Linguistics as a Historical Science

William Croft and Ian Maddieson
University of New Mexico

Evolutionary linguistics

Language and evolution

- The close relationship between language and evolution was a commonplace in the 19th century:

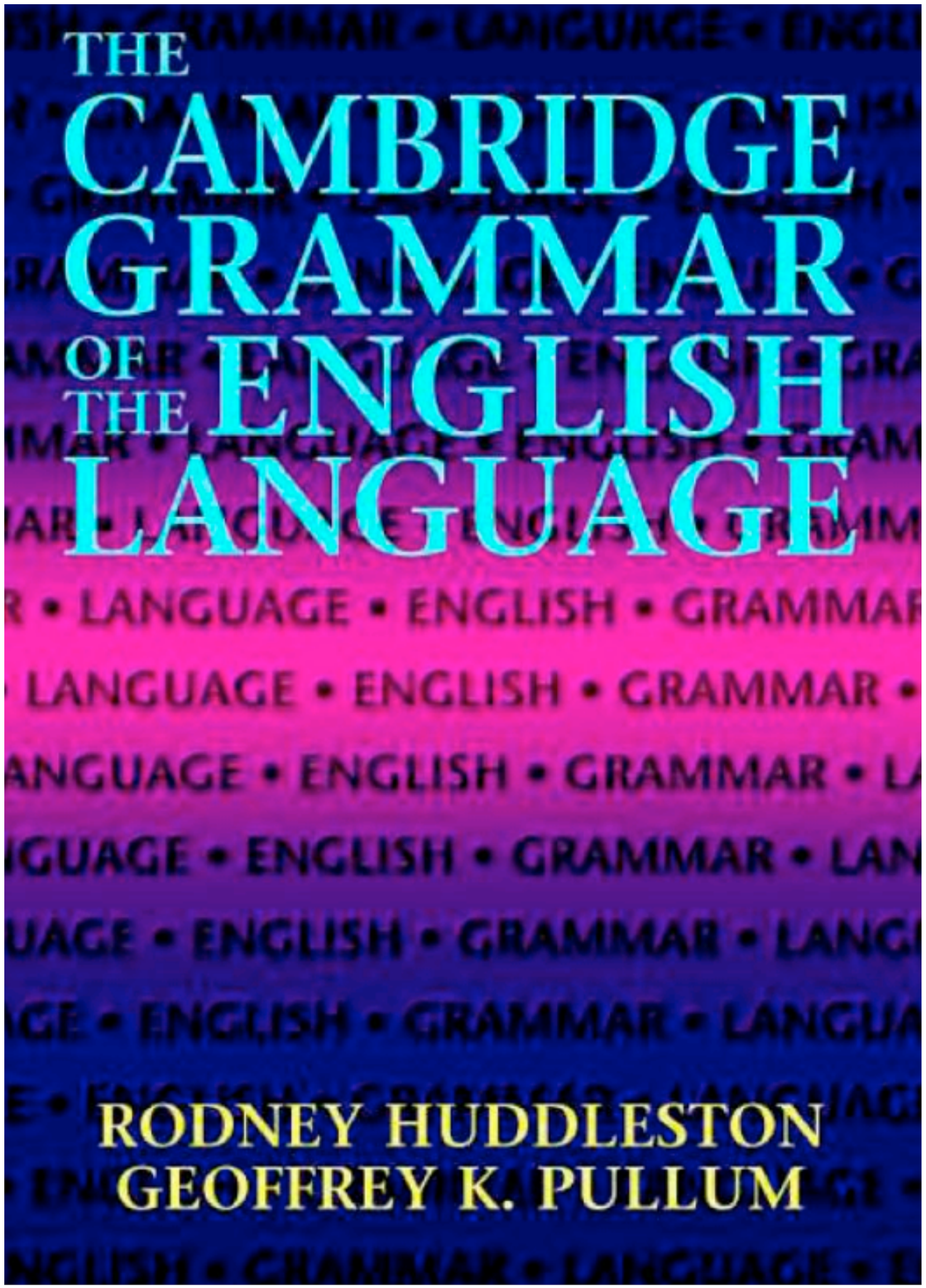
  The formation of different languages and of distinct species, and the proof that both have been developed through a gradual process, are curiously parallel…
  (Charles Darwin, *The Descent of Man*)

- In the past decade or so, evolutionary models of language have been proposed in linguistics

- What happened in the century in between?

Two views of language, I

- Beginning with Saussure (***structuralism***) and continuing with Chomsky (***generative grammar***), language was conceived of as a ***uniform system of synchronic rules***
- This approach, which has a long tradition, we call the "reference grammar" view of language



"Reference grammar" view = essentialism in biology

- Essentialism in biology is the idea that a species is a ***type*** defined by ***essential features***

Calocedrus decurrens (Torrey) Florin 1956. Resinous, aromatic tree 18-46(57) m tall and 90-150(360) cm dbh. Tapering, irregularly angled trunk and narrow, columnar crown, becoming open and irregular. Bark light or reddish-brown, thick, fibrous, deeply and irregularly furrowed into shreddy ridges. Twigs much-branched and flattish, with wedge-shaped joints longer than broad; composed of scalelike leaves. Leaves evergreen, shiny, opposite in 4 rows, 3-14 mm long, scalelike, including long-decurrent base, rounded abaxially, apex acute (often abruptly), usually mucronate, the side pair keeled, long-pointed, overlapping the next pair, extending down twig; aromatic when crushed. Pollen cones red-brown to light brown. Seed cones cones red-brown to golden brown, 14-25 mm long (including wings), oblong-ovate when closed, pendant at end of slender, leafy stalk, proximal scales often reflexed at cone maturity, median scales then widely spreading to recurved, distal scales erect. Seeds 4 or fewer in cone, paired with 2 unequal wings. 2n= 22 (Little 1980, Thieret 1993).

Problems with essentialism/ "reference grammar" view

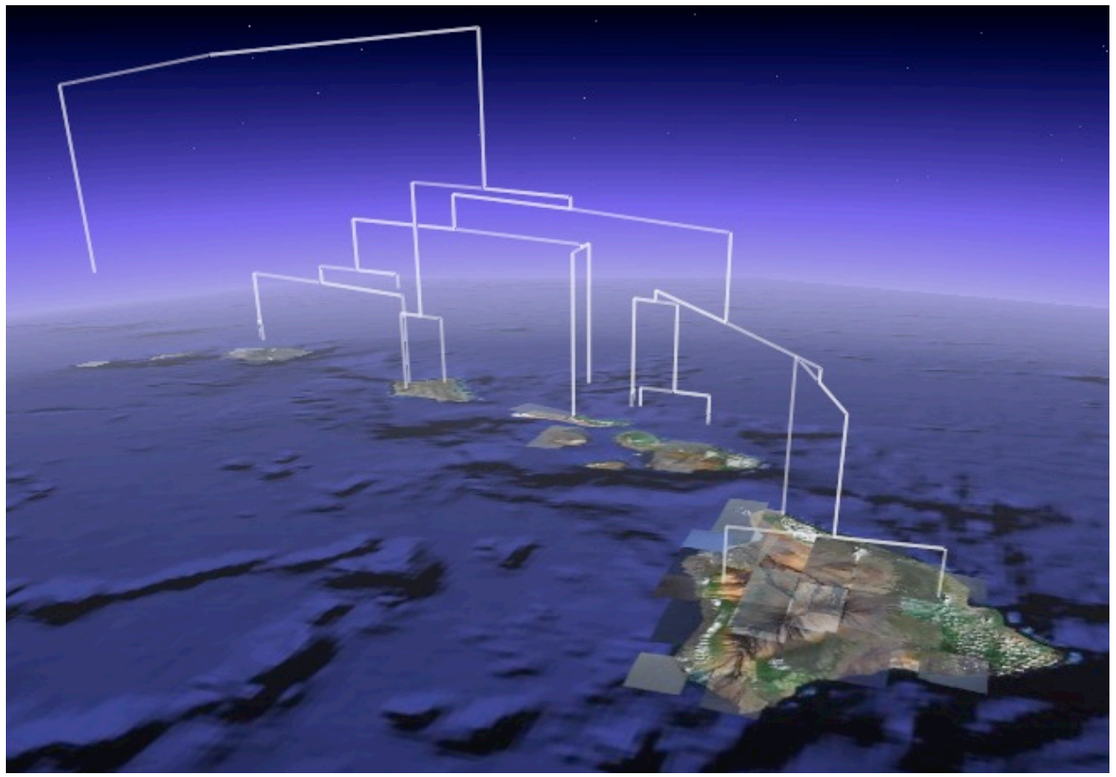
- Species vary tremendously across individuals in their "essential" features
- So do languages, as actually used
- Species evolve and change "essential" features, without a sharp dividing line
- So do languages: when does "Middle English" end and "Early Modern English" start?
- Essential/static types are *fictions*, i.e. an idealization that may be misleading

Population thinking

- In the neo-Darwinian synthesis, species are defined not as essential types but as ***populations***
  (Ernst Mayr, *The Growth of Biological Thought*)
- A population is a ***historical, spatiotemporally bounded entity***
  (David Hull, *Science as a Process*)

A species…

- Exists over a particular time and space
- And is bounded in time and space by its eventual end (breakup or extinction)



