pod-	mûs	phérei
F	duo	trēs	mē	tū	kwis	ne-	māter	pater	dent-	kord-	ped-	mūs	fert
G	twai	θreis	mik	θu	hwas	ni	aiθei	faðar	tunθus	haírtō	fōt		baíriθ
H	dó	trí	-m	tú	kía	ní-	máθir	aθir	dēt	kride	traig	lux	berid
I	iki	üč	ben-i	sen	kim	deyil	anne	baba	diš	kalp	ayak	sičan	tašıyor

Diagnostic evidence for genetic relatedness

Eine Klassifikationsübung nach der vergleichenden Methode à la Merritt Ruhlen:

Sprache	zwei	drei	ich	du	wer?	nicht	Mutter	Vater	Zahn	Herz	Fuß	Maus	er trägt
A	ʔiθn-	θalāθ-	-ni	-ka	man	lā	ʔumm-	abū	sinn	lubb	rijl-	fār	yaḥmil-
B	ʃn-	šaloš	-ni	-ka	mi	lo	ʔem	aβ	šen	leβ	regel	ʕaḵbɔr	nośeh
C	duvấ	tráyas	mấm	tuvám	kás	ná	mātár	pitár-	dant-	hṛd-	pád	muṣ-	bhárati
D	duva	θrāyō	mąm	tuvəm	čiš	naē-	mātar-	pitar-	dantan-	zərəd	paiðya		baraiti
E	duo	treîs	eme	sú	tís	ou(k)	māter	pater	odṓn	kardiā	pod-	mûs	phérei
F	duo	trēs	mē	tū	kwis	ne-	māter	pater	dent-	kord-	ped-	mūs	fert
G	twai	θreis	mik	θu	hwas	ni	aiθei	faðar	tunθus	haírtō	fōt		baíriθ
H	dó	trí	-m	tú	kía	ní-	máθir	aθir	dēt	kride	traig	lux	berid
I	iki	üč	ben-i	sen	kim	deyil	anne	baba	diš	kalp	ayak	sičan	tašıyor



Klassifizieren Sie die angegebenen neun Sprachen (von A bis I) in Familien und Unterfamilien und vergleichen Sie den Wortschatz für die 13 Wörter, die hier in phonetischer Umschrift geboten werden. Lösung: Sprache A und B (Arabisch und Hebräisch) gehören zur Familie der semitischen Sprachen. Die sechs Sprachen C bis H (Sanskrit, Awestisch, Altgriechisch, Latein, Gotisch und Altirisch) sind indogermanische Sprachen. I (Türkisch) läßt sich keiner Familie zuordnen. Mit einer längeren Wortliste kann man nach demselben Verfahren die Familien wieder in Überfamilien einteilen usw. Der Stammbaum, den man so erhält, würde dann beweisen, daß alle Sprachen von einer Muttersprache abstammen.

Example

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

- establish **regular**, i.e. recurrent, sound correspondences

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

- establish **regular**, i.e. recurrent, sound correspondences

  Greek, Latin, OCS, Russian [d] $\sim$ Dutch, Norse [t]

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zɔ̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

- establish **regular**, i.e. recurrent, sound correspondences

  Greek, Latin, OCS, Russian [d] ∼ Dutch, Norse [t]

  Greek, Latin, OCS, Russian [t] ∼ Dutch [d] ∼ Norse [θ]

Example

- identify cognates

	two	*three*	*tooth*	*foot*	*heart*	*skin*
Ancient Greek	dýo	treːs	odúːs	puːs	kardíaː	dérma
Dutch	twe	dri	tɑnt	vut	hɑrt	hœyt
Latin	'duo	treːs	dens	peːs	kor	'kutis
Old Church Slavonic	dŭva	trĭjɛ	zə̃bŭ	nɔgɑ	sr̩ʲdĭtsʲɛ	kɔʒa
Old Norse	tvɛir	θriːr	tɔnː	foːtr	'ɣjarta	huːð
Russian	dva	trʲi	zub	noga	sʲɛrdtsɛ	'kɔʐa

- establish **regular**, i.e. recurrent, sound correspondences

  Greek, Latin, OCS, Russian [d] ∼ Dutch, Norse [t]

  Greek, Latin, OCS, Russian [t] ∼ Dutch [d] ∼ Norse [θ]

  Greek, Latin [k] ∼ Dutch, Norse (?) [h] $\overset{?}{\sim}$ OCS, Russian [s]

Example

- reconstruct proto-forms and directionality of changes

Example

- reconstruct proto-forms and directionality of changes
  - PIE [*d] $\longrightarrow$ Germanic [t]

Example

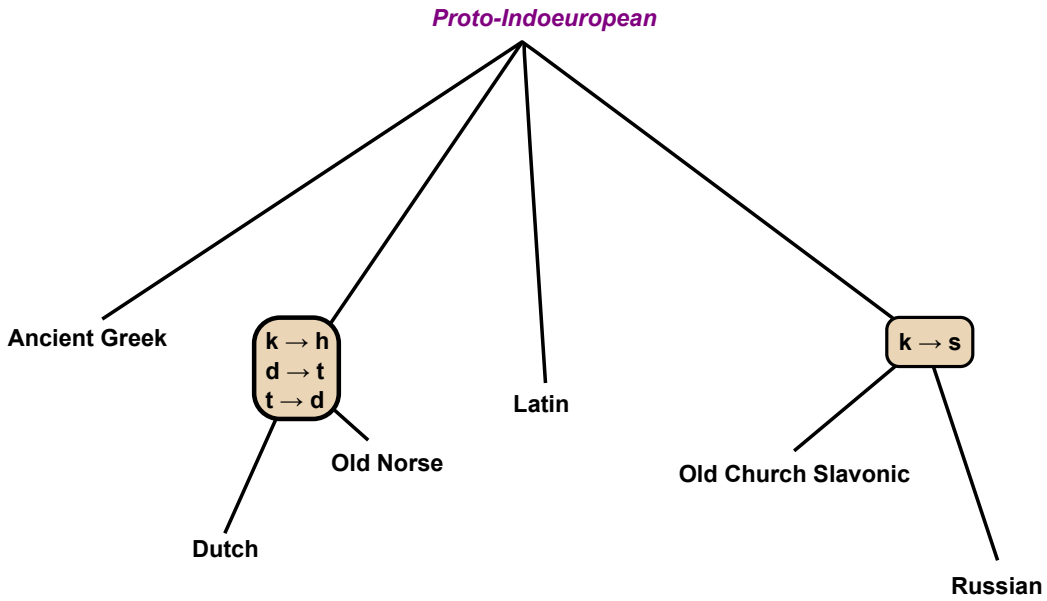
- reconstruct proto-forms and directionality of changes
  - PIE [*d] $\longrightarrow$ Germanic [t]
  - PIE [*t] $\longrightarrow$ Germanic [d/θ]

