Variability of languages in time and space

Lecture 1: Languages around the world – resources and classification

Magda Ševčíková

October 12, 2018

- Lecture 1 – October 12, 2018
  1. Intro
  2. Approaches to classification of the world's languages
  3. Linguistic resources on the world's languages
- Lecture 2 – October 19, 2018
  - Linguistic sign, linguistic disciplines
- Lecture 3 – October 26, 2018
  - Word formation
- Lecture 4 – November 2, 2018
  - Comparing word formation

Lecture 1 – October 12, 2018

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

2. Approaches to classification of the world's languages
   - Genetic classification
   - Areal classification
   - Linguistic typology

3. Linguistic resources on the world's languages
   - *Ethnologue: Languages of the World*
   - *WALS – The World Atlas of Language Structures*

Lecture 1 – October 12, 2018

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

2. **Approaches to classification of the world's languages**
   - Genetic classification
   - Areal classification
   - Linguistic typology

3. **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

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

According to *Ethnologue* (2018), which is the most extensive catalogue of world's languages,

> 7,097 living languages are used around the world

- **A living language**

  has at least one speaker for whom it is a first language.
    - e.g. English, Catalan, Japanese, Swahili, French Sign Language

  – it is distinguished from *dead* and *extinct* languages

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.

Extinct vs. dead languages

- in *Ethnologue* (2018),
  - an **extinct language** is defined as having no longer any speakers.
    - e.g. Old Church Slavonic, Gothic
  - a **dead language** is no longer the native language of any community but is still in use
    - e.g. Latin

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
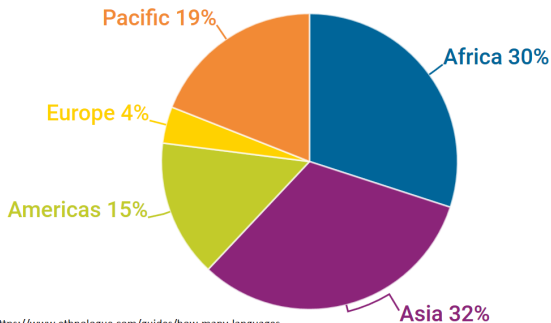   - Esperanto located in Poland

Regions of origin of languages

https://www.ethnologue.com/guides/how-many-languages
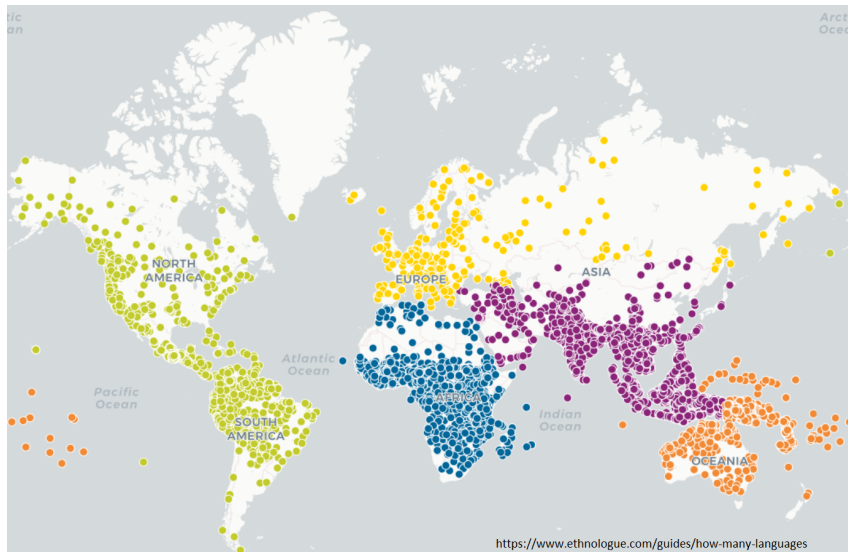
Languages by region of origin

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
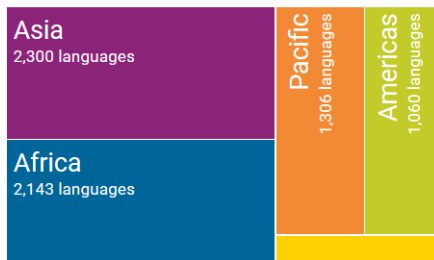
- **Language size**

  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

Asia
2,300 languages

Africa
2,143 languages

Pacific
1,306 languages

Americas
1,060 languages

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

Europe (yellow): 288 languages

Population

Asia
4.015B speakers

Europe
1.726B speakers

Africa
980M speakers

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.

- the language size is used as an important indicator of whether a language is endangered

- features defining **language endangerment**
  - less than 1,000 speakers remaining
  - age distribution of the speakers
  - trends in population size
  - use of a second language in the community
  - use of the language as a second language
  - attitudes to the language

Examples of endangered languages

Ethnologue (2018) estimates the number of endangered languages at about a third of the languages listed in the resource.

