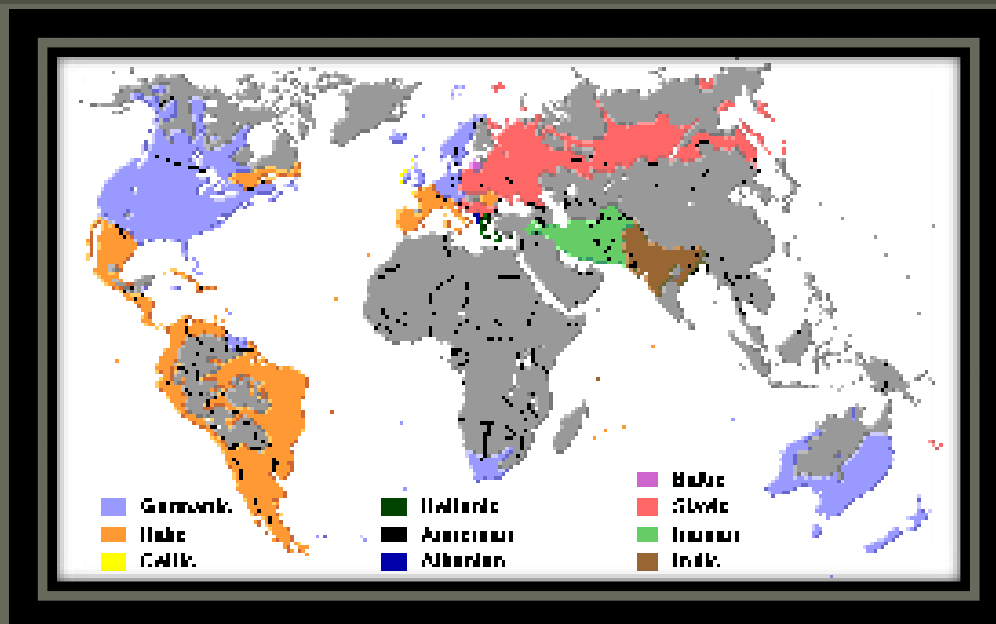


Language Classification in Encyclopedistics



- Ranko Matasović, University of Zagreb and Croatian Academy of Sciences and Arts
- E-mail: rmatasov@ffzg.hr

The problem

- 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 sources

- 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.



Number of language families by continents

◉ Source: www.glottolog.org

Africa: 59

Australia and “Papunesia” (roughly corresponding to our Australia and Oceania): 178

Eurasia: 46

North America: 82

South America: 117

What are language families?

- 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.

