SFI Complex Systems Summer School

Typological patterns in language and their relation to linguistic history

Ian Maddieson

ianm@berkeley.edu

and William Croft

wcroft@unm.edu

Genetic diversity puzzle

☀ Africa vs Americas: oldest vs newest populations — least vs most diverse genetic language families

☀ True or apparent difference?

☀ Language families in Americas more related than traditionally recognized?

☀ More language families in Africa than than traditionally recognized?

☀ Convergence between African languages?

☀ Replacement of earlier African families?

Genetic diversity puzzle

- ☀ Conventional historical/comparative linguistics seems unable to advance
- ☀ Sociology of science? Africanists accept 'lumping'; Americanists reject it
- ☀ New tools and methods include insights from *language typology*
- ☀ Typological patterns can provide prior probabilities as input to construction of family trees

3

Language typology

☼ Languages in constant process of change

☼ Changes convergent or divergent: increase or decrease similarity between given pair of languages

☼ Languages have similarities due to

— inheritance of traits (perhaps modified) from a common ancestor

— contact between languages leading to transfer of traits from one to another

☼ And they also have similarities because similar processes apply to them, or at least processes that produce similar outcomes (cf biology)

Homoplasy

homoplasy: biological term for a similarity of outcome not due to common origin, e.g. wings of birds and bats
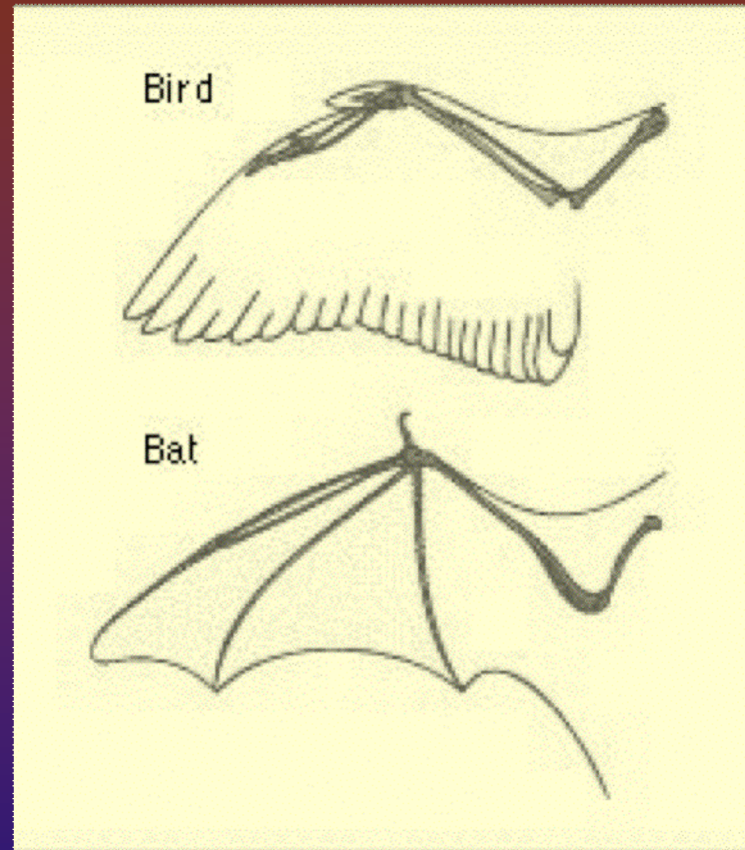
Bird and bat wings are structurally different: bird wing supported by digit number 2, bat wing by digits 2-5.

Homoplasy

Homoplasies can concern a complex of functional and behavioral patterns as well as physical properties

Especially shared between some insects (ants, bees, wasps; termites) and some mole rats (naked mole rat, Damara mole rat)

Naked mole rats

Honey bees

6

Homoplasy

- Homoplasies develop because similar structural and behavioral patterns are effective adaptations to particular ecological niches

- Likewise, similarities between languages can develop because similar solutions are effective responses to the situation of their use

- *Languages* can be metaphorically envisioned as "species" evolving in the ecosystem formed by the community of their users

- *Language* is also literally a major adaptive biological trait of *homo sapiens* as a species

Language typology

☀ Study of similarities/differences across languages is the field of linguistic typology

☀ Linguistic typology seeks to establish how languages vary along different parameters, how frequent the different variations are and what patterns of co-occurrence of different traits exist

☀ Some traits may be so widespread that they are considered *linguistic universals* — they are part of the basic design of language itself

8

Language universals

☼ Language universals may reflect:

    a) basic structures of human cognition or physiology

    b) the nature of children's language acquisition tools

    c) shared aspects of the social function of language

    d) or, possibly, inherited characteristics of a hypothetical proto-human language

