Language Sampling and Language Universals

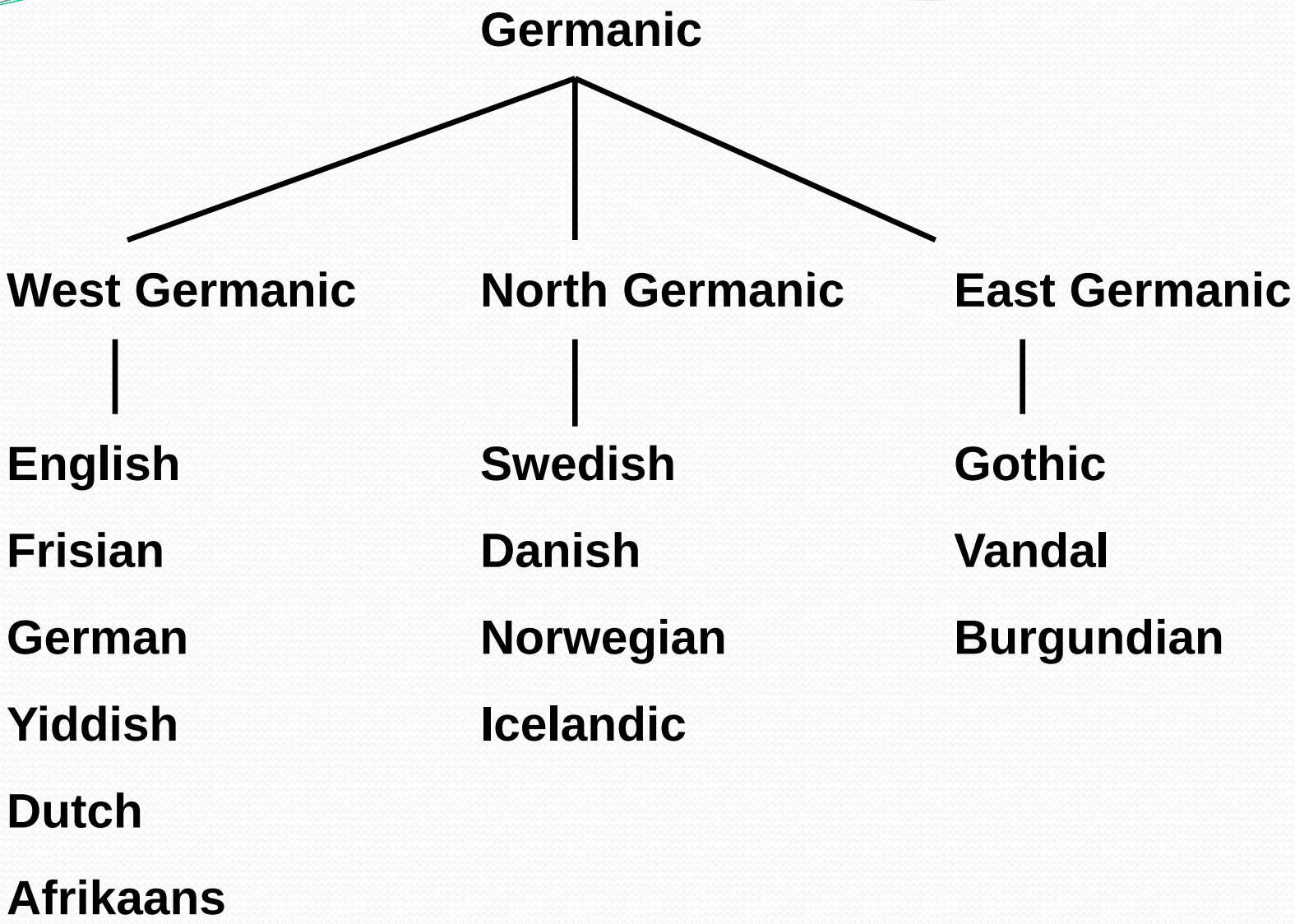
Holger Diessel
University of Jena

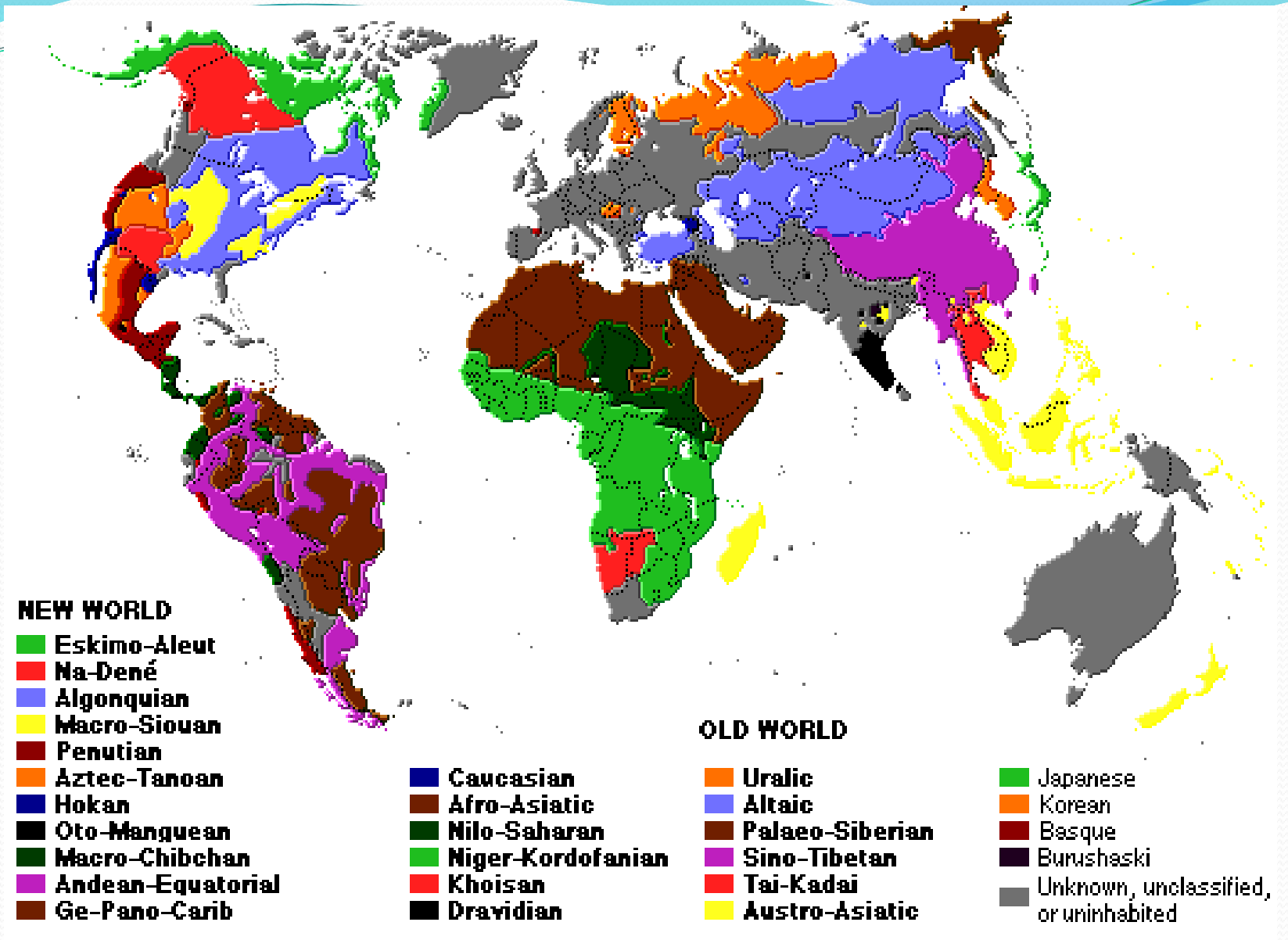
holger.diessel@uni-jena.de
http://www.holger-diessel.de/

Number of speakers (2007)

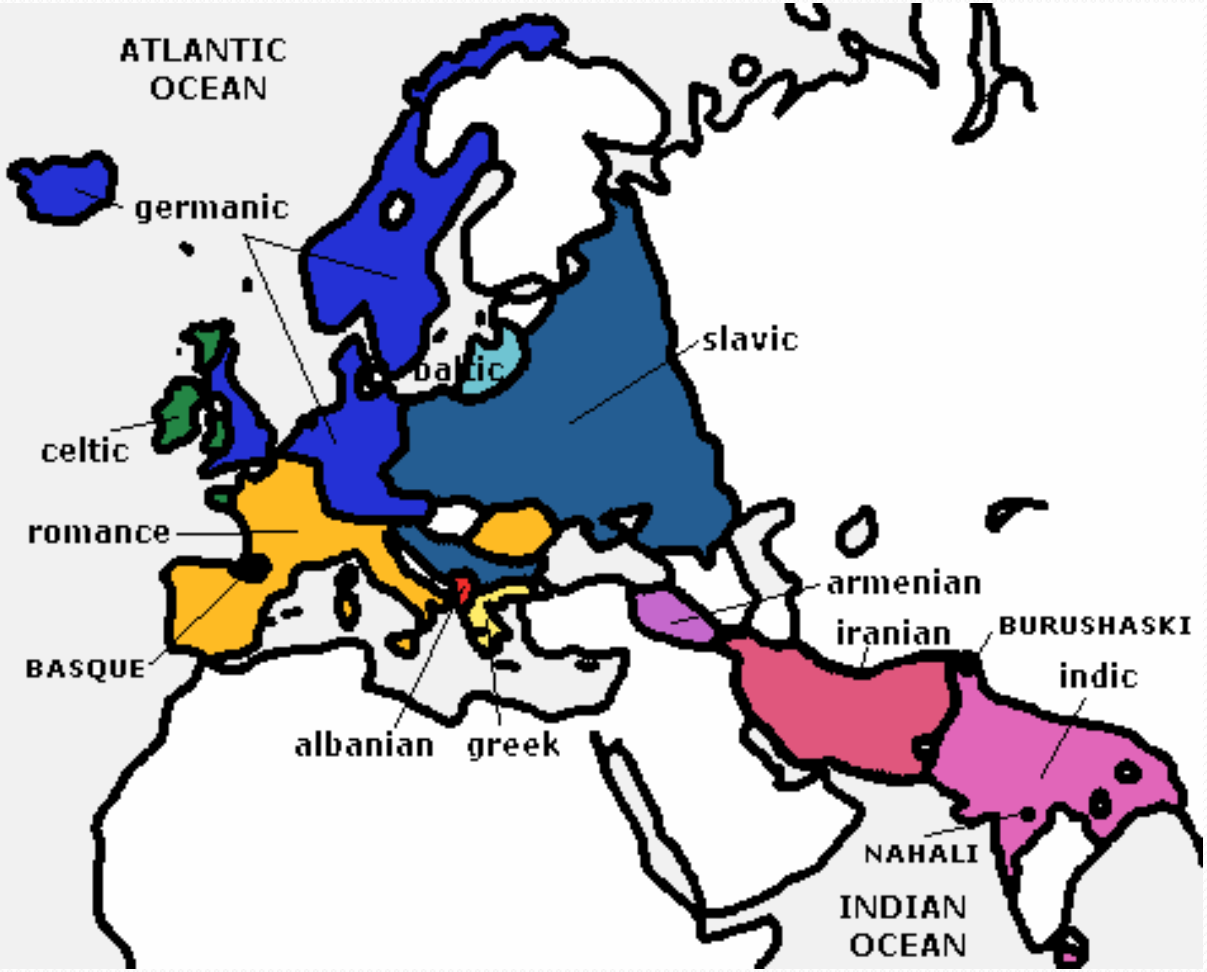
How many languages are there?