Population definition of a species

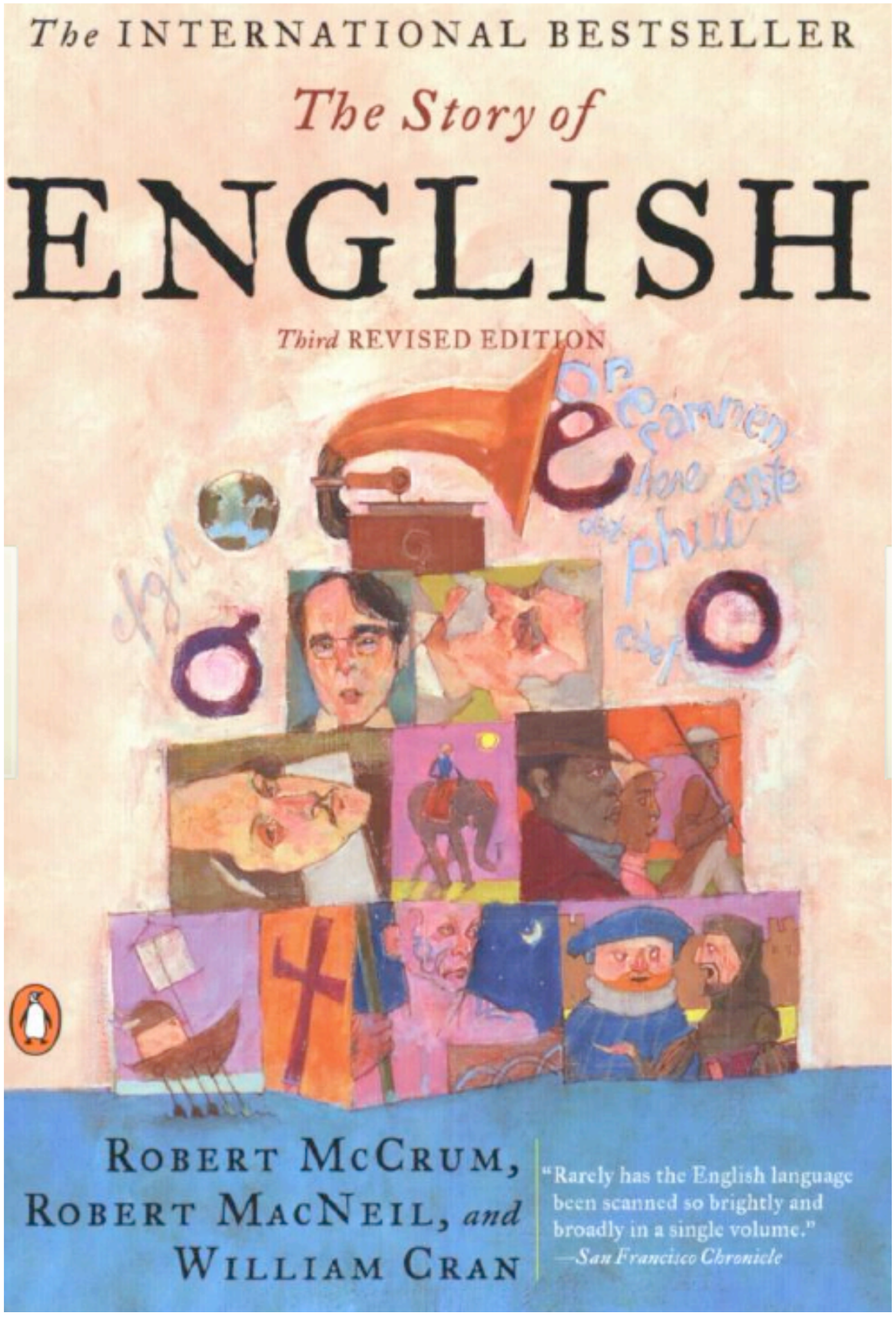
- A set of individuals that interbreed (i.e. they are defined by their relationship to each other)

- The population is (relatively) reproductively isolated from other populations
- The defining relationship is also the process by which replication occurs in the population

Two views of language, II

- There is another, informal view of language, which is more suited to the facts of language and is compatible with population thinking and the evolutionary framework

- Our goal is to formalize this "Story of English" view of language as a ***historical process***



A language…

- Exists over a particular time and space

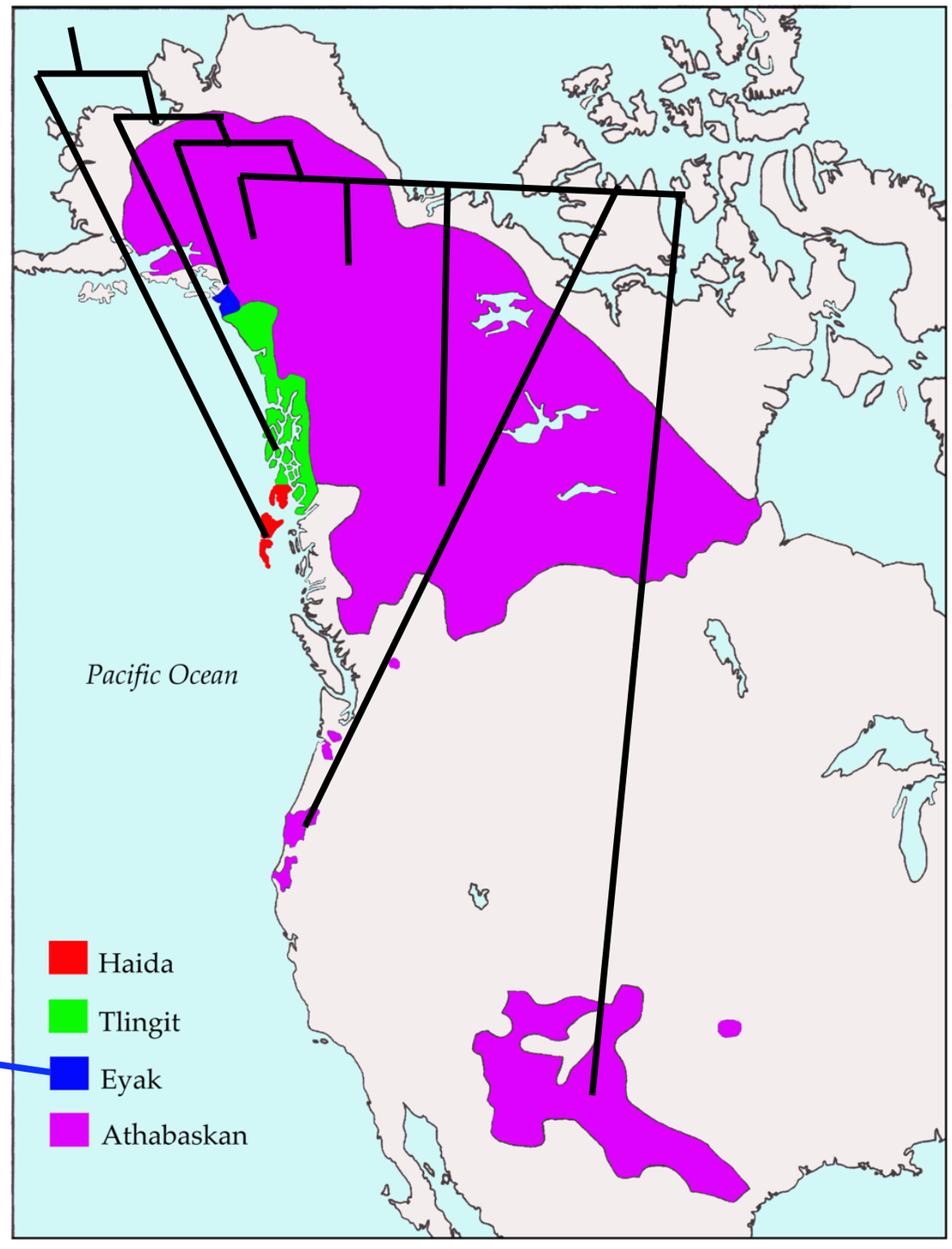
- And is bounded in time and space by its eventual end (breakup or extinction)

Eyak language dies with its last speaker

Tue, January 22, 2008
Posted in Alaska News, Top Stories

Chief Marie Smith Jones in 2004

The last traditional speaker of the Eyak language died yesterday, making the language extinct. Eyak Chief Marie Smith Jones was 89 years old. She was the last person to have learned the language the traditional way, taught as a child from her parents.



The Na-Dene Family

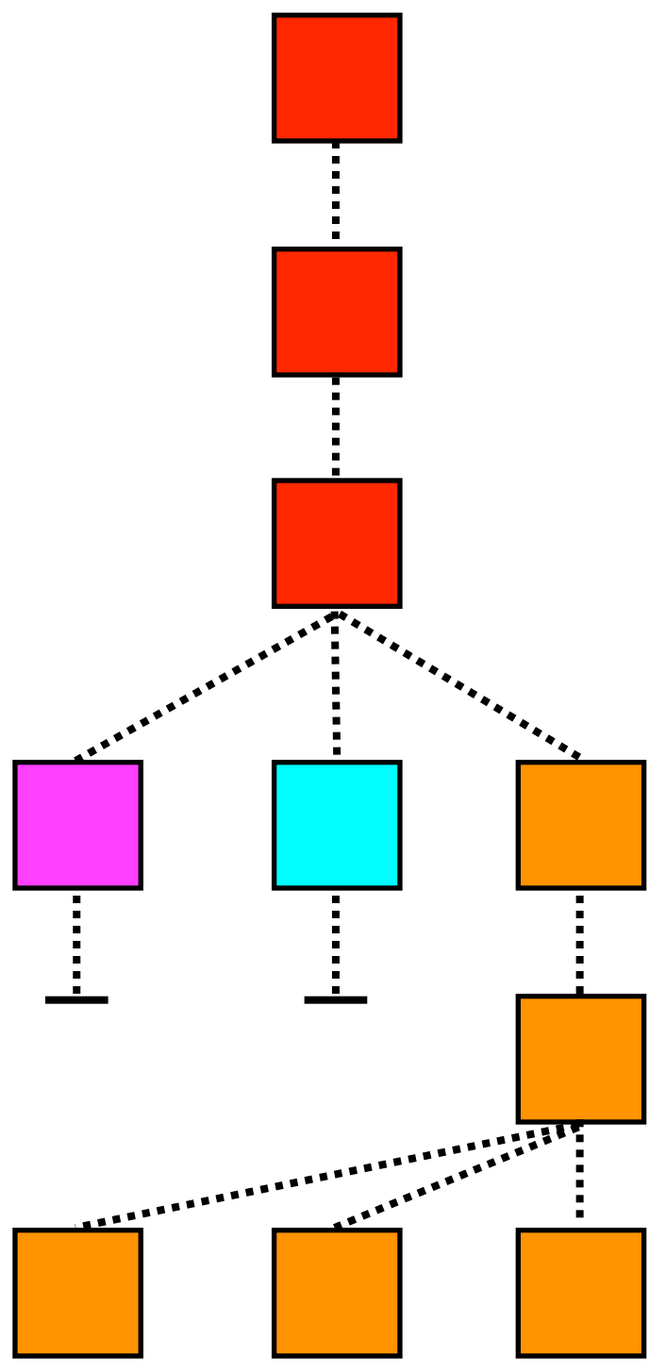
Population definition of a speech community

- A set of individuals that conversationally interact (i.e. they are defined by their relationship to each other)

- The population is (relatively) communicatively isolated from other populations
- The defining relationship is also the process by which replication occurs in the population