Example

- reconstruct proto-forms and directionality of changes
  - PIE [*d] $\longrightarrow$ Germanic [t]
  - PIE [*t] $\longrightarrow$ Germanic [d/θ]
  - PIE [*k] $\longrightarrow$ Germanic [h], PIE [*k] $\longrightarrow$ Slavic [s]

Example

- reconstruct proto-forms and directionality of changes
  - PIE [*d] $\longrightarrow$ Germanic [t]
  - PIE [*t] $\longrightarrow$ Germanic [d/θ]
  - PIE [*k] $\longrightarrow$ Germanic [h], PIE [*k] $\longrightarrow$ Slavic [s]
- construct family tree based on *shared innovations*



Example

- compile etymological dictionary
  here: Köbler, Gerhard, Indogermanisches Wörterbuch, (5. Auflage) 2014, `http://www.koeblergerhard.de/idgwbhin.html`

penkᵘe***, idg., Num. Kard.: nhd. fünf; ne. five; RB.: Pokorny 808 (1398/40), ind., iran., arm., gr., alb., ital., kelt., germ., balt., slaw., toch., heth.; Hw.: s. *penkᵘtos, *penkᵘēḱonta, *pn̥ksti-? (?), *penkᵘtos; W.: vgl. gr. πεντηκοστη (pentekoste), Num. Ord., fünfzigster Tag; ae. pent-e-costen, M., Pfingsten; W.: vgl. gr. πεντηκοστη (pentekoste), Num. Ord., fünfzigster Tag; afries. pink-ost-r-a 1 und häufiger?, pinxt-e-r-a, pinst-er, Sb. (Pl.), Pfingsten; W.: s. gr. πεντηκοστη (pentekoste), Num. Ord., fünfzigster Tag; as. pinkoston* 2, sw. F. (n), Pfingsten; W.: vgl. gr. πεντηκοστη (pentekoste), Num. Ord., fünfzigster Tag; mlat. pentecoste; mnd. pinkesten, pinxten: an. pīk-is-dag-r, pikk-is-dag-r, st. M. (a), „Pfingsten“, Weißer Sonntag; W.: lat. quīnque, Num. Kard., fünf; W.: s. lat. quīntus, quīnctus, Num. Ord., fünfte; W.: vgl. lat. Quīnctius, M.=PN, Quinctius (Name einer römischen Gens); W.: germ. *femf, *femfe, Num. Kard., fünf; got. fimf 23, krimgot. fynf*, fyuf, Num. Kard., indekl., fünf (, Lehmann F55); W.: germ. *femf, *femfe, Num. Kard., fünf; an. fim, fim-m, Num. Kard., fünf; W.: germ. *femf, *femfe, Num. Kard., fünf; ae. fīf, Num. Kard., fünf; W.: germ. *femf, *femfe, Num. Kard., fünf; afries. fīf 14, Num. Kard., fünf; W.: germ. *femf, *femfe, Num. Kard., fünf; as. fīf 17, Num. Kard., fünf; mnd. vīf, Num. Kard.; W.: germ. *femf, *femfe, Num. Kard., fünf; ahd. fimf 90, Num. Kard., fünf; mhd. vünf, Num. Kard., fünf; nhd. fünf, Num. Kard., fünf, DW 4, 548; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; got. *fimf-t-a, Num. Ord., fünfte; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; an. fim-t-i, Num. Ord., fünfte; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; ae. fīf-t-a, Num. Ord., fünfte; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; afries. fīf-ta 14, Num. Ord., fünfte; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; as. fīf-to* 1, Num. Ord., fünfte; mnd. vīfte, vifte, Num. Ord.; W.: s. germ. *femftō-, *femftōn, *femfta-, *femftan, Num. Ord., fünfte; ahd. fimfto 14, Num. Ord., fünfte; mhd. vünfte, Num. Ord., fünfte; nhd. fünfte, Num. Ord., fünfte, DW 4, 572; W.: vgl. germ. *femti-, *femtiz, st. F. (i), Fünfzahl; an. fim-t, st. F. (i), Fünfzahl; W.: vgl. germ. *fengra-, *fengraz, *fingra-, *fingraz, st. M. (a), Finger; got. figg-r-s* 1, st. M. (a), Finger (, Lehmann F47); W.: vgl. germ. *fengra-, *fengraz, *fingra-, *fingraz, st. M. (a), Finger; an. fing-r, st. M. (a), später st. N. (a), Finger; W.: vgl. germ. *fengra-, *fengraz, *fingra-, *fingraz, st. M. (a), Finger; ae. fing-er, st. M. (a), Finger; W.: vgl. germ. *fengra-,

Exercise

(from Crowley and Bowern 2010, 80)

	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*

	Tongan	Samoan	Rarotongan	Hawaiian	
19.	hau	asu	ʔau	hau	*dew*
20.	tafuafi	siʔa	ʔika	iʔa	*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*

1. Where do we find non-cognate words within the same row?
2. Which regular sound correspondences do we observe?
3. How do you reconstruct the proto-sounds?
4. What family tree best explains the observed patterns?