- 6000-7000 languages
- Many languages have few speakers
- Languages spoken by a few hundred or a few thousand speakers are not necessarily threatened by extinction.
- Still, if nothings happens, 90 percent of all languages will disappear within the next 100 years.
- Languages are diachronically related. How many language families are there?

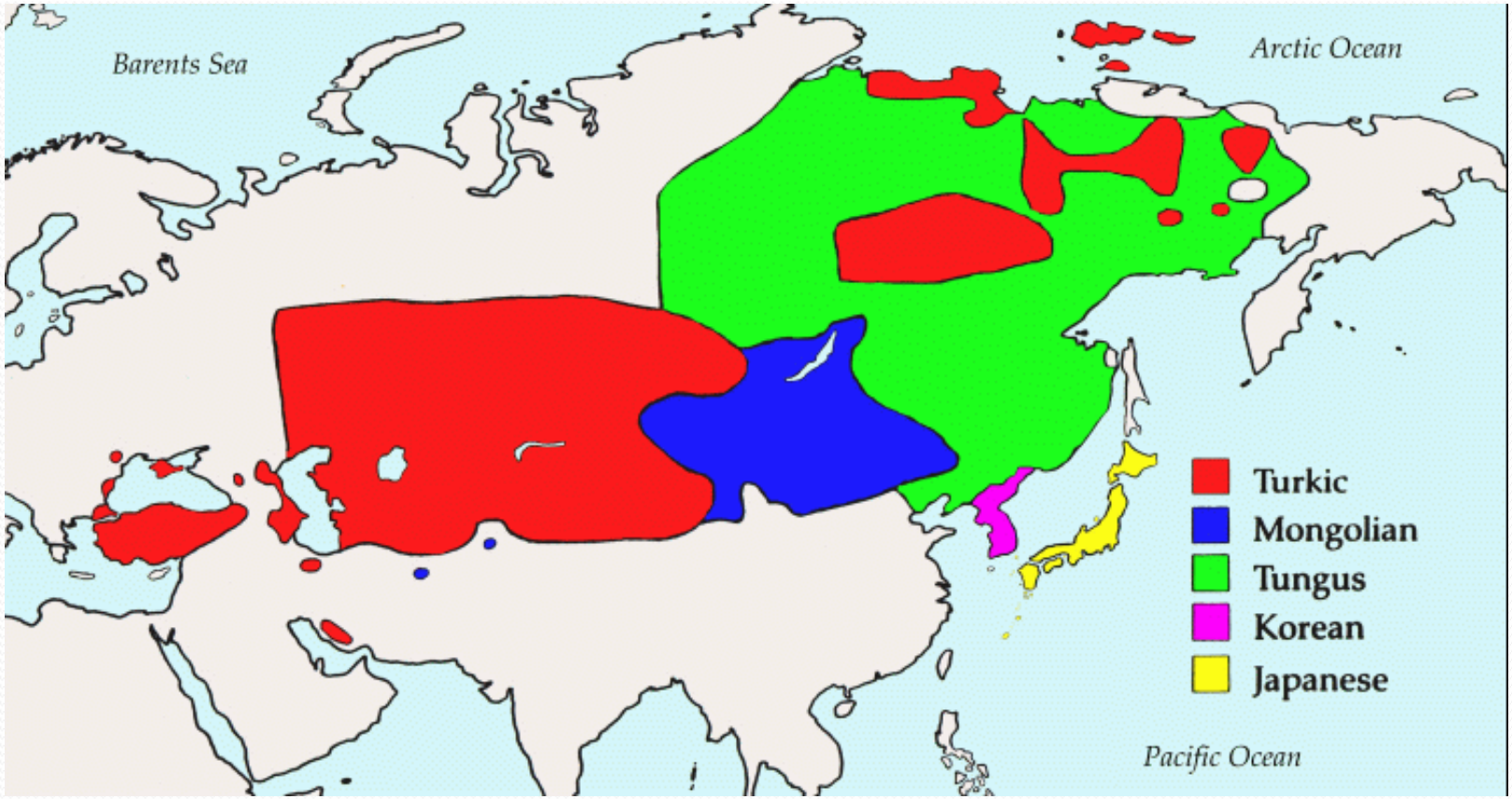

Germanic
West Germanic
North Germanic
East Germanic
English
Frisian
German
Yiddish
Dutch
Afrikaans
Swedish
Danish
Norwegian
Icelandic
Gothic
Vandal
Burgundian


NEW WORLD
Eskimo-Aleut
Na-Dené
Algonquian
Macro-Siouan
Penutian
Aztec-Tanoan
Hokan
Oto-Manguean
Macro-Chibchan
Andean-Equatorial
Ge-Pano-Carib
Caucasian
Afro-Asiatic
Nilo-Saharan
Niger-Kordofanian
Khoisan
Dravidian
OLD WORLD
Uralic
Altaic
Palaeo-Siberian
Sino-Tibetan
Tai-Kadai
Austro-Asiatic
Japanese
Korean
Basque
Burushaski
Unknown, unclassified, or uninhabited

