

Variability of languages in time and space

Intro

Magda Ševčíková & Anna Nedoluzhko & Zdeněk Žabokrtský

2 October 2024

<https://ufal.mff.cuni.cz/courses/npfl100>

Living languages around the world

- The number of languages in printed and online resources:
 - *Ethnologue* (27th ed., 2024): **7,164 languages**
 - *Ethnologue* (26th ed., 2023): 7,168 languages
 - *Ethnologue* (25th ed., 2022): 7,151 languages
 - *Ethnologue* (24th ed., 2021): 7,139 languages
 - *Ethnologue* (23rd ed., 2020): 7,117 languages
 - *Ethnologue* (22nd ed., 2019): 7,111 languages
 - *Ethnologue* (21st ed., 2018): 7,097 languages
 - ...
 - *Ethnologue* (16th ed., 2009): 6,909 languages
 - ...
 - *Ethnologue* (7th ed., 1969): 4,493 languages
 - ...
 - *Encyclopedia Britannica* (1911): app. 1,000 lang.

– not an increase in the number of languages,
but rather steadily improving documentation efforts

- **A first language (L1)**

is a language which a person acquires first in her/his life, usually naturally as a child.

- **A second language (L2)**

is a language which a person acquires next to her/his first language.

- **A living language**

has at least one speaker for whom it is a first language.

- e.g. English, Catalan, Japanese, Swahili, French Sign Language

– distinguished from *dead* languages

- a **dead language** is no longer the native language of any community but is still in use, e.g. Latin

– a more subtle insight provided by the

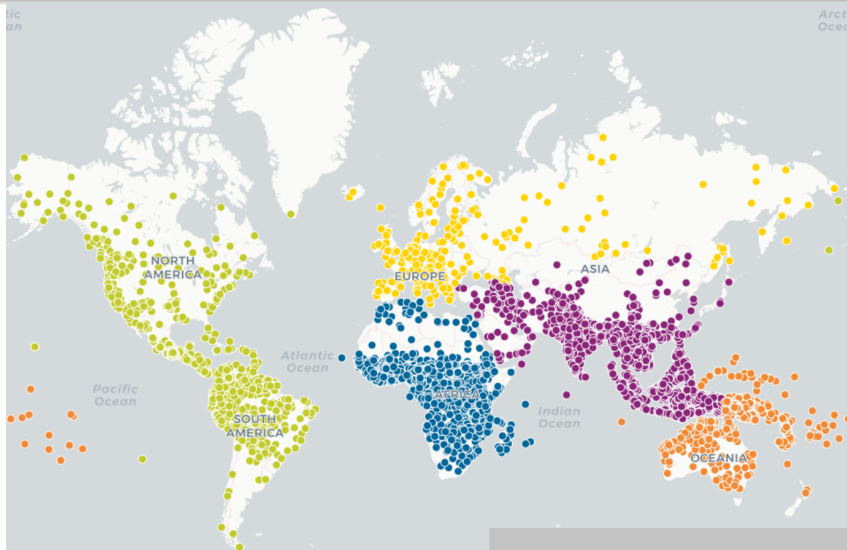
Expanded Graded Intergenerational Disruption Scale (EGIDS; Lewis & Simons 2010)

- from *internationally used* languages to *extinct* languages

Expanded Graded Intergenerational Disruption Scale (EGIDS)

| Level | Label | Description |
|-------|---------------------|--|
| 0 | International | The language is widely used between nations in trade, knowledge exchange, and international policy. |
| 1 | National | The language is used in education, work, mass media, and government at the national level. |
| 2 | Provincial | The language is used in education, work, mass media, and government within major administrative subdivisions of a nation. |
| 3 | Wider Communication | The language is used in work and mass media without official status to transcend language differences across a region. |
| 4 | Educational | The language is in vigorous use, with standardization and literature being sustained through a widespread system of institutionally supported education. |
| 5 | Developing | The language is in vigorous use, with literature in a standardized form being used by some though this is not yet widespread or sustainable. |
| 6a | Vigorous | The language is used for face-to-face communication by all generations and the situation is sustainable. |
| 6b | Threatened | The language is used for face-to-face communication within all generations, but it is losing users. |
| 7 | Shifting | The child-bearing generation can use the language among themselves, but it is not being transmitted to children. |
| 8a | Moribund | The only remaining active users of the language are members of the grandparent generation and older. |
| 8b | Nearly Extinct | The only remaining users of the language are members of the grandparent generation or older who have little opportunity to use the language. |
| 9 | Dormant | The language serves as a reminder of heritage identity for an ethnic community, but no one has more than symbolic proficiency. |
| 10 | Extinct | The language is no longer used and no one retains a sense of ethnic identity associated with the language. |

