THE BIG PERSPECTIVE

Mr Joseph Gavagai
- the typological bunny

Hedvig Skirgård
Young Scientists Stockholm
hedvig@lingolympiad.com
Who am I?

- Project leader at Unga Forskare Stockholm (Young Scientists Stockholm)
- Assistant at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands
- Master in Linguistics from Stockholm University (candiatet 3-4 yrs, MA 1-2 yrs, PhD >3-4 yrs)
- Focus on general linguistics, grammatical typology, contact linguistics, complexity
!QUIZ!

172 languages are spoken by more than 3.1 million each.
8 by more than 100 millions each

Guess the top-8!
Answer at the end of presentation

NB only native speakers
What is linguistics?

Scientific study of language as a phenomenon

Object of study

Ideally:

natural, spoken, informally

Often:

natural, written, formally

Comptutational

Field work

Psycholinguistics

“General”
What is language typology?
What is linguistic typology?

- classifications of languages into types
- cross-linguistic comparison

in order to find:
- patterns due to contact, genealogy and/or implicational biases
- Possible leads to mechanisms behind language change and the limits of language

What are the cognitive, areal, cultural, communicative and/or historical constraints on the language diversity, complexity and change?
[the aim of linguistic theory] must be to show which structures are possible, in general, and why it is just those structures, and not others, that are possible.

(Hjelmslev 1970 [1963]: 96)
Important to remember!

All languages *can* express all concepts

*grammar [...] determines those aspects of each experience that *must* be expressed*

(Boas 1938:132)
Other “truths”

However, for most researchers that in their research assume that the opposite is true these distinctions often don’t matter – i.e. they’re most often two small to be relevant. We think...
An example of a typological study
What is a language?
• Human & natural (i.e. not mathematics, codes, Klingon or Python)
• Linguistic criteria: mutual intelligibility (to a certain extent)
• Political issue rather than linguistic

Example

Danish: mit luftpudelfartøj er fyldt med ål
Swedish: min svävare är full av ål
Norwegian: min luftputefartøy er full av ål
Icelandic: svifnökkvinn minn er fullur af álum
Faroese: luftpútufar mítt er fult í álli!
(omniglot.com)
What is a language? cont.
You’re welcome to ask questions, nothing is too dumb! However, please remember to visit Ethnologue.com, Glottolog.org, WALS.info and the other sites on your own ^_^!
There are approximately 5,000–9,000 languages in the world today. (Ethnologue.com, 2014)

Summer Institute of Linguistics reports 7,105 languages in the latest edition.


How many languages do you think there are in the world today, approximately?

Want clues ^^?

- 193 member states of UN
- 203 states recognized by at least one UN member state
- 249 entities of ISO 3166-1
- +7 billion humans on Tellus
- We’ve had languages for at least 100,000 years
- (We’re not sure if language arose in one place or several)
- The distribution of speaker populations are Zipfian
Handy terms to deal with some of these problems

Ethnologue
- Language family
- (Macro language)
- language
- dialect

Cysouw & Good (2003)
- Languoid
  - language varieties (sociolect, dialect, etc)
- glossonym
- doculect
<table>
<thead>
<tr>
<th>Range</th>
<th># lgs</th>
<th>% of all lgs</th>
<th># speakers</th>
<th>% of all speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>100,000,000 to 999,999,999</td>
<td>8</td>
<td>0.1%</td>
<td>2,543,460,358</td>
<td>40.4%</td>
</tr>
<tr>
<td>10,000,000 to 99,999,999</td>
<td>80</td>
<td>1.1%</td>
<td>2,458,383,987</td>
<td>39.0%</td>
</tr>
<tr>
<td>1,000,000 to 9,999,999</td>
<td>305</td>
<td>4.3%</td>
<td>929,591,638</td>
<td>14.8%</td>
</tr>
<tr>
<td>100,000 to 999,999</td>
<td>937</td>
<td>13.2%</td>
<td>294,626,823</td>
<td>4.7%</td>
</tr>
<tr>
<td>10,000 to 99,999</td>
<td>1,811</td>
<td>25.5%</td>
<td>61,556,414</td>
<td>1.0%</td>
</tr>
<tr>
<td>1,000 to 9,999</td>
<td>1,978</td>
<td>27.8%</td>
<td>7,613,358</td>
<td>0.1%</td>
</tr>
<tr>
<td>100 to 99</td>
<td>1,062</td>
<td>14.9%</td>
<td>466,128</td>
<td>0.0%</td>
</tr>
<tr>
<td>10 to 9</td>
<td>338</td>
<td>4.8%</td>
<td>12,944</td>
<td>0.0%</td>
</tr>
<tr>
<td>1 to 9</td>
<td>137</td>
<td>1.9%</td>
<td>541</td>
<td>0.0%</td>
</tr>
<tr>
<td>0</td>
<td>204</td>
<td>2.9%</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Unknown</td>
<td>246</td>
<td>3.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>7,106</td>
<td>100%</td>
<td>6,295,712,191</td>
<td>100%</td>
</tr>
</tbody>
</table>
Useful sites

• Ethnologue
• WALS
• MultiTree
• Omniglot
• Glottolog
• Phoible
• The Universals Archive
How do typology?

