

# **Variability of languages in time and space**

## **Lecture 1: Basic facts and linguistic resources on the world's languages**

Magda Ševčíková

September 30, 2020

## ① Basic facts

- Number of the world's languages
- Living language
- Languages by region of origin and number of speakers

## ② Linguistic resources on the world's languages

- *Ethnologue: Languages of the World*
- *WALS – The World Atlas of Language Structures*

## ① Basic facts

- Number of the world's languages
- Living language
- Languages by region of origin and number of speakers

## ② Linguistic resources on the world's languages

- *Ethnologue: Languages of the World*
- *WALS – The World Atlas of Language Structures*

# Number of languages

- Printed and online resources differ considerably in estimating the number of languages
  - *Encyclopedia Britannica* (1911): app. 1,000 lang.
  - *Ethnologue* (7th ed., 1969): 4,493 languages
  - *Ethnologue* (16th ed., 2009): 6,909 languages
  - *Ethnologue* (21st ed., Simons – Fennig 2018): 7,097 languages
  - *Ethnologue* (22nd ed., Simons – Fennig 2019): 7,111 languages
  - *Ethnologue* (23rd ed., Simons – Fennig 2020): 7,117 languages

– the increasing figures are not due to any increase in the number of languages, but rather due to steadily improving documentation efforts

# Number of languages in *Ethnologue*

According to *Ethnologue* (2020), which is the most extensive catalogue of world's languages,

7,117 living languages are used around the world

# First language vs. second language

- **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)

<https://www.ethnologue.com/about/language-status:>

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.



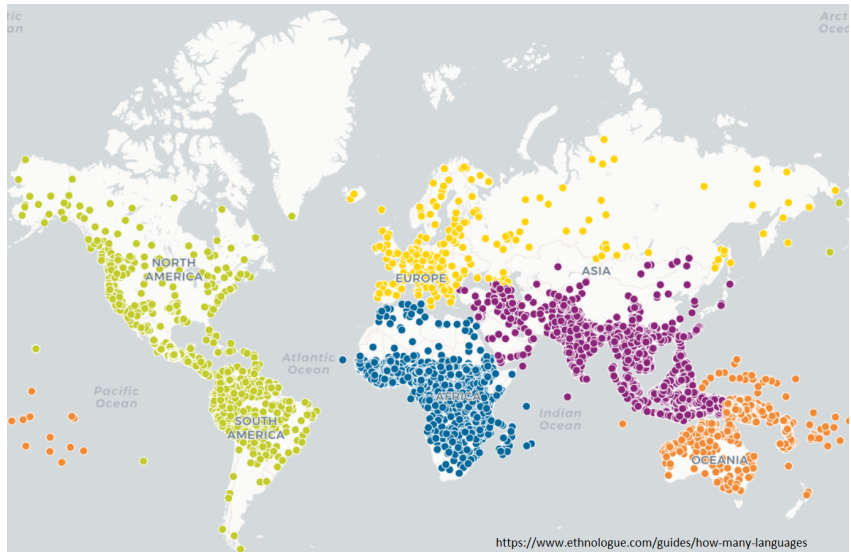
# EGIDS and language endangerment

- EGIDS levels 6b (Threatened) and 7 (Shifting)
  - seen as the first steps down the endangerment
  - 1,559 (22 %) out of 7,097 languages listed in *Ethnologue* (2018)
- EGIDS levels 8a (Moribund), 8b (Nearly Extinct), and 9 (Dormant)
  - referred to as “dying” languages
  - 922 (13 %) out of 7,097 languages listed in *Ethnologue* (2018)
  - e.g. Pipil (El Salvador; level 8a), Wymysorys (‘vilamovština’, Poland; level 8b), Krimchak (‘krymčačtina’, Ukraine; level 8b), Tirahi (Afghanistan; level 8b), Providencia Sign Language (Colombia; level 8b), Church Slavonic (Russia; level 9, “second language only”), Urali (India; level 9),  
**Latin** (Vatican State; level 9, “second language only”)
- EGIDS level 10 (Extinct)
  - in *Ethnologue* (2018), only languages that have become extinct since 1950 (370 languages)
  - e.g. Avestan (Iran)

# Region of origin of a language

- In maps at *ethnologue.com*, each language is located in its **primary country**
  - each language shown **just once**, e.g.
    - English located in the United Kingdom
    - Esperanto located in Poland