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

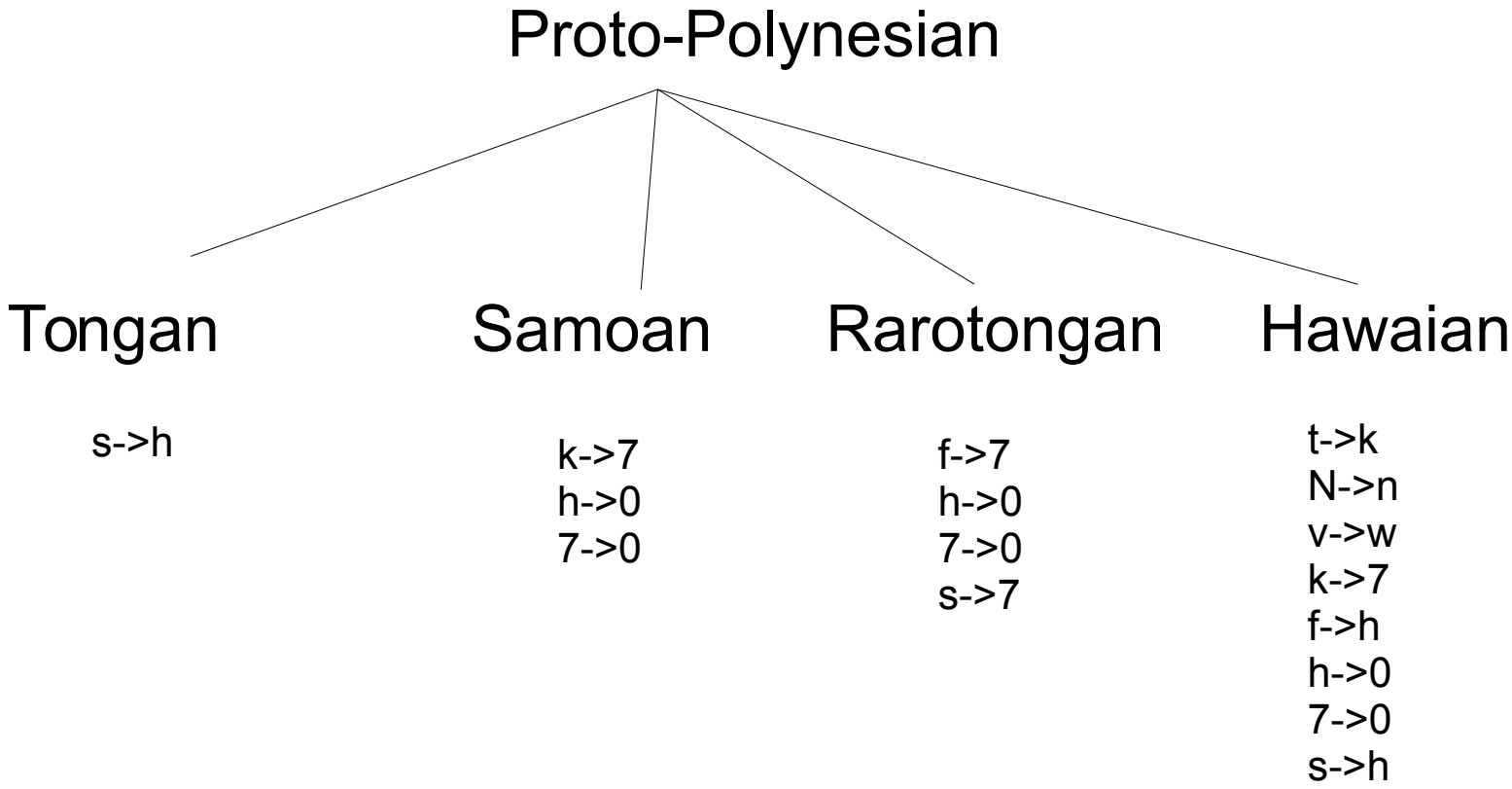
- Vowels in Proto-Polynesian are unchanged in daughter languages (otherwise we would stipulate unnecessary sound shift)
- Likewise, *p*, *m* and *n* are unchanged
- Majority rule:
  - pp. **t, *N, *v* → hw. *k, n, w*
- lenition is more likely than fortition
- also, Proto-Polynesian has *p* and *t*, so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