- Define a variable to be investigated
- sample languages
- go through the sources
- realize that the original variable is ill-defined
- become enough (lagom) satisfied and expand sample
## Databases of structural features

<table>
<thead>
<tr>
<th>Database</th>
<th># languages</th>
<th># features</th>
<th>datapoints</th>
<th>features with over 200 lgs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic Structures of the World's Languages (2009)</td>
<td>237</td>
<td>93</td>
<td>14 440</td>
<td>?</td>
</tr>
<tr>
<td>Nijmegen Typological Survey (2014)</td>
<td>274</td>
<td>279</td>
<td>48 435</td>
<td>145</td>
</tr>
<tr>
<td>Atlas of Pidgin and Creole Language Structures (2013)</td>
<td>76</td>
<td>130</td>
<td>TBA</td>
<td>N/A</td>
</tr>
<tr>
<td>Valency Patterns Leipzig Online Database (2013)</td>
<td>36</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Phonetics Information Base and Lexicon (2012)</td>
<td>1 010</td>
<td>1 680</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>South American Indigenous Language Structures (SAILS)</td>
<td>167</td>
<td>604</td>
<td>31794</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Sampling

- All languages: 5-9 000
- Unbiased with respect to genealogy or contact
How many language families are there?

Nichols (1992) \( \approx 300 \)
Campbell (2009) 250-300
WALS (2011) 212
### Language families cont.

**Genus**

Nichols (2003) about the cognate method: group of languages with a shared history at least 3500 – 4000 years ago

<table>
<thead>
<tr>
<th>Language Family</th>
<th>Approx Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto-Uraliska:</td>
<td>approx 7000-2000 BC</td>
</tr>
<tr>
<td>Proto-Indo-Europeiska:</td>
<td>approx 3700 BC</td>
</tr>
<tr>
<td>Proto-Dravidiska:</td>
<td>approx 500 BC</td>
</tr>
</tbody>
</table>

**Term from WALS**

4000-8000 BC

WALS includes 510 genera, e.g., slavic, germanic, semitic, bantoid,

- Slavic
- Germanic
- Semitic
- Bantu
More on language families

6 language families include 63% of all languages and are spoken by 85% of the world's population.

<table>
<thead>
<tr>
<th>Language families</th>
<th># languages</th>
<th>% of world population of languages</th>
<th># speakers</th>
<th>% of world population of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger-Congo</td>
<td>1,526</td>
<td>21.47%</td>
<td>435,432,213</td>
<td>6.92%</td>
</tr>
<tr>
<td>Austronesian</td>
<td>1,222</td>
<td>17.20%</td>
<td>346,489,508</td>
<td>5.50%</td>
</tr>
<tr>
<td>Trans-New-Guinea</td>
<td>476</td>
<td>6.70%</td>
<td>3,540,024</td>
<td>0.06%</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>455</td>
<td>6.40%</td>
<td>1,268,218,984</td>
<td>20.14%</td>
</tr>
<tr>
<td>Indo-European</td>
<td>437</td>
<td>6.15%</td>
<td>2,925,253,210</td>
<td>46.46%</td>
</tr>
<tr>
<td>Afro-Asiatic</td>
<td>367</td>
<td>5.16%</td>
<td>374,573,409</td>
<td>5.95%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,483</strong></td>
<td><strong>63.09%</strong></td>
<td><strong>5,353,507,348</strong></td>
<td><strong>85.03%</strong></td>
</tr>
</tbody>
</table>
How many isolates?

An isolate is a language with no known relatives.

Ethnologue (2014) 74
Glottolog (2014) 189
Campbell (2009) 129

What should we do about “very small families”?