Indo-European



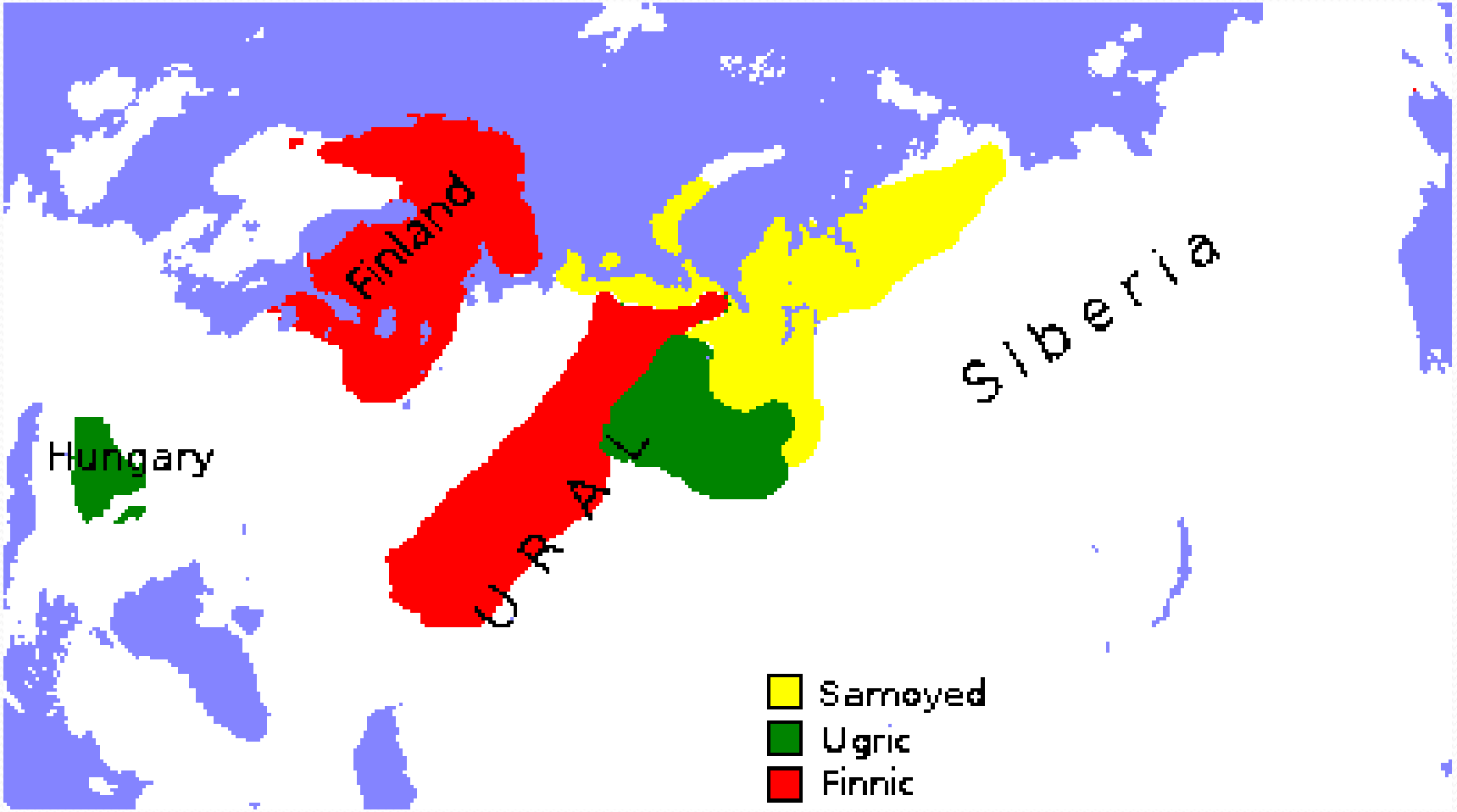
Altaic



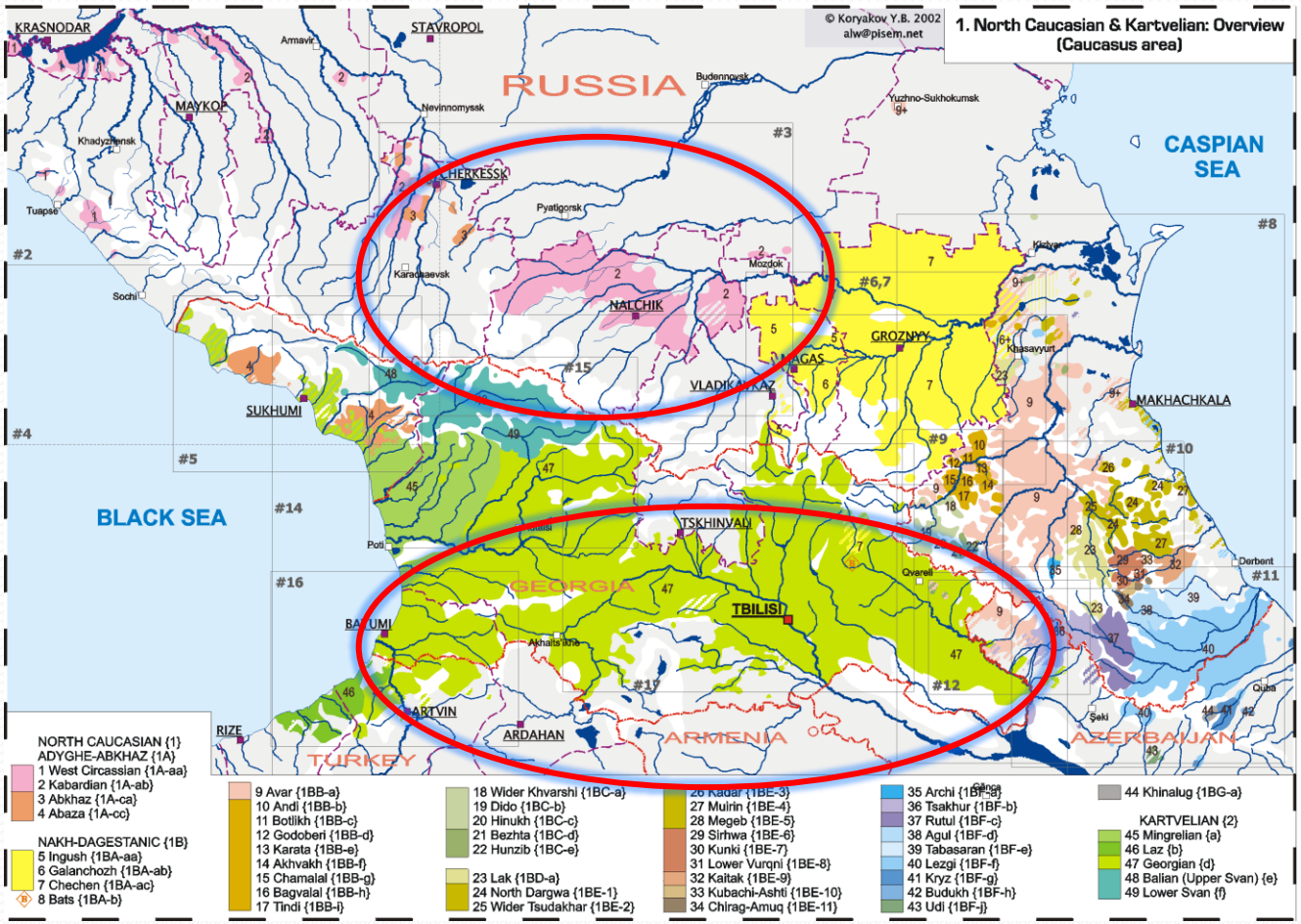
Turkish

Japanese

Uralic

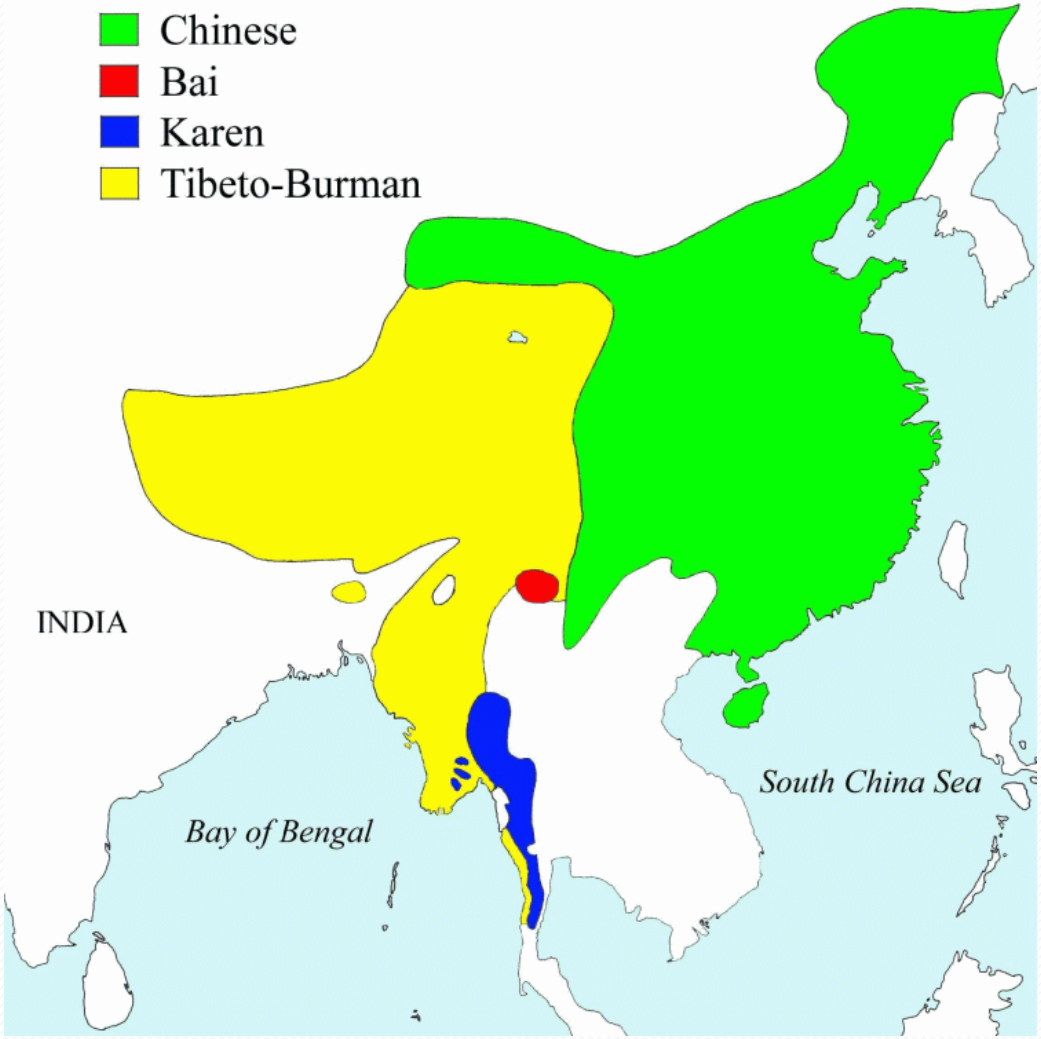


Caucasian languages



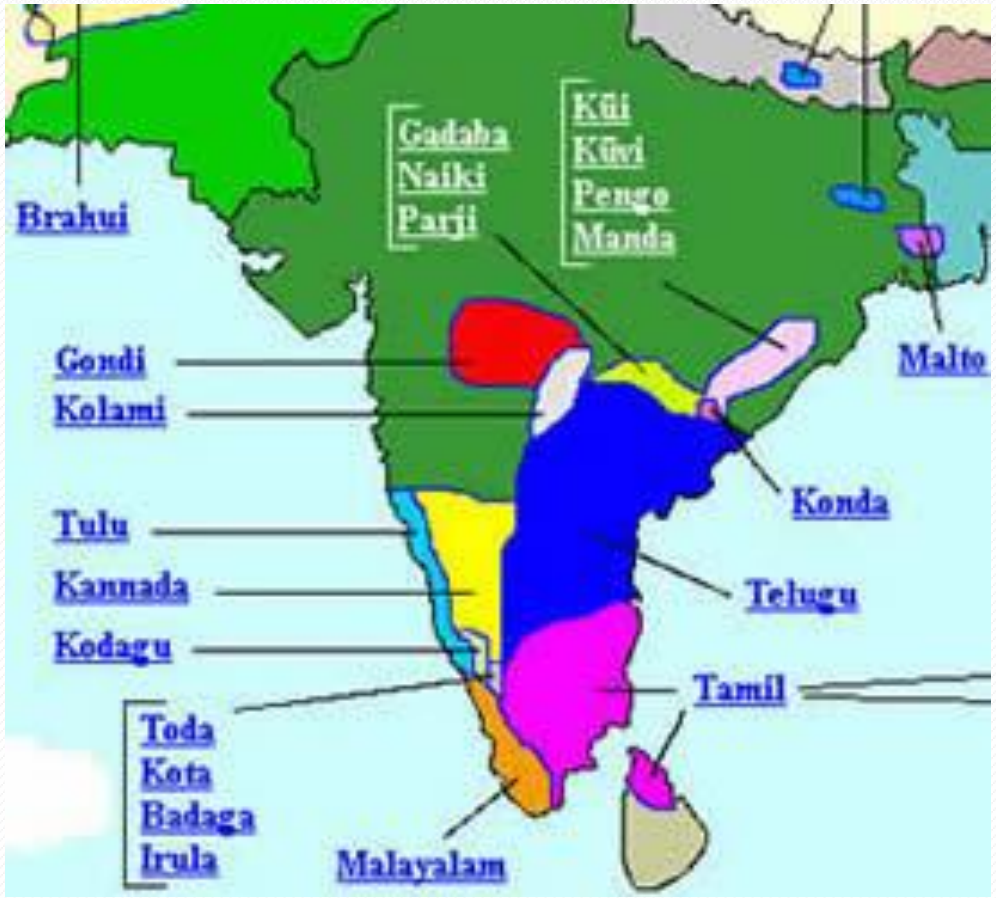
Georgian