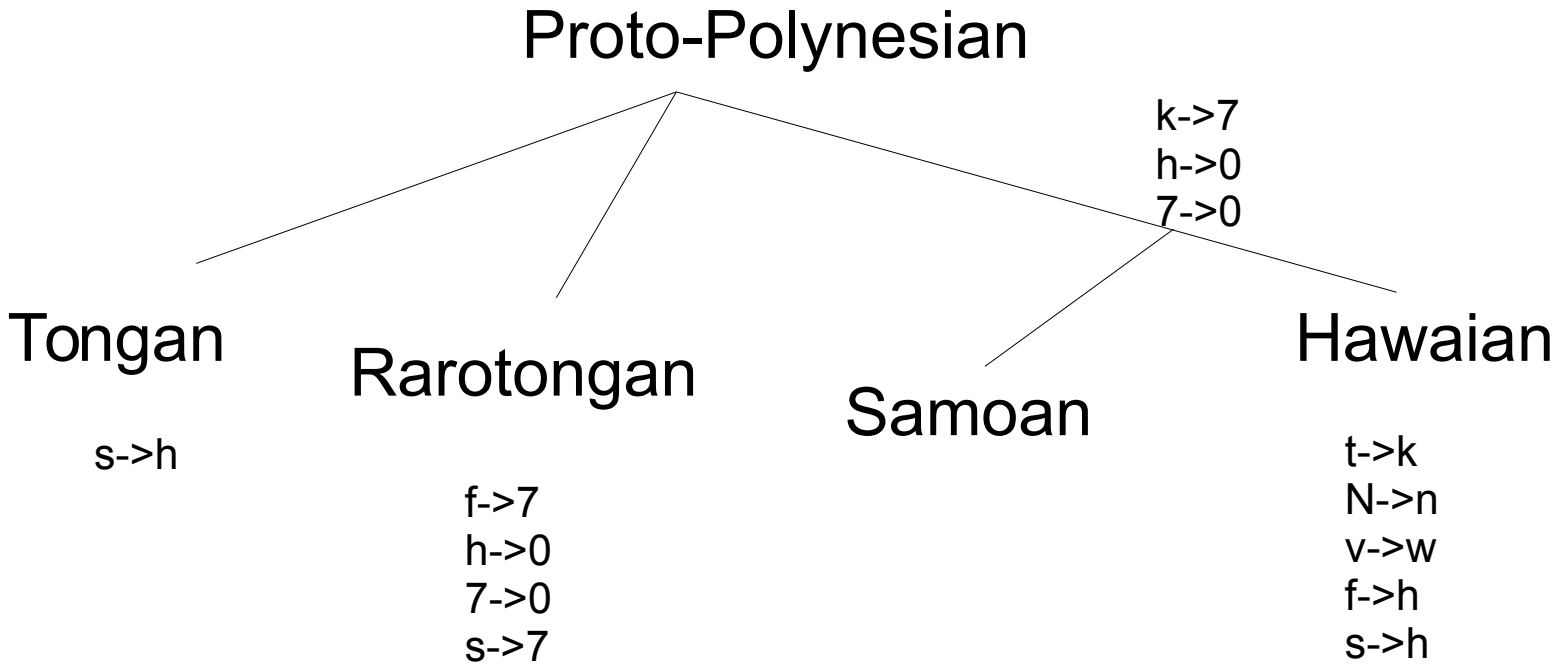
Polynesian example

- constructing a tree

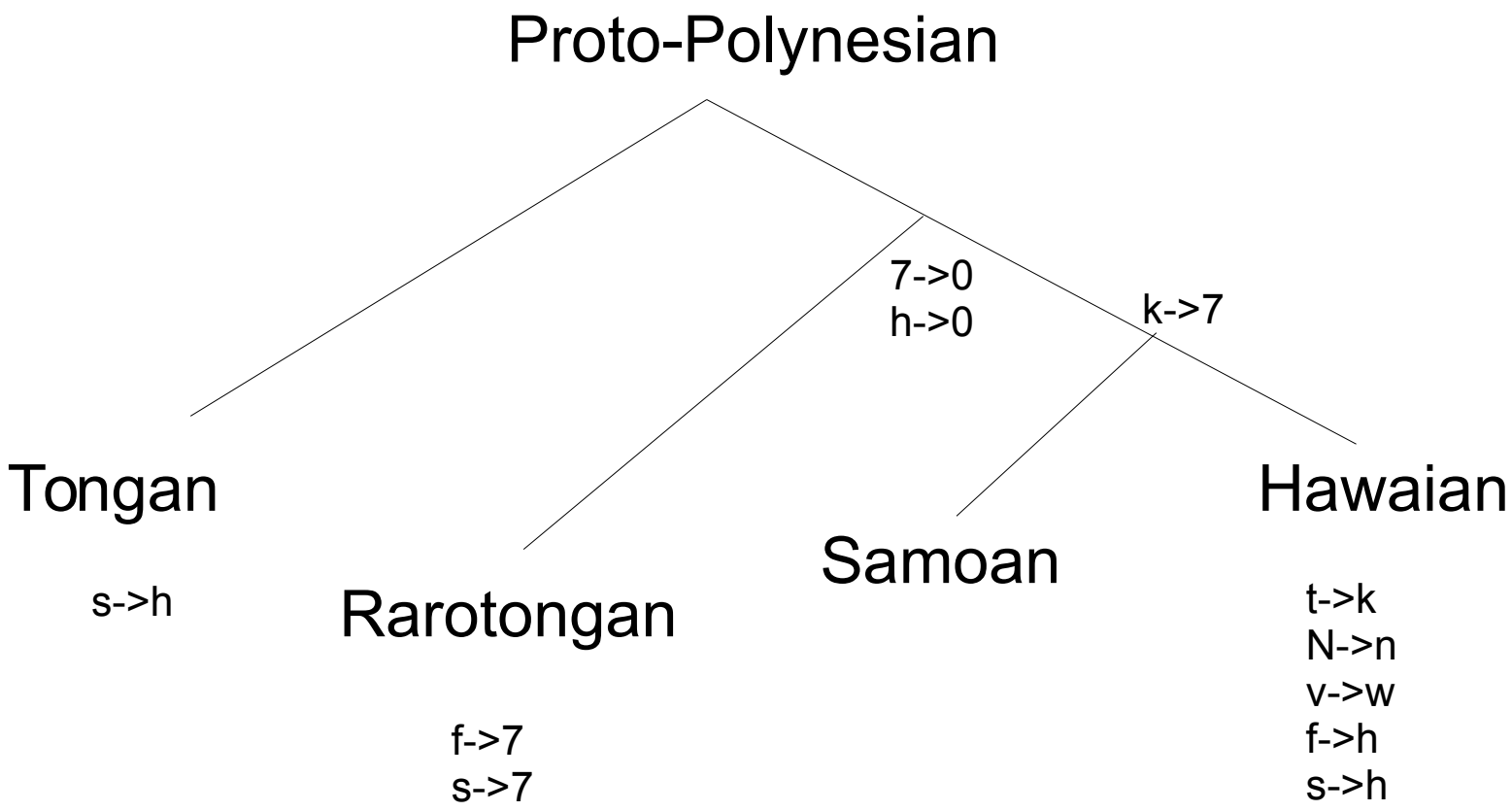


Polynesian example

- constructing a tree



Polynesian example

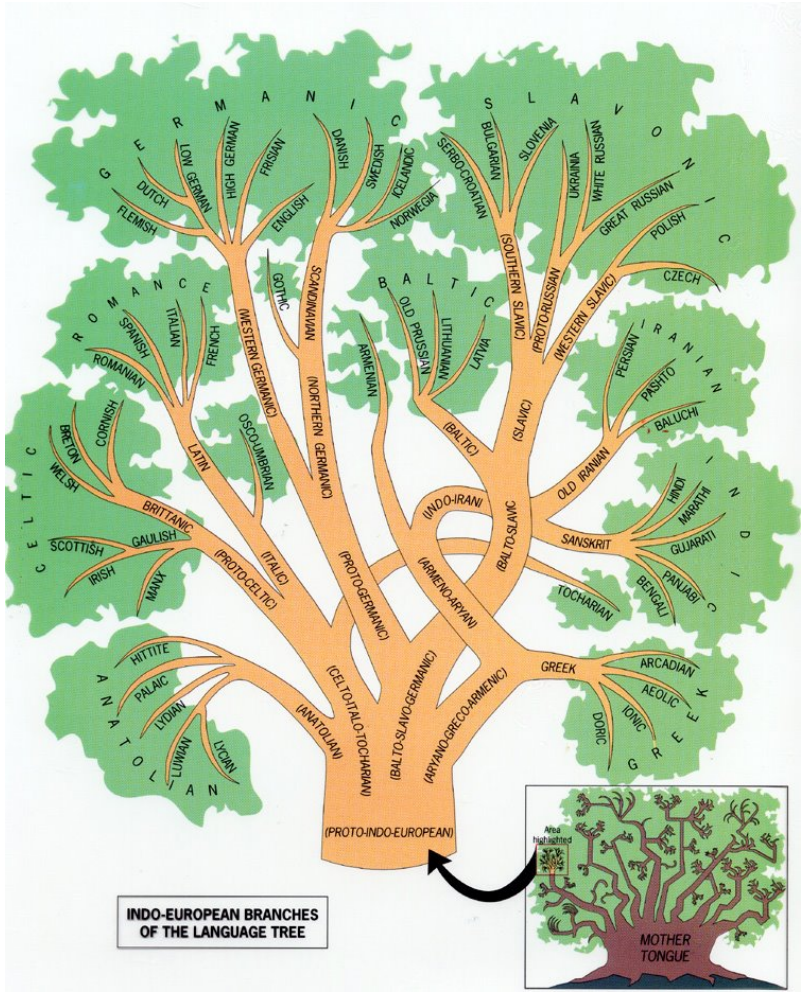


Polynesian example

- 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

Language trees

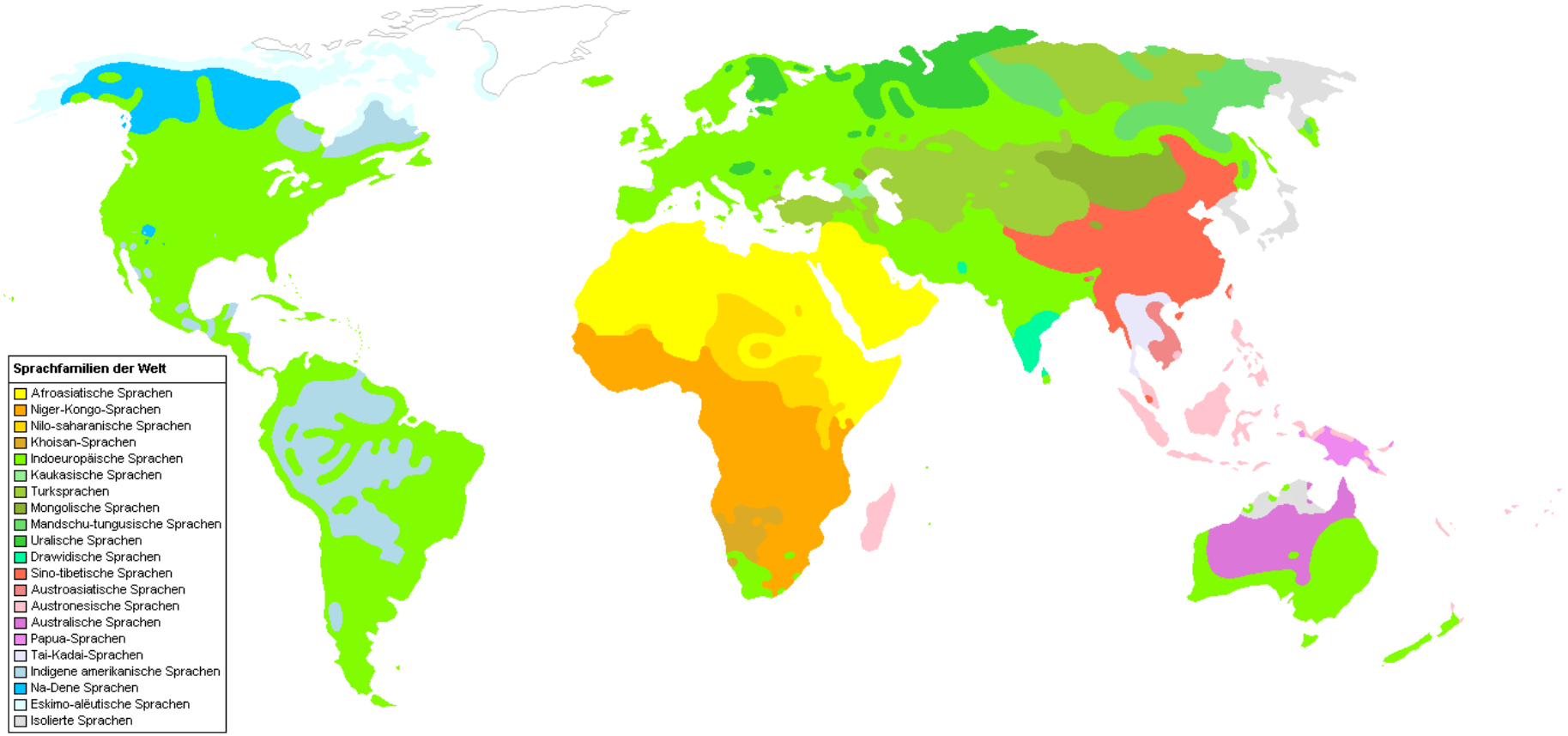
- comparative method gives rise to pyhlogenetic trees of historic development



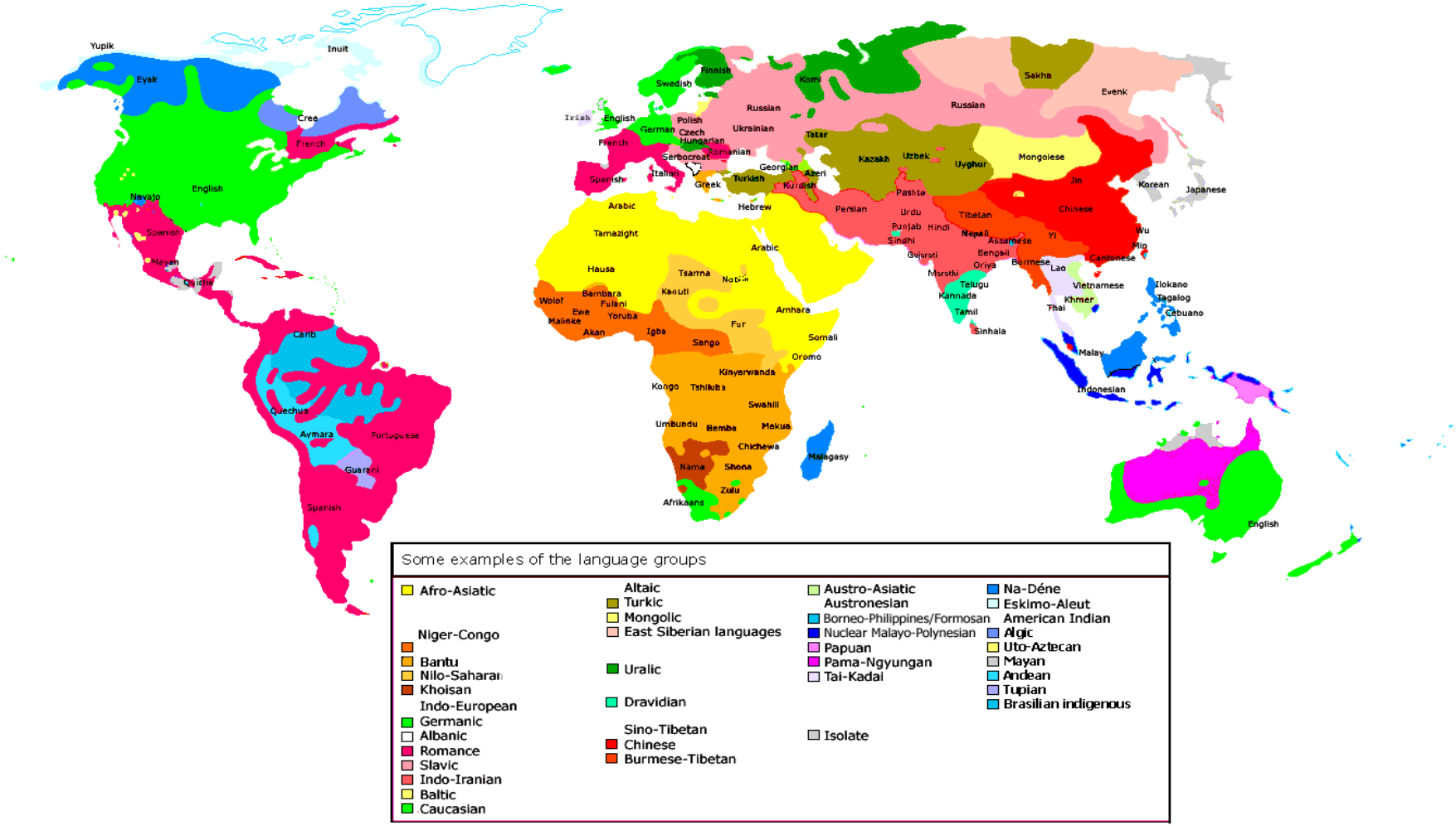
Language families

- Language family: group of genetically (i.e. historically) related languages
- Descent from a common proto language
- Descent has to be established via generally accepted methods
- Classification is (unavoidably) variable and sometimes subjective
- Ethnologue counts more then 100 language families

Language families



Language families



Language families

- **Afro-Asiatic**
  - Also called „Hamito-Semitic“ (obsolete)
  - subgroups:
    - Semitic (Arabic, Hebrew, Amharic, ...)
    - Berber (Tuareg, ...)
    - Egyptian (extinct)
    - Cushitic (Somali, Oromo, ...)
    - Chadic (Hausa, ...)

