

# Typology: Lecture II

## Language Diversity

Christian Bentz  
*University of Tübingen*

April 28, 2017



**WORDS BONES GENES TOOLS**  
Tracking Linguistic, Cultural, and Biological Trajectories of the Human Past

**EVOLAEMP**  
LANGUAGE EVOLUTION: THE EMPIRICAL TURN

# OVERVIEW

LANGUAGES

LANGUAGE DIVERSITY

LANGUAGE FAMILIES

DIFFERENT DIMENSIONS OF DIVERSITY

# LANGUAGES

## Definition

In linguistics, languages are generally defined based on mutual *intelligibility*. Speakers of different dialects can understand each other, speakers of different languages cannot.

## Note:

- ▶ There are (yet) no strict definitions of “intelligibility”
- ▶ Sometimes languages are defined by political rather than linguistic factors, e.g. Croatia/Serbian/Bosnian

# THE ISO 639-3 STANDARD

An **ISO 639-3** three character code is assigned to languages as an universally valid identification (e.g. German: deu, English: eng, Usila Chinantec: cuc)

**Note:**

There are other (emerging) standards such as glottocodes (German: stan1295, English: stan1293, Usila Chinantec: usil1237)



# THE ISO 639-3 STANDARD

- ▶ Two related varieties are normally considered varieties of the same language if speakers of each variety have **inherent understanding** of the other variety at a functional level (that is, can understand based on knowledge of their own variety without needing to learn the other variety).
- ▶ Where spoken intelligibility between varieties is marginal, the existence of a **common literature or of a common ethnolinguistic identity** with a central variety that both understand can be a strong indicator that they should nevertheless be considered varieties of the same language.
- ▶ Where there is enough intelligibility between varieties to enable communication, the existence of **well-established distinct ethnolinguistic identities** can be a strong indicator that they should nevertheless be considered to be different languages.

(<https://www.ethnologue.com/about/problem-language-identification>)

# HOW MANY LANGUAGES ARE THERE IN THE WORLD?

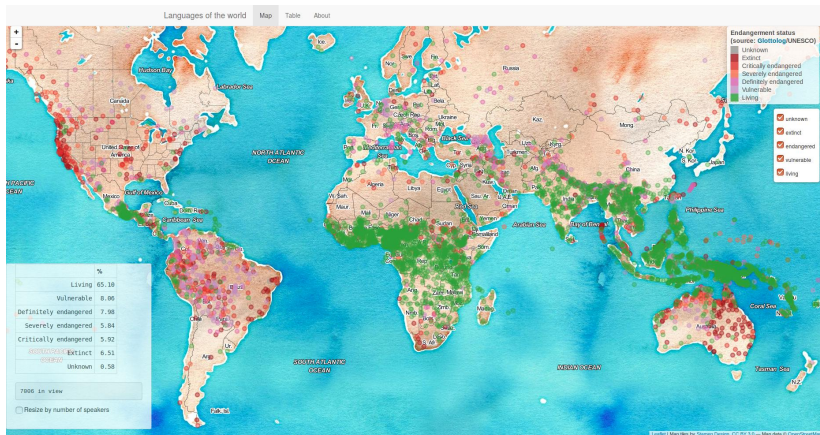
- ▶ Glottolog 3.0 lists **8444** languages, though these are not all “spoken L1 languages” (see next slide)
- ▶ Ethnologue (20th Edition) lists **7099** languages according to the ISO 639-3 criteria

# GLOTTOLOG 3.0 CLASSIFICATION

<b>Spoken L1 languages</b>	8175
<b>Unattested</b>	68
<b>Unclassifiable</b>	120
<b>Pidgin</b>	80
<b>Mixed Language</b>	23
<b>Artificial Language</b>	10
<b>Speech Register</b>	7
<b>Sign Language</b>	179
<b>All</b>	8444

(<http://glottolog.org/glottolog/glottologinformation>)