Regions of origin of languages

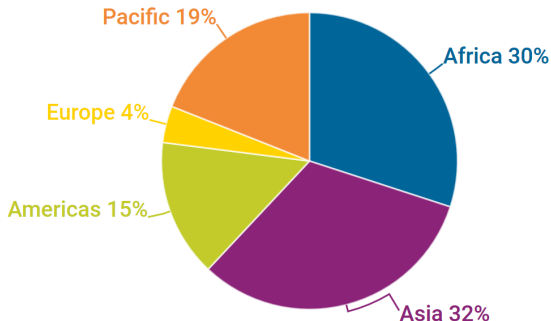


Each language is located in its **primary country**, i.e. it is shown **just once**; e.g. English in the United Kingdom, Esperanto in Poland.

Languages by region of origin

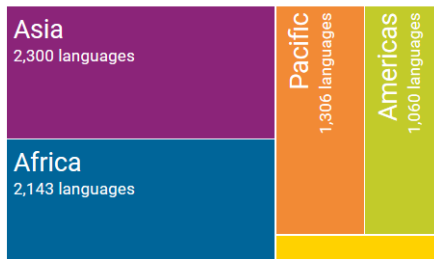
Ethnologue (2018):

- 2,300 languages (out of 7,097) are from Asia
- 2,143 languages are concentrated in Africa
- 288 languages belong to Europe



Languages by region and by size (number of L1 speakers)

Languages



Europe (yellow): 288 languages

Population



Americas (green): 49M
Pacific (orange): 7M speakers

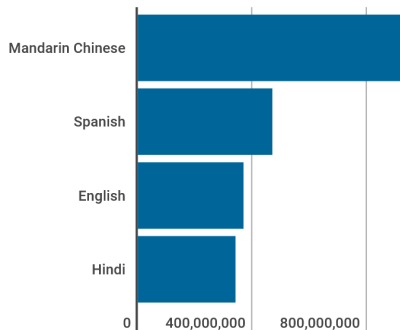
- English listed as a language belonging to Europe. Therefore, all L1 speakers of English (e.g. incl. English speakers living in the USA) categorized under Europe.

Languages by number of L1 speakers

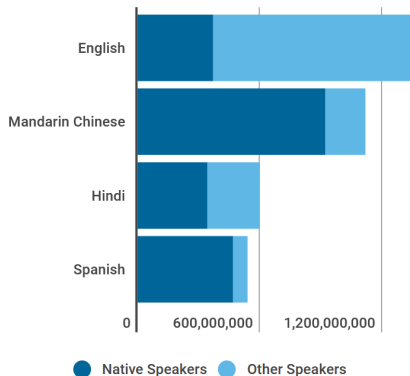
- nearly 7,900,000,000 speakers around the world
 - nearly 90 % of speakers use languages from Asia or Europe as L1
 - 40 % of speakers use one of just 5 languages as L1
 - Chinese, Spanish, English, Hindi, Arabic

Number of L1 speakers vs. of L2 speakers

Languages with the most native speakers



Languages with the most speakers



Variability of languages

Languages of the world share some features, but display differences at the same time – studied by **linguistic typology** (= language typology)

| En. | German | Czech | Slovene | Croatian | French | Spanish | Italian |
|---------------|--------------|-------------|------------|-------------|-------------|--------------|--------------|
| <i>father</i> | <i>Vater</i> | <i>otec</i> | <i>oče</i> | <i>otac</i> | <i>père</i> | <i>padre</i> | <i>padre</i> |

- WHY are they similar? WHY do they differ?
 - due to a common origin → genealogical (genetic) classification
 - due to language contact → areal linguistics
 - due to other factors

Genealogical classification

Some similarities can be explained by the **common origin** of languages.

| En. | German | Czech | Slovene | Croatian | French | Spanish | Italian |
|--------|--------|-------|---------|----------|--------|---------|---------|
| father | Vater | otec | oče | otac | père | padre | padre |

- languages that developed from the same ancestor language and belong to the same **language family** share common features
 - English and German belong to **Germanic** languages
 - Czech, Slovene, and Croatian to **Slavic** languages
 - French, Spanish and Italian to **Romance** languages

- different genealogical classification in different resources

Languages that display systematic similarities and differences are assumed to have descended from *a common source language*
= **ancestor language / proto-language**.