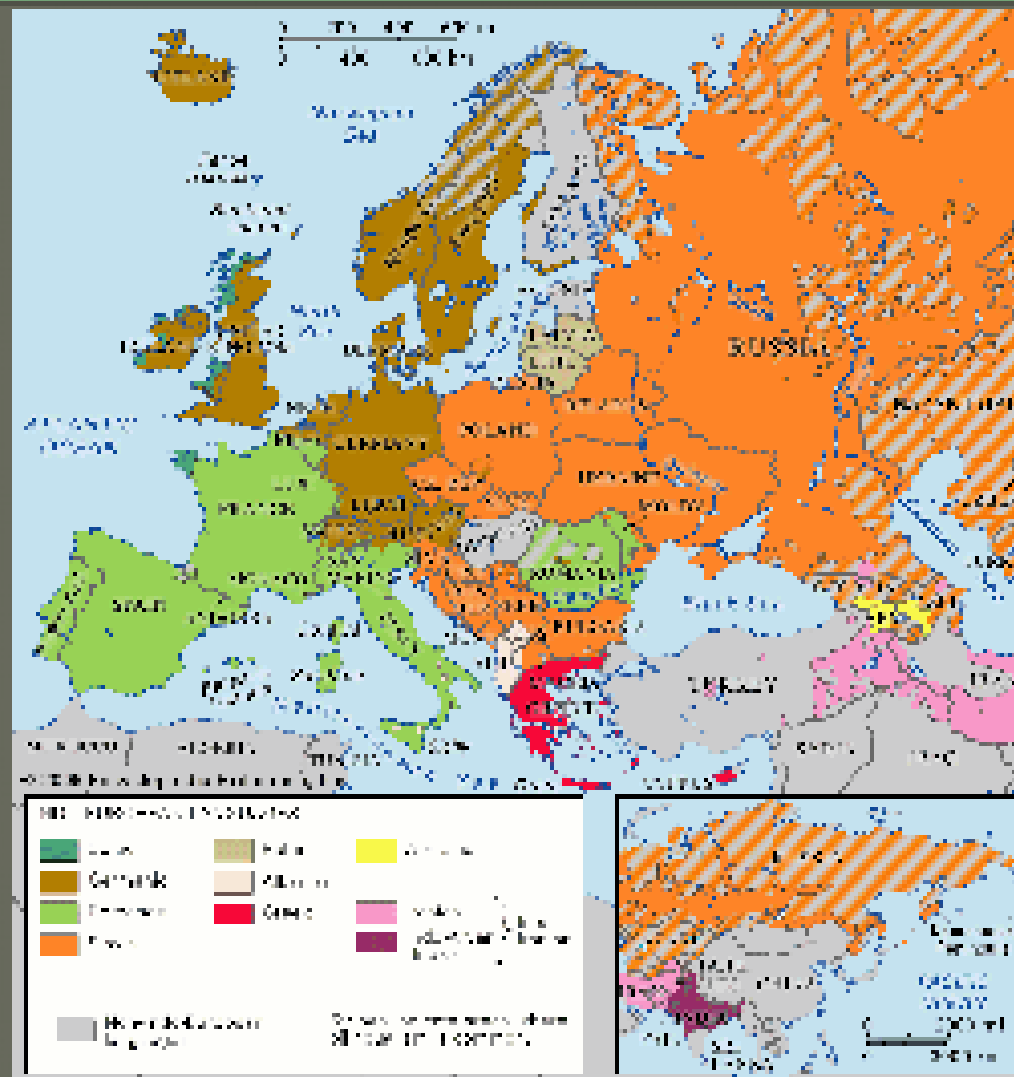
The proof of genetic relatedness

- 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.