Sino-Tibetan

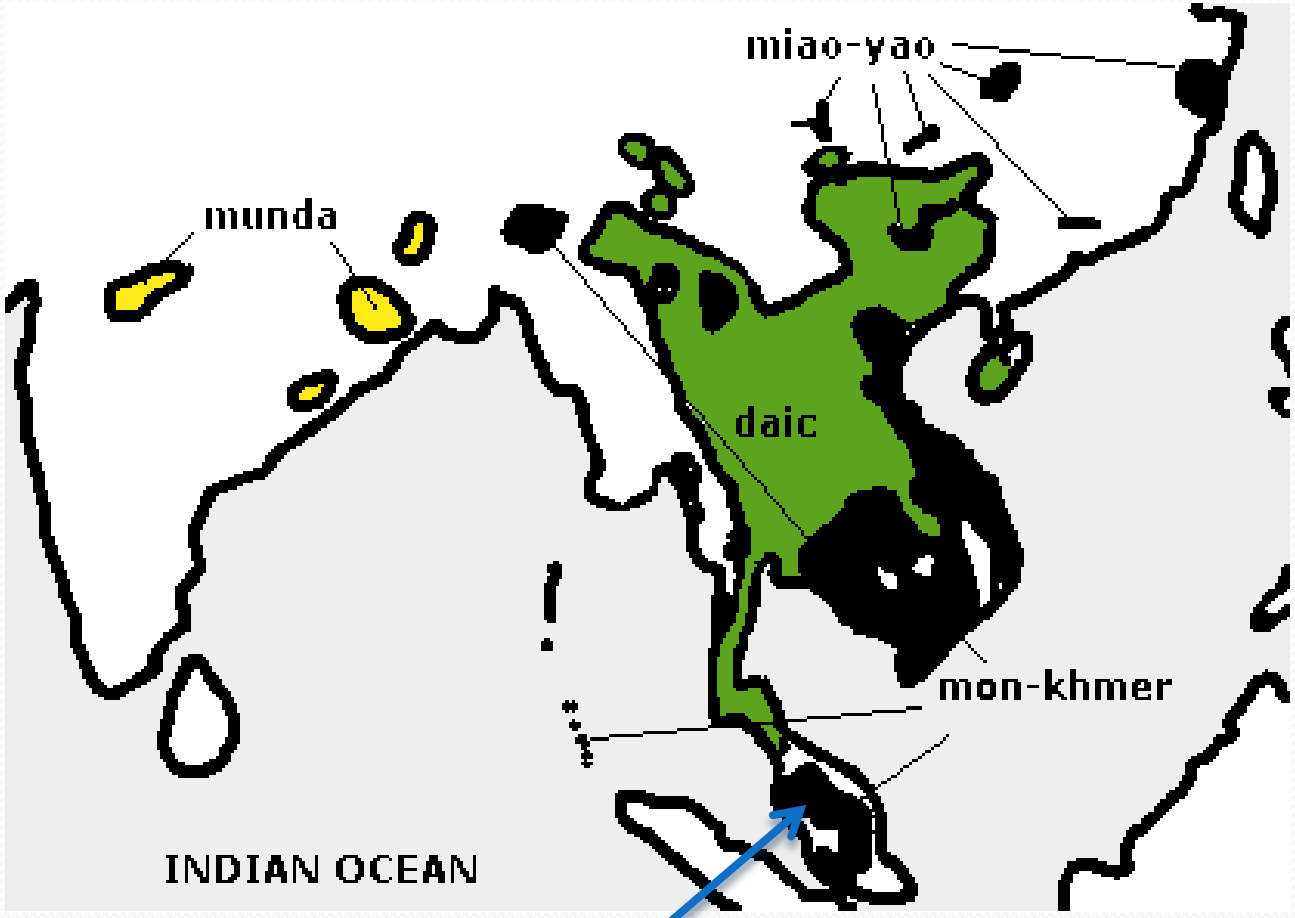


Mandarin Chinese

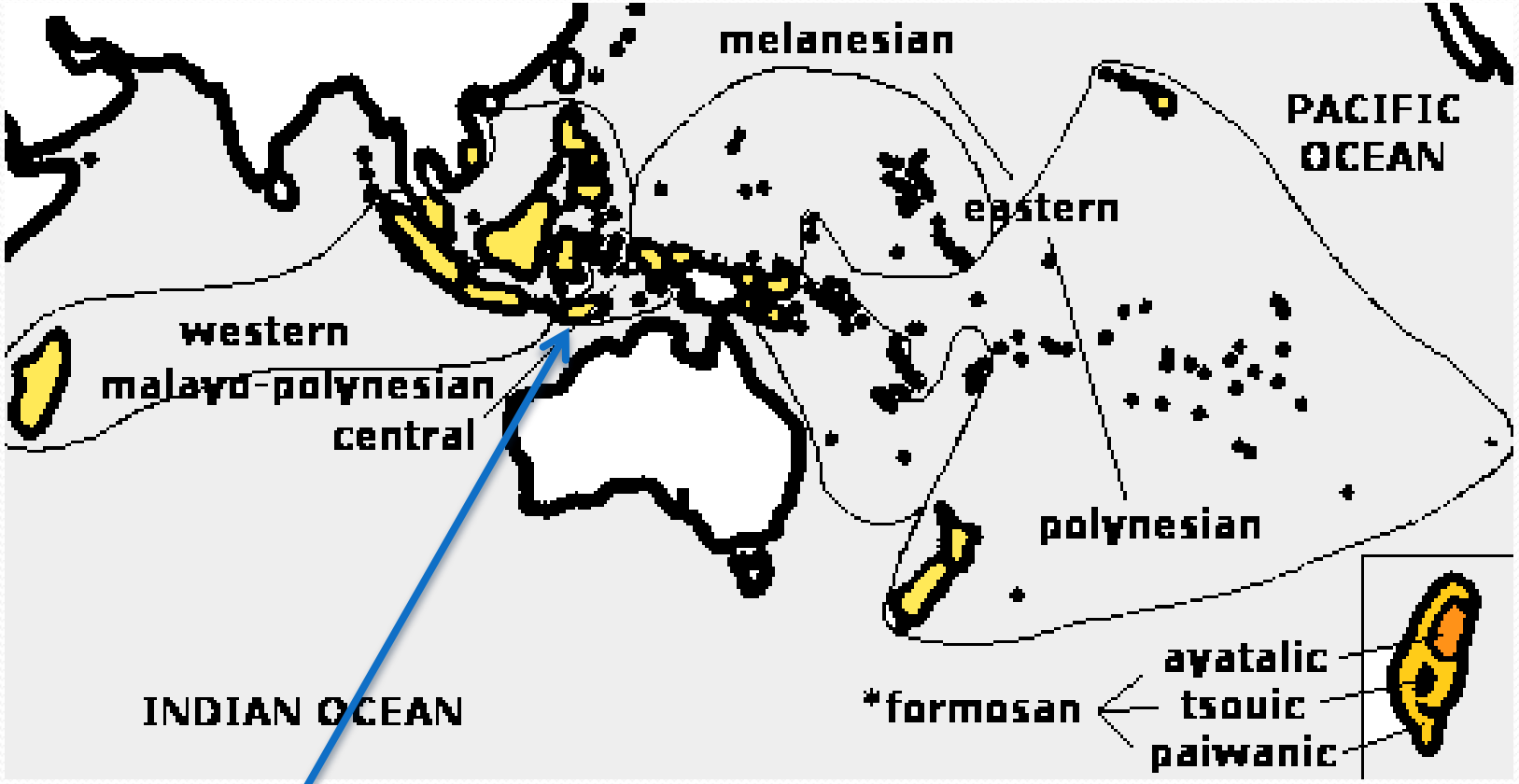
Dravidian


Brahui
Gadaba
Naiki
Parji
Kūi
Kūvi
Pengo
Manda
Malto
Gondi
Kolami
Konda
Tulu
Kannada
Telugu
Kodagu
Tamil
Toda
Kota
Badaga
Irula
Malayalam