# LANGUAGES ACROSS THE WORLD



# LANGUAGE DIVERSITY

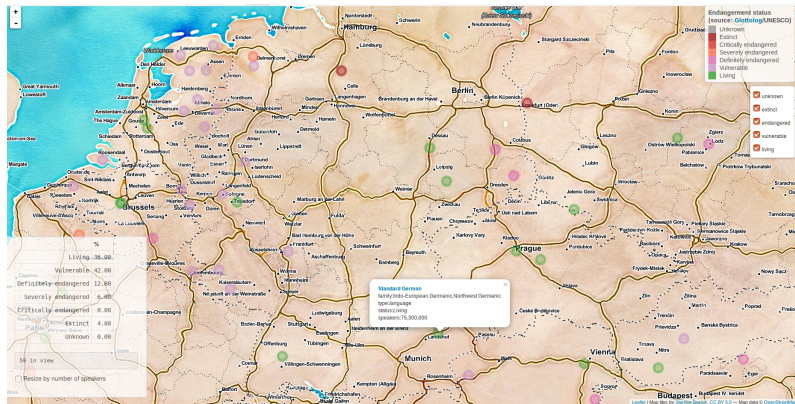
Example of high language diversity: **Papua New Guinea**



(*The Glottolog Explorer*, Caines et al. 2016)

# LANGUAGE DIVERSITY

Compare this to Germany (Indo-European languages)



Note that there are also dialects represented here, e.g. Bavarian, Swabian, etc.  
(*The Glottolog Explorer*, Caines at al. 2016)

# LANGUAGE FAMILIES

## Definition

A set of languages for which a common proto-language can be reconstructed, typically based on cognates, as well as lexical and morphological similarities.

## Note:

- ▶ whether a specific language belongs to a family or not can be controversial (e.g. Basque has been variously related to Celtic or Caucasian languages, but the consensus view is still that it is an isolate)
- ▶ the highest level of family grouping is often controversial too (e.g. is there a “Transeurasian” or “Altaic” family including Turkic, Mongolic, Japonic, and Koreanic languages?)

# HOW MANY LANGUAGE FAMILIES?

- ▶ Glottolog 3.0 lists **242** language families, and **188** isolates
- ▶ Ethnologue (20th Edition) lists **141** families (top-level genetic groups), and **53 unclassified** languages

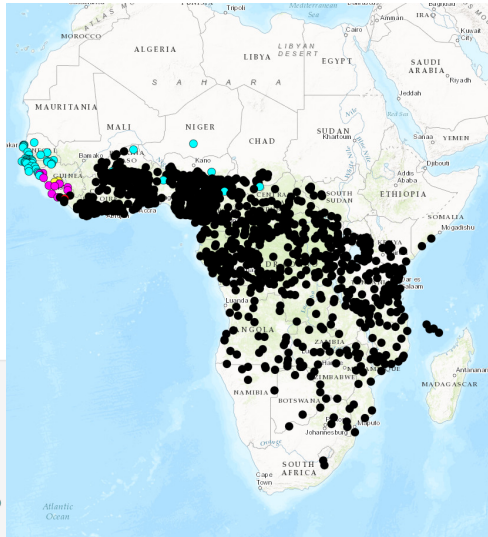
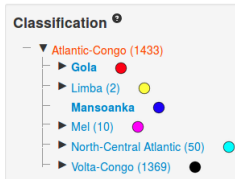