Evolutionary change

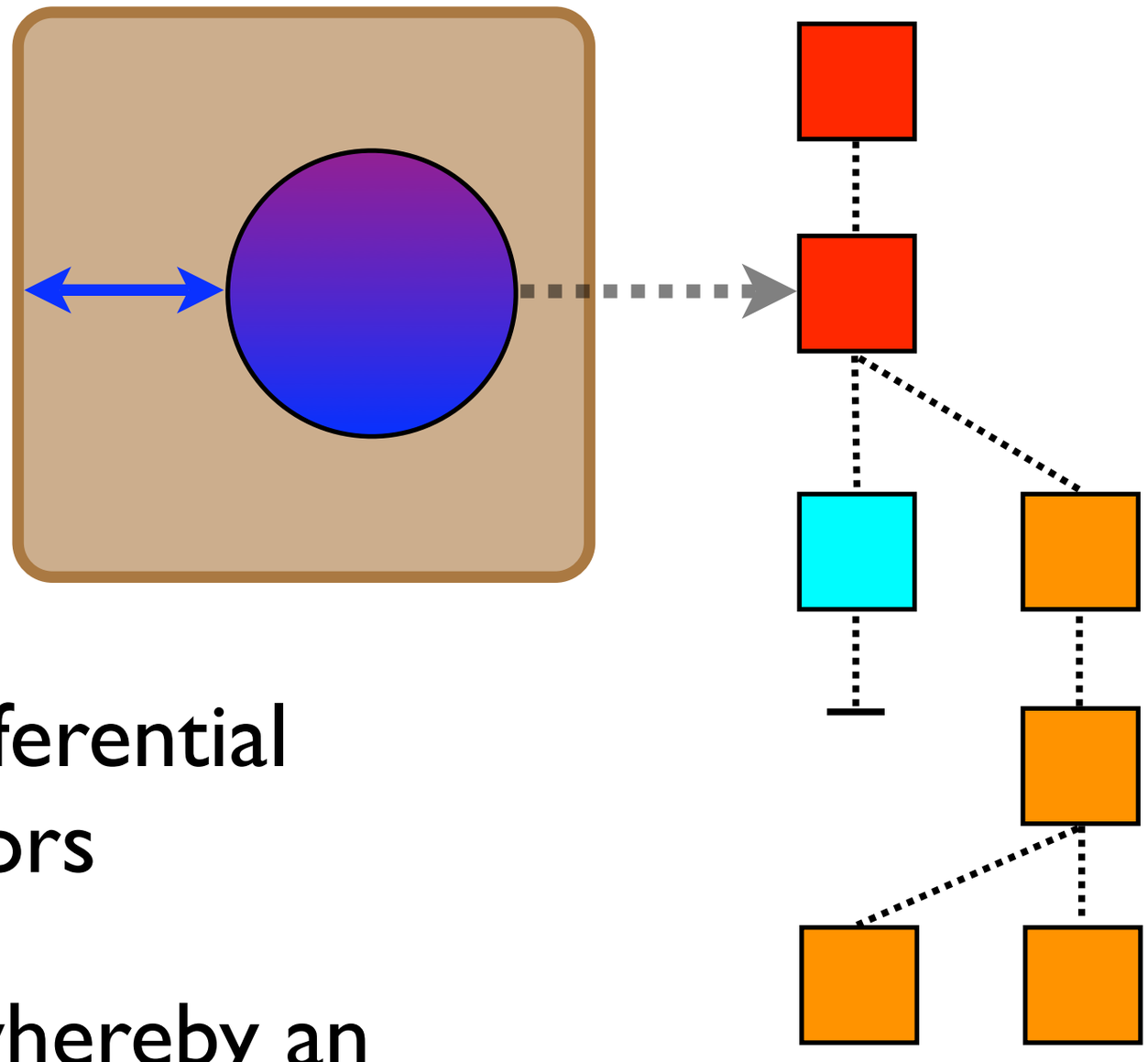
- Evolutionary change is change by **replication**
- Evolutionary change is a **two-step process**:
  - ⋆ ***Generation of variation*** *(creation of multiple variants)*
  - ⋆ ***Selection*** *(propagation/ extinction of variants)*



General Analysis of Selection (GAS)

- **Replicator** – an entity that preserves most of its structure in replication

- **Interactor** – an entity whose interaction with its environment *causes* differential replication of the replicators



- **Selection** – a process whereby an interaction with the environment *causes* differential replication of replicators

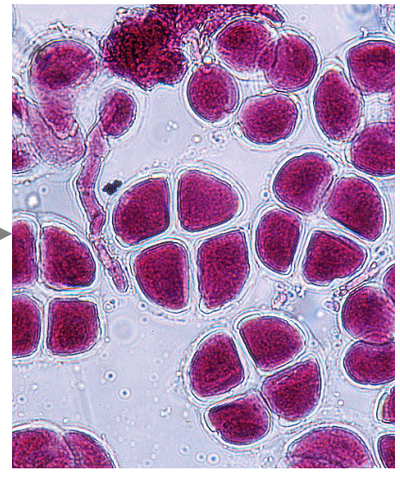
(David Hull, *Science as a Process*, 1988)

Instantiation of GAS in biology

Environment:
ecosystem, conspecifics

Interactor:
organism

Replicator:
gene



An evolutionary framework for language change

- Every time we talk, we replicate tokens of linguistic structure—***linguemes***—in utterances
- The process of communication, i.e. perception and production, generates variation in replication
- Speakers select variants to replicate, leading to differential replication of variants
- Speakers' knowledge of language changes over their lifetimes to reflect usage, memory

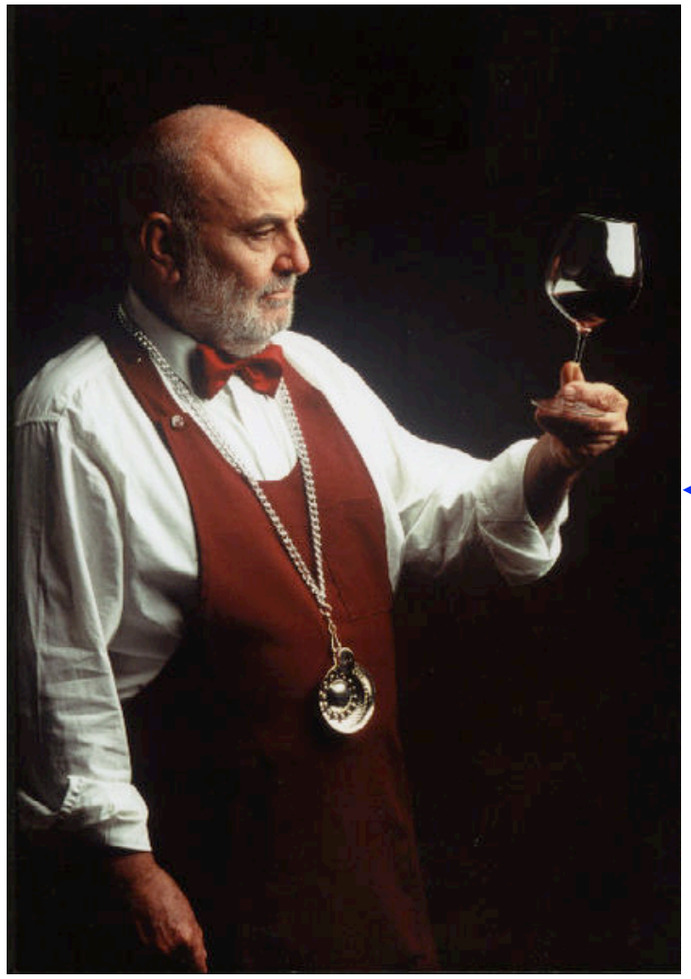
Croft, *Explaining Language Change*, 2000

Instantiation of GAS in language

Environment:
experience, interlocutors

Interactor:
speaker

Replicator:
lingueme



Nice Cab, eh?

Bored with genomics?

Too simple?

Try language!

- In genomics, a lot of headway can be made with simple sequence information
- Not so with language, despite possessing the same evolutionary processes to reconstruct:
  - ✦ The structure of an utterance is more complex than the structure of a genome
  - ✦ Individual replicators (sounds, words, constructions) change form: sister replicators are only similar, not identical

Anatomy of an utterance

- A language is not just "a bag of words", and its utterances are not just strings of words
- An utterance consists of a multiply-articulated structure of sounds, words and constructions
- Linguistic structure consists of two more or less hierarchically organized structures

The more you complain, the longer God lets you live.

Anatomy of an utterance

- One level of organization is the sound structure of an utterance
- A transcription represents a string of sound segments, but this is not the entire story:
  - ✦ Utterances result from a continuous flow of movement, creating a complex acoustic signal
  - ✦ The sound segments are organized into higher-level structures including syllables, words and intonation groups/units

The more you complain, the longer God lets you live.

SOUNDS [ðə moɹ juʷ kəmplɛʲn ðə laŋgɚ gad lɛts juʷ lɪv]

Anatomy of an utterance

- The other level of organization is the "lexico-grammatical" or ***morphosyntactic*** structure of an utterance
- The standard orthographic representation gives only the word string, but this is not the entire story either:
  - ✦ Words often may be analyzed into ***morphemes*** (*long-er, let-s*)
  - ✦ Words themselves are more or less hierarchically organized into ***constructions***

The more you complain, the longer God lets you live.

WORDS

Anatomy of an utterance

Comparative Conditional
[*The* X-*er*…, *the* Y-*er*…]

The more you complain, the longer God lets you live.

CONSTRUCTIONS

Anatomy of an utterance

Comparative Conditional
[*The* X-*er*…, *the* Y-*er*…]

Bare Infinitive Complement
[Sbj *let*-INFL Comp]

The more you complain, the longer God lets you live.