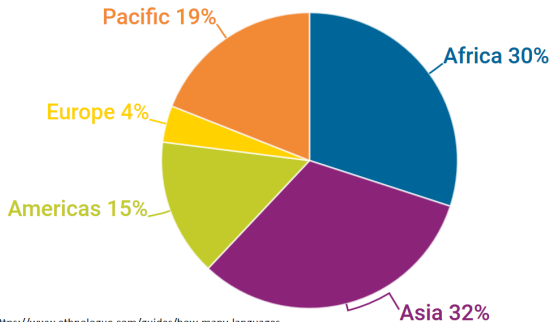
# Regions of origin of languages



# 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



<https://www.ethnologue.com/guides/how-many-languages>

- **Language size**

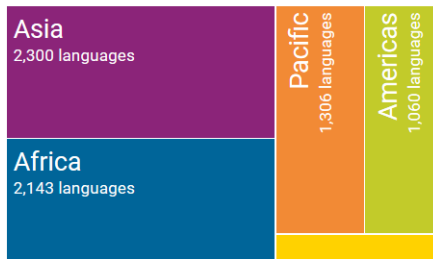
is the number of L1 speakers of a particular language

- **Average language size**

- calculated, for instance, for a group of languages from a particular region
- as the sum of speakers of the involved languages divided by the number of languages in question

# Languages by region and by number of speakers

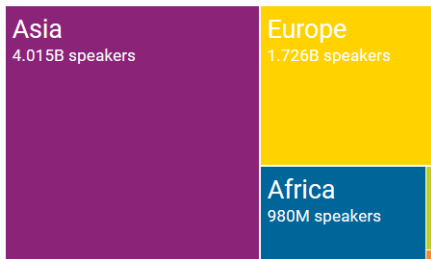
## Languages



<https://www.ethnologue.com/guides/how-many-languages>

Europe (yellow): 288 languages

## Population



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

Note: 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 6,778,000,000 speakers around the world
  - 86 % of speakers use languages from Asia or Europe as L1
  - 40 % of speakers use one of just 5 languages as L1
    - Chinese, Spanish, English, Arabic, Hindi

# Number of L1 speakers vs. of L2 speakers

rank	as L1	# of speakers	as L2	# of speakers
1	Chinese	1,299 mil.	English	660 mil.
2	Spanish	442 mil.	Hindi	215 mil.
3	English	378 mil.	Malay	204 mil.
4	Arabic	315 mil.	Chinese	193 mil.
5	Hindi	260 mil.	French	153 mil.



# A closer look: Languages in Spain

*Ethnologue* (2018) lists 16 living languages used in Spain:

- Aragonese [aragonština]
- Asturian [asturština]
- Basque [baskičtina]
- Caló
- Catalan [katalánština]
- Catalan Sign Language
- Erromintxela
- Extremaduran [extremadurština]
- Fala [falština]
- Galician [galicijština]
- Gascon, Aranese [gaskonština]
- Portuguese [portugalština]
- Quinqui
- Spanish [španělština]
- Spanish Sign Language
- Valencian Sign Language



<https://www.ethnologue.com/country/ES/languages>

# Languages in the Czech Republic

10 living languages acc. to *Ethnologue* (2018):

- Bavarian [bavorština]
- Czech [čeština]
- Czech Sign Language [český znakový jazyk]
- German, Standard [spisovná němčina]
- Polish [polština]
- Romani, Carpathian [karpatská romština]
- Romani, Sinte [sintská romština]
- Rusyn [rusínština]
- Silesian, Lower [dolní slezština]
- Slovak [slovenština]

<https://www.ethnologue.com/country/CZ/languages>

# Language status of languages spoken in the Czech Republic

- EGIDS level acc. to *Ethnologue* (2018)  
(<https://www.ethnologue.com/country/CZ/status>)
  - National lang. (level 1): Czech
  - Dispersed lang. (level 5): Standard German, Polish, Sinte Romani, Slovak
  - Developing lang. (level 5): Bavarian, Czech Sign Language, Carpathian Romani
  - Vigorous (level 6a): Rusyn, Lower Silesian

- A language can be used in more varieties = **dialects**
  - usually spoken in a specific part of a country or other geographical area
  - **not** counted as separate items in the total of living languages, but
  - subsumed under the particular language

# Distinguishing dialects from languages

- *Ethnologue* (2018) defines **dialects of a language** as “speech varieties which are functionally **intelligible** to each others’ speakers because of **linguistic similarity**”

- **intelligibility**

- how speakers of one variety of a language understand another variety
- difficulty in comprehension with less than 85 %

- linguistic similarity often specified as **lexical similarity**

- comparison of standardized lists of words
- for a dialect, similarity of at least 85 % expected

– however, cultural and/or linguistic traditions often important for whether a variety is considered a dialect

# Example: Dialects of Spanish

- in *Ethnologue* (2018), Spanish (alternate name: Castilian) listed as 1 of 16 languages used in Spain
- Spanish itself has 8 dialects:
  - Andalusian (Andalusí, Andaluz, Andalusí) [andaluzština]
  - Aragonese [aragonština]
  - Murcian [murcijština]
  - Navarrese [navarština]
  - Castilian [kastilština]
  - Canary Islands Spanish (Isleño)
  - American Spanish (Chicano)
  - Silbo Gomero

