Variability of languages in time and space

Lecture 1:
Basic facts and linguistic resources on the world’s languages

Magda Ševčíková

September 30, 2020
Basic facts

1. Number of the world’s languages
2. Living language
3. Languages by region of origin and number of speakers

Linguistic resources on the world’s languages

1. Ethnologue: Languages of the World
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*
Printed and online resources differ considerably in estimating the number of languages:

- *Encyclopedia Britannica* (1911): app. 1,000 lang.

– 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* (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:

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

- **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

Europe (yellow): 288 languages
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

<table>
<thead>
<tr>
<th>rank</th>
<th>as L1</th>
<th># of speakers</th>
<th>as L2</th>
<th># of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chinese</td>
<td>1,299 mil.</td>
<td>English</td>
<td>660 mil.</td>
</tr>
<tr>
<td>2</td>
<td>Spanish</td>
<td>442 mil.</td>
<td>Hindi</td>
<td>215 mil.</td>
</tr>
<tr>
<td>3</td>
<td>English</td>
<td>378 mil.</td>
<td>Malay</td>
<td>204 mil.</td>
</tr>
<tr>
<td>5</td>
<td>Hindi</td>
<td>260 mil.</td>
<td>French</td>
<td>153 mil.</td>
</tr>
</tbody>
</table>
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
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
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
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 [mucijština]
- 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
     - jihozápadocheská 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
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*
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

sevcikova@ufal.mff.cuni.cz  Variability of languages
Online resources on the world’s languages

- **Ethnologue: Languages of the World**
  - [http://www.ethnologue.com](http://www.ethnologue.com)

  - [http://wals.info](http://wals.info)

- **APiCS – Atlas of Pidgin and Creole Language Structures**
  - [http://apics-online.info](http://apics-online.info)

- **WOLD – The World Loanword Database**
  - [http://wold.clld.org](http://wold.clld.org)

- For other resources see e.g.
  - [https://linguistic-typology.org/databases/](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](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
# Dutch

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) 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 [en] (European Commission 2006), Standard German [deu] (European Commission 2006). Some also use French [fra] (European Commission 2006).


**Language Resources:** OLAC resources in and about Dutch

**Writing:** Braille script [Brail], used since 1951. Latin script [Latn], primary usage.

- 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](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:

<table>
<thead>
<tr>
<th>phonology</th>
<th>nominal syntax</th>
<th>lexicon</th>
</tr>
</thead>
<tbody>
<tr>
<td>morphology</td>
<td>simple clauses</td>
<td>sign languages</td>
</tr>
<tr>
<td>nominal categories</td>
<td>complex sentences</td>
<td>other</td>
</tr>
<tr>
<td>verbal categories</td>
<td>word order</td>
<td></td>
</tr>
</tbody>
</table>

- distribution of the feature values on the map
  >>>> see ex. on the next slide
WALS feature 33A: Coding of Nominal Plurality

Variability of languages
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
## Language Dutch

**WALS code:** dut

**Variability of languages**

### Showing 1 to 89 of 89 entries

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

### Alternative names

- **Ruhlen:** Dutch
- **Ethnologue:** Dutch

### Sources

- Kager 1989
- Trommelen and Zonneveld 1999a
- van der Hulst 1984
different approaches in individual resources
here, an example comparison of genetic classification with a focus on **Slavic languages** in

- *Ethnologue: Languages of the World*
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

<table>
<thead>
<tr>
<th>Ethnologue</th>
<th>WALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belarusian</td>
<td>Belorussian</td>
</tr>
<tr>
<td>Russian</td>
<td>Russian</td>
</tr>
<tr>
<td>Ukrainian</td>
<td>Ukrainian</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Bulgarian</td>
</tr>
<tr>
<td>Macedonian</td>
<td>Macedonian</td>
</tr>
<tr>
<td>Bosnian</td>
<td>Bosnian</td>
</tr>
<tr>
<td>Serbian</td>
<td>Serbian-Croatian</td>
</tr>
<tr>
<td>Slovene</td>
<td>Slovene</td>
</tr>
<tr>
<td>Czech</td>
<td>Czech</td>
</tr>
<tr>
<td>Slovak</td>
<td>Slovak</td>
</tr>
<tr>
<td>Kashubian</td>
<td>Kashubian</td>
</tr>
<tr>
<td>Polish</td>
<td>Polish</td>
</tr>
<tr>
<td>Lower Sorbian</td>
<td>Lower Sorbian</td>
</tr>
<tr>
<td>Upper Sorbian</td>
<td>Upper Sorbian</td>
</tr>
</tbody>
</table>

*Ethnologue only:*
- Croatian
- Montenegrin
- Church Slavonic
- Slavomolisano
- Silesian
- Rusyn

*WALS only:*
- Sorbian
- Polabian
- Slovincian
References

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