# THE BIGGEST FAMILIES

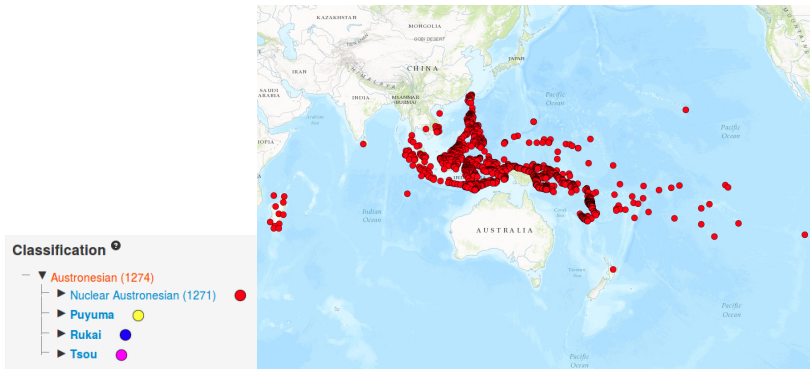
Name ▲	Level	Macro-area	Sub-families ▴	Child languages ▼
<input type="text" value="Search"/>	<input type="text" value="Top-level"/>	<input type="text" value="--any--"/>	<input type="text" value="Search"/>	<input type="text" value="Search"/>
<a href="#">Atlantic-Congo</a>	Top-level family	Africa, North America	851	1433
<a href="#">Austronesian</a>	Top-level family	Africa, Eurasia, Papunesia, South America	698	1274
<a href="#">Indo-European</a>	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	302	584
<a href="#">Sino-Tibetan</a>	Top-level family	Eurasia	279	486
<a href="#">Bookkeeping</a>	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	0	398
<a href="#">Afro-Asiatic</a>	Top-level family	Africa, Eurasia	210	374
<a href="#">Nuclear Trans New Guinea</a>	Top-level family	Papunesia	159	315
<a href="#">Pama-Nyungan</a>	Top-level family	Australia, Papunesia	140	243
<a href="#">Otomanguean</a>	Top-level family	North America, South America	70	180
<a href="#">Sign Language</a>	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	42	179
<a href="#">Austroasiatic</a>	Top-level family	Eurasia	95	158
<a href="#">Unclassifiable</a>	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	0	120
<a href="#">Tai-Kadai</a>	Top-level family	Eurasia	51	93
<a href="#">Dravidian</a>	Top-level family	Eurasia	38	81
<a href="#">Pidgin</a>	Top-level family	Africa, Australia, Eurasia, North America, Papunesia, South America	45	80
<a href="#">Arawakan</a>	Top-level family	North America, South America	45	77
<a href="#">Mande</a>	Top-level family	Africa	56	75
<a href="#">Tupian</a>	Top-level family	South America	36	71

(<http://glottolog.org/glottolog/family>)

# ATLANTIC-CONGO



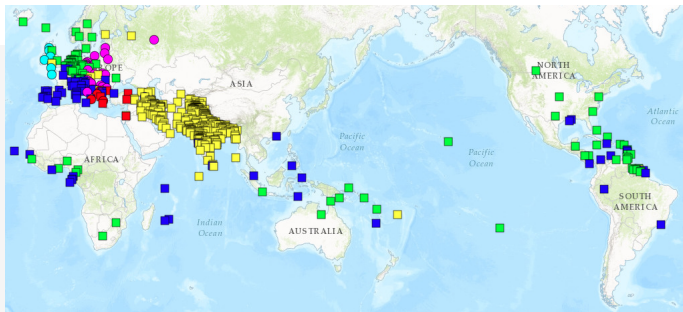
# AUSTRONESIAN



# INDO-EUROPEAN

## Classification <sup>9</sup>

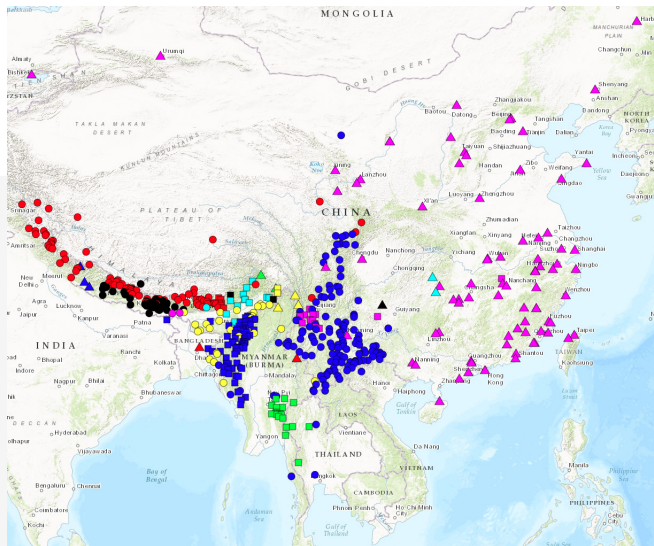
- ▼ Indo-European (584)
  - ▶ Albanian (4) ●
  - ▶ Anatolian (10) ●
  - ▶ Armenian (3) ●
  - ▶ Balto-Slavic (25) ●
  - ▶ Celtic (16) ●
  - ▶ Dacian
  - ▶ Germanic (106) ●
  - ▶ Graeco-Phrygian (10) ●
  - ▶ Indo-Iranian (318) ●
  - ▶ Italic (86) ●
  - ▶ Lusitanian
  - ▶ Messapic
  - ▶ Thracian
  - ▶ Tokharian (2) ▲