☼ Since many modern languages are poorly described or little known and many languages have gone extinct without trace, caution must be exercised in generalizing from currently available data — but it's our best guide

Language universals

☀ <u>Apparent</u> language universals which do not reflect anything truly basic about language may appear because

    a) only certain languages and language families have survived the accidents of history

    b) characteristics of dominant languages have been incorporated into neighboring languages

☀ We must also take care that we don't infer universals from the patterns found in a few familiar languages — hence concern to base universals on large samples of genetically and geographically diverse languages

Language typology

☼ *Any* characteristic or set of characteristics can be used to establish a typology and there is an infinite number of possible typologies, but only those that we believe show us something interesting about language structure are worth constructing

☼ For example we could divide languages into two types — those which have words meaning 'fog' and 'dog' which rhyme, and those which don't. But this is not interesting

Language typology

☀ On the other hand if we construct a typology of languages according to the occurrence of oral and nasalized vowels in their phonemic inventory we find something interesting:

☀ In a sample of 556 languages
   116 (21%) have both oral and nasalized vowels
   440 (79%) have only oral vowels
   0 (0%) have only nasalized vowels

☀ This reveals a language universal that tells us something interesting about language design: all languages have oral vowel phonemes

Language typology

☀ Further, there is an ***implicational*** relationship:
— a language only has nasalized vowel phonemes if it has oral vowel phonemes

☀ Nasalized vowels (almost) always arise from the spread of nasality from a consonant to an adjoining vowel, followed by loss of the nasal consonant (e.g. VN → v̨), as in Latin *panem → pan →* French *pE)* 'bread')

☀ No language has only nasal <u>consonants</u>, so no language will have only vowels flanked by nasal consonants; hence none will develop nasalized vowels in every word and eliminate all oral vowels

Language typology

☼ But if the only process concerning nasalized vowels, was VN → ṽ) then the proportion of languages with nasalized vowels would be constantly increasing

☼ However, other processes can result in denasalization of vowels

☼ Simple loss of nasalization ṽ) → v
Old Church Slavonic znajõ > Russian znaju "I know"

☼ Restoration of VN structure
In some varieties of Portuguese, earlier word-final
ṽ) → ṽ)N (Sampson 1999)
Latin unum > un > ũ) > ũN   "one"

14

Language typology

☀ If processes that create and destroy nasalized vowels are in balance, the proportion of languages with them will remain in (approximate) equilibrium

☀ Question: Is the estimated proportion of current languages with v)'s (21%) a good estimate of the probability that any given language (at any time period) will have nasalized vowels?

☀ Needed: a model to estimate whether equilibrium exists (more on this later — note that this issue concerns *any* typological property)

Co-occurrence patterns

☼ Much that is interesting to linguists in the study of typology concerns such patterns of co-occurrence — what characteristics go along with other characteristics — either as strict *implicational hierarchies* (as above) or as general patterns of association

☼ Typological properties can be studied for all aspects of language — semantics, syntax, morphology, phonology, information structure, etc

A semantic pattern

- A well-studied topic — color terms
- Many studies using a standard set of 'color chips' to determine a 'map' of the range of basic color terms in different languages (reproduction below — but colors not faithful to original!)
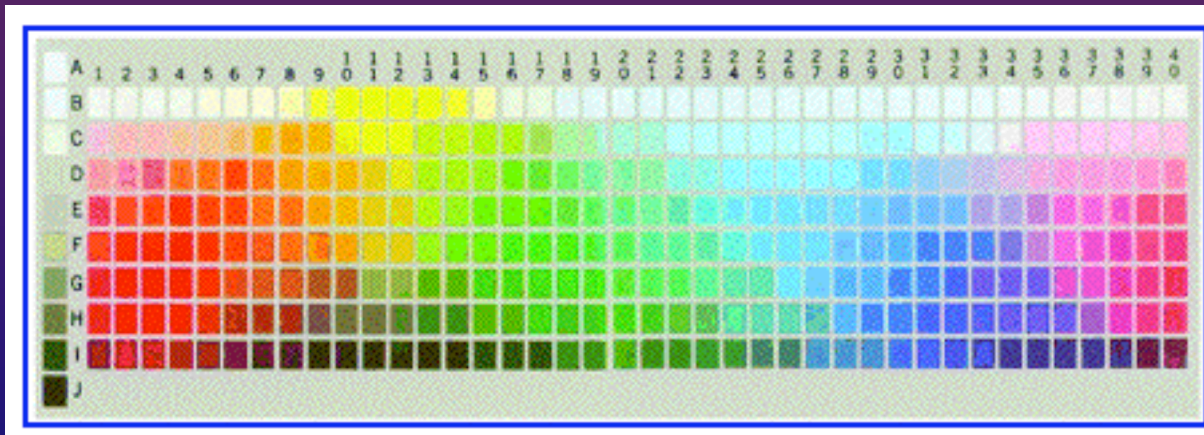
Basic color terms