Isolate
Basque
Korean
Burushaski

Language family A
Language family B

Campbell (2009)

74
189
129
## Sampling cont: geography

<table>
<thead>
<tr>
<th>Macroarea</th>
<th>Living languages</th>
<th>Speaker population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent of total</td>
</tr>
<tr>
<td>Africa</td>
<td>2,146</td>
<td>30.2 %</td>
</tr>
<tr>
<td>Americas</td>
<td>1,060</td>
<td>14.9 %</td>
</tr>
<tr>
<td>Asia</td>
<td>2,303</td>
<td>32.4 %</td>
</tr>
<tr>
<td>Europe</td>
<td>285</td>
<td>4.0 %</td>
</tr>
<tr>
<td>Pacific</td>
<td>1,312</td>
<td>18.5 %</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7,106</strong></td>
<td><strong>100 %</strong></td>
</tr>
</tbody>
</table>

*Lewis (2009)*
### Participating countries in IOL 2014

<table>
<thead>
<tr>
<th>Country</th>
<th># Lgs</th>
<th>GDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>449 (442)</td>
<td>0.913</td>
</tr>
<tr>
<td>Pakistan</td>
<td>76 (71)</td>
<td>0.795</td>
</tr>
<tr>
<td>Singapore</td>
<td>32 (24)</td>
<td>0.756</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>44 (41)</td>
<td>0.699</td>
</tr>
<tr>
<td>Latvia</td>
<td>14 (7)</td>
<td>0.628</td>
</tr>
<tr>
<td>Canada</td>
<td>172 (88)</td>
<td>0.601</td>
</tr>
<tr>
<td>China</td>
<td>301 (298)</td>
<td>0.511</td>
</tr>
<tr>
<td>Taiwan</td>
<td>26 (21)</td>
<td>0.489</td>
</tr>
<tr>
<td>Estonia</td>
<td>10 (4)</td>
<td>0.445</td>
</tr>
<tr>
<td>Ukraine</td>
<td>42 (24)</td>
<td>0.420</td>
</tr>
<tr>
<td>Turkmenistan</td>
<td>27 (4)</td>
<td>0.385</td>
</tr>
<tr>
<td>United States</td>
<td>420 (215)</td>
<td>0.334</td>
</tr>
<tr>
<td>Netherlands</td>
<td>38 (15)</td>
<td>0.301</td>
</tr>
<tr>
<td>Spain</td>
<td>22 (15)</td>
<td>0.277</td>
</tr>
<tr>
<td>Russia</td>
<td>137 (105)</td>
<td>0.251</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>17 (11)</td>
<td>0.217</td>
</tr>
<tr>
<td>Australia</td>
<td>244 (213)</td>
<td>0.211</td>
</tr>
<tr>
<td>Slovenia</td>
<td>10 (8)</td>
<td>0.160</td>
</tr>
<tr>
<td>Sweden</td>
<td>23 (13)</td>
<td>0.153</td>
</tr>
<tr>
<td>UK</td>
<td>55 (12)</td>
<td>0.144</td>
</tr>
<tr>
<td>Greece</td>
<td>23 (16)</td>
<td>0.134</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>18 (10)</td>
<td>0.096</td>
</tr>
<tr>
<td>Hungary</td>
<td>17 (15)</td>
<td>0.086</td>
</tr>
<tr>
<td>Ireland</td>
<td>5 (5)</td>
<td>0.067</td>
</tr>
<tr>
<td>Poland</td>
<td>19 (14)</td>
<td>0.065</td>
</tr>
<tr>
<td>Brazil</td>
<td>224 (211)</td>
<td>0.054</td>
</tr>
<tr>
<td>Japan</td>
<td>13 (12)</td>
<td>0.036</td>
</tr>
<tr>
<td>Isle of Man</td>
<td>2 (2)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

GDI stands for the Greenbergs Diversity Index
High GDI = unlikely that two people in that country speak the same language natively.
Low GDI = likely
Universals

Joseph Greenberg (1915-2001)
• founder of modern typology
• genealogical classification

The Universals Archive
Absolute universals (297/2029)
“all languages have X”

Dependent universals (1476/2029)
“if a language has X, if also has Y”

Hierarchies
singular>plural>dual>trial

(Greenberg 1963, universalie 511)
Exemple of universals

Universal 1175 (Brown 1993)

IF there are terms for 'north' and/or 'south', THEN there are terms for 'east' and/or 'west'

Implicational hierarchy

Dryer (1992) VO languages tend to have prepositions and OV tends to have postpositions.

"Languages with normal SOV order are postpositional"
Chapter 83: Order of Verb and Object
av Matthew Dryer (2011)
How many do we know anything about?

Common size of typological investigations: 200-450 languages

<table>
<thead>
<tr>
<th>Type</th>
<th># Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; Grammar</td>
<td>2580</td>
</tr>
<tr>
<td>&lt; Grammar sketch</td>
<td>1161</td>
</tr>
<tr>
<td>&lt; Dictionary</td>
<td>212</td>
</tr>
<tr>
<td>&lt; Phonology</td>
<td>106</td>
</tr>
<tr>
<td>&lt; Text collection</td>
<td>26</td>
</tr>
<tr>
<td>&lt; Wordlist</td>
<td>1073</td>
</tr>
<tr>
<td>&lt; Ethnographic</td>
<td>78</td>
</tr>
<tr>
<td>&lt; Minimal</td>
<td>21</td>
</tr>
<tr>
<td>Unknown</td>
<td>263</td>
</tr>
<tr>
<td>All</td>
<td>5520</td>
</tr>
</tbody>
</table>

Glottolog
UNESCO
vulnerable 600
definitely endangered 646
severely endangered 527
critically endangered 577
extinct 231
**total** 2581
(Revitalized: 3. Cornish Livonian, Manx)
(Moseley 2010)
How many are dying?