Formal (morphological) correspondences

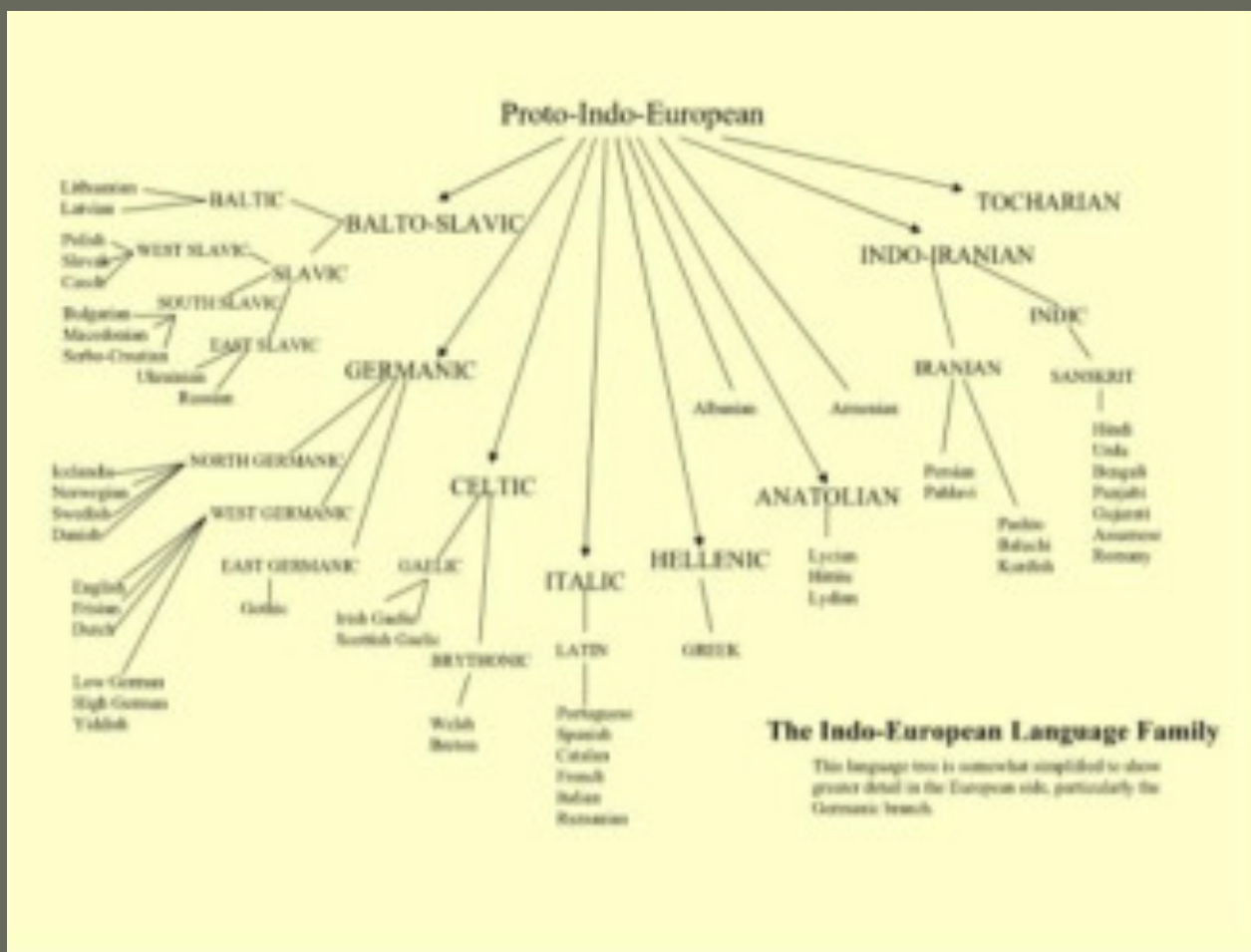
Lat.	Skt.	Hitt.	Lith.
● <i>sum</i>	<i>asmi</i>	<i>ešmi</i>	<i>esu</i>
● <i>es</i>	<i>asi</i>	<i>eši</i>	<i>esi</i>
● <i>est</i>	<i>asti</i>	<i>ešzi</i>	<i>yra</i>
● <i>sumus</i>	<i>smas</i>	<i>ešweni</i>	<i>esame</i>
● <i>estis</i>	<i>stha</i>	<i>ešteni</i>	<i>esate</i>
● <i>sunt</i>	<i>santi</i>	<i>ašanzi</i>	<i>yra</i>

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 rest (the „Indo-Hittite hypothesis” of E. Sturtevant).

The structure of the IE family



An alternative sub-classification



The treatment of Indo-European in encyclopedias

	HR	Britannica	BRE	Brockhaus	Larousse	Treccani	Wiki
Proof explained	NO	YES	NO	NO	NO	NO	NO
Exhaustive list of branches	YES	YES	YES	YES	YES	YES	YES
Exhaustive list of languages	NO	NO	NO	NO	NO	NO	YES
Problems of sub-classification	YES	NO	NO	NO	NO	NO	YES

<https://www.britannica.com/topic/Indo-European-languages>

[Home](#) > [Geography & Travel](#) > [Languages](#)

Indo-European languages

WRITTEN BY

[Jay H. Jasanoff](#) [See All Contributors](#)

Diebold Professor of Indo-European Linguistics and Philology, Department of Linguistics, Harvard. Author of *Hittite and the Indo-European Verb*.