☀ ***Basic*** color terms are
morphologically simple (e.g. green, not greenish)
less specialized in meaning (e.g. red, not scarlet)
not recently borrowed (e.g. not taupe)

☀ Suggested universal structures and evolution of color terms as indicated in next chart — original/simplest state is division into 'warm' and 'cool' colors

☀ White separates out first from 'warm'

☀ 'cool' splits into green/blue vs black; or green vs blue/black; or remains whole while red and yellow split

Ejagham
(Nigeria/Cameroun)
e!ba!re! *'white'*
e!bi! *'red/yellow'*
e!nyàgà *'black/green/blue'*

Mura-Pirahã (Brazil)
bi!opa@ia!i *'white'*
bi!i~sa!i *'red/yellow'*
ko!biai! *'black'*
a!hoa!saa!gà
'green/blue'

Kwerba (Papua)
´siram *'white'*
nokonim *'red'*
kainanesEnum *'yellow'*
icEm *'black/green/blue'*

Figure 3

English

'white'   'green'

'red'   'blue'

'yellow'   'black'

Figure 3

Morphology: Patterns in pronouns

☀ All languages have pronouns

☀ All languages make person distinctions in pronoun reference — distinguishing at least speaker (*first*) and spoken to (*second*)

☀ Most languages make a distinction between *singular* and *plural* reference in pronouns

☀ Some languages make a further distinction between reference to two and to more than two persons/objects (*dual* vs *plural*)

Morphology: Patterns in pronouns

☼ Some languages have a further distinction between reference to a few more than two and to many more than two persons/objects (*paucal* vs *plural*)

☼ Implicational universals:
— If a language has dual pronoun forms, it also has plural forms
— If a language has paucal pronoun forms, it also has dual forms

Mandarin subject pronouns:

	singular	plural
first	wo#	wo#m´n
second	ni‹	ni‹m´n
third	ta@	ta@m´n

Swahili independent pronouns:

	singular	plural
first	mimi	sisi
second	wewe	ninyi
third	yeye	wao

Dyirbal subject pronouns:

	singular	dual	plural
first	Naᶑa	Naliᶞi	NaNaᶞi
second	Ninda	≠ubalaᶞi	≠uraᶞi

Comanche subject pronouns:

	singular	dual	plural
first incl	nu/, nuu	tahu, takWu	tanu
first excl	— —	nukWu	nunu
second	unu	muhu, mukWu	munu
third	i/	ituhu, itukWu	ituu

23

Universals of pronouns

☼ Some languages, e.g. Comanche, make a distinction between first-person plural pronouns referring to the speaker and others <u>including</u> the addressee and ones referring to the speaker and others but <u>excluding</u> the addressee (*inclusive* vs *exclusive*)

☼ This (and all typologies) can be usefully viewed geographically — to suggest, e.g., regions where languages may be converging due to contact