CONSTRUCTIONS

Anatomy of an utterance

Comparative Conditional
[*The* X-*er*…, *the* Y-*er*…]

Intransitive
Clause
[Sbj Verb-INFL]

Bare Infinitive Complement
[Sbj *let*-INFL Comp]

Intransitive
Clause
[Sbj Verb-INFL]

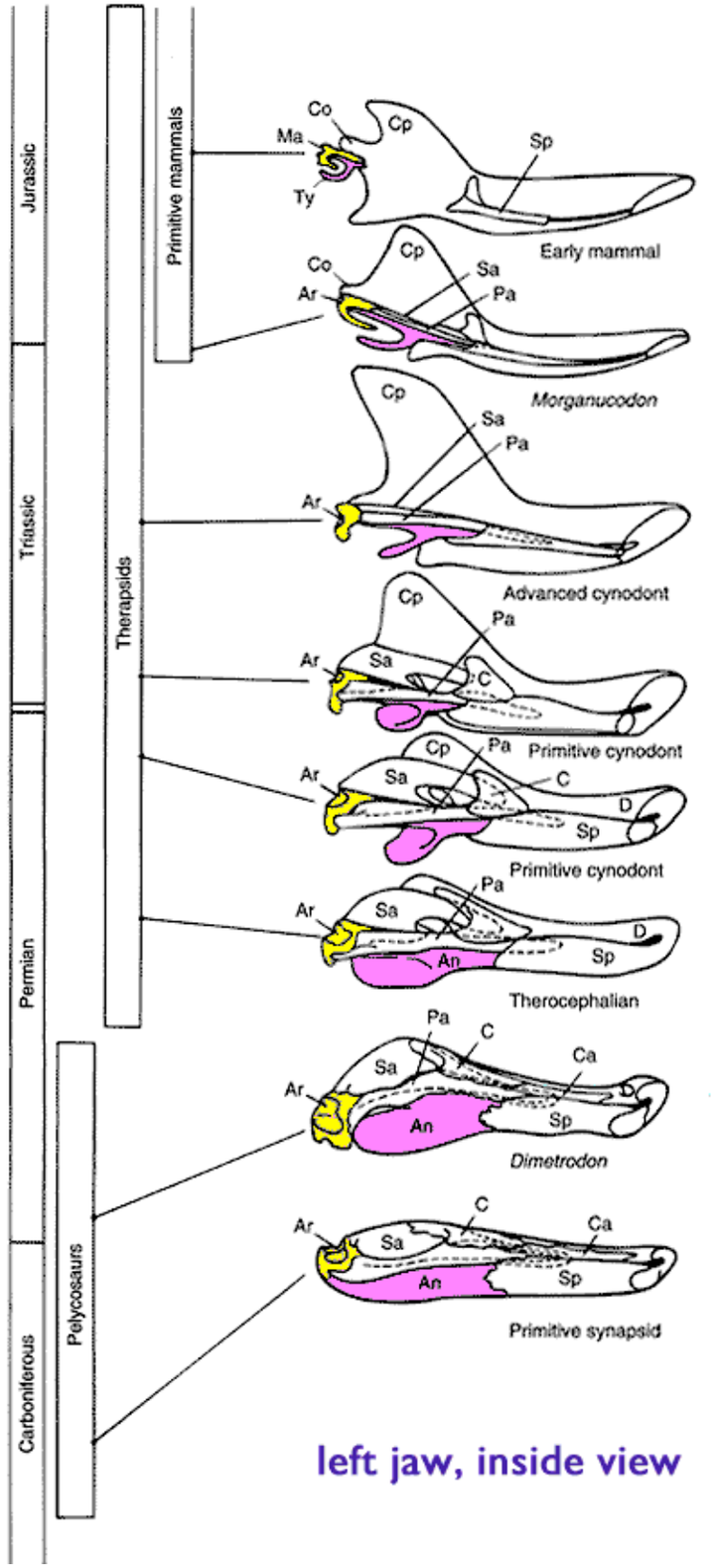
The more you complain, the longer God lets you live.

CONSTRUCTIONS

Replication lineages

Biological lineages

Lineage from reptiles to mammals



Lingueme lineages

Sounds (sound changes)

Late Latin ałtĕrum

Old French autr

Middle French ōtr

Modern French otʁ

Lingueme lineages

Words (etymologies)

Late Latin *Cyprium* ‘(copper) from Cyprus’

Common Germanic **kupur* ‘copper’

Modern English *copper* ‘copper’

Torres Strait Creole *kapa* ‘corrugated iron’

Lingueme lineages

Constructions (grammaticalization chains)

Classical Latin *haec habeo cantare* ‘I have these things to sing’

Late Latin *cantare habeo* ‘I will sing’

French *chanter-ai* ‘I will sing’

Language as a process

- A ***language*** is a population of actually occurring utterances (more precisely, linguemes) defined by the communicative isolation of their speakers—not an infinite set of possible sentences

- A ***grammar*** is an actual speaker's actual knowledge about their language, which evolves through exposure to utterances—not an idealized, homogeneous, static structure

- Grammatical ***categories and constructions*** are defined by their replication lineages—not by "essential" grammatical features

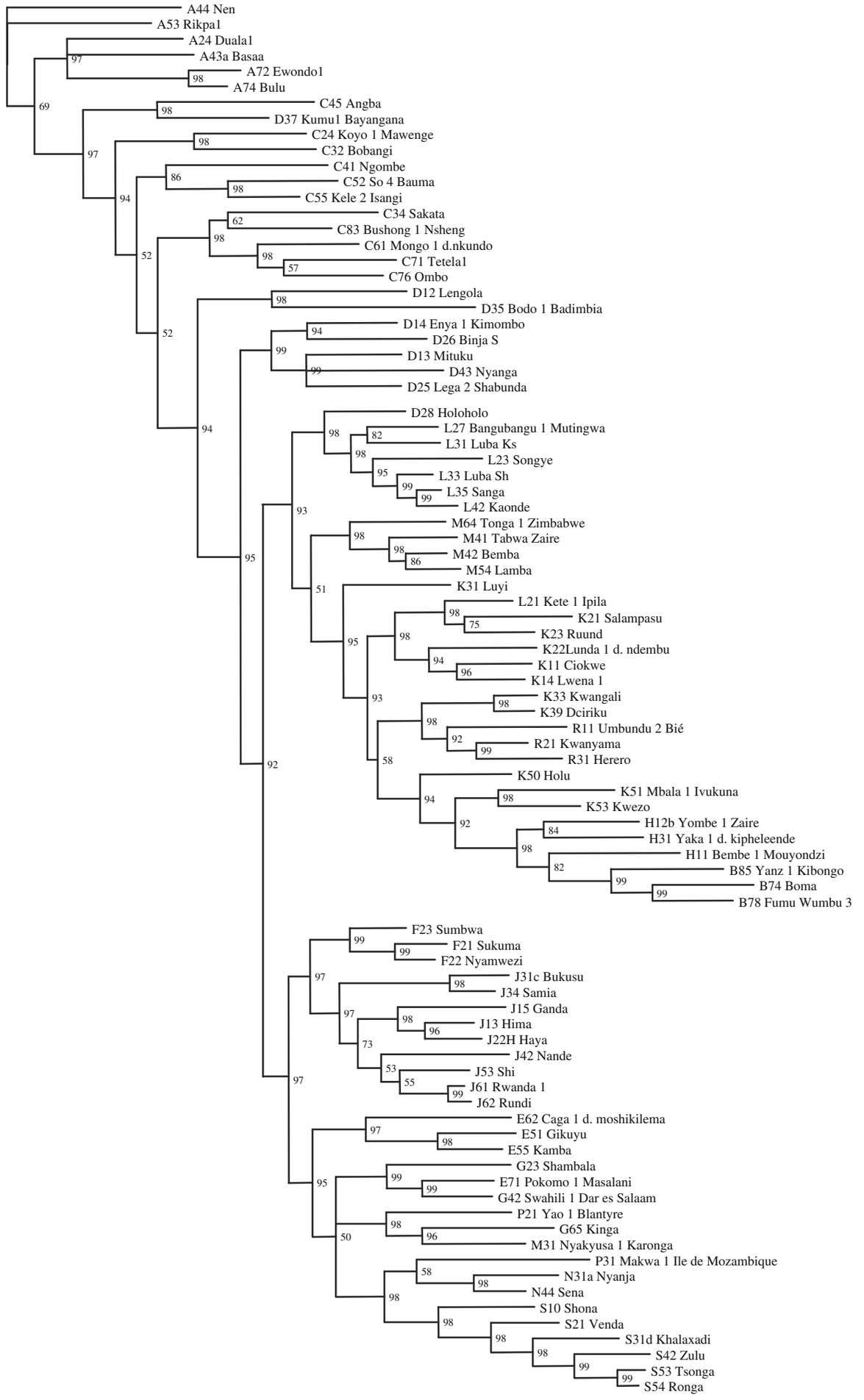
Language phylogeny

Language change and language phylogeny

- Speech communities split and lingueme lineages diverge ("the languages change")
- The result is global linguistic diversity
- Linguists reconstruct language phylogeny ("language families"), not unlike biologists
- But it's a lot harder than using genomes