<https://www.ethnologue.com/language/spa>

# Dialects of Czech

- *Ethnologue* (2018):

- 1 Central Bohemian
- 2 Czecho-Moravian
- 3 Hanak
- 4 Lach (Yalach)
- 5 Northeast Bohemian
- 6 Southwest Bohemian

- vs. ÚJČ AV ČR:

- 1 česká nářečí v užším smyslu
  - severovýchodočeská nářeční podskupina
  - středočeská nářeční podskupina
  - jihozápadočeská nářeční podskupina
  - českomoravská nářeční podskupina
- 2 středomoravská nářeční skupina
  - centrální středomoravská podskupina
  - jižní středomoravská podskupina
- 3 východomoravská nářeční skupina
  - severní (valašská) podskupina
  - jižní (slovácká) podskupina
- 4 slezská nářečí
  - západní (opavská) podskupina
  - východní (ostravská) podskupina
  - jižní (frenštátská) podskupina
  - přechodová nářečí česko-polská

<https://www.ethnologue.com/language/ces>

<http://www.ujc.cas.cz/miranda2/export/sitesavcr/ujc/sys/galerie-obrazky/publikace-obalky/>



## ① Basic facts

- Number of the world's languages
- Living language
- Languages by region of origin and number of speakers

## ② Linguistic resources on the world's languages

- *Ethnologue: Languages of the World*
- *WALS – The World Atlas of Language Structures*

# Available resources on the world's languages

- linguistic description of languages for cross-linguistic research available
  - in printed publications
    - see selected titles in References
  - in online resources

# Online resources on the world's languages

- *Ethnologue: Languages of the World*
  - <http://www.ethnologue.com>
- *WALS – The World Atlas of Language Structures*
  - <http://wals.info>
- *APiCS – Atlas of Pidgin and Creole Language Structures*
  - <http://apics-online.info>
- *WOLD – The World Loanword Database*
  - <http://wold.clld.org>
- for other resources see e.g.
  - <https://linguistic-typology.org/databases/>

# A closer look: *Ethnologue: Languages of the World*

- a catalogue of languages
    - print and web publication
    - published by SIL International
      - “Summer Institute of Linguistics”
      - a Christian linguistic service organization studying and supporting lesser-known languages
  - work in progress
    - 1st edition 1951
    - 21st edition 2018
    - 22nd edition 2019
    - 23rd edition 2020
- <http://www.ethnologue.com>

- 7,117 descriptions of living languages
- for each language
  - multi-level genetic classification (cf. previous slides)
  - number of L1 and L2 speakers
  - list of dialects
  - list of countries where the language is used
  - basic typological information
    - inventory of phonemes, selected grammatical categories, basic syntactic information

>>> cf. the entry for Dutch on the next slide

LANGUAGE

MAP

FEEDBACK

## A language of Netherlands

ISO 639-3	<a href="#">nld</a>
Autonym	Hollands, Nederlands
Population	16,000,000 (European Commission 2012). Total users in all countries: 23,025,430.
Language Maps	<a href="#">Belgium</a> , <a href="#">Luxembourg</a> and <a href="#">Netherlands</a>
Language Status	1 (National). De facto national language.
Classification	<a href="#">Indo-European</a> , <a href="#">Germanic</a> , <a href="#">West</a> , <a href="#">Low Saxon-Low Franconian</a> , <a href="#">Low Franconian</a>
Dialects	Northern North Hollandish (Westfries). Belgian variety of Dutch (not Vlaams [ <a href="#">vls</a> ]) only slightly different from that of the Netherlands.
Typology	SVO (SOV in subordinate clauses); prepositions; noun head final; gender (common/neuter); definite and indefinite articles; passives (middle, active, passive); tense and aspect; comparatives; 21 consonants, 13 vowels, 3 diphthongs; non-tonal; stress on first syllable.
Language Use	Most speakers of other languages are also fluent in Dutch. Most also use English [ <a href="#">eng</a> ] (European Commission 2006), Standard German [ <a href="#">deu</a> ] (European Commission 2006). Some also use French [ <a href="#">fra</a> ] (European Commission 2006).
Language Development	Fully developed. Bible: 1522–2004.
Language Resources	<a href="#">OLAC resources in and about Dutch</a>
Writing	Braille script [ <a href="#">Bra</a> ], used since 1951. Latin script [ <a href="#">Latn</a> ], primary usage.

# A closer look: *WALS – The World Atlas of Language Structures*

- database of phonological, grammatical and lexical properties of languages
  - obtained from reference grammars and other descriptive material
- 55 authors
  - Greville G. Corbett, Martin Haspelmath, Bernard Comrie, Matthew S. Dryer et al.