Daic, Mon-Khmer, Mio-Yao, Munda

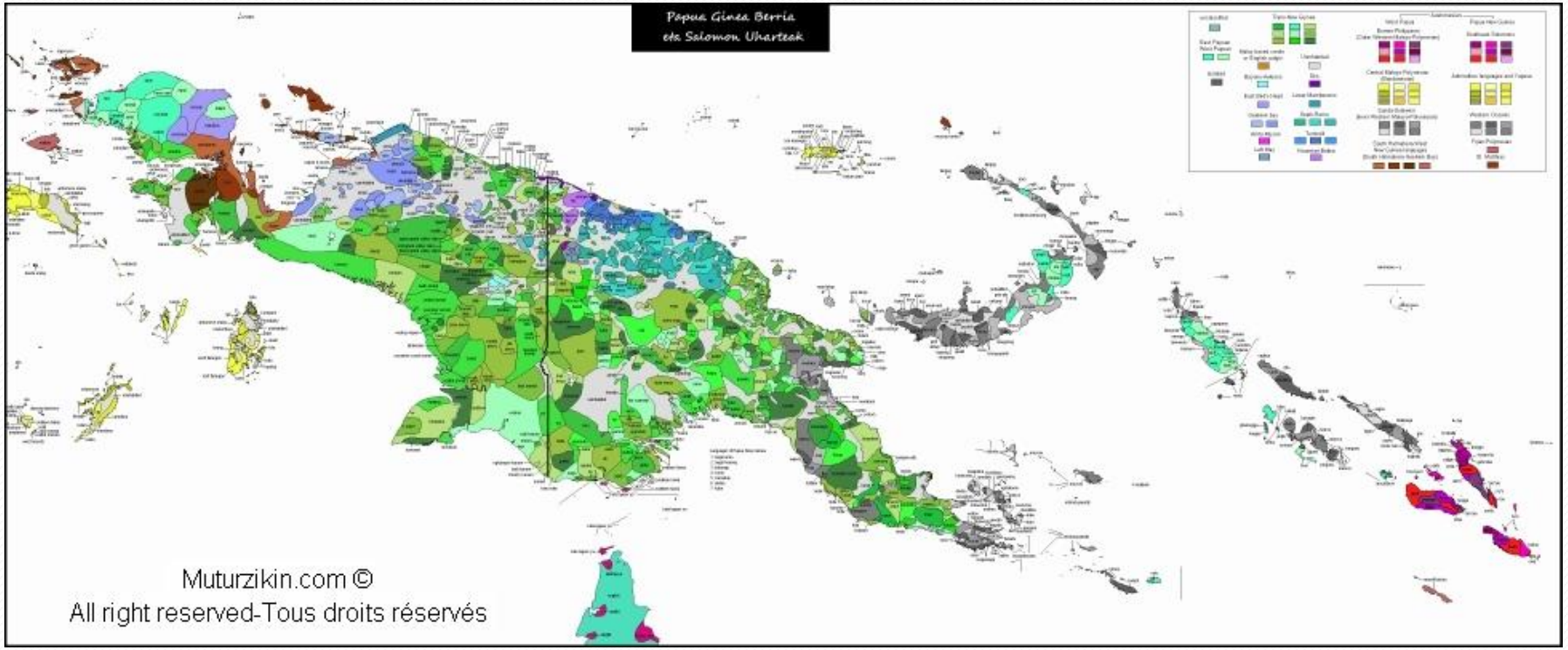


Austronesian



Tetun

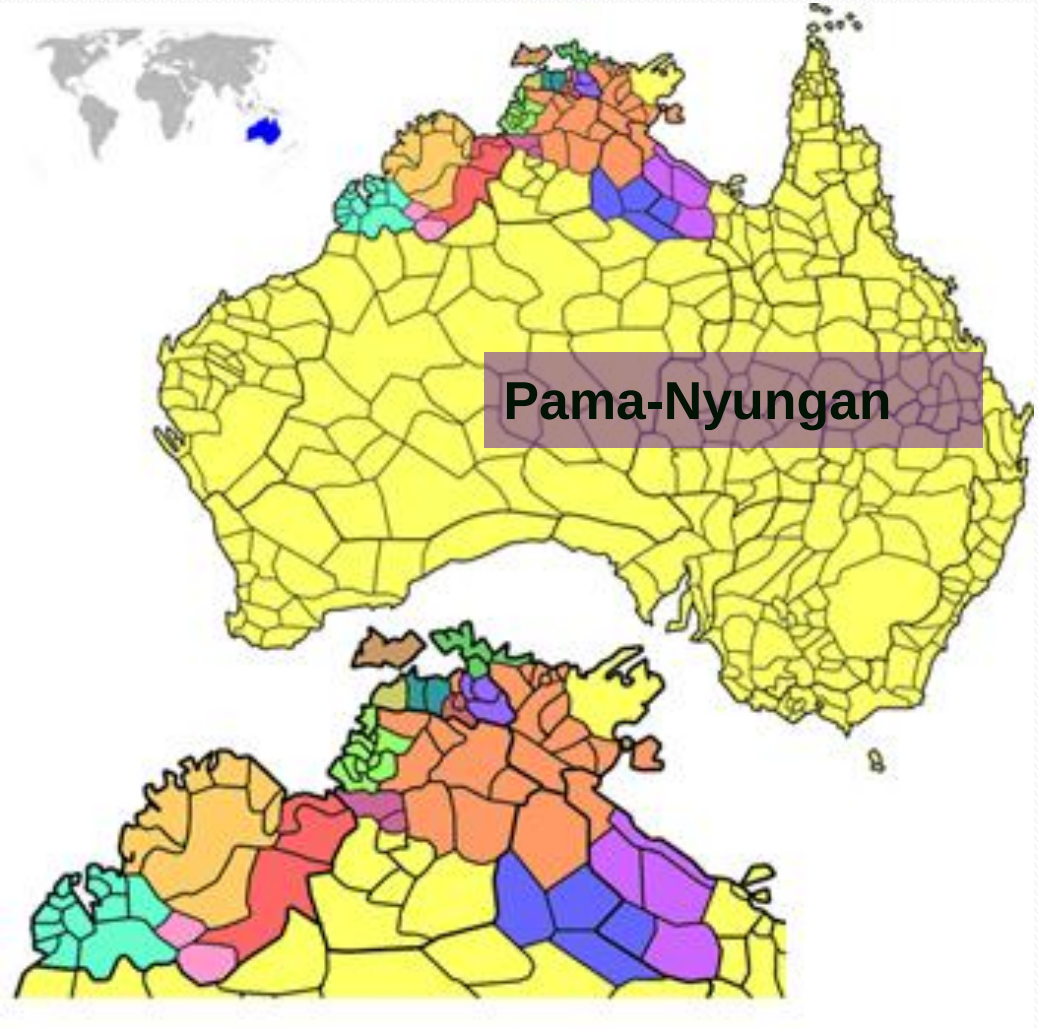
New Guinea



Abun

Lavukaleve

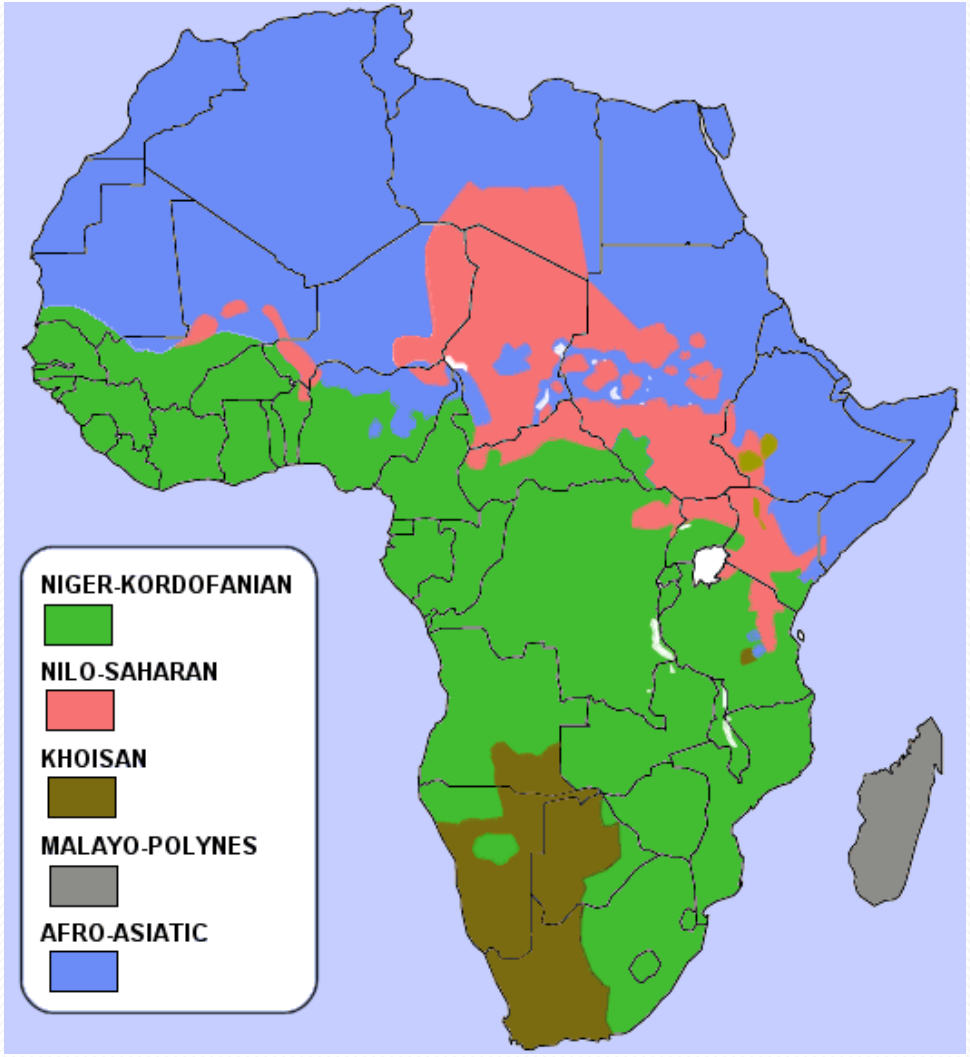
Australia



Wambaya

Kayardild

Africa



Arabic

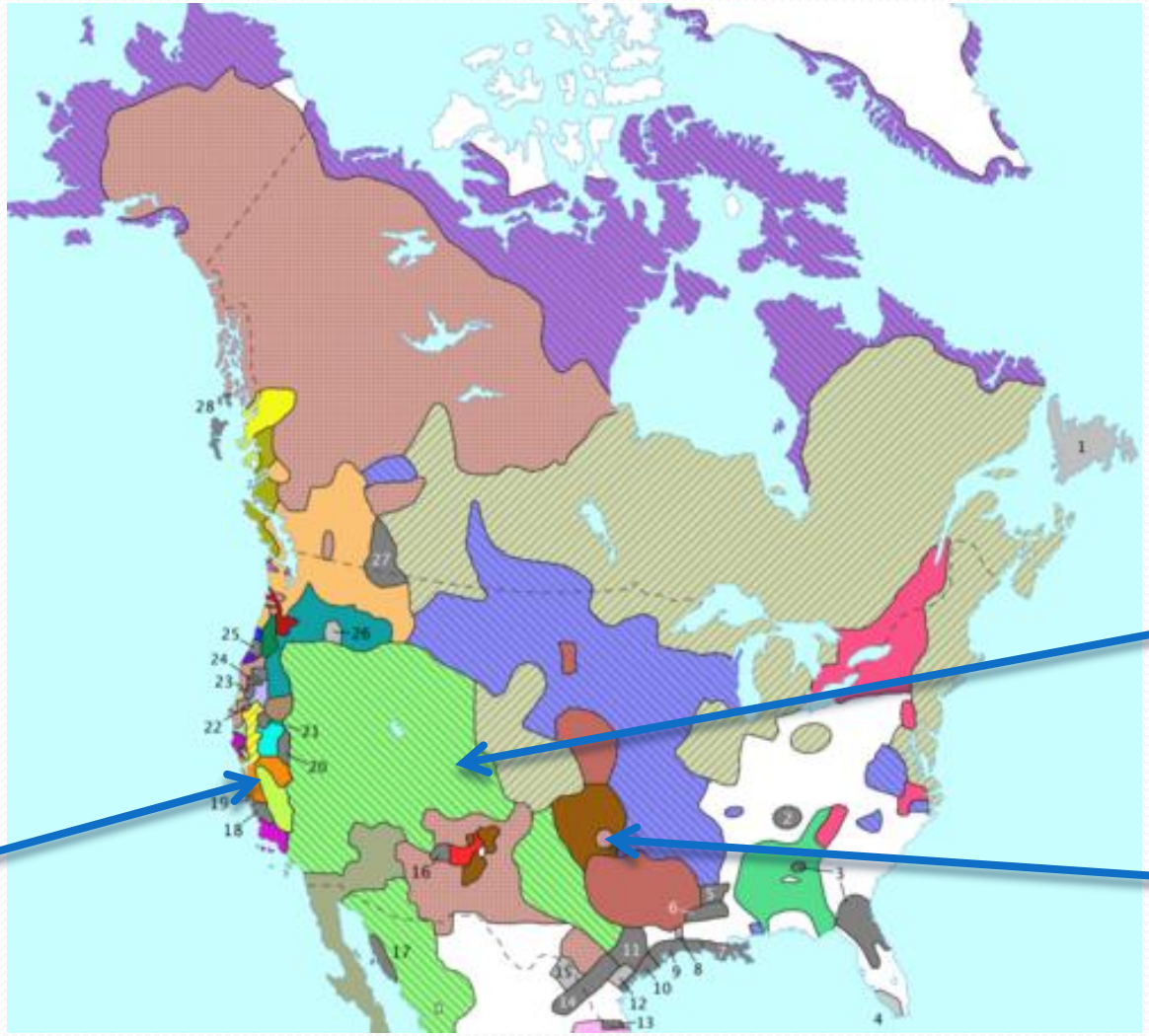
Koyra Chinii

Babungo

Khwe

Bantu

North America

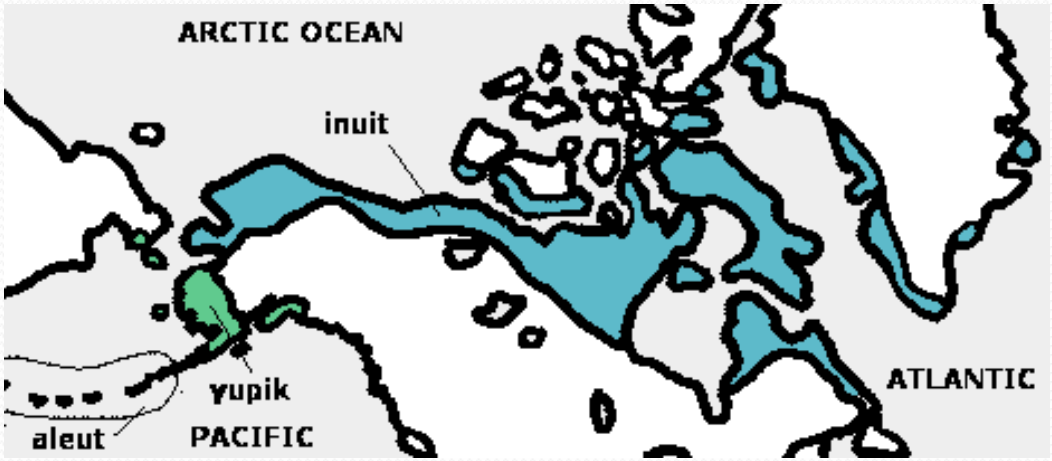


Tümpisa Shoshone

Wappo

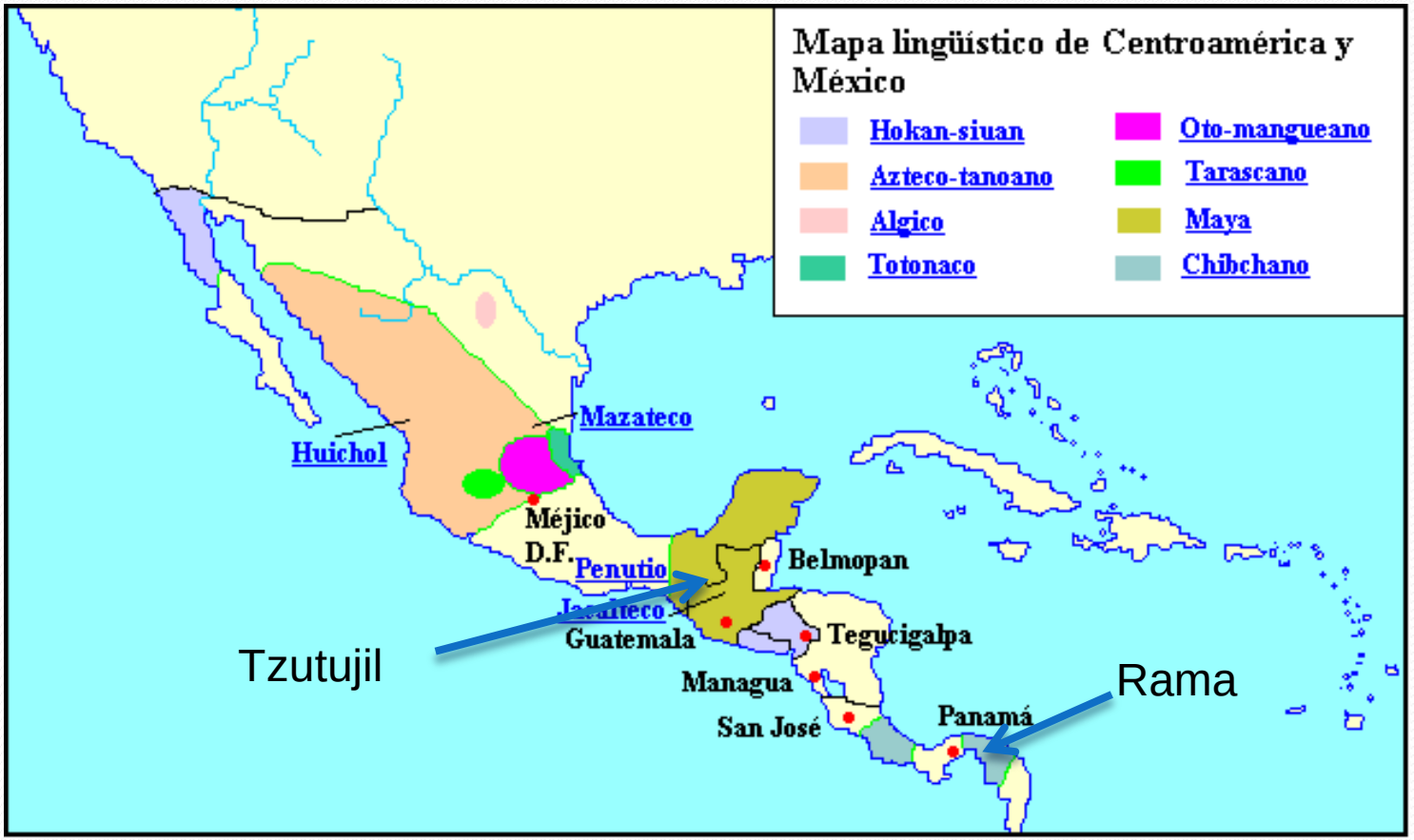
Choctaw

Eskimo-Aleut



Na-Dene

Central America



South America

Yuracaré

Hup

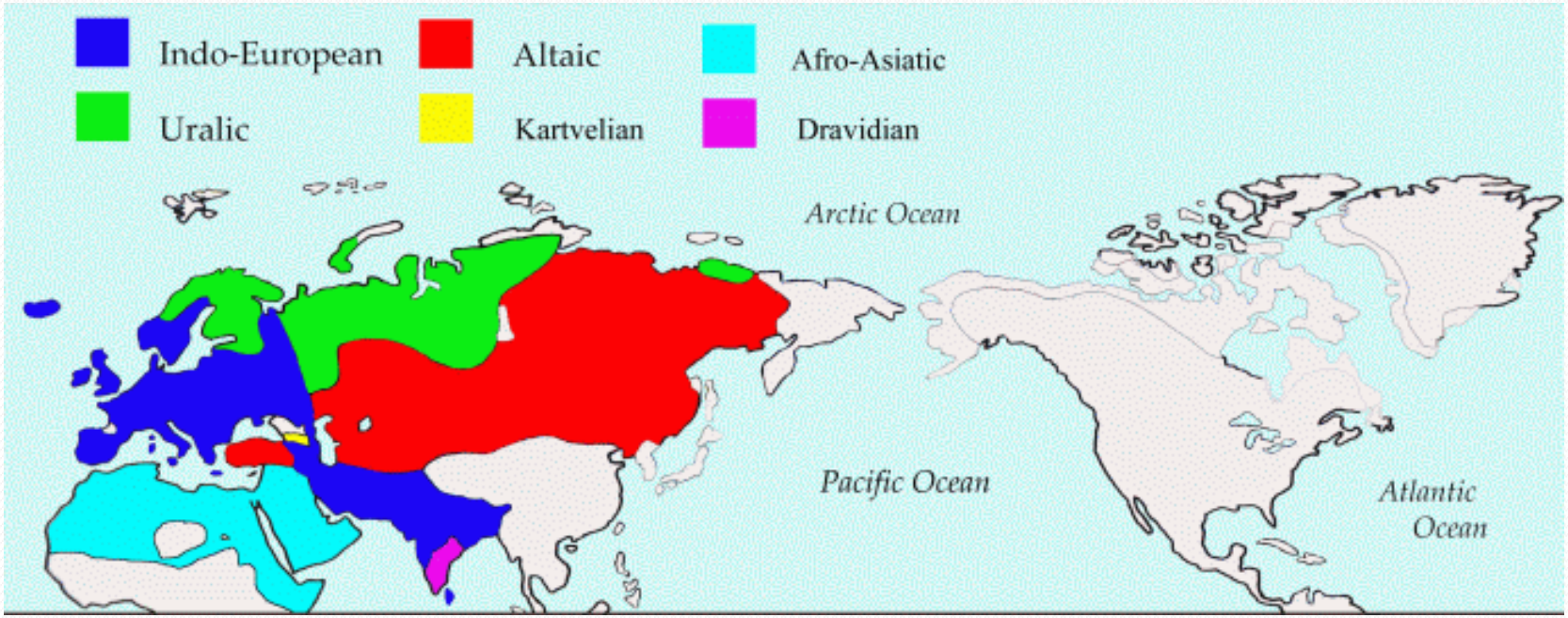
Origin of language families

Are the various language families related?

Amerind hypothesis

Greenberg: The native American languages fall into three major language families: (1) Eskimo, (2) Na-Dene, (3) Amerind.

Macro language families



Nostratic (hypothesis): Indoeuropean, Altaic, Uralic, Afro-Asiatic, Kartvelian.

Number of speakers (2007)

	Native	Percentage
Mandarin	955	14.4%
Spanish	407	6.2%
English	359	5.4%
Hindi	311	4.7%
Arabic	293	4.4%