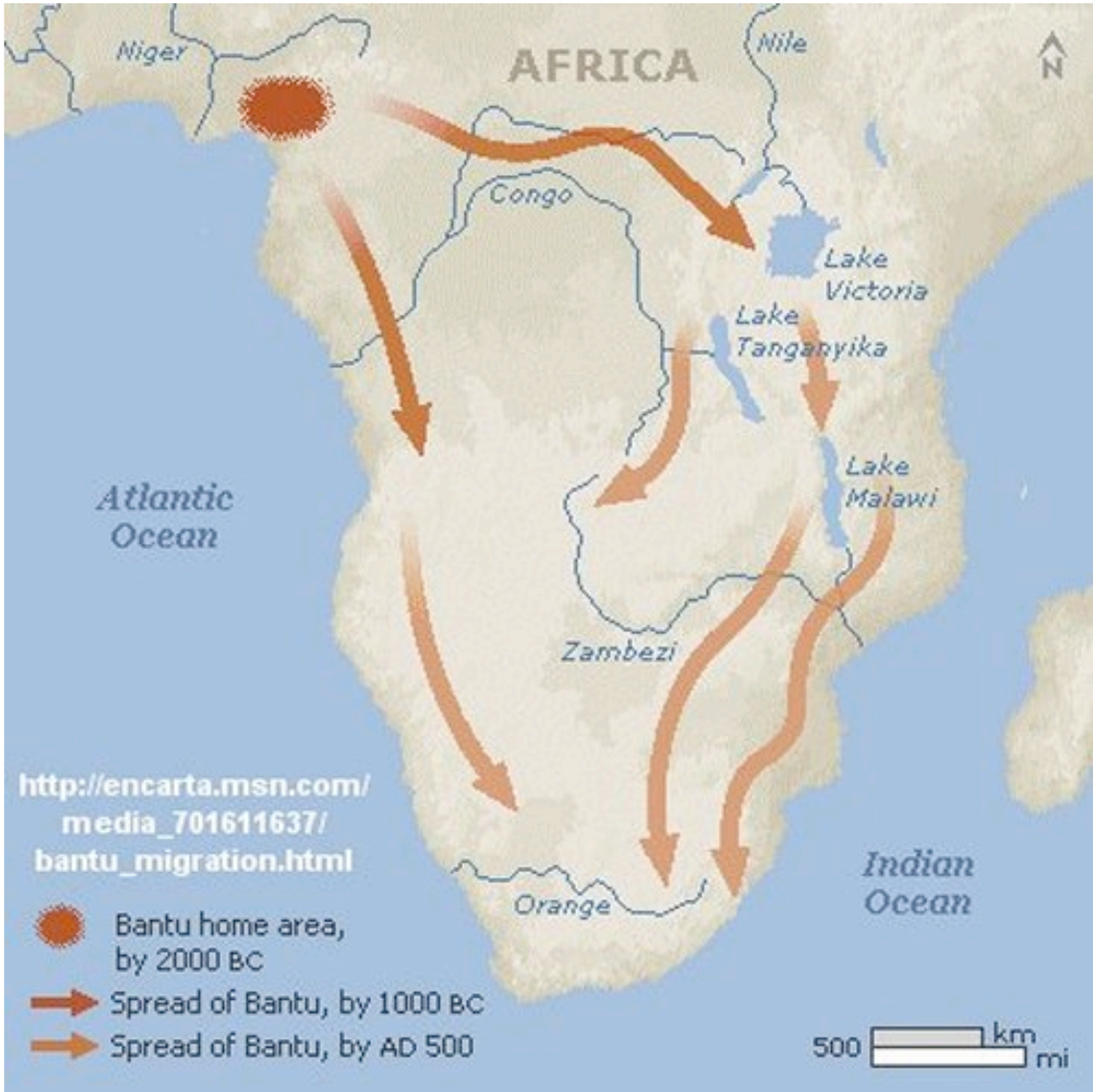
Language and human prehistory

- Bantu is a shallow language family, dating back not much more than 3000 years



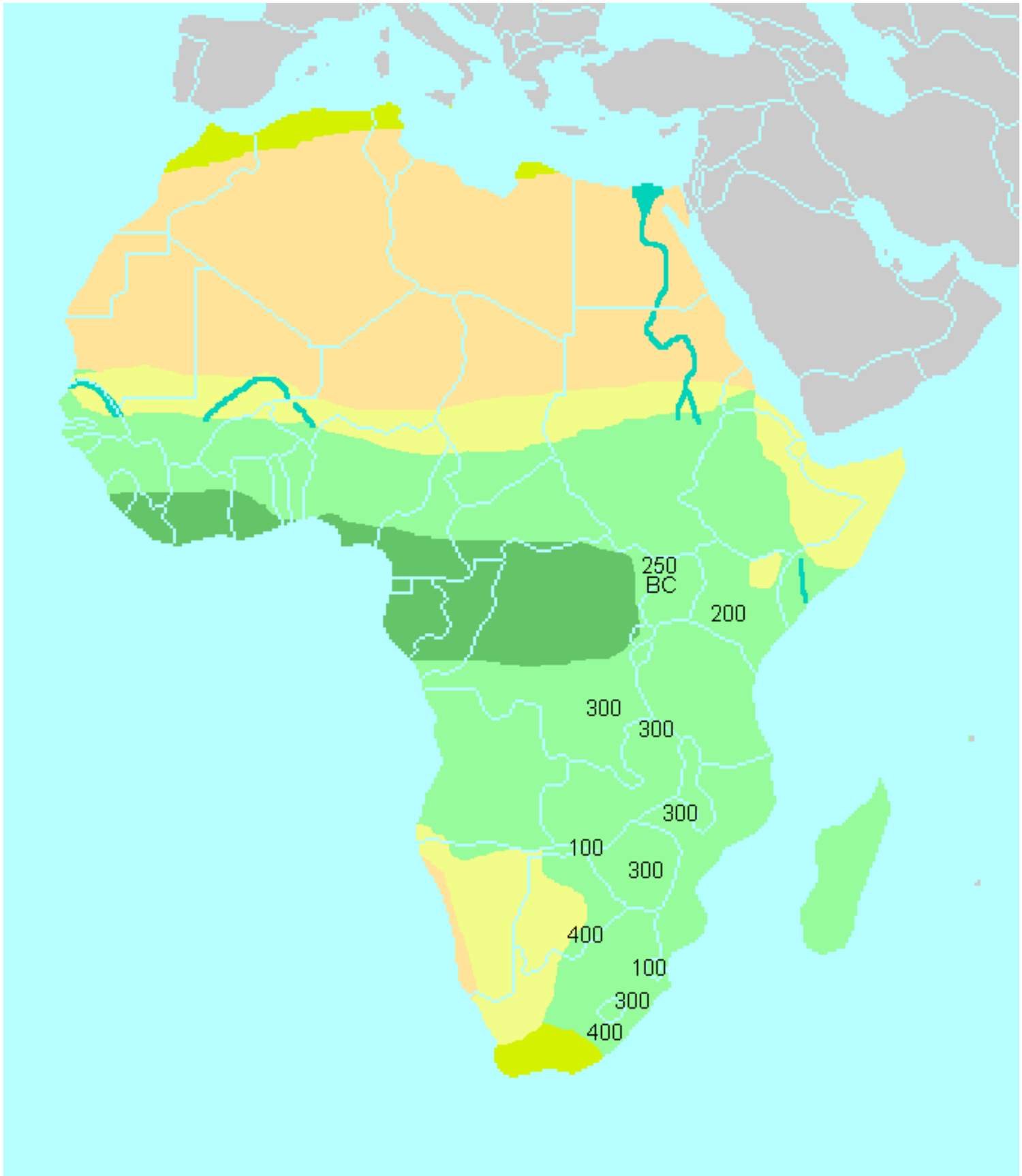
Language and human prehistory

- The phylogeny of the Bantu family indicates a rapid migration through central and southern Africa



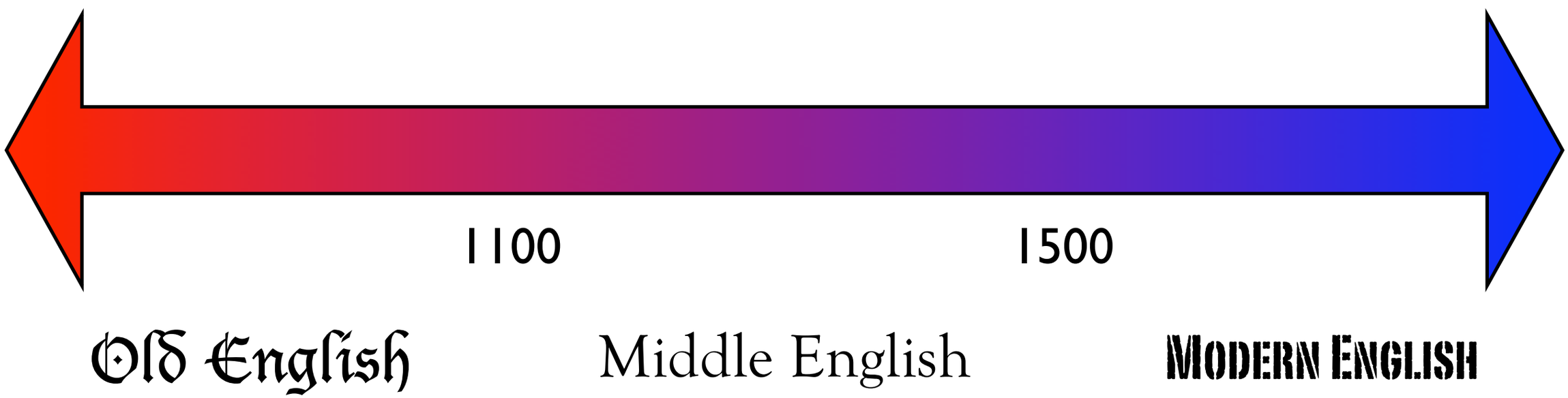
Language and human prehistory

- The prehistory of Bantu is matched by the archaeological record (spread of ironworking)



The data for language change, I

- Documents from the past (but what time and place of the language do they document?)
- An example: English has been documented as early as 600CE, up to the present



Historical change in English: Luke XV,12

990 þa dælde hi him his æhte,

1175 Đa dælde he him his ehte.

1395 And he departide to hem the catel.

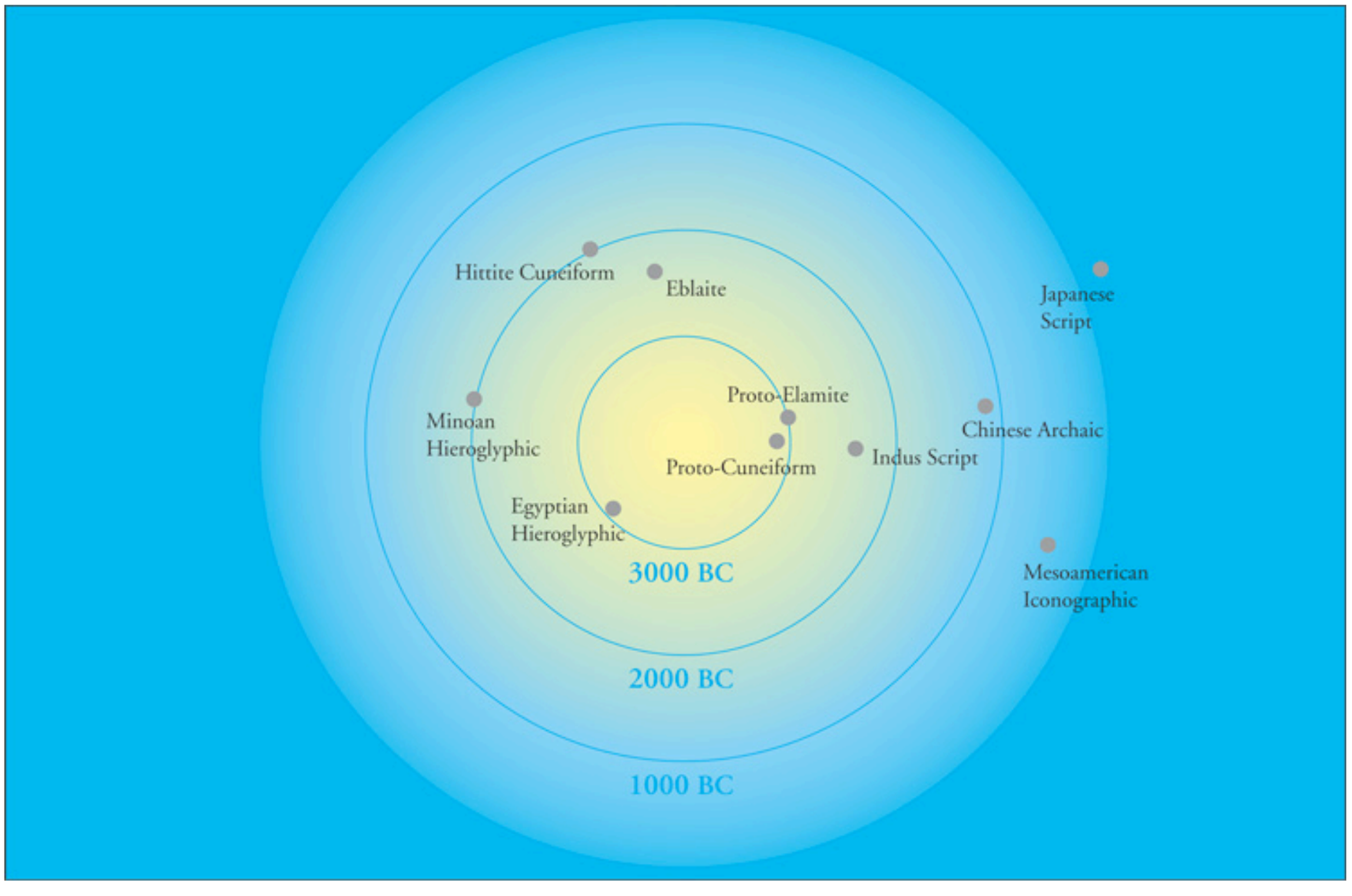
1540 And he deuided vnto them his substance.