Language families

- **Nilo-Saharan**
  - Comprises about 200 African languages
  - Nubian, Fur, ...

Language families

- **Niger-Congo languages**
  - Most important subgroup: Bantu languages
  - Swahili, Rwanda, Zulu, Yoruba

Language families

- **Khoisan languages**
  - Languages of the bushmen in Southern Africa
  - Use click sounds (which are typologically uncommon)

Language families

- **Uralic**
  - subgroups
    - Finno-ugric: Hungarian, Estonian, Sami, Karelian
    - Samoyedic (< 30,000 speaker in Nothern Eurasia)

Language families

- **Altaic**
  - subgroups
    - Turkic: Turkish, Turkmen, Kyrgyz, Kazakh
    - Mongolic
    - Tungusic (Northern China, East Siberia)
    - Korean
    - Japanese
  - Partially controversial, especially the inclusion of Korean and Japanese

Language families

- **Dravidian**
  - Telugu, Tamil, Kannada, ...
  - Spoken mainly in Southern India and Sri Lanka

Language families

- **Sino-Tibetan**
  - subgroups
    - Sinitic (chinese languages)
    - Tibeto-Burman (spoken in Myanmar, Northern Thailand, Nepal, Bhutan, parts of China, India and Pakistan): Tibetan, Brahmaputran, ...

Language families

- **Austro-Asiatic**
  - Vietnamese, Khmer, Santali
  - Spoken in South-East Asia and Northern India

Language families

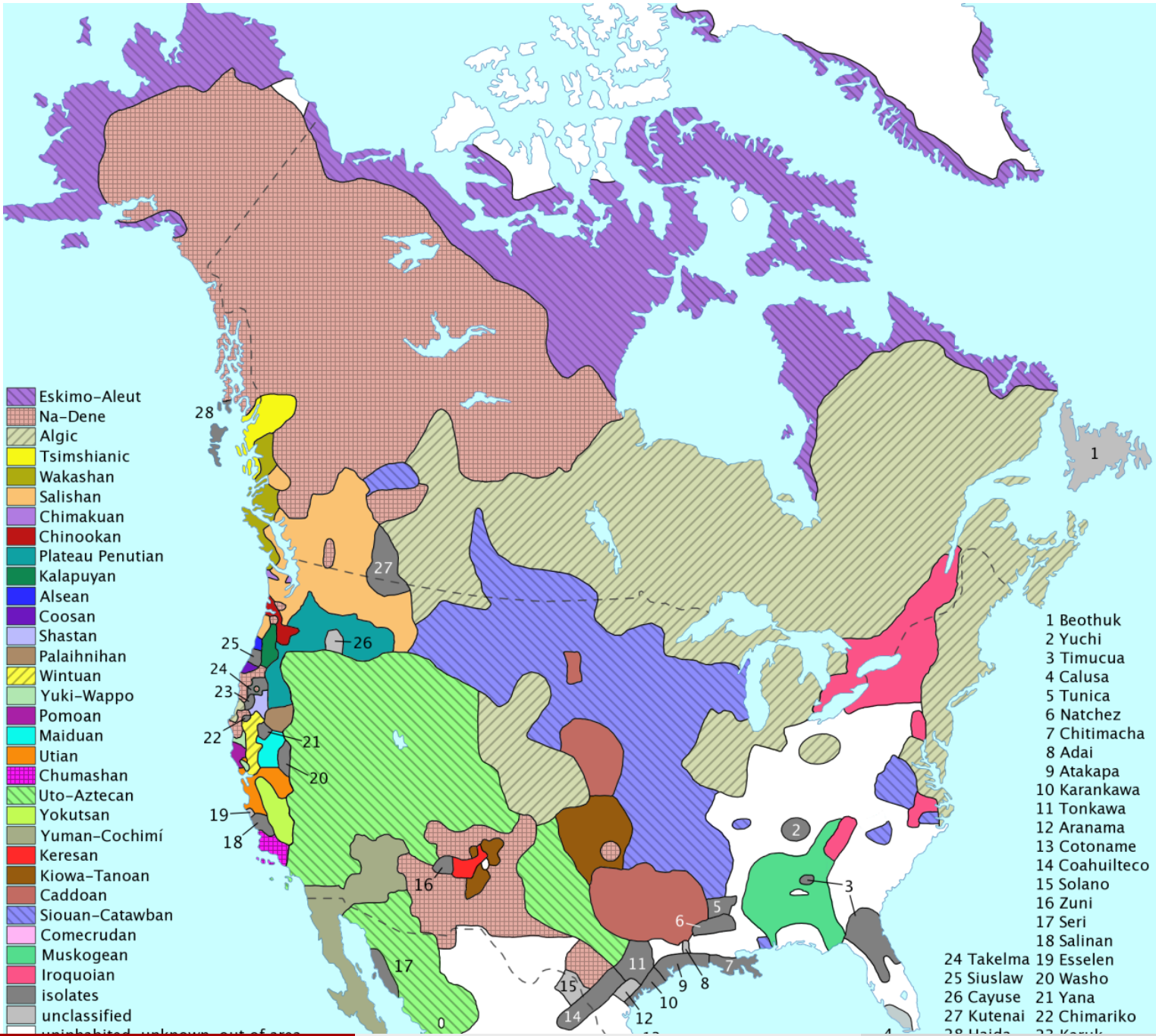
- **Austronesian**
  - Family with the largest geographical expansion (from Madagaskar in the West until Hawaii in the East)
  - Malagasy, Javanese, Bahasa Indonesian, Tagalog, Taiwanese languages, Maori (language of the aborigines of New Zealand), polynesian languages, ...