Portuguese	216	3.3%
Bengali	216	3.1%
Russian	154	2.3%
Japanese	126	1.9%
Punjabi	102	1.5%
German	89	1.4%
Javanese	82	1.3%
Wu (Shanghainese)	80	1.2%
Malay	77	1.2%

Language Sampling

Language sampling

SVO	VOS
SOV	OVS
VSO	OSV

Language sampling

Order	Greenberg (1966)
SVO	43%
SOV	37%
VSO	20%
VOS	0%
OVS	0%
OSV	0%

Language sampling

Order	Greenberg (1966)	Tomlin (1986)
SVO	43%	42%
SOV	37%	45%
VSO	20%	9%
VOS	0%	3%
OVS	0%	1%
OSV	0%	0%

Language sampling

- Convenient language sample
- Balanced language sample

WALS

World Atlas of Language Structures

- Front rounded vowels (feature 11a)
- Tone (feature 13a)
- Distance contrasts in demonstratives (feature 41a)

Language Universals

Types of universals

- All languages have vowels and consonants.
- All languages have nouns and verbs.
- All languages have demonstratives.

Absolute universals vs. statistical universals

Absolute universal:

All languages have vowels and consonants.

Statistical universal:

Most languages place the subject before the object.

Implicational universals

(1) Peter saw **himself** (in the mirror).

(2) Peter saw **him** (in the mirror).

If a language has reflexive pronouns for first and second person, it also has reflexive pronouns for third person.

Implicational universals

	1.+2.+3. person	3. only person
Reflexive Pronouns	x	x
No reflexive pronouns	x	

Implicational universals

English		German	
me	**myself**	mich	mich
you	**yourself**	dich	dich
him/her/it	**him/her/itself**	ihm/ihr/es	**sich**

Old English		Non existent	
mē	mē	1	**x**
Þē	þē	2	**x**
hine/hiē/hit	hine/hiē/hit	3	x

Universal hierarchies

Noun phrase accessibility hierarchy (Keenan and Comrie 1977)

(1) The man who likes me.
(2) The man who I like.
(3) The man who I gave the book to.
(4) The man who I went to.
(5) The man whose book I read.

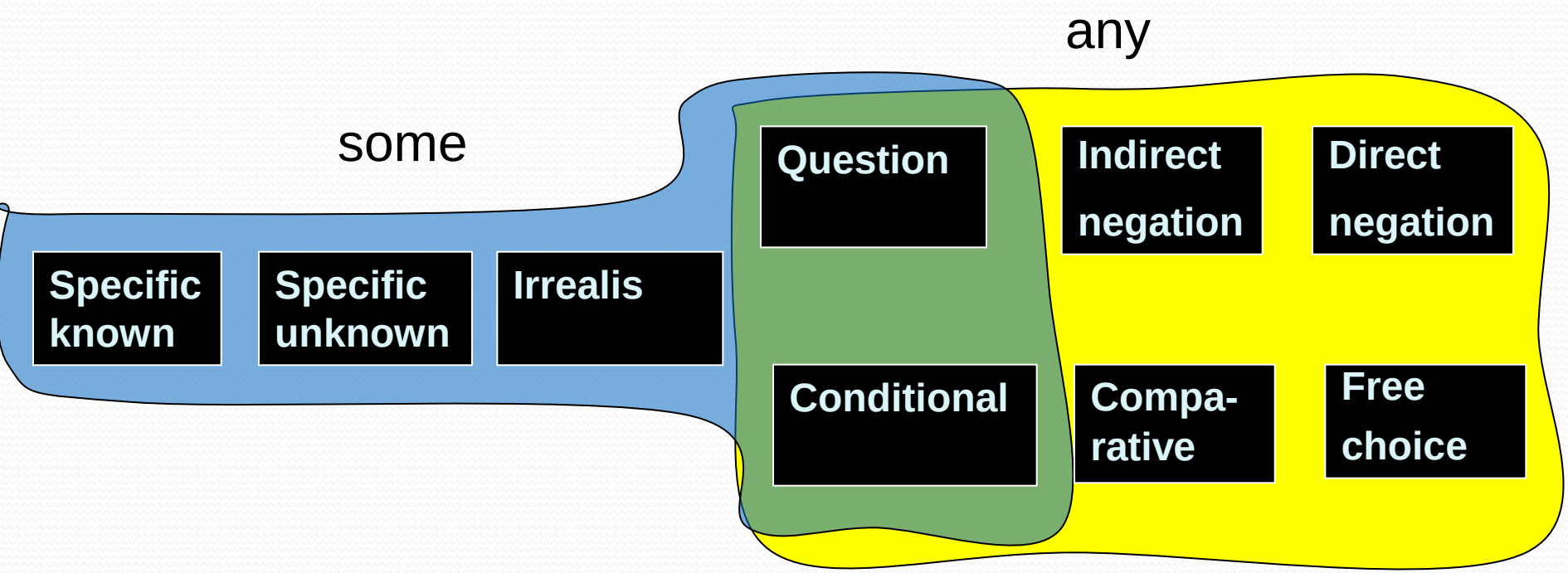
Noun phrase accessibility hierarchy

SUBJ > OBJ > OBL > GEN

If a language has object RCs it also has subject RCs.

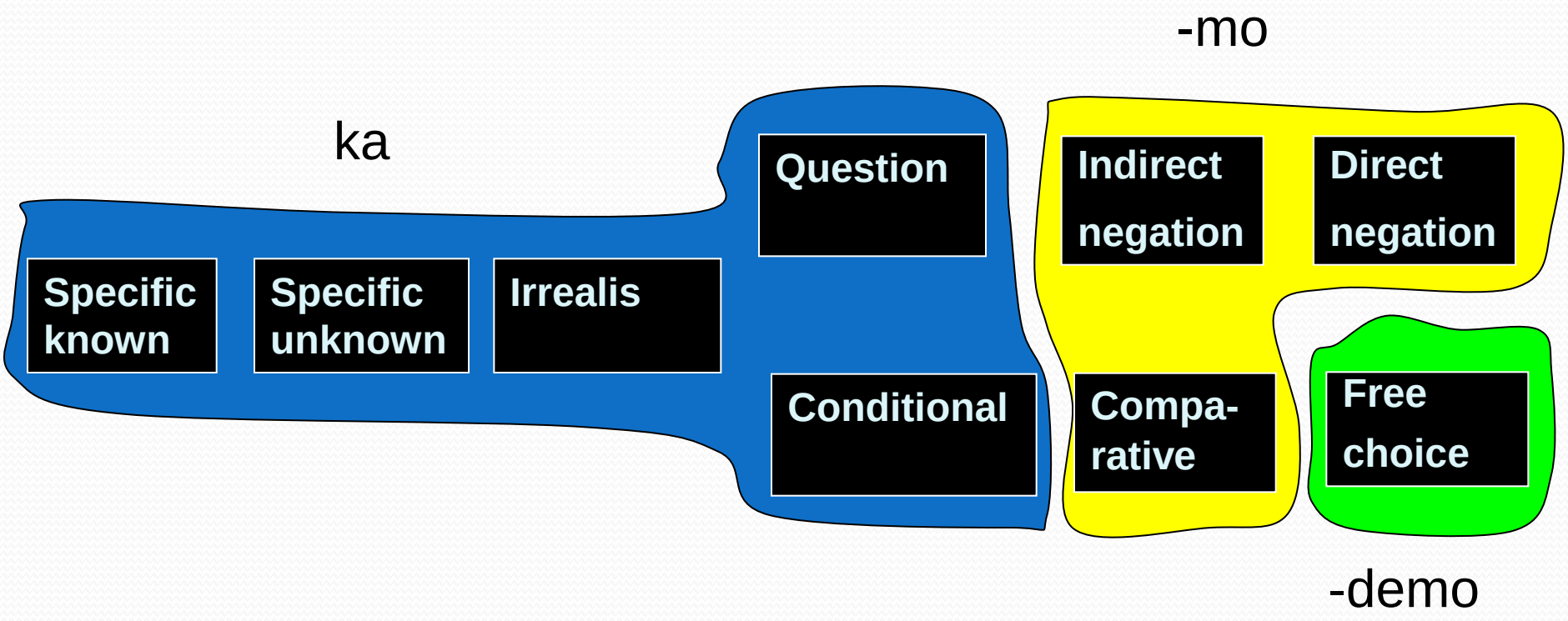
If a language has oblique RCs it also has subject + object RCs.