- examples of languages classified as endangered
    - Tirahi (Afghanistan)
    - Urali (India)
    - Wymysorys (Poland) [vilamovština]
    - Krimchak (Ukraine) [krymčačtina]
    - Pipil (El Salvador)
    - Providencia Sign Language (Colombia)

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/guides/

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

- 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

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

Dialects: intelligibility and 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

A closer look: 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, Andalú) [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/mapa-nareci.jpg

Similarities and differences between languages

Languages of the world do share some features, but display differences at the same time.

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

- WHY are they similar?
- WHY do they differ?

Seeking explanations

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

Language contact

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

En.	Czech	Basque	Swahili	Maori
pepper	*pepř*	*piperra*	*pilipili*	*pepa*

"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)

Classification of languages

Similarities and differences among languages explained

- by the origin of languages
  - $\rightarrow$ **genetic classification**
- by language contact
  - $\rightarrow$ **areal classification**

Both perspectives are taken into account within classification based on linguistic analysis of structural features of languages

$\rightarrow$ **linguistic typology**

Genetic / genealogical classification of lang.

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

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

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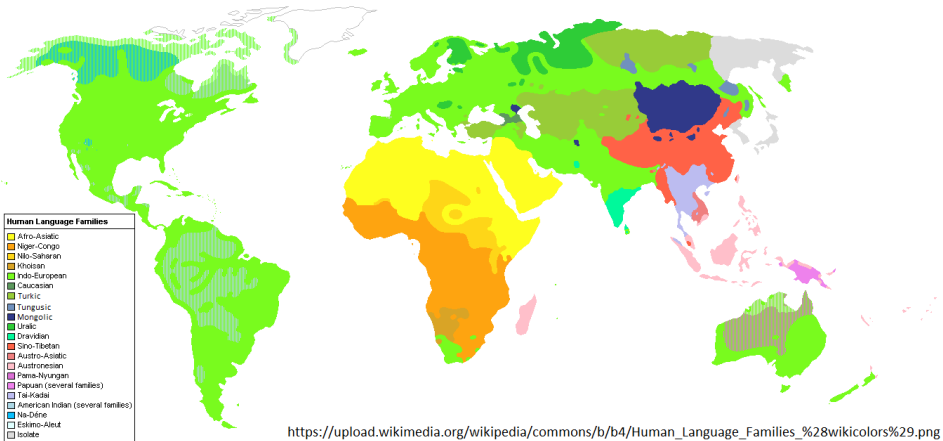is a group of languages that have developed from the same proto-language.

- language families represented as trees

Major language families

lang. family	lang.	mil. 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 146 language families

Human Language Families

- Afro-Asiatic
- Niger-Congo
- Nilo-Saharan
- Khoisan
- Indo-European
- Caucasian
- Turkic
- Tungusic
- Mongolic
- Uralic
- Dravidian
- Sino-Tibetan
- Austro-Asiatic
- Austronesian
- Pama-Nyungan
- Papuan (several families)
- Tai-Kadai
- American Indian (several families)
- Na-Déne
- Eskimo-Aleut
- Isolate

https://upload.wikimedia.org/wikipedia/commons/b/b4/Human_Language_Families_%28wikicolors%29.png

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
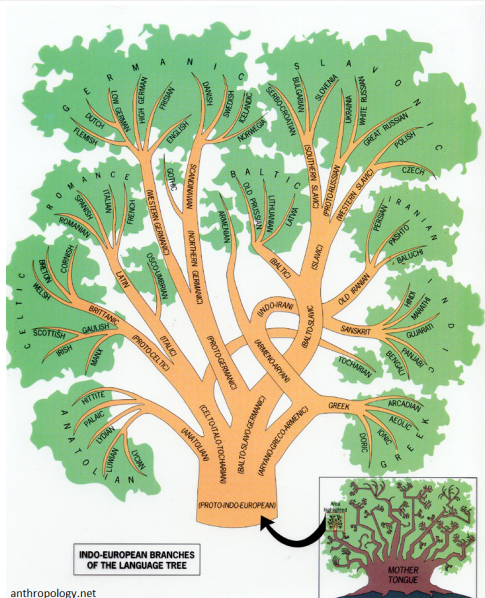
INDO-EUROPEAN BRANCHES OF THE LANGUAGE TREE

anthropology.net

MOTHER TONGUE

Sub-classification of language families

Language families are subdivided into smaller groups. However, there is consensus neither on granularity (number of levels) nor on terminology.

- for instance:

  language family

  — language group (branch)

  —— language subgroup (subbranch)

  ——— individual language

Sub-classification example

- from Indo-European family to Czech:
  - the Indo-European language family subsumes several groups, e.g. Germanic, Italic, Indo-Iranian, **Slavic** languages
  - the group of Slavic languages is diveded into the subgroups of East, South, and **West** Slavic languages
  - the subgroup of West Slavic Languages involves **Czech**, Slovak, Polish etc.
- the other way round:
  - —— *individual language*: Czech
  - —— *language subgroup*: West Slavic language
  - — *language group*: Slavic language
  - *language family*: Indo-European

Slavic languages

Language isolates

- natural languages that cannot be assigned to any language family
- assumed to be remnants of families
- e.g. Basque, Korean, Japanese

Areal classification

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

Sprachbund

is a group of languages that are spoken in the same region and have developed similar features.