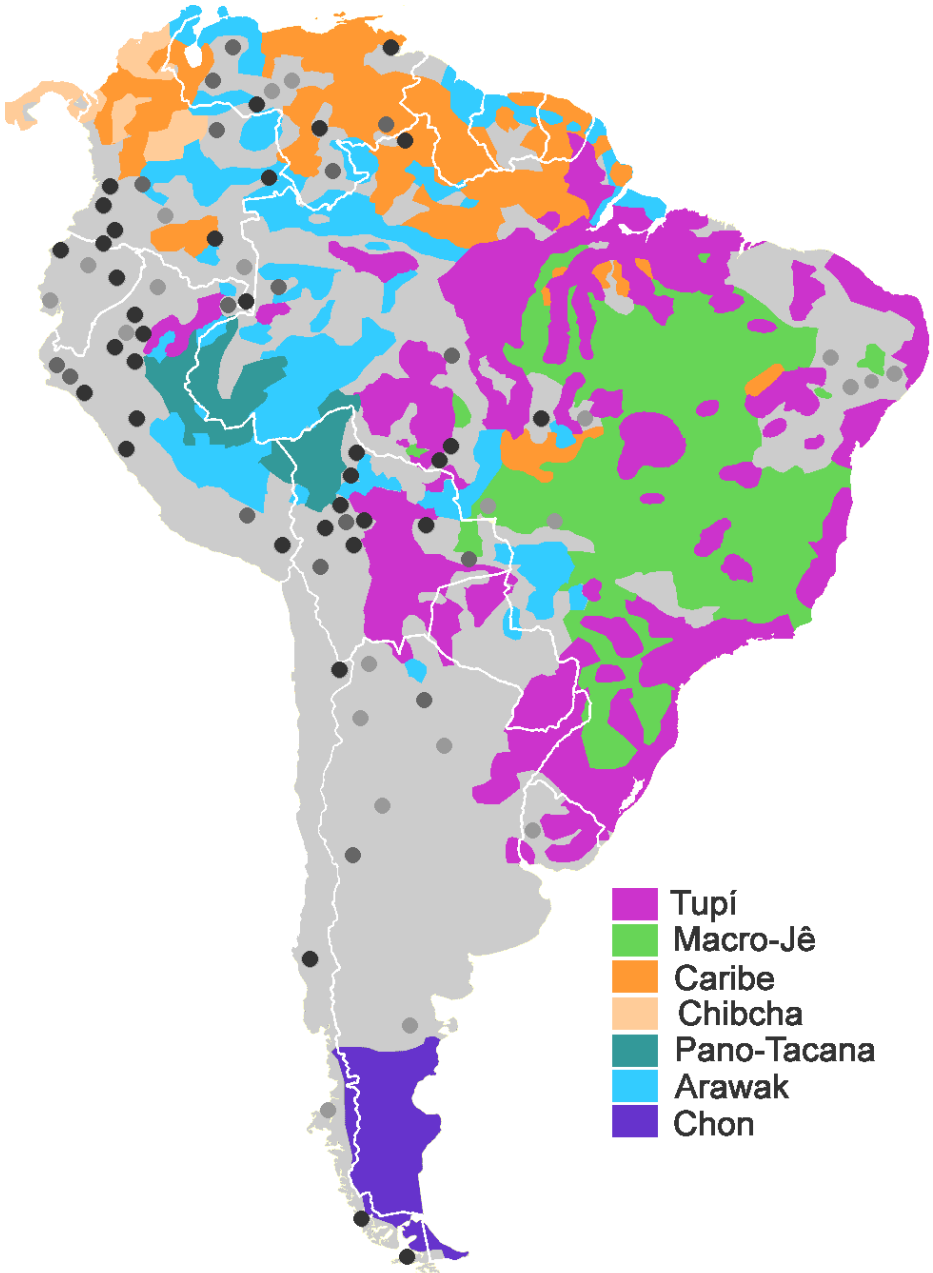
Language families

- **Tai-Kadai languages**
  - Thai, Isan, Lao, ...
  - Speculations, that Austronesian and Tai-Kadai form a single family („Austro-Thai“)

Paleo-American language families

- Classification according to Greenberg:
  - Eskimo-Aleut
  - Na-Dene (Northern and Western North-America)
  - Amerindian (rest of North-America and South-America)
- „Amerindian“ is heavily contested
- Using traditional methods, only many much smaller families can be established


Eskimo-Aleut
Na-Dene
Algic
Tsimshianic
Wakashan
Salishan
Chimakuan
Chinookan
Plateau Penutian
Kalapuyan
Alsean
Coosan
Shastan
Palaihnihan
Wintuan
Yuki-Wappo
Pomoan
Maiduan
Utian
Chumashan
Uto-Aztecan
Yokutsan
Yuman-Cochimí
Keresan
Kiowa-Tanoan
Caddoan
Siouan-Catawban
Comecrudan
Muskogean
Iroquoian
isolates
unclassified
1 Beothuk
2 Yuchi
3 Timucua
4 Calusa
5 Tunica
6 Natchez
7 Chitimacha
8 Adai
9 Atakapa
10 Karankawa
11 Tonkawa
12 Aranama
13 Cotoname
14 Coahuilteco
15 Solano
16 Zuni
17 Seri
18 Salinan
19 Esselen
20 Washo
21 Yana
22 Chimariko
24 Takelma
25 Siuslaw
26 Cayuse
27 Kutenai


Tupí
Macro-Jê
Caribe
Chibcha
Pano-Tacana
Arawak
Chon

Language families

- In many cases, it is impossible to come up with a clear classification
  - 700 languages in Papua-New Guinea, often unrelated to each other
  - Several hundred languages of Australian aborigines; genetic classification is unclear
  - Many „isolated“ language (i.e. no genetic relationship to any other language can be established), for instance Basque

Limits of the comparative method

- Similarities between languages may be due to horizontal transfer (loans)
- limited time depth ($\leq$ 10,000 years)

Hock & Joseph (1996):

Let us pursue this issue a little further by taking a closer look at the relationship between Modern Hindi and English – pretending that we do not yet know that they are related, and trying to establish their relationship by vocabulary comparison. This is actually more difficult than it appears. It is all too easy to be influenced by one's knowledge of the historical relationship between the two languages and therefore to notice the genuine cognates, or even to underestimate the effects of linguistic change on the recognizability of genuine cognates.

Limits of the comparative method

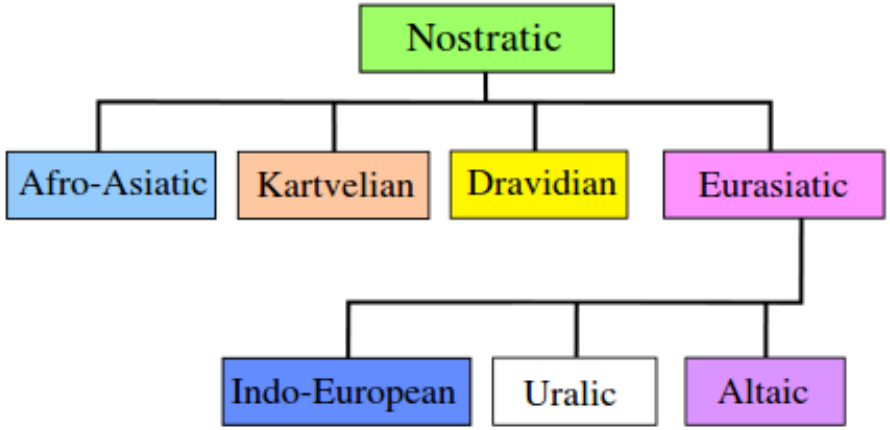
- Similarities between languages may be due to horizontal transfer (loans)
- limited time depth ($\leq$ 10,000 years)

Hock & Joseph (1996):

Clearly, one correspondence is not enough; nor are twenty. And just as clearly, a thousand correspondences with systematic recurrences of phonetic similarities and differences would be fairly persuasive. Are 500 enough, then? And if not, are 501 sufficient? Nobody can give a satisfactory answer to these questions. And this is no doubt the reason that linguists may disagree over whether a particular proposed genetic relationship is sufficiently supported or not.