1770 And he divided his substance between them.

1970 So he divided his estate between them.

Limits of written documents

- Only for a tiny fraction of languages does writing extend back more than a few centuries or even decades, if at all



The data for language change, II

- The data that can be used in all languages for phylogeny is comparative data
- The comparative method examines characters of a set of languages for evidence of phylogeny

	English	***Spanish***	***Welsh***	***Russian***	***Greek***	***Farsi***	***Urdu***
sound	two	dos	dau	dva	dúo	do	dō
meaning	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’

Characters used in the comparative method

- The primary characters used for language phylogeny are sound-meaning pairings: words or grammatical elements
  - In most (but not all) cases, the sound-meaning pairing is arbitrary, so it is a good diagnostic for phylogeny

	English	***Spanish***	***Welsh***	***Russian***	***Greek***	***Farsi***	***Urdu***
sound	two	dos	dau	dva	dúo	do	dō
meaning	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’

Characters used in the comparative method

- What properties of the sound-meaning pairing are used in reconstructing language phylogenies?

	English	***Spanish***	***Welsh***	***Russian***	***Greek***	***Farsi***	***Urdu***
sound	two	dos	dau	dva	dúo	do	dō
meaning	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’	‘2’

Western European languages

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	Mutter	mor	mère	madre	madre
nose	Nase	näsa	nez	naso	nariz
lip	Lippe	läpp	lèvre	labbro	labio
six	sechs	sex	six	sei	seis
father	Vater	far	père	padre	padre
foot	Fuss	fot	pied	piede	pie
for	für	för	pour	per	por
fish	Fisch	fisk	poisson	pesce	pez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Western European languages

- The similarity in sound and meaning indicates that all six languages form a taxon (a part of Indo-European)
- The initial consonants are identical in all six languages

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	**M**utter	**m**or	**m**ère	**m**adre	**m**adre
nose	**N**ase	**n**äsa	**n**ez	**n**aso	**n**ariz
lip	**L**ippe	**l**äpp	**l**èvre	**l**abbro	**l**abio
six	**s**echs	**s**ex	**s**ix	**s**ei	**s**eis
father	Vater	far	père	padre	padre
foot	Fuss	fot	pied	piede	pie
for	für	för	pour	per	por
fish	Fisch	fisk	poisson	pesce	pez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Western European languages

- Sound-meaning similarities/differences indicate that there a two subtaxa
- These are traditionally named Germanic and Romance

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	**M**utter	**m**or	**m**ère	**m**adre	**m**adre
nose	**N**ase	**n**äsa	**n**ez	**n**aso	**n**ariz
lip	**L**ippe	**l**äpp	**l**èvre	**l**abbro	**l**abio
six	**s**echs	**s**ex	**s**ix	**s**ei	**s**eis
father	Vater	far	père	padre	padre
foot	Fuss	fot	pied	piede	pie
for	für	för	pour	per	por
fish	Fisch	fisk	poisson	pesce	pez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Western European languages

- Sound-meaning similarity supports the larger taxon, but the initial consonants are not identical (*f* in Germanic, *p* in Romance)

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	**M**utter	**m**or	**m**ère	**m**adre	**m**adre
nose	**N**ase	**n**äsa	**n**ez	**n**aso	**n**ariz
lip	**L**ippe	**l**äpp	**l**èvre	**l**abbro	**l**abio
six	**s**echs	**s**ex	**s**ix	**s**ei	**s**eis
father	**V**ater	**f**ar	**p**ère	**p**adre	**p**adre
foot	**F**uss	**f**ot	**p**ied	**p**iede	**p**ie
for	**f**ür	**f**ör	**p**our	**p**er	**p**or
fish	**F**isch	**f**isk	**p**oisson	**p**esce	**p**ez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Western European languages

- In other words, similarity not just identity is useful, indeed necessary, to reconstruct linguistic phylogenies

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	**M**utter	**m**or	**m**ère	**m**adre	**m**adre
nose	**N**ase	**n**äsa	**n**ez	**n**aso	**n**ariz
lip	**L**ippe	**l**äpp	**l**èvre	**l**abbro	**l**abio
six	**s**echs	**s**ex	**s**ix	**s**ei	**s**eis
father	**V**ater	**f**ar	**p**ère	**p**adre	**p**adre
foot	**F**uss	**f**ot	**p**ied	**p**iede	**p**ie
for	**f**ür	**f**ör	**p**our	**p**er	**p**or
fish	**F**isch	**f**isk	**p**oisson	**p**esce	**p**ez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Western European languages

- Reconstruction of proto-form: did *p* become *f*, *f* become *p*, or did both come from another sound?
- Evidence from other languages indicates *p* > *f*

English	*German*	*Swedish*	*French*	*Italian*	*Spanish*
mother	**M**utter	**m**or	**m**ère	**m**adre	**m**adre
nose	**N**ase	**n**äsa	**n**ez	**n**aso	**n**ariz
lip	**L**ippe	**l**äpp	**l**èvre	**l**abbro	**l**abio
six	**s**echs	**s**ex	**s**ix	**s**ei	**s**eis
father	**V**ater	**f**ar	**p**ère	**p**adre	**p**adre
foot	**F**uss	**f**ot	**p**ied	**p**iede	**p**ie
for	**f**ür	**f**ör	**p**our	**p**er	**p**or
fish	**F**isch	**f**isk	**p**oisson	**p**esce	**p**ez
winter	Winter	vinter	hiver	inverno	invierno
hand	Hand	hand	main	mano	mano
milk	Milch	mjölk	lait	latte	leche
mouth	Mund	mun	bouche	bocca	boca

Greek
pater
Sanskrit
pitṛ
Tokharian B
pacer

Language contact and borrowing: phylogeny and reticulation

- The analysis of the data in the preceding slides assumes that the word lineages all bundle together and define a single phylogeny
- In fact, many words are ***borrowed*** from other languages, that is, their lineages jump from one phylogeny to another
- Reticulation of phylogenies may happen (dialect mixture, maybe some so-called "mixed languages"), but the contribution of each parent language is usually asymmetrical

Borrowing in Western European

- English, a supposedly Germanic language, also shows form-meaning similarities with Romance languages
- These similarities are due to language contact, in particular the Norman invasion and prestige of French

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla

Borrowing in Western European

- Hence form-meaning similarities in themselves are not a foolproof guide to phylogeny
- But the ***distribution*** of form-meaning pairings can be a clue to common ancestry vs. contact

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla

Borrowing in Western European

- Common ancestry gives rise to similarities that are distributed across languages, in basic vocabulary, with less obvious similarities (depending on depth of phylogeny)

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla

Borrowing in Western European

- Borrowing gives rise to similarities that are often associated with just one language (English in this case), with less basic vocabulary, and with a higher degree of similarity, especially between the languages in contact

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla

Borrowing in Western European

- Having inferred a borrowing relationship between English and French, one can make inferences about other patterns of form-meaning similarity

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla
Treue	tro	faith	foi	fede	fé

Borrowing in Western European

- Having inferred a borrowing relationship between English and French, one can make inferences about other patterns of form-meaning similarity

German	*Swedish*	*English*	*French*	*Italian*	*Spanish*
Schloss	slott	castle	château	castello	castillo
Tal	dal	valley	vallée	valle	valle
Dorf	by	village	village	villaggio	aldea
Bauernhof	bondgård	farm	ferme	fattoria	finca
Stuhl	stol	chair	chaise	sedia	silla
Treue	tro	faith	foi	fede	fé
Alkohol	alkohol	alcohol	alcool	alcool	alcohol

Accidental similarity

- Not all sound-meaning similarities have a historical explanation (common ancestry or borrowing)
- But a broader distributional perspective demonstrates chance vs. phylogeny

	French	***Italian***	***Spanish***	***Hawaiian***	***Samoan***	***Tongan***
honey			miel	meli		
eye	yeux	occhio	ojo	maka	mata	mata
fish	poisson	pesce	pez	iʔa	iʔa	ika
blood	sang	sanguo	sangre	koko	toto	toto
man	homme	uomo	hombre	kanaka	taŋata	taŋata
cry (out)	crier	gridare	gritar	kani	taŋi	taŋi

Accidental similarity

- The problem is ***selection bias***: select only the words and languages that happen to have sound-meaning similarity
- Selection bias leads to bizarre claims: Bantu and Sumerian; Zuñí and Japanese; Basque and just about anything

	French	***Italian***	***Spanish***	***Hawaiian***	***Samoan***	***Tongan***
honey			miel	meli		
eye	yeux	occhio	ojo	maka	mata	mata
fish	poisson	pesce	pez	iʔa	iʔa	ika
blood	sang	sanguo	sangre	koko	toto	toto
man	homme	uomo	hombre	kanaka	taŋata	taŋata
cry (out)	crier	gridare	gritar	kani	taŋi	taŋi

Traditional historical linguistics doesn't have selection bias, does it?

