Introduction

- How many languages are spoken today?
- Ethnologue (2005): 6,912: table 1
- Number of speakers varies substantially
How many languages?

many languages in Europe
How many languages?

ca. 150 languages in Europe, 40 the Caucasus alone
Around 7,000 languages world-wide

high diversity around the equator
(data from 1999 edition of Ethnologue)
<table>
<thead>
<tr>
<th>language</th>
<th>number of native speakers (Mill.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mandarin</td>
<td>873</td>
</tr>
<tr>
<td>Spanish</td>
<td>322</td>
</tr>
<tr>
<td>English</td>
<td>309</td>
</tr>
<tr>
<td>Hindi</td>
<td>181</td>
</tr>
<tr>
<td>Portuguese</td>
<td>177</td>
</tr>
<tr>
<td>Bengali</td>
<td>171</td>
</tr>
<tr>
<td>Russian</td>
<td>145</td>
</tr>
<tr>
<td>Japanese</td>
<td>122</td>
</tr>
<tr>
<td>German</td>
<td>95</td>
</tr>
<tr>
<td>Wu (China)</td>
<td>77</td>
</tr>
<tr>
<td>Javanese</td>
<td>76</td>
</tr>
<tr>
<td>Telugu</td>
<td>70</td>
</tr>
<tr>
<td>Marathi</td>
<td>68</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>67</td>
</tr>
<tr>
<td>Korean</td>
<td>67</td>
</tr>
<tr>
<td>Tamil</td>
<td>66</td>
</tr>
<tr>
<td>French</td>
<td>65</td>
</tr>
<tr>
<td>Urdu</td>
<td>61</td>
</tr>
<tr>
<td>Yue (Kantonese)</td>
<td>55</td>
</tr>
<tr>
<td>Turkish</td>
<td>51</td>
</tr>
</tbody>
</table>

More recent data source
Quantitative distribution
Quantitative distribution

- Zipfian distribution
- Number of speakers is inversely proportional to rank of a language
- Frequent distribution in linguistics/social sciences
## Language diversity in past, present, future

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,000 BCE</td>
<td>20,000 languages</td>
</tr>
<tr>
<td>1000 CE</td>
<td>9,000 languages</td>
</tr>
<tr>
<td>1500</td>
<td>7,500 languages</td>
</tr>
<tr>
<td>2000</td>
<td>6,500 languages</td>
</tr>
<tr>
<td>2050</td>
<td>4,500 languages</td>
</tr>
<tr>
<td>2100</td>
<td>3,000 languages</td>
</tr>
<tr>
<td>2200</td>
<td>100 languages</td>
</tr>
</tbody>
</table>

source: Martin Haspelmath
What counts as „speaker“?

- 1996 edition of Ethnologue: 266 million speaker of Spanish
- 1999 edition: 322 million
- Does not correspond to population growth
- Data sources are sometimes unreliable
What counts as a language?

- Arabic does not belong to „top twenty“
  - Arabic (including all variants): 202 mill. speaker (would amount to 4th rank)
  - Ethnologue treats different variants of Arabic as different languages
  - Justification: variants are mutually unintelligible. Algerian and Egyptian Arabic are as different as Spanish and Portuguese.
What counts as a language?

- Hindi and Urdu are the same language
  - History/politics: different writing systems, different strata of loan words
  - Regular speakers understand each other fairly well
  - If counted as one language, Hindi/Urdu would be on 4th place.
What counts as a language?

- Depending on how you count, Turkish might have higher number of speakers
  - 51 millionen speakers (46 million in Turkey)
  - However, more than 80 million people speak a language that is mutually intelligible with Turkish
  - Counting them in would bring Turkish to 10\textsuperscript{th} rank
What counts as a language?

- Serbo-Croatian
  - Before Balkan wars of the nineties:
    - Serbo-Croatian counted as one language
    - Two writing systems – Latin alphabet in Croatia, kyrillic alphabet in Serbia
    - Continuum of dialectal variants
  - Now:
    - Three languages – Serbian, Croatia, Bosnian
What counts as a language?

- **Scandinavian**
  - Norwegian and Swedish – and, up to a point, also Danish, are mutually intelligible
  - Count as different languages though, because they are associated with different countries
What counts as a language?

- **Chinese**
  - Is frequently considered a single language
  - Consists of at least seven different languages (with considerable internal dialectal variation)
  - Chinese is considered as a unit for cultural and political reasons, like the common writing system
What counts as a language?