- 1st version 2005
  - book with CD-ROM, Oxford University Press
- 1st online version (*WALS Online*) 2008
  - a separate publication of the Max Planck Institute for Evolutionary Anthropology
- current version from 2013  
<http://wals.info>



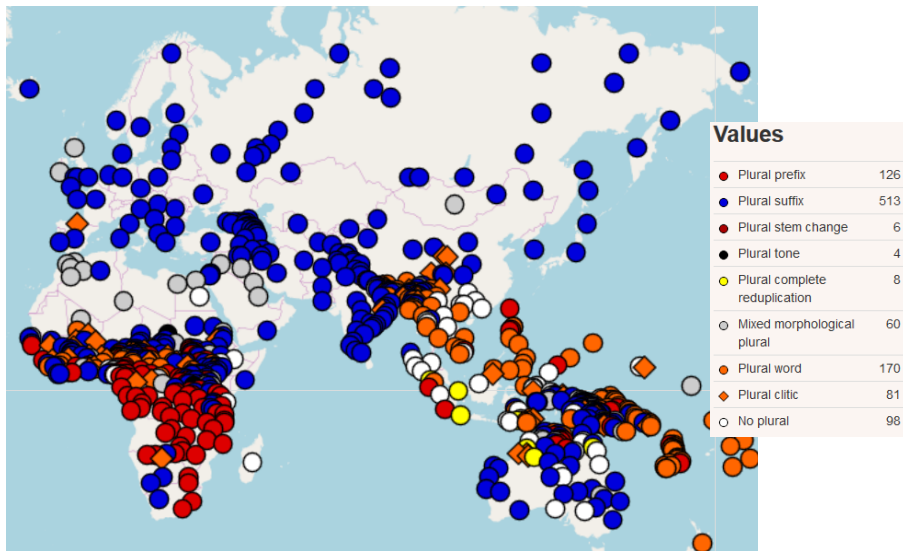
# Linguistic features in *WALS*

- 144 features in total
  - structural properties of language that describe “one aspect of linguistic diversity”
  - concise linguistic description of each feature
  - 2 to 28 values of the feature
  - assigned to one of 11 areas:

phonology	nominal syntax	lexicon
morphology	simple clauses	sign languages
nominal categories	complex sentences	other
verbal categories	word order	

- distribution of the feature values on the map  
>>> see ex. on the next slide

# WALS feature 33A: Coding of Nominal Plurality



# Description of languages in *WALS*

- an entry for each language
  - name
  - geographical info
  - two-level genetic classification
    - language family and genus  
e.g. Dutch: Indo-European, Germanic
  - list of relevant features and their values

>>> cf. the entry for Dutch on the next slide



Family: Indo-European / Genus: Germanic

Glottocode: dutc1256

ISO 639-3: nld

## Language Dutch

WALS code: dut

Showing 1 to 89 of 89 entries

Fid	Value	Feature	Source	Area
	<input type="text" value="Search"/>	<input type="text" value="Search"/>		
14A	No fixed stress	Fixed Stress Locations	Kager 1989; Trommelen and Zonneveld 1999a; van der Hulst 1984	Phonology
15A	Right-oriented: One of the last three	Weight-Sensitive Stress	Kager 1989; Trommelen and Zonneveld 1999a; van der Hulst 1984	Phonology
16A	Coda consonant	Weight Factors in Weight- Sensitive Stress Systems	Kager 1989; Trommelen and Zonneveld 1999a; van der Hulst 1984	Phonology
17A	Trochaic	Rhythm Types	Kager 1989; Trommelen and Zonneveld 1999a; van der Hulst 1984	Phonology
26A	Strongly suffixing	Prefixing vs. Suffixing in	Koolhoven 1968:	Morphology

Coordinates WGS84 52°30'N, 6°E  
52.50, 6.00

Spoken in: Netherlands

### Alternative names

Ruhlen: Dutch  
Routledge: Netherlandic  
Ethnologue: Dutch

### Sources

# Example comparison: *Ethnologue* vs. *WALS*

- different approaches in individual resources
- here, an example comparison of genetic classification with a focus on **Slavic languages** in
  - *Ethnologue: Languages of the World*
  - *WALS – The World Atlas of Language Structures*

# Classification of languages in *Ethnologue*

- 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

# Classification of languages in *WALS*

- 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 lang. 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



## 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>

- other online resources:

- *APiCS – Atlas of Pidgin and Creole Language Structures*. Available online: <http://apics-online.info>
- *WOLD – The World Loanword Database*. Available online: <http://wold.clld.org>
- <https://www.linguisticsociety.org>
- <http://www.linguistic-typology.org>
- <https://www.omniglot.com>
- [www.uis.unesco.org](http://www.uis.unesco.org)