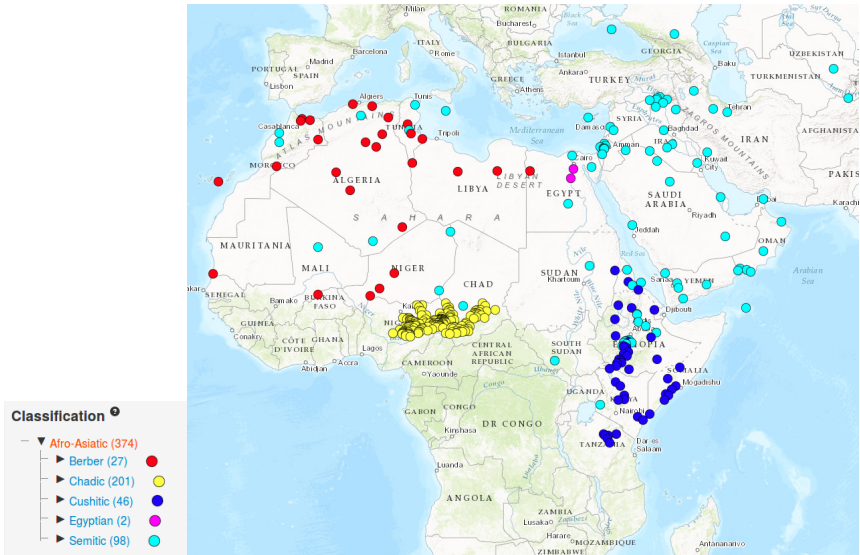
## SINO-TIBETAN

## Classification

- ▼ Sino-Tibetan (486)
  - ▶ Bodic (82)
  - ▶ Brahmaputran (36)
  - ▶ Burmo-Qiangic (150)
  - ▶ Dhimlish (2)
  - ▶ Gongduk (46)
  - ▶ Himalayish (46)
  - ▶ Karenic (20)
  - ▶ Kho-Bwa (7)
  - ▶ Kman-Meyor (2)
  - ▶ Kuki-Chin-Naga (81)
  - ▶ Macro-Bai (5)
  - ▶ Macro-Tani (12)
  - ▶ Miji (2)
  - ▶ Mishmic (2)
  - ▶ Mruic (2)
  - ▶ Nungish (3)
  - ▶ Raji-Raute (3)
  - ▶ Sinitic (26)
  - ▶ Tujia (2)
  - ▶ Unclassified Sino-Tibetan (2)