Inclusive/exclusive pronouns: absent in W. Eurasia/North Africa; dominant in Australian languages
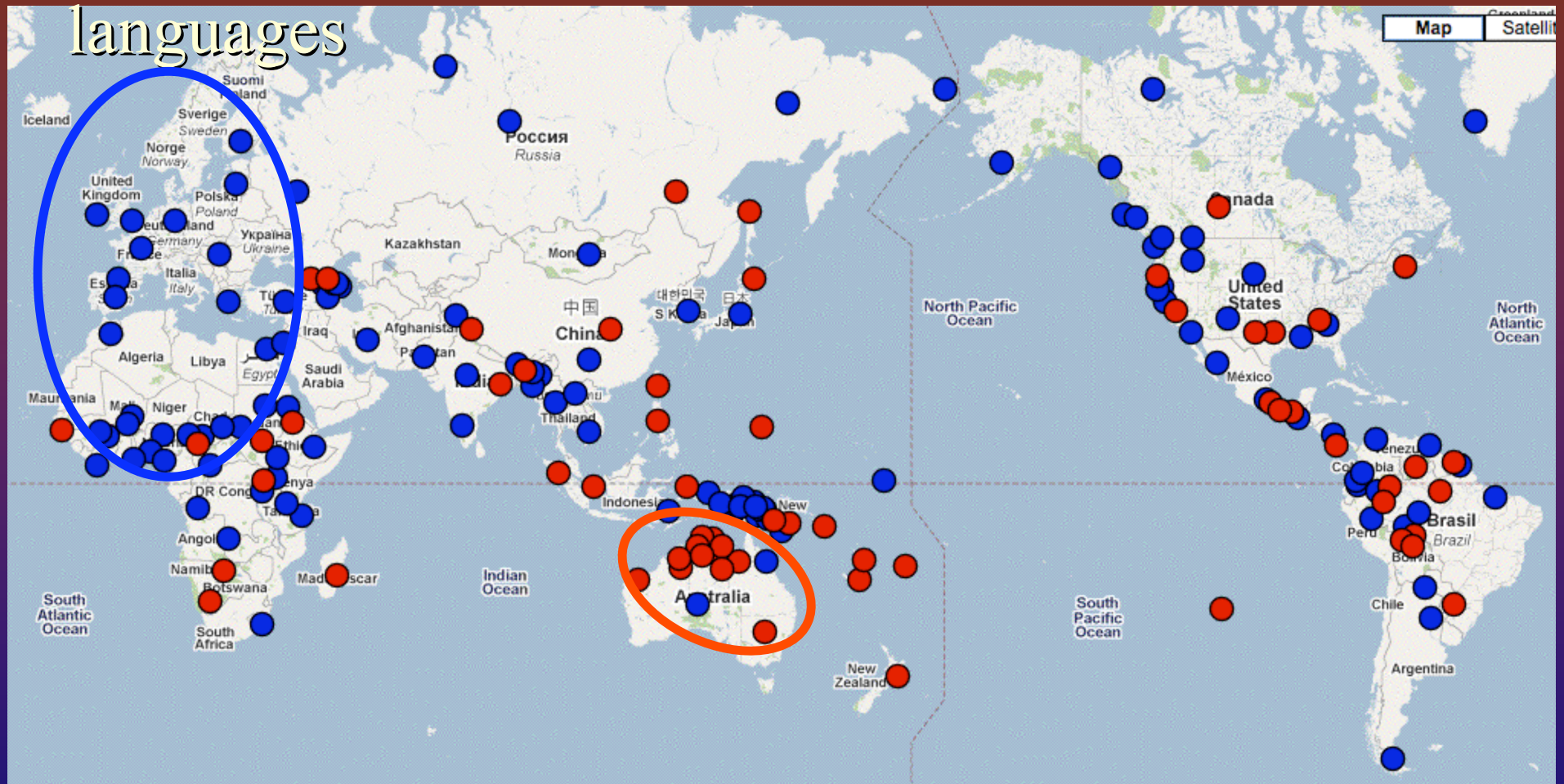
Red - Inclusive/exclusive distinction in pronouns (63)
Blue - No distinction (120)   *from Cysouw (2005)*

Map from the WALS project

Syntax: word order patterns

☀ **Basic word order** typology:
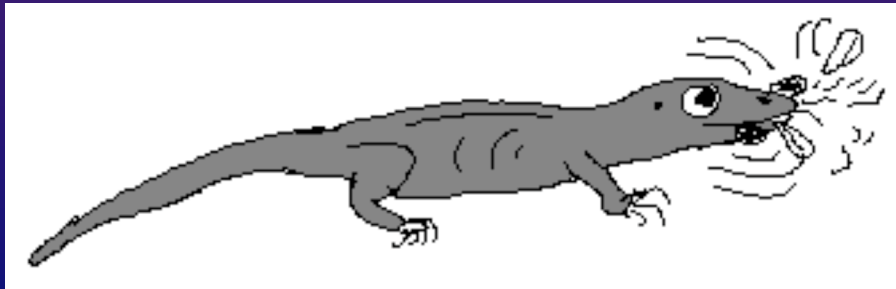3 common orders of <u>S</u>ubject, <u>V</u>erb and <u>O</u>bject
— SVO (English), SOV (Korean), VSO (Tagalog):
E.   the lizard caught the fly
K.   tomapemi pÓalilÈl tSapatta   *"lizard fly caught"*
T.   nakahuli aN butiki N laNaw   *"caught the lizard the fly"*

☀ The other three logical possibilities are either rather unusual – VOS (Malagasy, Fijian); or very rare – OVS (Hixkaryana) and OSV (Nadëb)

26

Syntax: word order patterns

☀ In the great majority of languages <u>the subject precedes the object</u> in basic sentences — SVO (English, Thai, Bole), SOV (Korean, Turkish, Navajo), VSO (Tagalog, Welsh):