- regular correspondences between languages studied by comparative methods
- proto-languages not attested, they are **reconstructed** from available documents of the descendent languages

Language families

Language family is a group of languages that have developed from the same proto-language.

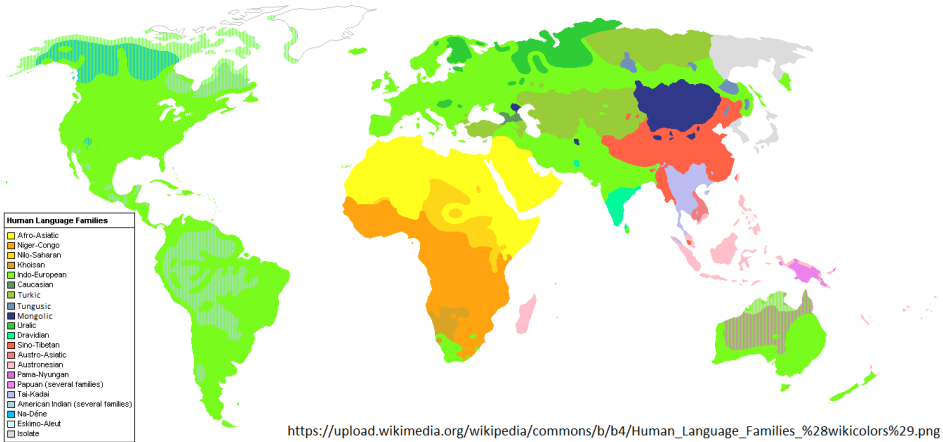
| language family | languages | million of speakers |
|------------------|-----------|---------------------|
| Niger-Congo | 1,527 | 511 |
| Austronesian | 1,224 | 324 |
| Trans-New Guinea | 479 | 3.5 |
| Sino-Tibetan | 453 | 1,370 |
| Indo-European | 444 | 3,090 |
| Afro-Asiatic | 366 | 496 |

- plus another 150 language families in *Ethnologue* 2023:
<https://www.ethnologue.com/browse/families/>

Language isolates are natural languages that cannot be assigned to any language family; e.g. Basque (= Euskera; spoken in Spain), Ainu (Japan).

- <https://www.ethnologue.com/subgroup/79/>

Language families around the world

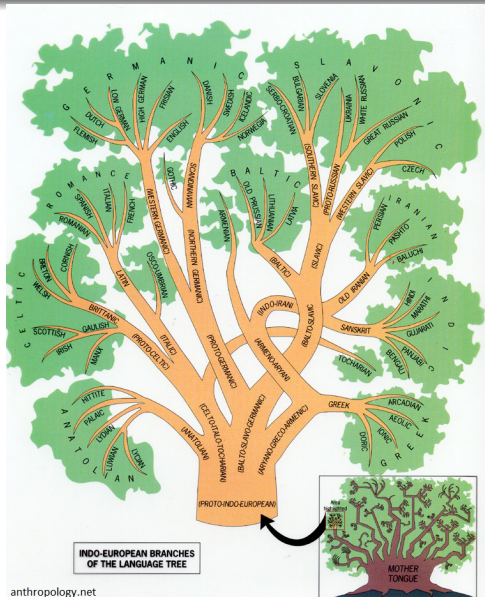


Indo-European language family tree

proto-language: Proto-Indo-European (PIE)

- reconstructed on the basis of Sanskrit, Ancient Greek, Latin, and others
- assumed existence: 4,500–2,500 B.C.E.

| | | | |
|----------|------|------|------|
| Sanskrit | asmi | asi | asti |
| Greek | eimi | essi | esti |
| Latin | sum | es | est |
| Gothic | im | is | ist |



- multiple-level classification of languages

language family

— **sub-family**

— **group**

— **subgroup**

— **sub-subgroup**

— **individual language**

- classification of **Slavic** languages

Indo-European lang. family

— Balto-Slavic sub-family

— group: Slavic lang.

— subgroups: East, South, West

— e.g. sub-subgroups of South Slav. lang.: Eastern, West.