- The data supporting a phylogeny is published in the form of an ***etymological dictionary***

Proto-Eskimo **nakaʀ 'stalk'**
Central Alaskan Yupik: nakaag 'wild rhubarb'
Naukan Siberian Yupik: nakayiaXat 'grass variety', nakanɬat 'grass'
Seward Peninsula Inuit: naɣaaq 'flower of sourgrass'
North Alaskan Inuit: naka(a)a, nakaun, nakautaq 'stem of edible plant, stalk'
Western Canadian Inuit: nakaq 'stem'
Eastern Canadian Inuit: nakaq 'stem'

Traditional historical linguistics doesn't have selection bias, does it?

- An etymological dictionary presents only the words which are believed to be cognate, in only the languages which are believed to belong to the family

Proto-Eskimo **nakaʀ 'stalk'**
Central Alaskan Yupik: nakaag 'wild rhubarb'
Naukan Siberian Yupik: nakayiaXat 'grass variety', nakanɬat 'grass'
Seward Peninsula Inuit: naɣaaq 'flower of sourgrass'
North Alaskan Inuit: naka(a)a, nakaun, nakautaq 'stem of edible plant, stalk'
Western Canadian Inuit: nakaq 'stem'
Eastern Canadian Inuit: nakaq 'stem'

Traditional historical linguistics doesn't have selection bias, does it?

- In other words, a outside observer could argue that the data in an etymological dictionary reflects selection bias (and they have, in many cases)

Proto-Eskimo **nakaʀ 'stalk'**
Central Alaskan Yupik: nakaag 'wild rhubarb'
Naukan Siberian Yupik: nakayiaXat 'grass variety', nakanɬat 'grass'
Seward Peninsula Inuit: naɣaaq 'flower of sourgrass'
North Alaskan Inuit: naka(a)a, nakaun, nakautaq 'stem of edible plant, stalk'
Western Canadian Inuit: nakaq 'stem'
Eastern Canadian Inuit: nakaq 'stem'

Traditional historical linguistics doesn't have selection bias, does it?

- The data that really could demonstrate a valid taxon is a ***comparative dictionary***, showing all the forms and all the meanings in a word list, for all the languages in a region

Proto-Eskimo **nakaʀ 'stalk'**
Central Alaskan Yupik: nakaag 'wild rhubarb'
Naukan Siberian Yupik: nakayiaXat 'grass variety', nakanɬat 'grass'
Seward Peninsula Inuit: naɣaaq 'flower of sourgrass'
North Alaskan Inuit: naka(a)a, nakaun, nakautaq 'stem of edible plant, stalk'
Western Canadian Inuit: nakaq 'stem'
Eastern Canadian Inuit: nakaq 'stem'

Distributional patterns and sound correspondences

- Distributional patterns of assumed cognates across languages forms the basis of phylogeny reconstruction (and computational techniques from biology can be used to refine the phylogeny)
- Distributional patterns can also help to identify borrowings and accidental sound-meaning similarities
- But a stronger diagnostic is the occurrence of regular sound correspondences
- Regular sound correspondences form the basis of the comparative method

The Kuhnian exemplar for the comparative method

- In an earlier slide, we observed that *p* > *f* in Germanic; also *k* > *h*, *t* > *θ* (Grimm's Law)

unconditioned sound change

Latin	**p**iscis	***Gothic***	**f**isks	'fish'
Latin	**c**anis	***Old English***	**h**und	'dog'

The Kuhnian exemplar for the comparative method

- But not every Germanic word with the sounds undergo the sound change

unconditioned sound change

conditioned sound change

Latin	**p**iscis	***Gothic***	**f**isks	‘fish’
Latin	**c**anis	***Old English***	**h**und	‘dog’
Latin	s**p**uere	***Old High German***	s**p**iwan	‘spit’
Latin	pis**c**is	***Gothic***	fis**k**s	‘fish’

The Kuhnian exemplar for the comparative method

- Still other words undergo a different change: *p* > *b*, *t* > *d*, *k* > *g* (Verner's Law)

unconditioned sound change

conditioned sound change

conditioned sound change - conditioning environment lost

Latin	p**p**iscis	***Gothic***	f**f**isks	'fish'
Latin	**c**anis	***Old English***	**h**und	'dog'
Latin	s**p**uere	***Old High German***	s**p**iwan	'spit'
Latin	pis**c**is	***Gothic***	fis**k**s	'fish'
Greek	pa**t**ér	***Old English***	fæ**d**er	'father'
Greek	he**k**urá	***Old High German***	swi**g**ur	'mother-in-law'

Principles of the Comparative Method

- Decide by inspection that certain languages form a taxon
- Place side by side words with similar meanings to compare
- Identify regular sound correspondences
- Posit a plausible sound in the ancestral language for each correspondence
- Reconstruct the word form in the language

Problem (1): sound changes are not always apparently(?) regular

- A textbook example of a universally accepted taxon: Romance languages

Problem (1): sound changes are not always apparently(?) regular

- A example from a textbook of sound correspondences in Western Romance

Portuguese	*Spanish*	*Catalan*	*French*	
bar**b**a	**b**arba	**b**ar**b**ə	**b**ar**b**	‘beard’
brãku	**b**lanko	**b**laŋ	**b**lã	‘white’
bajšu	**b**axo	**b**aš	**b**a	‘low’
bej	**b**jen	**b**e	**b**jẽ	‘well (adv.)’
verdi	**b**erde	**b**ɛrt	**v**ɛr	‘green’
vĩñu	**b**ino	**b**i	**v**ẽ	‘wine’
kal**v**u	kal**b**o	kal**p**	šo**v**	‘bald’
nɔ**v**u	nwe**b**o	nɔ**w**	nø**f**	‘new’

Problem (1): sound changes are not always apparently(?) regular

- A Portuguese historical grammar on Latin sē**p**ĭam > Port. si**b**a vs. Latin ă**p**ĭum > Port. ai**p**o:

It has been suggested that yod [*ĭ*] lengthened *p* to *pp* after a Classical Latin short vowel but not after a long vowel: *ăpĭum* > **appiu* > *aipo*, but *sēpĭam* > *siba*; and accordingly, the *b* of *caibo* (from *căpĭo*), etc., and the *b* of *saiba* (from *săpĭat*), etc. have been explained as analogical with other forms of these verbs in which there was no yod. But the form *aipo* is probably semi-learned; cf. Spanish *apio*.

undocumented conditioned sound change

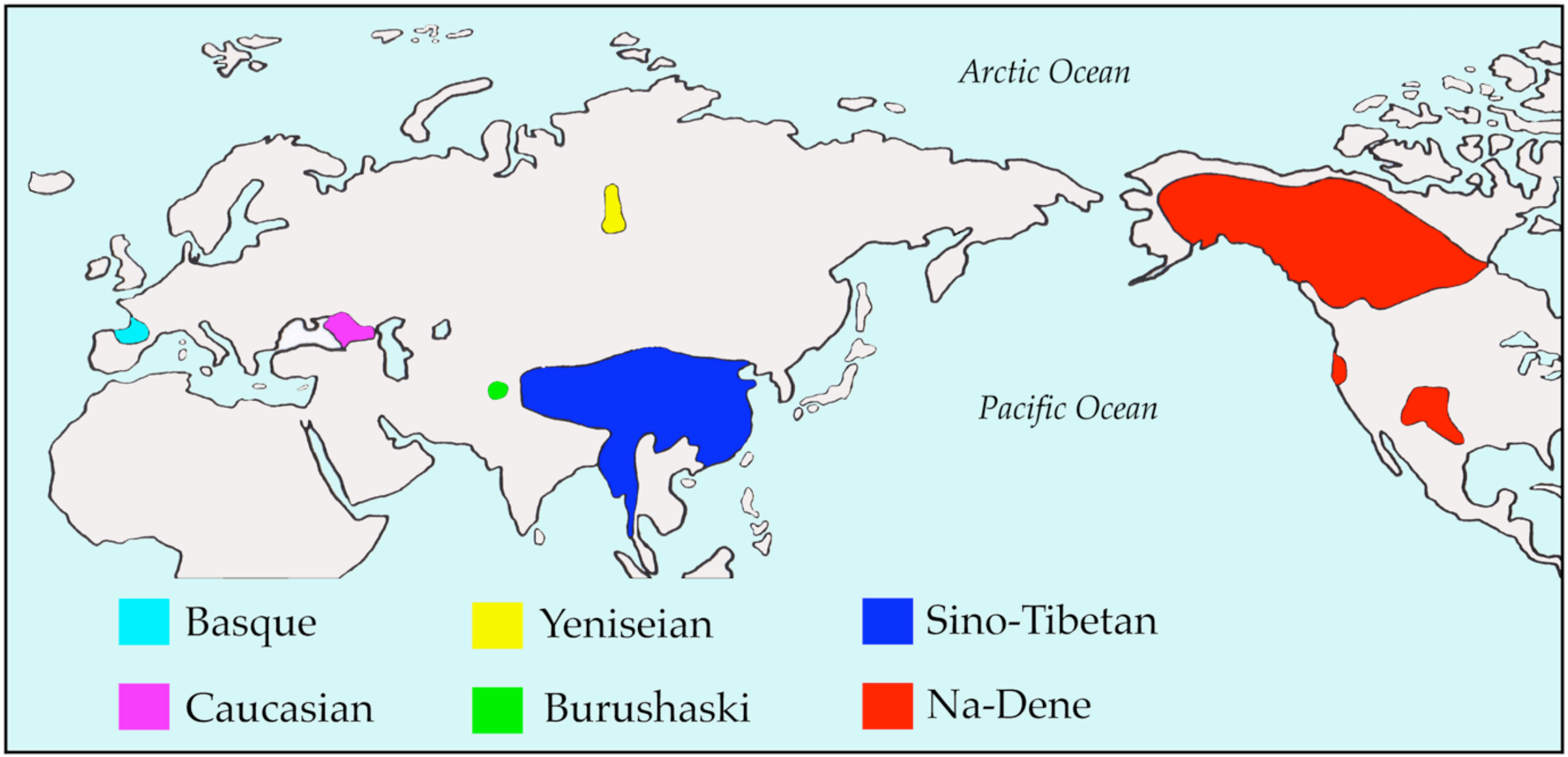
analogical sound change

borrowing from learned register

Problem (II): borrowed words display regular correspondences

“Turko-Arabic”		
Arabic	***Turkish***	
ʕaji**b**a	aji**b**e	‘marvel (n.)’
ray**b**iya	rey**b**iye	‘skepticism’
tāli**b**	tali**p**	‘student’
dirham	**d**irhem	‘dirham (coin)’
ʕā**b**i**d**	a**b**i**t**	‘worshipper’

Problem (III): What happens when "inspection" isn't so easy?



The Dene-Caucasian Family

Problem (III): What happens when "inspection" isn't so easy?