Eng. the liz

Kor. tomap   *ard fly caught"*

Tag. nakah   *ught the lizard the fly"*
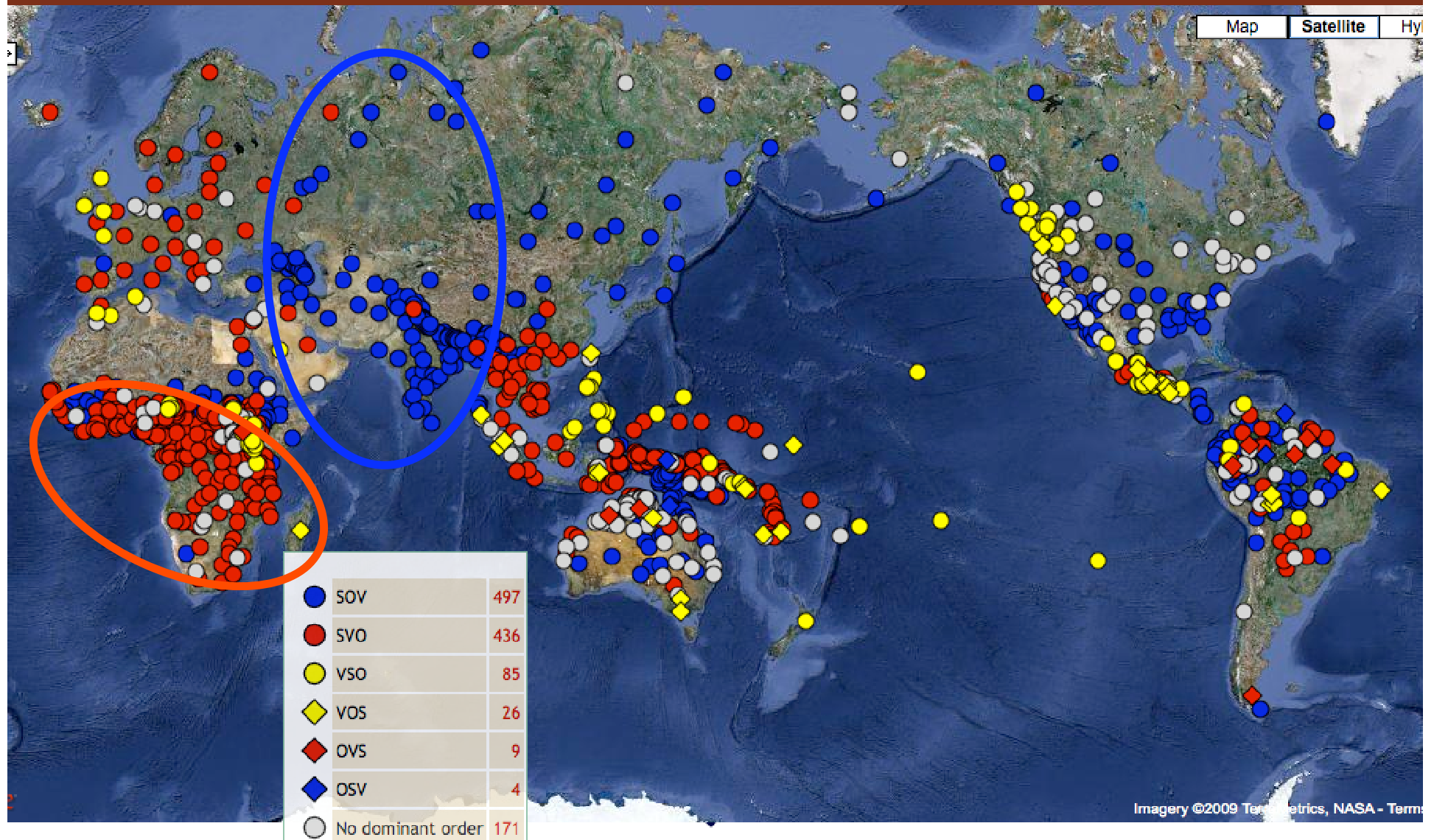
☀ In the three uncommon basic word orders <u>the object precedes the subject</u> – VOS (Malagasy, Fijian); – OVS (Hixkaryana); – OSV (Nadëb)

☀ Each of the common basic word orders is quite widely distributed around the world, but with concentrations in some particular areas:

WALS map for Order of Subject, Verb and Object (1228 languages)

http://wals.info/feature/81

Data and chapter by Matthew Dryer

🔵	SOV	497
🔴	SVO	436
🟡	VSO	85
🔶	VOS	26
🔴	OVS	9
🔷	OSV	4
⚪	No dominant order	171

SVO (red) especially dominant in subSaharan Africa; SOV (blue) in Central/South Asia

Accounting for typological patterns

- Possible paths of development
- Physiological/perceptual constraints
- Cognitive explanations (including acquisition)
- Iconicity
- Social explanations
- Single origin of language

