
Bernard Comrie
comrie@eva.mpg.de

University of California Santa Barbara and
Max Planck Institute for Evolutionary Anthropology, Leipzig
The World Atlas of Language Structures (WALS)
edited by
Martin Haspelmath, Matthew S. Dryer, David Gil,
and Bernard Comrie

Interactive Reference Tool by
Hans-Jörg Bibiko


a five-year project at the Department of Linguistics,
Max Planck Institute for Evolutionary Anthropology
(MPI-EVA), Leipzig (1999-2004)
Genealogical (genetic) diversity

English belongs to the
  Germanic branch of the
  Indo-European language family.
Arabic belongs to the Semitic branch of the Afroasiatic language family.
Source: http://en.wikipedia.org/wiki/Afroasiatic_languages (2012 March 06)
Structural diversity
Tone

Chinese (Mandarin): 4 tones

mā   ‘mother’
má   ‘hemp’
mǎ   ‘horse’
mà   ‘reproach’
Gender

Spanish: 2 genders (masculine \textit{el}, feminine \textit{la} ‘the’)

- el chico ‘the boy’
- la chica ‘the girl’
- el árbol ‘the tree’
- la casa ‘the house’

German: 3 genders (masculine \textit{der}, feminine \textit{die}, neuter \textit{das} ‘the’)

- der Mann ‘the man’
- die Frau ‘the woman’
- der Löffel ‘the spoon’
- die Gabel ‘the fork’
- das Messer ‘the knife’
Order of object and verb

English: V O

The student bought the book.

Japanese: O V

Gakusei ga hon o katta.
Words for ‘arm’/‘hand’

English: differentiation

    arm
    hand

Russian: identity

    ruká
Representation of languages in WALS
Preliminary remarks

About 6-7,000 languages are spoken in the world today; most are endangered, in 100 years we will be lucky to find half this number.

Most of these languages have very few speakers: fewer than 400 have over a million speakers; nearly 4,000 have fewer than 10,000 speakers.

WALS presents data on a sample of these languages, using an ideally fixed sample of 200 languages, supplemented by other languages according to convenience.
Most of these languages are spoken in small communities, in traditional locations, and can be readily represented by a dot on a map.

We extend this practice to ALL languages, taking various practical decisions, e.g. “undoing” recent migrations; so: English is spoken in England. Spanish is spoken in Spain. Zoogocho Zapotec is spoken in San Bartolomé Zoogocho (Oaxaca).
Feature values are represented by the color of the dot
[Main areas: Southeast Asia; equatorial Africa]
[Black dots in Africa mainly Niger-Congo languages.]
Order of Object and Verb

1. OV [640]
2. VO [639]
3. No dominant order [91]

1370 languages

Author: Matthew S. Dryer
Hand and Arm

Author: Cecil H. Brown

617 languages

1. Identical [228]
2. Different [389]
Correlations Between Features
Relative clauses

English: N Rel

(the) book [that the student bought] N Rel

Japanese: Rel N

[gakusei ga katta] hon Rel N
Order of Relative Clause and Noun

Author: Matthew S. Dryer

1. Noun-Relative clause [507]
2. Relative clause-Noun [117]
3. Other [81]
Hypotheses:

a) VO order tends to correlate with NRel order.

b) OV order tends to correlate with RelN order
Relationship between the Order of Object and Verb and the Order of Relative Clause and Noun

Author: Matthew S. Dryer
Correlation between VO/OV and NRel/RelN order

<table>
<thead>
<tr>
<th>Combination</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>VO &amp; NRel</td>
<td>370</td>
</tr>
<tr>
<td>VO &amp; RelN</td>
<td>5</td>
</tr>
<tr>
<td>OV &amp; RelN</td>
<td>109</td>
</tr>
<tr>
<td>OV &amp; NRel</td>
<td>96</td>
</tr>
</tbody>
</table>
Hypotheses:

a) VO order tends to correlate with NRel order

370 : 5 strongly confirmed
only 5 exceptions, all in China

b) OV order tends to correlate with RelN order

109 : 96 not confirmed
confirmatory instances concentrated in Asia
Hand and arm
Hypothesis:

Non-differentiation of ‘hand’ and ‘arm’ terms is more likely closer to the equator, differentiation more likely further from the equator.

<table>
<thead>
<tr>
<th>Latitude</th>
<th>Differentiation</th>
<th>Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>70°N–36°N, 10°S–55°S</td>
<td>231</td>
<td>60</td>
</tr>
<tr>
<td>35°59´N–9°59´S</td>
<td>158</td>
<td>168</td>
</tr>
</tbody>
</table>

\[ \text{gamma} = .61 \]

\[ p < .001 \]

\[ N = 617 \]
Relatives and Neighbors

To what extent do languages tend to be similar structurally

a) to their relatives?
   genealogical basis

b) to their neighbors?
   areal basis

Test case: (Mainland) Southeast Asia
Order of Object and Verb

1. OV [640]
2. VO [639]
3. No dominant order [91]

1370 languages

Author: Matthew S. Dryer
Numeral classifiers

English    two dogs

Vietnamese hai con chó
    two CL dog
20 diagnostic features of Southeast Asian languages

Percentage of features shared by selected languages
(NB: Not all data complete)
<table>
<thead>
<tr>
<th>Language</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mongolian</td>
<td>5</td>
</tr>
<tr>
<td>Mandarin</td>
<td>42</td>
</tr>
<tr>
<td>Hmong Njua</td>
<td>64</td>
</tr>
<tr>
<td>Burmese</td>
<td>50</td>
</tr>
<tr>
<td>Thai</td>
<td>95</td>
</tr>
<tr>
<td>Vietnamese</td>
<td>89</td>
</tr>
<tr>
<td>Khmer</td>
<td>85</td>
</tr>
<tr>
<td>Malay-Indonesian</td>
<td>65</td>
</tr>
</tbody>
</table>

Bernard Comrie
University of California Santa Barbara and Max Planck Institute for Evolutionary Anthropology, Leipzig