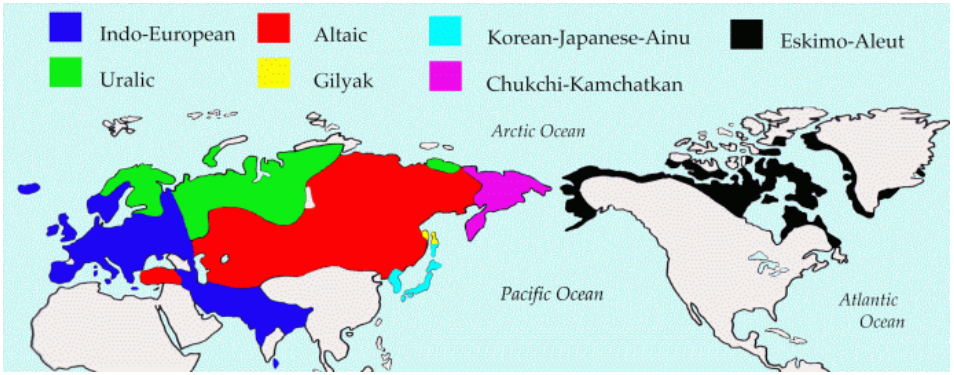
Deep genetic relationships

- Plethora of proposals beyond well-established families:
  - Nostratic:
    - proposed by Pedersen (1903)
    - original proposal: Indo-European, Finno-Ugric, Samoyed, Turkish, Mongolian, Manchu, Yukaghir, Eskimo, Semitic, and Hamitic
    - revived by "Moscow school" in 1960
    - traditional comparative method, including reconstruction of proto forms



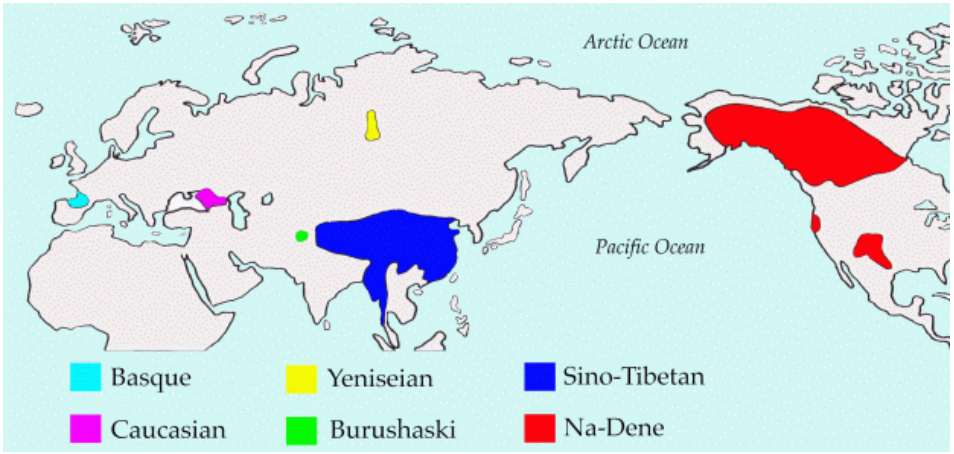
Deep genetic relationships

- Plethora of proposals beyond well-established families:
  - Eurasiatic
    - proposed by Greenberg (2000)
    - comprises Indo-European, Uralic–Yukaghir, Altaic, Chukotko-Kamchatkan, Eskimo–Aleut, Korean-Japanese-Ainu, Gilyak, Etruscan
    - multitude of arguments, mostly from morphology and phonology



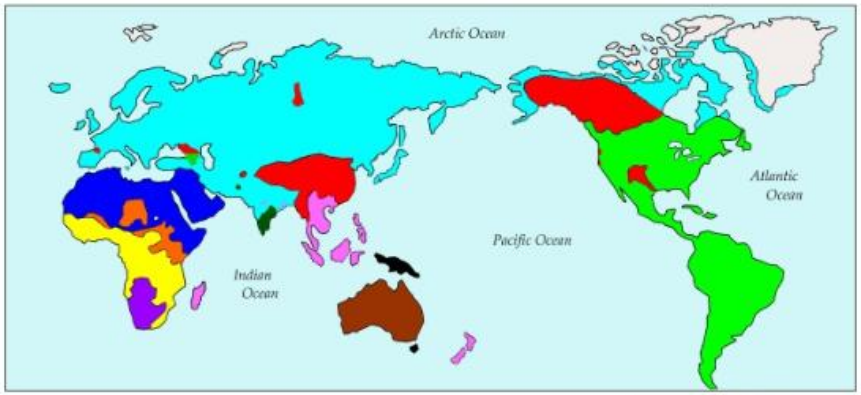
Deep genetic relationships

- Plethora of proposals beyond well-established families:
  - Dene-Caucasian
    - based on work by Sapir, Starostin, Swadesh and others
    - comprises Ne-Dene, Caucasian, Sino-Tibetan, Yeniseian, Burushaski, perhaps Basque and other languages
    - also multitude of arguments, mostly from morphology and phonology



Deep genetic relationships

- Plethora of proposals beyond well-established families:
  - Amerind
    - proposed by Greenberg (1987)
    - comprises all American languages except Na-Dene and Eskimo-Aleut
    - arguments based on mass lexical comparison



Language Families of the World (after Greenberg)

Deep genetic relationships

- Merritt Ruhlen, a student of Greenberg, even claims to have reconstructed a few words of "Proto-World" (for instance the word *aqua* for water, which miraculously didn't change from the dawn of time till Cicero)
- such deep connection are mostly based on suggestive salient features of the languages involved, like pronoun forms
- Nostratic pronouns
- Amerind pronouns
- generally, these approaches neither quantify the probability of chance resemblances nor do they take negative evidence into account

Clackson, J. (2007). *Indo-European Linguistics. An Introduction*. Cambridge University Press, Cambridge, UK.

Crowley, T. and C. Bowern (2010). *An introduction to historical linguistics*. Oxford University Press, Oxford.

Ross, M. and M. Durie (1996). Introduction. In M. Durie and M. Ross, eds., *The Comparative Method Reviewed. Regularity and Irregularity in Language Change*, pp. 3–38. Oxford University Press, New York and Oxford.