Greek, Bulgarian, Macedonian, Romanian, Albanian, Turkish etc.

- e.g. in spite of being genetically not related, the languages have developed postponement of articles as a common feature:

language	'woman' – '**the** woman'	'man' – '**the** man'
Bulgarian	жена – жена**та**	мъж – мъж**ът**
Romanian	femeie – fememi**a**	bărbat – bărbat**ul**
Albanian	grua – grua**ja**	burrë – burr**i**

Linguistic typology

- Typological classification is based on the study of structural and functional features of languages.
  - phonological features
  - morphological features
  - syntactic features
  - word formation
  - lexicon

Linguistic typology: Phonology

- Typological approach to phonology covers the following questions:
    - 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?

- Typological approach to morphology includes, for instance, the following questions:
  - Are grammatical categories conveyed by morphemes, auxiliaries, and/or by word order?
  - How many morphological cases are available in the language?

- Typological approach to syntax:
  - 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

Linguistic typology: Word formation

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

Linguistic typology: Lexicon

- Examples of questions asked:
    - How are the same objects or actions named in particular languages?
    - – esp. body parts, colour terms, weather terms, motion verbs
    - What meanings can be expressed by a single word in the language?

1. Intro
   - Number of the world's languages
   - Living language
   - Languages by region of origin and number of speakers
2. Approaches to classification of the world's languages
   - Genetic classification
   - Areal classification
   - Linguistic typology
3. Linguistic resources on the world's languages
   - *Ethnologue: Languages of the World*
   - *WALS – The World Atlas of Language Structures*

- linguistic description of languages for cross-linguistic research available
    - in printed publications
      $\rightarrow$ 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.
    - `http://www.linguistic-typology.org/resources.html`

- 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 lang. 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

- 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

Ethnologue	WALS
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

- 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
  - current edition: 21st edition 2018
    http://www.ethnologue.com

Ethnologue: Languages of the World (2018)

- 7,097 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

Dutch

LANGUAGE | MAP | FEEDBACK

A language of Netherlands

ISO 639-3	nld
Autonym	Hollands, Nederlands
Population	16,000,000 (European Commission 2012). Total users in all countries: 23,025,430.
Language Maps	Belgium, Luxembourg and Netherlands
Language Status	1 (National). De facto national language.
Classification	Indo-European, Germanic, West, Low Saxon-Low Franconian, Low Franconian
Dialects	Northern North Hollandish (Westfries). Belgian variety of Dutch (not Vlaams [vls]) 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 [eng] (European Commission 2006), Standard German [deu] (European Commission 2006). Some also use French [fra] (European Commission 2006).
Language Development	Fully developed. Bible: 1522–2004.
Language Resources	OLAC resources in and about Dutch
Writing	Braille script [Brai], used since 1951. Latin script [Latn], 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
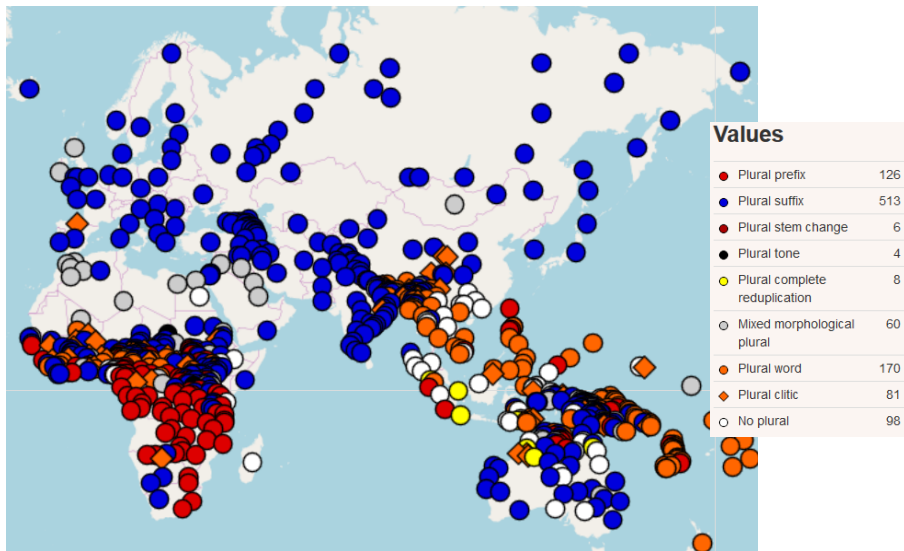 - Greville G. Corbett, Martin Haspelmath, Bernard Comrie, Matthew S. Dryer et al.

WALS editions

- 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

Values

● Plural prefix	126
● Plural suffix	513
● Plural stem change	6
● Plural tone	4
● Plural complete reduplication	8
● Mixed morphological plural	60
● Plural word	170
◆ Plural clitic	81
○ No plural	98

- 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

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

- Körtvélyessy, L. (2017): *Essentials of Language Typology*. Košice: UPJŠ. Available online:
  https://unibook.upjs.sk/sk/anglistika-a-amerikanistika/365-essentials-of-language-typology

- 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`