“Critical limit”: 20 000 speakers

45% of the world's languages today have less than 20 000 speakers each

Reasonable estimation

>50% will die in 100 years or even
90% dead or endangered

6% of the population of the world today are speaking 90% of the languages

(Crystal 2000:1)
Why is language typology interesting?

- Same species=same brains = same capacity
- Yet, great diversity and disparity!
- Is something impossible?
- What is complexity?
- Is complexity even?
- What influences ”complexity”?
- What does this say about our cognitive capacity and categorize our reality?
  (cf psychology, anthropology, neurology and biology)
Answers to the quiz!
(Ethnologue 2014)

1. "Chinese" 848 millions
2. Spanish 414 millions
3. English 335 millions
4. Hindi 260 millions
5. "Arabic" 237 millions
6. Portuguese 203 millions
7. Bengali 193 millions
8. Russian 167 millions

BUT!

1. Mandarin 848 millions
2. Spanish 414 millions
3. English 335 millions
4. Hindi 260 millions
5. Portuguese 203 millions
6. Bengali 193 millions
7. Russian 167 millions
8. Japanese 122 millions
In summary

• Linguistics is a very broad field
• It’s tricky defining what a language is
• There are many languages, families, and speakers
• Few speakers speak a lot of languages
• There is a great variety of types
• We know about 2500 languages
• Most of them are dying :(
We’ve talked about a lot of things...

:) or :( ?
Hedvig Skirgård
Unga Forskare Stockholm
hedvig@lingolympiad.org
Campbell, Lyle (1998) Historical linguistics - an introduction
Campell, Lyle (unpublished) Language Isolates and Their History, or, What’s Weird, Anyway?
Referenser 2


Hammarström (2009). chapter 5: Counting Languages in Dialect Continua Using the Criterion of Mutual Intelligibility
Referenser 3


EXTRA-OST?
Kapitel 18: Absence of Common Consonants
av Ian Maddieson 2011 (LÄNK)

<table>
<thead>
<tr>
<th>Bilabial</th>
</tr>
</thead>
<tbody>
<tr>
<td>All present</td>
</tr>
<tr>
<td>No bilabials</td>
</tr>
<tr>
<td>No fricatives</td>
</tr>
<tr>
<td>No nasals</td>
</tr>
<tr>
<td>No bilabials or nasals</td>
</tr>
<tr>
<td>No fricatives or nasals</td>
</tr>
</tbody>
</table>

- All present: 503
- No bilabials: 4
- No fricatives: 48
- No nasals: 10
- No bilabials or nasals: 1
- No fricatives or nasals: 1
How many languages per country?

<table>
<thead>
<tr>
<th>Languages per country</th>
<th>Indigenous languages per country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 45.6</td>
<td>Mean 38.3</td>
</tr>
<tr>
<td>Median 18</td>
<td>Median 10</td>
</tr>
<tr>
<td>Max 830</td>
<td></td>
</tr>
<tr>
<td>(Papua New Guinea)</td>
<td></td>
</tr>
<tr>
<td>Min 1</td>
<td>40 countries have 100% indigenous</td>
</tr>
<tr>
<td>(St Helena, Vatican, No. Korea, Falklands &amp; British Indian Ocean Territory)</td>
<td>830/830 (Papua New Guinea)</td>
</tr>
</tbody>
</table>
How many has there ever been?

If we assume the following:

- At each point in time there is 5,000 languages
- Each language has a lifetime of 1,000 years
- We’ve spoken languages for at least 100,000 years

\[
\frac{100,000}{1,000} \times 5,000 = 500,000
\]

This means that <2% are alive today.

At least half a million!!
1 language = 1 name?

Ethnologue 2005 (Hammarström 2008)
39,419 unique names for languages
42,768 occurrences of these names

- mean 5.86 names per language
  - max the language Armenian has 77 names
- mean 1.08 languages per name
  - max the name Miao refers to 14 languages
### Kombowals

_(Länk)_

<table>
<thead>
<tr>
<th>129A</th>
<th>Identical (228)</th>
<th>Different (389)</th>
</tr>
</thead>
<tbody>
<tr>
<td>130A</td>
<td>Identical (72)</td>
<td>Different (521)</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td>57</td>
<td>309</td>
</tr>
</tbody>
</table>

129: hand & arm  
130: finger & hand  

Identical = samma ord  
Different = olika ord