Language Classification in Encyclopedistics

- Ranko Matasović, University of Zagreb and Croatian Academy of Sciences and Arts
- E-mail: rmatasov@ffzg.hr
Language classification is a very important subject for encyclopedias since Diderot’s „Encyclopédie” (1751-1772)
All the major encyclopedias contain articles about the most important language families (e.g. Indo-European, Uralic, Austronesian).
However, there is often no consensus (even among specialists) which language families are valid unit of language classification, or which languages belong to which families.
How is this uncertainty to be represented in encyclopedias, which should contain only uncontroversial information?
The following encyclopedias were consulted: Croatian Encyclopedia (HE), Encyclopedia Britannica, Great Russian Encyclopedia (BRE), Brockhaus, Larousse, Treccani, Wikipedia.

Both internet and paper editions were used; unless otherwise stated, the data in this presentation are from the internet editions.
Main language families
Number of language families by continents

- Africa: 59
- Australia and “Papunesia” (roughly corresponding to our Australia and Oceania): 178
- Eurasia: 46
- North America: 82
- South America: 117

Source: www.glotolog.org
Maximal sets of languages for which we can prove that they are genetically related, i.e. that they developed from a single proto-language (informally — in the way that Romance languages, such as French, Portuguese and Italian, developed from Latin).

Very few encyclopedias (e.g. HE and British Encyclopedia) contain a precise definition of what a language family is.
What constitutes a proof of genetic relatedness? Shared features that cannot be attributed to chance or contact (borrowing).

Languages differ to a large extent, so there is no single set of grammatical features that all languages must have.

Most languages have at least some *morphological paradigms* (sets of words with the same basic meaning, but different grammatical function).

Correspondences in morphological paradigms can usually be considered as proof of genetic relatedness.
<table>
<thead>
<tr>
<th></th>
<th>Lat.</th>
<th>Skt.</th>
<th>Hitt.</th>
<th>Lith.</th>
</tr>
</thead>
<tbody>
<tr>
<td>sum</td>
<td>asmi</td>
<td>ešmi</td>
<td>esu</td>
<td></td>
</tr>
<tr>
<td>es</td>
<td>asi</td>
<td>eši</td>
<td>esi</td>
<td></td>
</tr>
<tr>
<td>est</td>
<td>asti</td>
<td>ešzi</td>
<td>yra</td>
<td></td>
</tr>
<tr>
<td>sumus</td>
<td>smas</td>
<td>ešweni</td>
<td>esame</td>
<td></td>
</tr>
<tr>
<td>estis</td>
<td>stha</td>
<td>ešteni</td>
<td>esate</td>
<td></td>
</tr>
<tr>
<td>sunt</td>
<td>santi</td>
<td>ašanzi</td>
<td>yra</td>
<td></td>
</tr>
</tbody>
</table>
Indo-European languages
The validity of Indo-European family is not controversial, and neither is the set of languages that constitute it. What is controversial is its internal structure (branching).

One widespread classification divides Indo-European into Anatolian and the reset (the „Indo-Hittite hypothesis” of E. Sturtevant).
The structure of the IE family
An alternative sub-classification
# The treatment of Indo-European in encyclopedias

<table>
<thead>
<tr>
<th></th>
<th>HR</th>
<th>Britannica</th>
<th>BRE</th>
<th>Brockhaus</th>
<th>Larousse</th>
<th>Treccani</th>
<th>Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proof explained</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Exhaustive list of branches</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Exhaustive list of languages</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Problems of subclassification</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
Indo-European languages, family of languages spoken in most of Europe and areas of European settlement and in much of Southwest and South Asia. The term Indo-Hittite is used by scholars who believe that Hittite and the other Anatolian languages are not just one branch of Indo-European but rather a branch coordinate with all the rest put together; thus, Indo-Hittite has been used for a family consisting of Indo-European proper plus Anatolian. As long as this view is neither definitively proved nor disproved, it is convenient to keep the traditional use of the term Indo-European.
Language families of Africa

Although this classification is still found in most textbooks of linguistics (including my own), as well as encyclopedias, it has been cast into doubt.

Today most experts agree that Khoisan is not a family (but an areal grouping).

Similar doubts have been expressed with respect to Niger-Kordofanian (now usually called Niger-Congo) and Nilo-Saharan.
Several languages previously classified to one of the families are now considered isolates (e.g. Laal in Chad).

Glottolog.org (a reliable internet source) lists 59 language families in Africa!