	'I'	*'you'*	*'we'*	*'one'*	*'two'*	*'three'*
Basque	ni	hi, zu	gu	bat	bi(ga)	hiru(r)
Ubykh	sə-	wə-	ɕə-	za	$tq^{w\prime}a$	ɕə
Chechen	suo	ħuo	tχuo	tsħaʔ	šiʔ	qoʔ
Tsakhur	zə	vu, ʁu	ši	sa-	$q'o^{ʕ}$-	xeb-
Burushaski	dže	ūn, ūŋ	mi	hi(k)	ālti, ālto	īski
Tibeto-Burman	*ŋa	*naŋ		*kat, *it, *tik	*g-nis	*g-sum
Arch. Chinese	ŋo	ño		jět	ńjər, ńźi	səm
Ket	āˈt, āb-	ūˈ, ūk-	ɤ̄ˈt, ɤ́tn	qūˈs	ūˈn	dɔʔŋ
Haida	ɬaa, dii	daa, daŋ	t'allŋ, ʔiitɬ'	sɢwaan	sdiŋ	ɬɢunʔuɬ
Koyukon	see	neɳ	χʌn	k'eeɬ	neteeχ	toχ

Tackling/evading the problem

- Greenberg: distributional patterns of sound-meaning similarity without establishing regular sound correspondences is sufficient for deeper families
  - ✦ Rejected by almost all traditional historical linguists
- Propose regular sound correspondences among languages claimed to be deeply related
  - ✦ Accused of selection bias by, and explanations of irregular correspondences rejected by, most traditional historical linguists
- Traditional historical linguists: restrict one's attention to shallow families which really are identifiable as taxa by "inspection"

Comparative historical linguistics is– or should be–crying out for a complex systems model

- Information emerges in distribution patterns of sound-meaning similarities and in patterns of phonological (and semantic) similarity
- Multiple independent factors, of which regular sound change is only one, account for observed patterns
- We must posit a probability distribution of different histories leading to the observed patterns

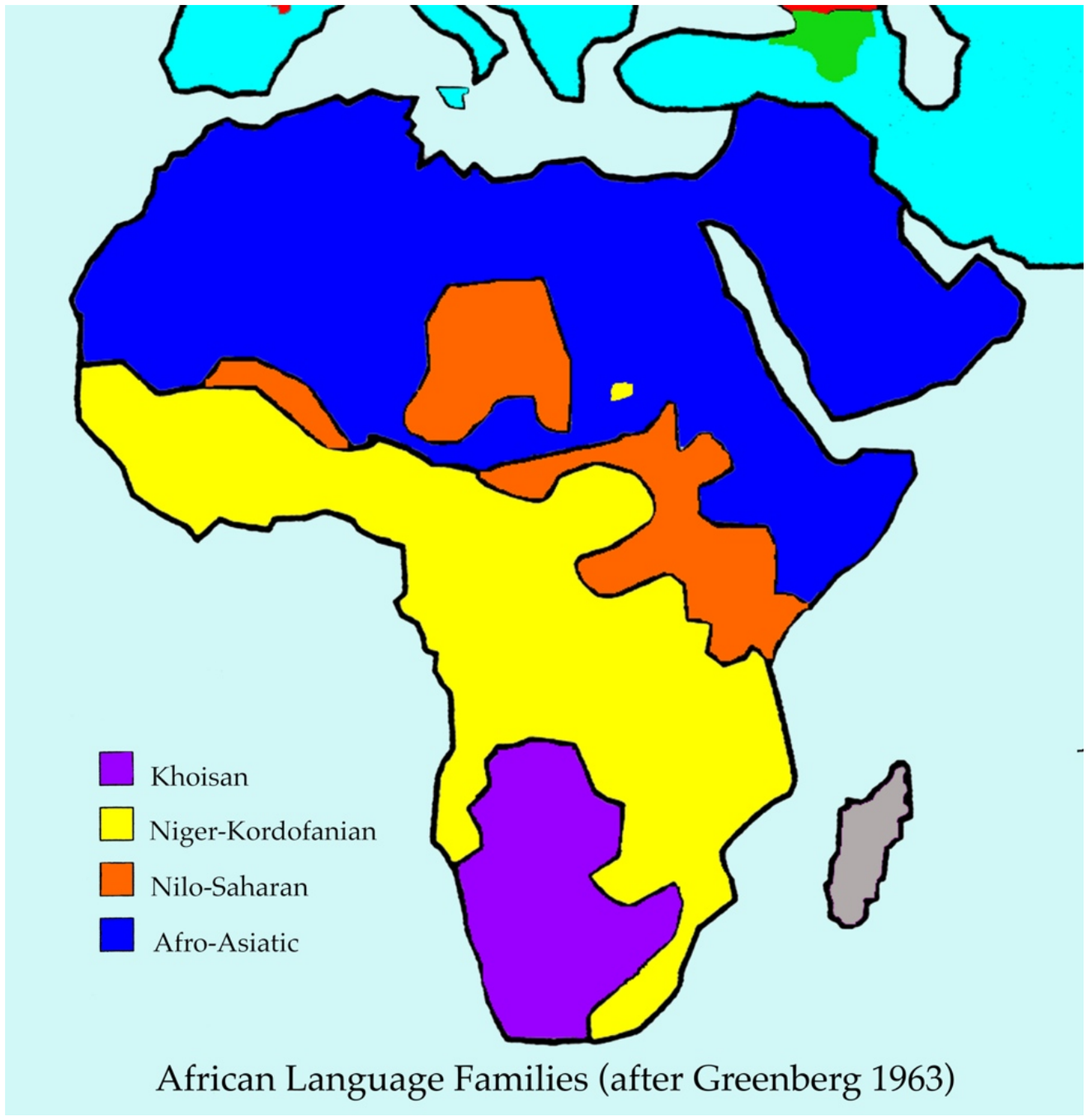
Linguistic Diversity

Linguistic Diversity

- ***Genetic diversity***: the diversity of languages and their phylogenies
  - ✦ Often related to geography, because genetic diversity reflects human demographic history
- ***Typological diversity***: the diversity of lingueme structures (sounds, words, constructions) found in the world's languages
  - ✦ Generally related to language processing (articulatory/auditory for sound patterns, semantic/discourse for grammatical patterns)

A puzzle of genetic linguistic diversity

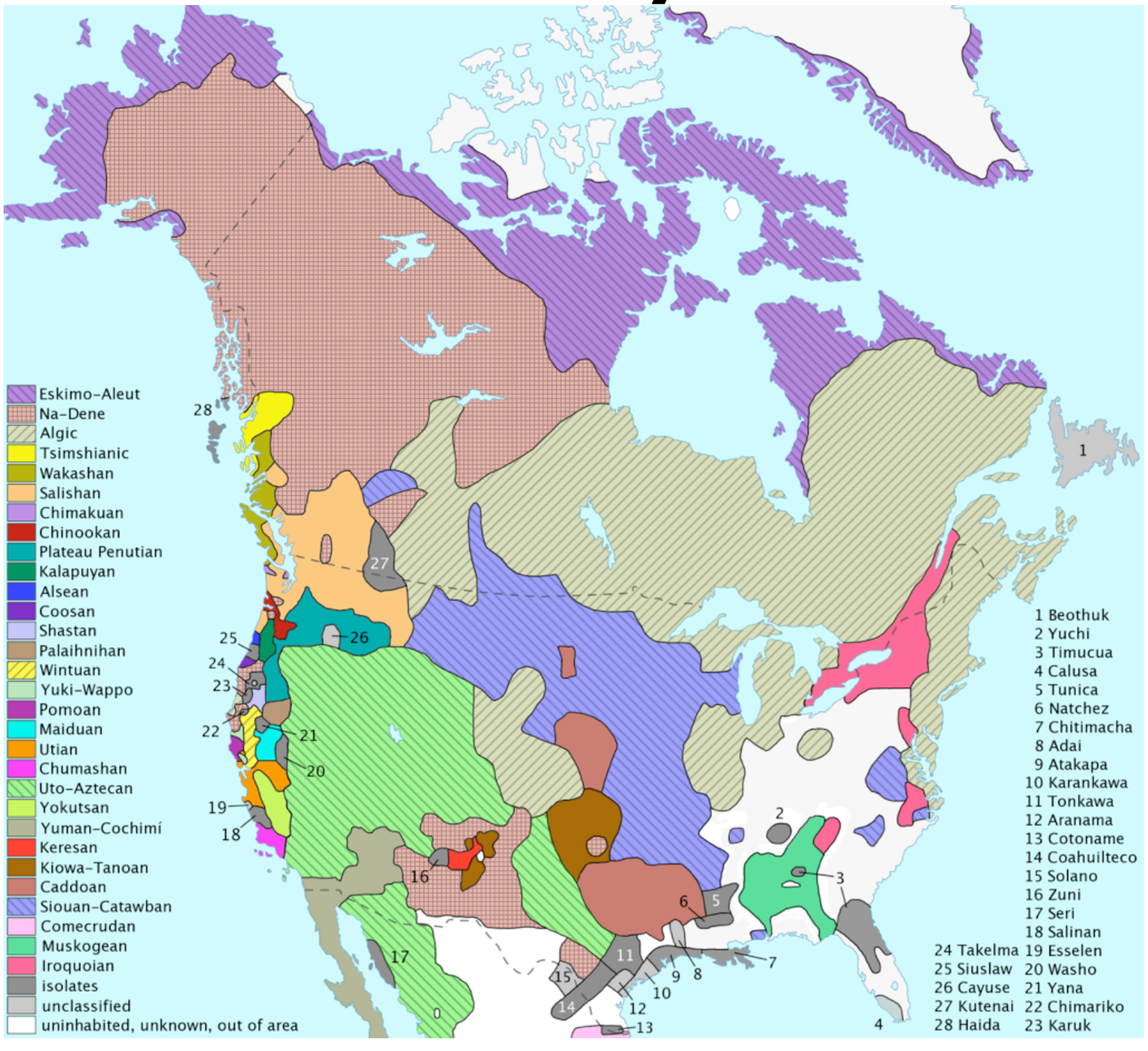
- Africa is the origin place of modern humans
- Yet Africa has only four major families, with a scattering of isolated languages



African Language Families (after Greenberg 1963)

A puzzle of genetic linguistic diversity

- The Americas were the last continents to be colonized
- Yet the Americas appear to be extremely diverse in genetic linguistic terms



A puzzle of genetic linguistic diversity

- Within the Americas, some regions are genetically very diverse…
- …while other regions are genetically much more uniform
- Why?