— 20 Slavic languages in total

- three-level genetic classification of languages:

language family

— **genus**

— **individual language**

- classification of **Slavic** languages:

Indo-European family

— genus: Slavic languages

— 17 individual Slavic languages in total

20 Slavic languages in *Ethnologue* vs 17 Slavic languages in *WALS*

| <i>Ethnologue</i> | <i>WALS</i> |
|-------------------|------------------|
| Belarusian | Belorussian |
| Russian | Russian |
| Ukrainian | Ukrainian |
| Bulgarian | Bulgarian |
| Macedonian | Macedonian |
| Bosnian | Bosnian |
| Serbian | Serbian-Croatian |
| Slovene | Slovene |
| Czech | Czech |
| Slovak | Slovak |
| Kashubian | Kashubian |
| Polish | Polish |
| Lower Sorbian | Lower Sorbian |
| Upper Sorbian | Upper Sorbian |

Ethnologue only:

- Croatian
- Montenegrin
- Church Slavonic
- Slavomolisano
- Silesian
- Rusyn

WALS only:

- Sorbian
- Polabian
- Slovincian

But what about similarities between unrelated languages? They can result from **language contact**.

| En. | Czech | Basque | Swahili | Maori |
|---------------|-------------|----------------|-----------------|-------------|
| <i>pepper</i> | <i>pepř</i> | <i>piperra</i> | <i>pilipili</i> | <i>pepa</i> |

“The country of origin of black pepper is south India. The origin of the word can be found in Sanskrit pippali. The merchants from India distributed to other countries not only spice but also its name. In this way languages got in touch, in contact.” (Körvélyessy 2017:80)

Languages that are geographically in contact may influence each other

- based on the regional location of languages
- without respect to their genetic classification
- speakers of the languages interact and influence each other
 - pronunciation
 - grammatical constructions
 - borrowing of words

| language | 'woman' – 'the woman' | 'man' – 'the man' |
|-----------|-----------------------|-------------------|
| Bulgarian | жена – жената | мъж – мъжът |
| Romanian | femeie – fememia | bărbat – bărbatul |
| Albanian | grua – gruaja | burrë – burri |

Körtvélyessy (2017:2):

“**Language typology** is a system or study that divides languages into smaller groups according to similar properties they have. [...] These smaller groups are called **language types**.”

- a **holistic approach** to language typology

“The classification of languages into language types attempts to ‘match’ the complete language system with one language type.”

- a **partial approach** to language typology

“the classification is based on the analysis of a selected language construction and/or phenomenon (not the entire language), for example the size of the consonantal inventory, the presence vs. absence of articles in language, the order of words in a sentence etc.”

- Typological classification is based on the study of structural and functional features at particular levels of the language system.
- Typological approach to phonology:
 - How many vowels and consonants are used in the language?
 - Which features do they have?
 - How are syllable formed in the language?
 - Which prosodic features are employed in the language?

- Are grammatical categories conveyed by morphemes, auxiliaries, and/or by word order?
 - How many morphological cases are available in the language?
- In holistic approaches to morphological typology, several language types discerned:
- 1 agglutinating
 - 2 fusional
 - 3 polysynthetic
 - 4 isolating

Linguistic typology of word formation

- Which word-formation processes are available in a language to coin new words?
- How frequent are particular word-formation processes in a language?
- How are things/abstract phenomena/actions named in a language?

- How are sentence elements ordered in a linear sequence (sentence)?
 - focus on three main elements: *subject*, *verb*, *object*
 - the order of the elements analysed in simple indicative sentences

A language universal

is a feature that all languages have, or a statement that holds for all languages, or *less strictly* (more commonly), a feature or statement that holds for a majority of languages.

- morphological and syntactic universals listed in the *Universals Archive*

<https://typo.uni-konstanz.de/rara/category/universals-archive/>

References

- Comrie, B., Matthews, S. & Polinsky, M.: *The Atlas of Languages: The Origin and Development of Languages Throughout the World*. 1997.
- Comrie, B. (ed.): *The World's Major Languages*. Oxford: Oxford University Press, 1987.
- Dryer, M. S. & Haspelmath, M. (eds., 2013): *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online: <http://wals.info>
- Haspelmath, M. et al. (eds.; 2001): *Language Typology and Language Universals*. Mouton de Gruyter.
- Körtvélyessy, L. (2017): *Essentials of Language Typology*. Košice: UPJŠ. Available online: <https://unibook.upjs.sk/sk/anglistika-a-amerikanistika/365-essentials-of-language-typology>
- Lewis, M. P. & Simons, G. F. (2010): Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*, 55, pp. 103–120.
- Matthews, P. H. (2007): *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.
- Simons, G. F. & Fennig, C. D. (eds., 2018): *Ethnologue: Languages of the World*, Twenty-first edition. Dallas, Texas: SIL International. Online version: <http://www.ethnologue.com>