- Chinese
What counts as a language

• Dialect continua
  – Portugese, Spanish, French and Italian are counted as different languages
  – Nonetheless, local dialects changes only gradually if you travel from town to town from Portugal to Italy.
  – The same holds for German and Dutch.
What counts as a language
What counts as a language

- Cynically speaking: A *language is a dialect with an army and a navy*.
- Distinction between language and dialect cannot be done by purely linguistic criteria
- In the end, it is a political and cultural decision of a linguistic community about its identity
- *Criteria from Ethnologue*
Language families

• Languages: no clearly separated unites, rather a hierarchy/tree structure.
  - Categories can be split into ever smaller units, until the level of the single speaker
  - Assumption of a meta-unit is justified if there is evidence for a common origin
Language families

- German belongs to the family of Indo-European
- Sometimes also called (obsolete now) „Indo-Germanic“
- It is the language family that was discovered first and is best studied
The Indo-European language family

• Ancient times: little interest in comparative linguistic research

• Middle ages:
  – Written documents from many European languages
  – Wide-spread assumption that all languages originate from Hebrew
  – No real concept of language change

• Real starting point of comparative linguistics was the discovery of Sanskrit
The Indo-European language family

- William Jones 1786:
  „The Sanskrit Language, whatever be its antiquity, is of wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either; yet bearing to both of them a stronger affinity both in the roots of verbs and the forms of grammar, than could possibly have been produced by accident; so strong indeed that no philologer could examine them at all without believing them to have sprung from some common source, which perhaps no longer exists: there is similar reason, so not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a different idiom, had the same origin with the Sanskrit; and the old Persian might be added to the same family, if this were the place for discussing any question concerning the antiquities of Persia.“
The Indo-European language family

- Cœurdoux 1767

<table>
<thead>
<tr>
<th>Sanskrit</th>
<th>devah</th>
<th>„god“</th>
<th>Latin</th>
<th>deus</th>
<th>Greek</th>
<th>theós</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>padam</td>
<td>„foot“</td>
<td></td>
<td>pes, ped-is</td>
<td></td>
<td>poús, podo-ós</td>
</tr>
<tr>
<td></td>
<td>maha</td>
<td>„large“</td>
<td></td>
<td></td>
<td>mégas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>viduva</td>
<td>„widow“</td>
<td></td>
<td>viduva</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Also grammatical similarity between Greek and Sanskrit

- Partially incorrect according to modern insights (for instance, the Greek cognate to lat. deus is Zeus, not theos)
## The Indo-European language family

<table>
<thead>
<tr>
<th>Sanskrit</th>
<th>Latin</th>
</tr>
</thead>
<tbody>
<tr>
<td>as-mi</td>
<td>I am</td>
</tr>
<tr>
<td>as-i</td>
<td>you(sg.) are</td>
</tr>
<tr>
<td>as-ti</td>
<td>he is</td>
</tr>
<tr>
<td>s-mas</td>
<td>we are</td>
</tr>
<tr>
<td>s-tha</td>
<td>you(pl) are</td>
</tr>
<tr>
<td>s-anti</td>
<td>they are</td>
</tr>
</tbody>
</table>
The Indo-European language family

• Sanskrit as- and lat. es- both mean „to be“
• Both have allomorph s-
• Inflectional paradigm comprises both variants
• Sanskrit has additional suffix -i; otherwise the suffixes are virtually identical

→ Sufficient evidence to establish genetic relatedness
The Indo-European language family

- Reconstructed paradigm of the Indo-European proto language

\[(V)s-(V)m(i)\]
\[Vs-(i)\]
\[Vs-t(i)\]
\[s-(V)mVs\]
\[s-(V)mVt\]
\[s-Vnt(i)\]
The Indo-European language family

- Middle of 19\textsuperscript{th} century: discovery of sound laws
- Phonological change is not arbitrary, but applies essentially to all words of a language
- For instance \textit{Grimm's Law} (applies to all Germanic languages), \textit{High German consonant shift} (applies to all High-German dialects)
Sound laws and the reconstruction of language families

- Applicable to other languages as well (example from Austronesian)
- Reconstruction is usually possible at most until 8,000 years into the past
The Indo-European language family

- Modern Indo-European languages are
  - All European languages except Hungarian, Finnish, Estonian, and Basque
  - Many West Asian and South Asian languages
Branches of the IE family
Branches of the IE family

• 8 living branches
  – Celtic
  – Germanic
  – Romance
  – Balto-Slavic
  – Greek
  – Albanian
  – Indo-Iranian
  – Armenian

• 2 well-documented extinct branches
  – Tocharian
  – Anatolian

• several poorly documented extinct branches
Branches of the IE family

- **Indo-Iranian**
  - **Indo-Aryan**: Sanskrit, Hindi, Urdu, Bengali, Marathi, Sinhala, …
  - **Iranian**: Avestan, ancient Persian (cuneiform documents), Farsi, Pashto, Kurdish, Balochi, …
  - **Nuristani**: Kati, Prasuni, Ashkunu, Waigali, Gambiri, … (small languages, mostly spoken in Pakistan/Afghanistan)
Branches of the IE family

• Armenian:
  – Old Armenian, Eastern Armenian, Western Armenian
Branches of the IE family

• Balto-Slavic:
  – Slavic:
    • **East Slavic:** Russian, Belarussian, Ukrainian, Ruthenian
    • **West Slavic:** Sorbian (Upper Sorbian, Lower Sorbian), Polabian (extinct), Polish, Pomeranian (Kashubian, Slovincian (extinct)), Czech, Slovak
    • **South Slavic:** Burgenland Croatian, Bosnian, Croatian, Molise Croatian, Macedonian, Montenegrin, Serbian, Slovenian
Branches of the IE family