Semantic maps



(1) I saw somebody/*anybody. specific unknown
(2) Did you see somebody/anybody. question
(3) I didn't see *somebody/anybody. indirect negation
(4) *Somebody/anybody can win. free choice

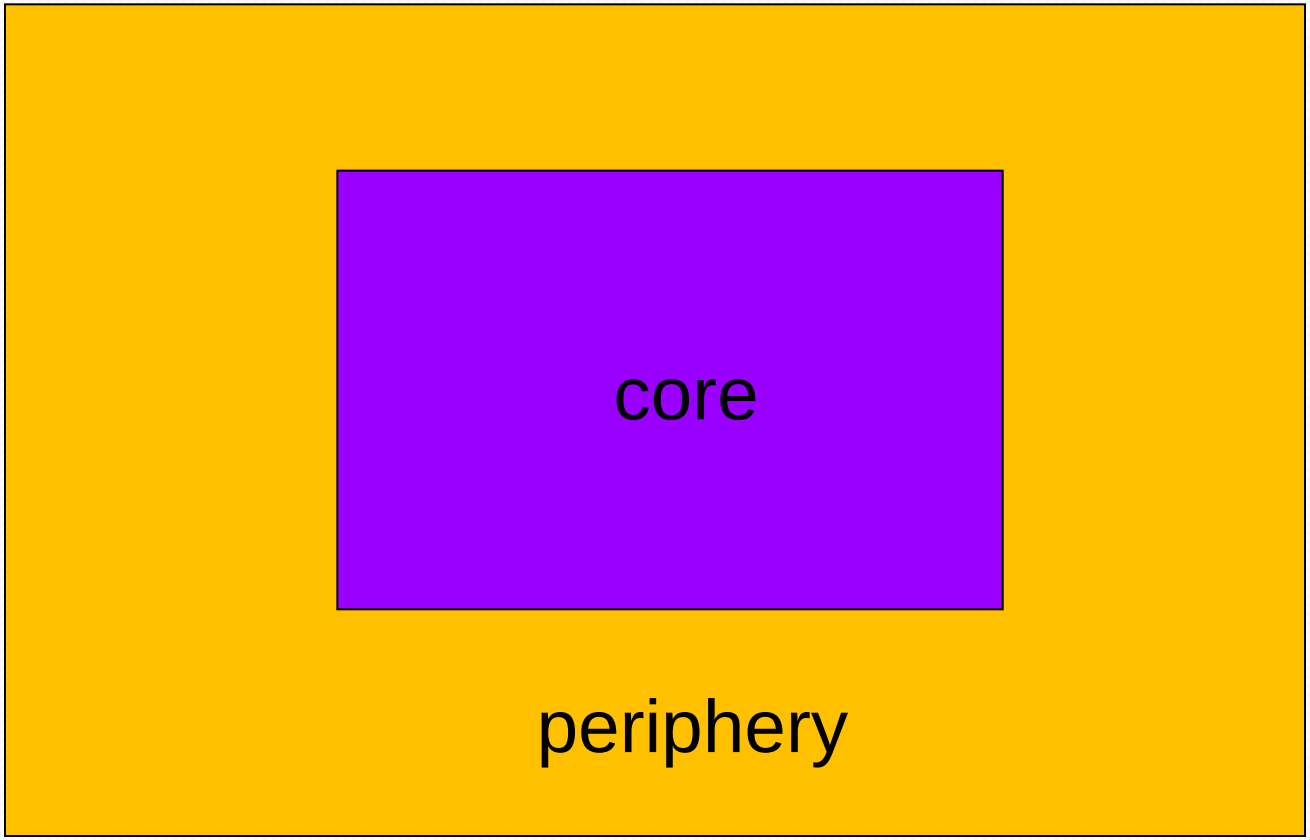
Semantic maps



Japansese

How do we account for the existence of language universals?

Nativist theory



The core principles of human grammar are innate.

Functional/cognitive explanations

- Discourse
- Processing
- Economy
- Iconicity

Discourse pressure

(1) The police officer saw the **woman$_i$**. He probably knew **her$_i$** but …

(2) The police officer saw **her$_i$**. He probably knew the **woman$_i$** but …

Preposed RCs are rare:

(1) Der von Peter berarbeitete Fall ist gelöst.

(2) Der Fall, den Peter bearbeitet hat, ist gelöst.

Sentence processing

(1) The man who Peter who was tired saw was sick.

Sentence processing

(1) The man [**who Peter** [**who was tired**] **saw**] was sick.

Economy

lexical word > grammatical word > affix > zero

is going to → ‘s gonna

talk did → talk-ed

Today’s morphology is yesterday’s syntax. (Givón 1971)

Frequently used words/structures tend to be short. [Zipf’s law]

Iconicity

(1) a. We went home before Mary left.

b. Before Mary left we went home.

(2) a. We went home after Mary left.

b. After Mary left we went home.

Principle 1: Iconic clause orders are easier to process than non-iconic clause orders.

Principle 2: Postposed subordinate clauses are easier to process than preposed subordinate clauses.

Competing motivations

	Iconic	Non-iconic
MAIN-SUB SUB-MAIN		

Competing motivations

	Iconic	Non-iconic
MAIN-SUB SUB-MAIN	x, before y	

Competing motivations

	Iconic	Non-iconic
MAIN-SUB SUB-MAIN	x, before y	y, after x

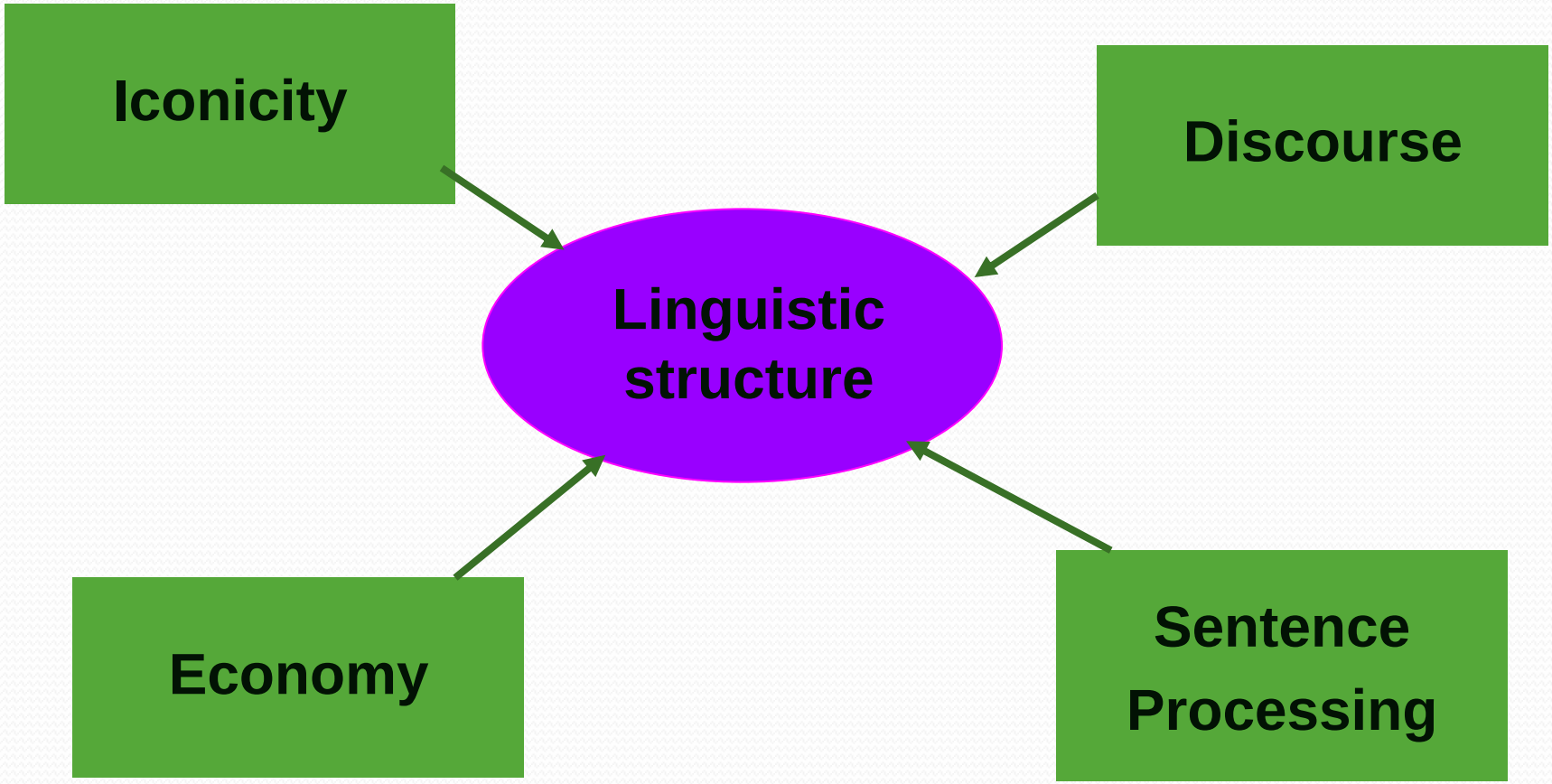
Competing motivations

	Iconic	Non-iconic
MAIN-SUB SUB-MAIN	x, before y after x, y	y, after x

Competing motivations

	Iconic	Non-iconic
MAIN-SUB	x, before y	y, after x
SUB-MAIN	after x, y	before y, x

Competing motivations



Conclusion

Linguistic structure is shaped by competing motivations. It is a dynamic system that is constantly changing. Since the various forces are often in conflict, there is no optimal language.