Paths of development

☼ Some universals in language may derive from the fact that there are a particular paths through which the phenomena in question arise in individual languages

☼ previous example: no language has only nasalized vowels

☼ Nasalized vowels from the spread of nasality from consonant to adjoining vowel and loss of the consonant (VN → ṽ); Latin *panem* → *pan* → French *pE)* 'bread')

☼ Since no language has only nasal consonants, no language will have only vowels flanked by nasal consonants, hence none will develop nasalized vowels in every word and eliminate all oral vowels

Physiological/Perceptual constraints

☀ Physiological constraints — in speech production, sounds that are more difficult to produce will be less common that those that are easier to produce (needs careful definition!).

☀ Limiting case — impossible to produce → non-existent. e.g. there are pharyngeal stops, fricatives and approximants, but no pharyngeal nasals, since air cannot reach the nasal cavity if it is blocked from flowing through the pharynx
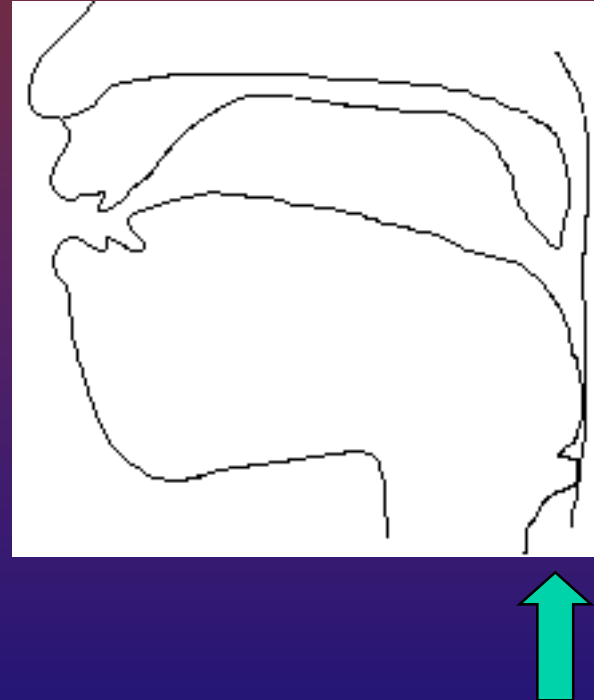
Velar nasal

Air can pass into nasal cavity behind location of closure
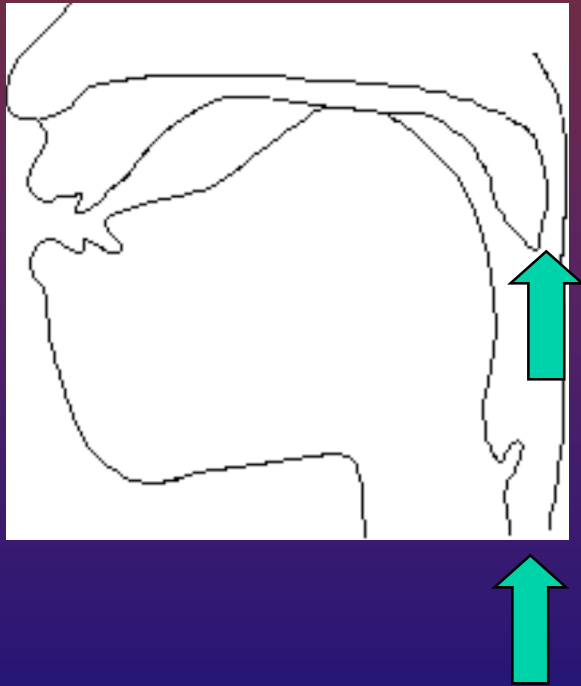

*Pharyngeal nasal

Air cannot reach nasal cavity because flow is blocked by closure in the pharynx

Voiced plosives are less common than voiceless plosives (and only occur if voiceless plosives occur)