Much comparative work has been done on several uncontroversial low level groupings (e.g. Mande, North and South Atlantic)

Several languages and groups of languages are known by different names (e.g. Peul, Fulbe, Fulfulda, Pulari)

How much of this has found its way to the encyclopedias?
### Evidence for Niger-Congo: noun classes (genders)

<table>
<thead>
<tr>
<th>Class no. 1</th>
<th>3</th>
<th>4 (pl of 3)</th>
<th>5</th>
<th>6 (pl of 5)</th>
<th>6a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kordofanian</td>
<td>gu-</td>
<td>gu-</td>
<td>i-</td>
<td>li-</td>
<td>ngu-</td>
</tr>
<tr>
<td>Atlantic</td>
<td>gu-</td>
<td>gu-</td>
<td>Ci-</td>
<td>de-</td>
<td>ga-</td>
</tr>
<tr>
<td>Oti-Volta</td>
<td>U-</td>
<td>-bu</td>
<td>-Ci</td>
<td>-di</td>
<td>-a</td>
</tr>
<tr>
<td>Togo Remnant</td>
<td>o-</td>
<td>o-</td>
<td>i-</td>
<td>li-</td>
<td>a-</td>
</tr>
<tr>
<td>Benue-Congo</td>
<td>u-</td>
<td>u-</td>
<td>(t)i-</td>
<td>li-</td>
<td>a-</td>
</tr>
<tr>
<td>Bantu</td>
<td>mu-, ju-</td>
<td>mu-, gu-</td>
<td>mi-, gi-</td>
<td>li-</td>
<td>ma-, ga-</td>
</tr>
<tr>
<td><strong>Typical referents</strong></td>
<td>people</td>
<td>trees</td>
<td>egg, name</td>
<td>liquids</td>
<td></td>
</tr>
</tbody>
</table>

Source: Schadeberg (1989)
Note: “Kwa” is polysemous; Atlantic, Adamawa-Ubangi and Kordofanian are not considered to be single branches by most africanists; Mande may not be a NC language after all.
## Treatment of Niger-Congo in encyclopedias

<table>
<thead>
<tr>
<th></th>
<th>HR</th>
<th>Britannica</th>
<th>BRE</th>
<th>Brockhaus</th>
<th>Larousse</th>
<th>Treccani</th>
<th>Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proof explained</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>Branches surveyed</td>
<td>11</td>
<td>9</td>
<td>12</td>
<td>No data</td>
<td>No data</td>
<td>7</td>
<td>9 + 5 (possible)</td>
</tr>
<tr>
<td>Lgs. surveyed</td>
<td>No data</td>
<td>1400</td>
<td>1500</td>
<td>No data</td>
<td>No data</td>
<td>No data</td>
<td>1540</td>
</tr>
<tr>
<td>Problems of subclassification</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Typological characteristics</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
Niger-Congo languages, a family of languages of Africa, which in terms of the number of languages spoken, their geographic extent, and the number of speakers is by far the largest language family in Africa. The area in which these languages are spoken stretches from Dakar, Senegal, at the westernmost tip of the continent, east to Mombasa in Kenya and south to Cape Town, South Africa. Excluding northern Africa (Mauritania to Egypt and Sudan) and the Horn of Africa (Ethiopia to Somalia), some 85 percent of the population of Africa—at least 600 million people—speak a Niger-Congo language. In two countries, Niger and Chad, Niger-Congo languages are spoken by a minority. In northern Nigeria, northern Uganda, and Kenya there are substantial populations speaking other languages, but even in these countries the majority of the population speaks a Niger-Congo language.
Nilo-Saharan languages
<table>
<thead>
<tr>
<th>Language</th>
<th>Pl. markers</th>
<th>Interr. pron.</th>
<th>Negation</th>
<th>„water”</th>
<th>„spit”</th>
<th>„Moon”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Songhai</td>
<td></td>
<td></td>
<td>hari (Zerma)</td>
<td>tufa</td>
<td>-dar- (Gao)</td>
<td></td>
</tr>
<tr>
<td>Sakharan</td>
<td></td>
<td></td>
<td>orui „river“</td>
<td>tefa  (Kanuri)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maban</td>
<td>-ng (Maban, pl. Pron.)</td>
<td>ng</td>
<td>are „lake“</td>
<td>tuf (Masalit)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fur</td>
<td>k- (pl. Dem.)</td>
<td></td>
<td>roo „river“</td>
<td></td>
<td>dul-</td>
<td></td>
</tr>
<tr>
<td>East Sudanic</td>
<td>-gu (dem. Pl., Nera)</td>
<td>nd, ng (Nera)</td>
<td>ka (pres. Nera)</td>
<td></td>
<td>tuf</td>
<td></td>
</tr>
<tr>
<td>Central Sudanic</td>
<td>-ge (pl. Pron. Bagirmi)</td>
<td>nanga (Bagirmi)</td>
<td>-ku (Kara)</td>
<td>Ar-(a) ? „rainy season“</td>
<td>tibi</td>
<td>TVr</td>
</tr>
<tr>
<td>Berta</td>
<td>-gu (dem. Pl.)</td>
<td>naano (Undu dial.)</td>
<td></td>
<td>(r)ro</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Treatment of Nilo-Saharan in Encyclopedias