• Balto-Slavic:
  – Baltic:
    • **Eastern Baltic**: Lithuanian, Latvian, Curonian, Selonian (extinct), Semigallian (extinct)
    • **Western Baltic (extinct)**: Old Prussian, Sudovian, Galindian, Skalvian
Branches of the IE family

• **Celtic:**
  
  – **Continental Celtic (extinct):** Gaulish, Galatian, Lepontian, Celtiberian
  
  – **Insular Celtic:**
    
    • British languages: Cumbric (extinct), Welsh, Cornish (extinct), Breton
    
    • Goidelic languages: Irish, Scottish Gaelic, Manx
Branches of the IE family

- **Germanic:**
  - **East Germanic** (extinct): Burgundian, Vandalic, Gothic
  - **North Germanic:** Norwegian, Faroese, Jamtlandic, Norn (extinct), Swedish, Danish, Gutnish
  - **West Germanic:** English, Scots, Frisian, Dutch, Low German, German, Swiss German, Yiddish, ...
Branches of the IE family

• Romance (Italic):
  – Latino-Faliscan: Latin (extinct), Faliscan (extinct), Spanish, Portuguese, French, Italian, Romanian, Moldovan, Catalan, Galician, Occitan, Sardinian, Ladin, Romansh
  – Osco-Umbrian (extinct)
Branches of the IE family

- Greek
- Albanian
- Illyric (extinct)
- Venetic (extinct)
- Lusitanian (extinct)
Branches of the IE family

• **Tocharian** (extinct):
  – Was spoken in second half of the first millennium in present day China
  – About 5,000 written documents survive
Branches of the IE family

- **Anatolian languages** (extinct):
  - Hittite, Lydian, Palaic, Luwian, Lycian, Carian, Pisidian, Sidetic

- **Phrygian** (extinct)

- **Thracian** (extinct)

- **Macedonian** (extinct; was spoken during antiquity, unrelated to modern Macedonian, which is a Slavic language)
Language families

- Language family: group of genetically (i.e. historically) related languages
- Descent from a common proto-language
- Descent has to be established via generally accepted methods
- Classification is (unavoidably) variable and sometimes subjective
- Ethnologue counts more than 100 language families
Language families
Language families

- **Afro-Asiatic**
  - Also called „Hamito-Semitic“ (obsolete)
  - subgroups:
    - Semitic (Arabic, Hebrew, Amharic, ...)
    - Berber (Tuareg, ...)
    - Egyptian (extinct)
    - Cushitic (Somali, Oromo, ...)
    - Chadic (Hausa, ...)
Language families

- **Nilo-Saharan**
  - Comprises about 200 African languages
  - Nubian, Fur, ...
Language families

- **Niger-Congo language**
  - Most important subgroup: Bantu languages
  - Swahili, Rwanda, Zulu, Yoruba
Language families

- **Khoisan languages**
  - Languages of the bushmen in Southern Africa
  - Use click sounds (which are typologically uncommon)
Language families

- **Uralic**
  - subgroups
    - Finno-ugric: Hungarian, Estonian, Sami, Karelian
    - Samoyedic (< 30,000 speaker in Northern Eurasia)
Language families

- **Altaic**
  - subgroups
    - Turkic: Turkish, Turkmen, Kyrgyz, Kazakh
    - Mongolic
    - Tungusic (Northern China, East Siberia)
    - Korean
    - Japanese
  - Partially controversial, especially the inclusion of Korean and Japanese
Language families

• Dravidian
  – Telugu, Tamil, Kannada, ...
  – Spoken mainly in Southern India and Sri Lanka
Language families

• **Sino-Tibetan**
  – subgroups
    • Sinitic (Chinese languages)
    • Tibeto-Burman (spoken in Myanmar, Northern Thailand, Nepal, Bhutan, parts of China, India and Pakistan): Tibetan, Brahmaputran, ...
Language families

- **Austro-Asiatic**
  - Vietnamese, Khmer, Santali
  - Spoken in South-East Asia and Northern India
Language families

• **Austronesian**
  - Family with the largest geographical expansion (from Madagascar in the West until Hawaii in the East)
  - Malagasy, Javanese, Bahasa Indonesian, Tagalog, Taiwanese languages, Maori (language of the aborigines of New Zealand), polynesian languages, ...
Language families

- **Tai-Kadai languages**
  - Thai, Isan, Lao, ...
  - Speculations, that Austronesian and Tai-Kadai form a single family („Austro-Thai“)
Paleo-American language families

- Classification according to Greenberg:
  - Eskimo-Aleut
  - Na-Dene (Northern and Western North-America)
  - Amerindian (rest of North-America and South-America)
- „Amerindian“ is heavily contested
- Using traditional methods, only many much smaller families can be established
Language families

• In many cases, it is impossible to come up with a clear classification
  – 700 languages in Papua-New Guinea, often unrelated to each other
  – Several hundred languages of Australian aborigines; genetic classification is unclear
  – Many „isolated“ language (i.e. no genetic relationship to any other language can be established), for instance Basque
Language families

Number of languages per family also follow Zipfian distribution.