To produce vocal fold vibration (voicing) air must flow from the lungs into the oral cavity. If there is (1) a closure in the oral cavity and (2) air is flowing into the cavity, (3) the air pressure will build up behind that closure, increasing the resistance to continued flow. Hence voicing is harder to produce in a plosive than in a vowel or nasal
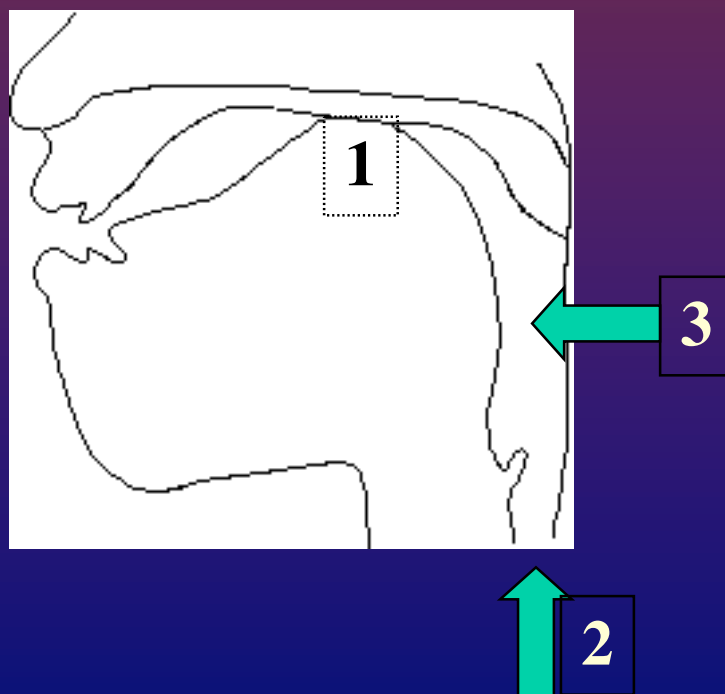
Physiological/Perceptual constraints

☀ Perceptual constraints — in phonology, sounds that are more difficult to recognize will be less commonly used than those that are easier to recognize.

☀ Again, nasalized vowels: no language has more nasalized than oral vowels (most have two less)

☀ Adding nasal resonance reduces the acoustic distinctions between vowels with given differences in their articulatory positions compared to their oral counterparts (e.g. nasalized [a̰), E)] less distinct than oral [a, E])

☀ Increased probability that nasalized vowel distinctions will merge (French /U)/, /{)/ → /{)/)

Cognitive explanations

❀ Cognitive constraints (memory); psychological salience

❀ Subjects (in active not passive sentences) are generally the most salient objects in the discourse situation: this may account for why all languages allow relative clauses to modify subjects, and so on down the 'relativization hierarchy'

Language	Subject	D. O.	I. O.	Oblique	Possessor
Malagasy	+	-	-	-	-
Kinyarwanda	+	+	-	-	-
Basque	+	+	+	-	-
Catalan	+	+	+	+	-
English	+	+	+	+	+

Iconicity

☀ Iconicity — not Platonic ideas of natural associations of form and meaning (except in very small sets of onomatopes, etc, such as 'miaow')

☀ More 'marked' values are generally expressed in phonologically longer forms: e.g. in Comanche pronoun system plural and dual forms are longer (and morphologically more complex) than singulars

Comanche subject pronouns:

	singular	*dual*	*plural*
first incl	nu/, nuu	tahu, takWu	tanu
first excl	— —	nukWu	nunu
second	ᵾnu	muhu, mukWu	munu
third	i/	ituhu, itukWu	ituu

36

Social explanations

☼ Social explanations derive from the fact that languages share common social purposes

☼ E.g. the most basic language use is face-to-face conversation, hence salience of speaker and addressee:

— universality of first and second person pronouns as shorthand for encoding these roles (rather than using full names)

— inclusive/exclusive first person to denote if the addressee is included or excluded

Single origin

☀ If there was a single origin of language some 'accidental' properties of that first language would have been perpetuated in all descendants (cf. 'founder effects' in populations)

☀ *Traits of this kind would not be informative about the phylogeny of modern languages*

Typology in diachronic linguistics

☀ Most knowledge of language history is indirect — inferred from similarities/differences between present-day languages

☀ Hypotheses of common origin of traits

☀ Reconstruction of prior state
  — one of present states is original, others derived
  — all present states are derived

☀ Relative probability of the states can guide choice

Typology in diachronic linguistics

☀ Classic example: Indo-European stops

☀ Both the structure of segment inventories and the relative frequency of particular segment types invoked in discussion of competing ideas of the nature of Proto-Indo-European (PIE) phonology

☀ Original idea: Indo-European parent language had four sets of stops, like ancient Sanskrit and modern Hindi

Typology in diachronic linguistics

Hindi stops

Examples at post-alveolar place of articulation

voiceless unaspirated	Êal	'postpone'	🔊
voiceless aspirated	ÊÓal	'lumber shop'	🔊
plain voiced	∂al	'branch'	🔊
breathy voiced	∂ᴴal	'shield'	🔊

Typology in diachronic linguistics

☀ Later evidence suggested voiceless aspirated series are not single segments, but C + /h/

☀ Standard reconstruction thus became 3 series (e.g. Meier-Brügger 2003), at five places of articulation

voiceless unaspirated p   t1   c   k   kW
plain voiced   b   d1   Ô   g
      gW
breathy voiced   b^H d1^H Ô^H g^H
gW^H

Typology in diachronic linguistics

◆ Roman Jakobson (1957) among first to question this reconstruction on typological grounds

◆ He pointed to the rarity of languages having breathy voiced stops without <u>also</u> having a voiceless aspirated series

"To my knowledge, no language adds to the pair /t/ – /d/ a voiced aspirate /d^h/ without having its voiceless counterpart /t^h/, while /t/, /d/, /t^h/ frequently occur without the comparatively rare /d^h/ ....; therefore theories operating with the three phonemes /t/ – /d/ – /d^h/ in Proto-Indo-European must reconsider the question of their phonemic essence."