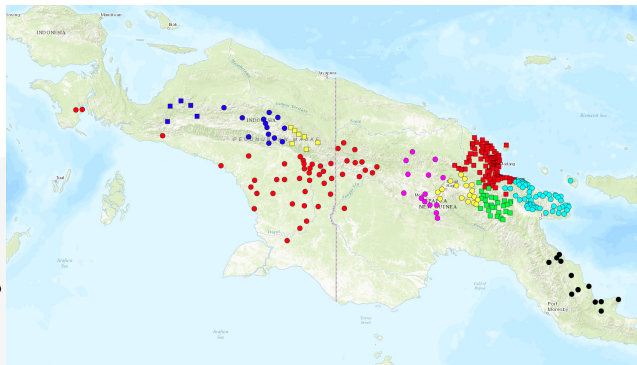
## AFRO-ASIATIC



# NUCLEAR TRANS NEW GUINEA

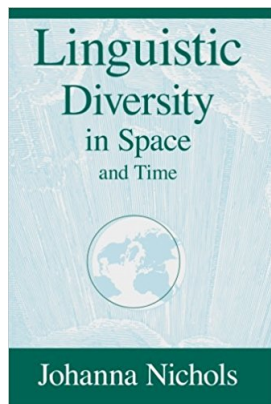
## Classification<sup>9</sup>

- ▼ Nuclear Trans New Guinea (315)
  - ▶ Asmat-Awyu-Ok (49) ●
  - ▶ Chimbu-Wahgi (17) ●
  - ▶ Dani (13) ●
  - ▶ Enga-Kewa-Huli (14) ●
  - ▶ Finisterre-Huon (51) ●
  - ▶ Greater Binanderean (13) ●
  - ▶ Kainantu-Goroka (28) ●
  - ▶ Madang (107) ●
  - ▶ Mek (8) ●
  - ▶ Paniai Lakes (5) ●



# DIFFERENT DIMENSIONS OF DIVERSITY

1. **Language diversity:** the **number of languages** in a geographic area (country, continent, or any measure of area size, e.g.  $100 \text{ km}^2$ )
2. **Phylogenetic diversity:** the number of **language families** or branches of families per area size
3. **Structural diversity:** refers to diversity in the **typological features/parameters** found in an area



Nichols (1992)



# LANGUAGE DIVERSITY AND PHYLOGENETIC DIVERSITY

Intuitively, we would expect that **more languages in an area** (higher language diversity) imply **more language families** for the same area (higher phylogenetic diversity). This seems to hold on a global scale. However, there are also systematic counter examples.

# PAPUA NEW GUINEA

high language diversity, high phylogenetic diversity

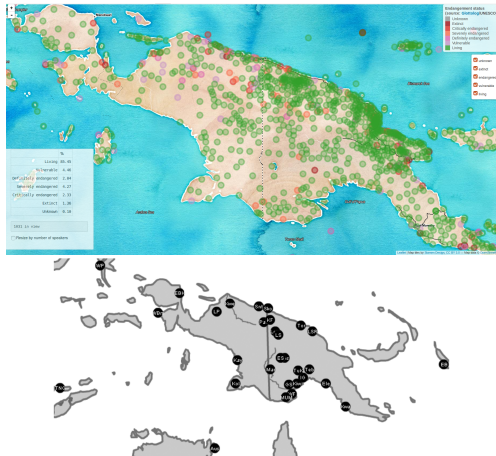
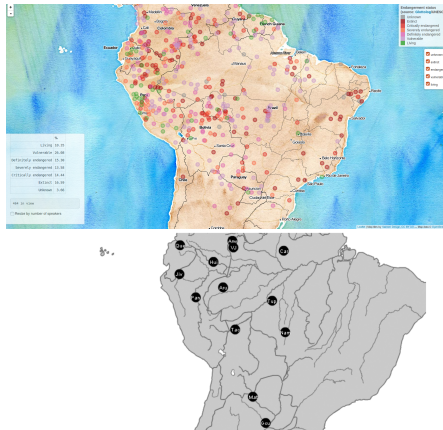


Figure 5. Homelands of New Guinea and Australian families in the sample

Legend: Aus: Australian; Bos: Bosavi; EB: East Bougainville; EBH: East Bird's Head; Ele: Eleman; ES: East Strickland; GS: Gogodala-Suki; IG: Inland Gulf; Kay: Kayagar; KF: Kwomtari-Fas; Kiv: Kiwaian; Kol:

# SOUTH AMERICA

**low** language diversity, **high** phylogenetic diversity



**Figure 8.** Homelands of Central South American language families in the sample  
 Legend: Arw: Arawakan; Arsc: Araucan; Car: Cariban; Hui: Huaitoan; Jiv: Jivaroan; Guc: Guaicuruan; Mat: Matacoan; Nam: Nambikuaran; Pan: Panoan; Que: Quechuan; Tac: Tacanan; Tup: Tupian; VJ: Vaupés-Japurá.