<table>
<thead>
<tr>
<th></th>
<th>HR</th>
<th>Britannica</th>
<th>BRE</th>
<th>Brockhaus</th>
<th>Larousse</th>
<th>Treccani</th>
<th>Wiki</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proof explained</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Branches surveyed</td>
<td>6</td>
<td>10</td>
<td>11 + 4 isolates</td>
<td>6</td>
<td>NO</td>
<td>6</td>
<td>Different theories are presented</td>
</tr>
<tr>
<td>Lgs. surveyed</td>
<td>No data</td>
<td>No data</td>
<td>200</td>
<td>No data</td>
<td>No data</td>
<td>No data</td>
<td>15 (only languages above 1 mil. Speakers)</td>
</tr>
<tr>
<td>Problems of sub-classification</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Typological characteristic</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>
Nilo-Saharan languages, a group of languages that form one of the four language stocks or families on the African continent, the others being Afro-Asiatic, Khoisan, and Niger-Congo. The Nilo-Saharan languages are presumed to be descended from a common ancestral language and, therefore, to be genetically related. The family covers major areas east and north of Lake Victoria in East Africa and extends westward as far as the Niger valley in Mali, West Africa. Its genetic unity was first proposed in a classificatory study dating from 1963 by the American linguist and anthropologist Joseph H. Greenberg.
Internet resources

- WALS (wals.info)
- Glottolog (glottolog.org)
- Ethnologue (www.ethnologue.com)

Standard codes for languages and families (used by internet sources) are not explained in most encyclopedias (except Wiki). Wikipedia contains more information and uses more of the constantly updated internet-based sources than national encyclopedias. It also uses international codes (ISO and Glottolog) to identify languages.
Welcome to WALS Online

The World Atlas of Language Structures (WALS) is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of 55 authors.

The first version of WALS was published as a book with CD-ROM in 2005 by Oxford University Press. The first online version was published in April 2006.

The 2013 edition of WALS corrects a number of coding errors especially in Chapters 1 and 3. A full list of changes is available here.

Starting with the 2013 edition of WALS, we will release and publish sets of corrections periodically. Thus, any citation of WALS Online 2013 should include the particular version, as listed on Zenodo.

WALS Online is a publication of the Max Planck Institute for the Science of Human History. It is a separate publication, edited by Dryer, Matthew S. & Haspelmath, Martin (Jena: Max Planck Institute for the Science of Human History, 2013). The main programmer is Robert Forkel.

How to use WALS Online

Using WALS Online requires a browser with Javascript enabled.

You find the features or chapters of WALS through the items “Features” and “Chapters” in the navigation bar.

You can also browse and search for languages through the item “Languages” on the navigation bar.

You can search for references through the item “References”, and once you have navigated to a particular feature, you see a second navigation bar with citation information and various export options.

A description of changes from previous editions is available through the item “Changes”.

How to cite WALS Online

It is important to cite the specific chapter that you are taking your information from, not just the general work “The World Atlas of Language Structures Online” (Dryer, Matthew S. & Haspelmath, Martin 2013), unless you are citing data from more than 25 chapters simultaneously.
Welcome to Glottolog 4.3

Comprehensive reference information for the world’s languages, especially the lesser known languages.

Information about the different languages, dialects, and families of the world (‘languages’) is available in the Languages and Families sections. The References section contains bibliographic information. You can query the bibliographical database by filtering the table view or using a complex query involving genealogical affiliation, document type, and macro-area.