43

Typology in diachronic linguistics

◆ No modern languages known with Modern Hindi-like /b^H, d1^H, g^H/, without also voiceless aspirates

◆ Further oddity in 'traditional' PIE system —words with reconstructed */b/ extremely rare. In languages with a /b/ phoneme it is typically quite common

In Rix et al (1998) *Lexicon der Indogermanischen Verben*

      139 roots with */p/

       14 roots with */b/

      119 roots with */b^H/

cited by James Clackson (2007). *Indo-European linguistics: An introduction*. Cambridge University Press.

Typology in diachronic linguistics

◆ Suggested solution: use typologicial data to reinterpret the phonetic nature of the three series.

◆ In languages with ejective stops, bilabials are the least common. Hence, reinterpret traditional 'voiced' series as ejective. Traditional breathy voiced series can be interpreted as plain voiced. (Gamkrelidze & Ivanov 1973, Hopper 1973)

◆ 'Ejective' series become deglottalized in Germanic, but are voiced stops in Slavic, Baltic, Iranian and Celtic branches

Trad.	Glottalic (1)	Latin	English	Russian
*/p/	*/p/	pes, pedem	foot	peSkom (on foot)
*/b/	*/p'/	—	apple	jabloko
*/bH/	*/b/	fero	bear (verb)	bratΔ

Typology in diachronic linguistics

◆ The change /p', t', k'/ > /b, d, g/ seems a little strange

◆ Hence variant of glottalic theory with implosives /ʃ, Î, ƒ/ instead of ejectives (e.g. Beekes 1995)

◆ 'Implosive' series become devoiced and deglottalized in Germanic, but simple voiced stops in Slavic, Baltic, Iranian and Celtic branches

◆ Problem with this variant — bilabial implosive is most frequent cross-linguistically, but the reconstructed */b/ or */ʃ/ element is very rare (slide 44)

Trad.	Glottalic (2)	Latin	English	Russian
*/p/	*/p/	pes, pedem	foot	peSkom (on foot)
*/b/	*/ʃ/	—	apple	jabloko
*/bᴴ/	*/b/	fero	bear (verb)	bratΔ

Typology in diachronic linguistics

◎ Typological factors involved in this debate
— a breathy voiced series of stops does not occur unless a voiceless aspirated one is also present
— in a voiced stop series /b/ is typically frequent in a language
— in an ejective stop series /p'/ is typically rare or absent
— in an implosive stop series /ʃ/ is typically the most frequent

◎ Reconsider these questions in terms of processes and associated rates of change

Linguistic typology as process

◎ Why would there be a relationship between the breathy voiced and voiceless aspirated stop series?
— breathy series occurs with voiceless aspirates either
(a) because a sound change producing breathy stops must also produce aspirated stops, *whereas the reverse is not true.*
(b) because there are more diachronic pathways to aspiration than to breathiness and therefore it is likely that a language which has breathy stops will also have aspirated ones.
In the first case the two are causally related; in the second they are not.
_Note also: if traditional PIE reconstruction is correct there must be many pathways leading away from the PIE system since so few IE daughter languages have breathy voiced series

Linguistic typology as process

- in a voiced stop series /b/ is typically frequent in a language
    Process implications: /b/ is unaffected by change, or many/rapid processes introduce /b/ to a language, or few/slow processes remove /b/ from a language

- in an ejective stop series /p'/ is typically rare or absent
    few/slow processes introduce /p'/ to a language, or many/rapid processes remove /p'/ from a language

- in an implosive stop series /ʃ/ is typically the most frequent
    many/rapid processes introduce /ʃ/ to a language, or few/slow processes remove /ʃ/ from a language

49

Linguistic typology as process

- This perspective puts the focus on understanding the processes of change

- Our problem: in traditional linguistic scholarship, <u>process</u> has been inferred after the historical reconstruction has been done, rather than guiding the reconstruction in the first place

- The next two talks will suggest how we might advance beyond this roadblock in understanding the very complex question of language history