## AFRICA

high language diversity, low phylogenetic diversity

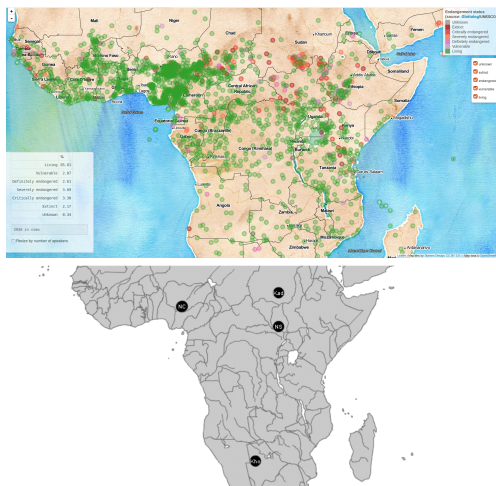


Figure 3. Homelands of African language families in the sample

Legend: AA: Afro-Asiatic, NC: Niger-Congo, Kad: Kadugli, NS: Nilo-Saharan; Kho: Khoisan. Note: NC is superimposed on AA.

# PHYLOGENETIC DIVERSITY AND STRUCTURAL DIVERSITY

We would expect that **more language families in an area** (higher phylogenetic diversity) imply **more structural diversity** in the same area. Again, this might hold true globally, but not necessarily for all areas of the world.

# SOUTH AMERICA

**high** phylogenetic diversity, **high** structural diversity (of word orders)

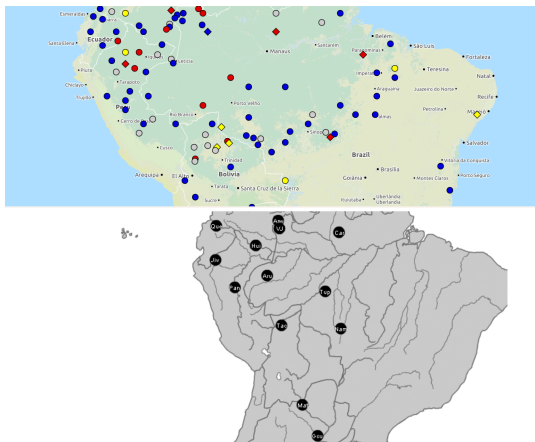


Figure 8. Homelands of Central South American language families in the sample  
 Legend: *Arw*: Arawakan; *Aru*: Arauan; *Car*: Cariban; *Hui*: Huitotoan; *Jiv*: Jivaroan; *Gcu*: Guaicuruan; *Mat*: Matacoan; *Nam*: Nambikuaran; *Pan*: Panoan; *Que*: Quechuan; *Tac*: Tacanan; *Tup*: Tupian; *VJ*: Vaupés-Japurá.

# PAPUA NEW GUINEA

**high** phylogenetic diversity, **low** structural diversity (of word orders)

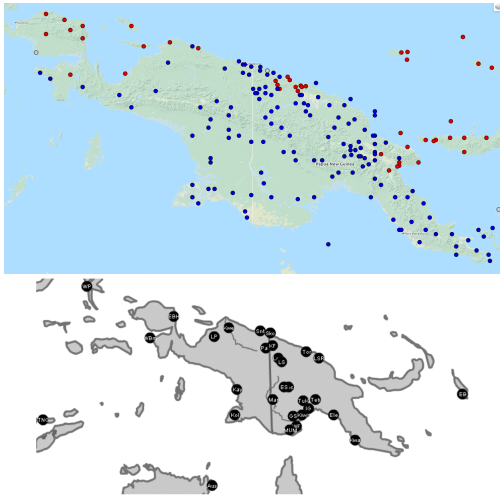


Figure 5. Homelands of New Guinea and Australian families in the sample

Legend: *Aus*: Australian; *Bos*: Bosavi; *EB*: East Bougainville; *EBH*: East Bird's Head; *Ele*: Eleman; *ES*: East Strickland; *GS*: Gogodala-Suki; *IG*: Inland Gulf; *Kay*: Kayagar; *KF*: Kwomtari-Fas; *Kiv*: Kiwaian; *Kol*:

# SUMMARY

- ▶ Languages and Language Families
  - ▶ between ca. 7000 and 8500 languages depending on classification
  - ▶ between ca. 150 and 250 language families
- ▶ Different Dimensions of Diversity
  - ▶ Language diversity and phylogenetic diversity
  - ▶ Structural diversity and phylogenetic diversity