Catalogue of languages and families

Glottolog provides a comprehensive catalogue of the world’s languages, language families and dialects. It assigns a unique and stable identifier (the ‘Glottocode’) to (principle) all languages, i.e., all families, languages, and dialects. Any variety of a language which should eventually get its own entry. The languages are organised via a genealogical classification (the Glottlog tree) that is based on available historical-comparative research (see also the Languages information section).

Bibliography

The References section (‘languages’) provides a comprehensive collection of bibliographic data for the world’s lesser known languages. It provides access to 541,597 references of descriptive works such as grammars, dictionaries, wordlists, texts etc. Search criteria include author, year, title, country, and genealogical affiliation. References can be downloaded as txt, bib, html, or with the Zotero-Firefox plugin.

New Grammars

Brendan Yooger 2020

A grammar of Mafu, a Lolan Plate language of Papua, Indonesia

Yameo, Aistrahi 2020

Farauto SSP

Nell Alexander Walker 2020

A Grammar of Southern Pono

New Languages

Ganichare

Hilo (Bomberai Peninsula)

Ramanos

New Dictionaries

Aung Si 2020

Sigeja English Dictionary

Margaret Sharpe 2020

Gogun Mileyth, Togadenkugao andisgaw, Togadonkugaw. A dictionary and grammar of Miley language varieties from the Territory to the Lena-Ivens

Thiothuss Adranius Bodl 2020

Dukumbi Dictionary
Fula language

From Wikipedia, the free encyclopedia

Fula (French: Ful, Wolof: Fuul, Hausa: Fulo, Mandinka: Fula, Wolof: Fulye) is a language spoken as a set of various dialects in a continuum that stretches across some 20 countries in West and Central Africa by more than 65 million people. Along with other related languages such as Serer and Wolof, it belongs to the Senegambian branch within the Niger–Congo family, which does not have tones, unlike most other Niger–Congo languages. More broadly, it belongs to the Atlantic geographic grouping within Niger–Congo. It is spoken as a first language by the Fula people (“Fulani”, “Fula”, “Fu”), from the Senegambia region and Guinea to Cameroon, Nigeria, and Sudan and by related groups such as the Toucouleur people in the Senegal River Valley. It is also spoken as a second language by various peoples in the region, such as the Kori of northern Cameroon and northeastern Nigeria.

Contents

1. Nomenclature
2. Morphology
   2.1 Noun classes
   2.2 Voice
   2.3 Consonant mutation
   2.4 Prefixes
3. Varieties
4. Status
5. Writing systems and phonology
   5.1 Latin alphabet
     5.1.1 Sample Fula alphabet
   5.2 Phonology
     5.2.1 Consonants
     5.2.2 Vowels
   5.3 Arabic script
   5.4 Arabic script
6. References
7. Notes
8. External links

Nomenclature

Several names are applied to the language, just as to the Fula people. They call their language Pulaar or Pular in the western dialects and Pufufe in the central and eastern dialects. Fula, Fulah and Pular in English come originally from Manding (esp. Mandinka, but also Malinke and Bamana) and Hausa, respectively. Peul in French, also occasionally found in literature in English, comes from Wolof.

Morphology
# Wikipedia Fula/Fulani/Peul

<table>
<thead>
<tr>
<th>Language codes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 639-1</td>
<td>ff – Fula</td>
</tr>
<tr>
<td>ISO 639-2</td>
<td>ful – Fula</td>
</tr>
<tr>
<td>ISO 639-3</td>
<td>ful – inclusive code – Fula</td>
</tr>
</tbody>
</table>

**Individual codes:**
- fuc – Pulaar (Senegambia, Mauritania)
- fuf – Pular (Guinea, Sierra Leone)
- ffm – Maasina Fulfulse (Mal, Ghana)
- fue – Borgu Fulfulse (Benin, Togo)
- fuh – Western Niger (Burkina, Niger)
- fuq – Central–Eastern Niger (Niger)
- fuv – Nigerian Fulfulse (Nigeria)
- fub – Adamawa Fulfulse (Cameroon, Chad, Sudan)

**Glottochronology:**
- fula1264 is [1]
A good encyclopedia article about a language family should:

1. give a reasonable assessment of its validity
2. discuss problems of its sub-classification
3. give a list of culturally/typologically/politically most important languages of that family
4. inform the reader how this family is treated in other reliable (including internet-based) sources

Thanks!

- E-mail: rmatasov@ffzg.hr
- Homepage: www.ffzg.hr/~rmatasov