[See Article History](#)

Alternative Titles: [Indogermanic Indo-European language](#), [Indogermanisch Indo-European language](#)

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.

Tools to get ready for back to school



[LEARN MORE](#)

Language families of Africa

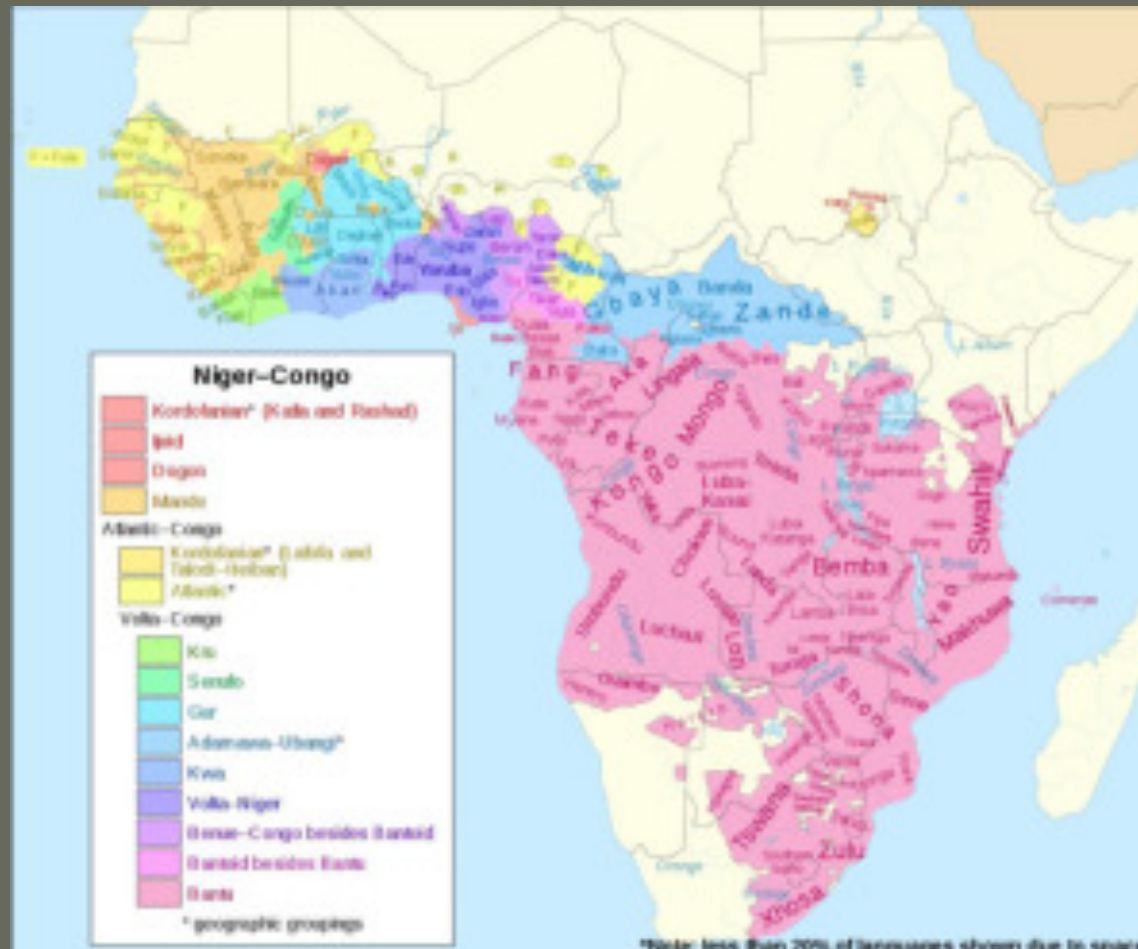


„Lumpers” and „Splitters”

- Greenberg (1955, 1963): only four indigenous language families in Africa (Afro-Asiatic, Nilo-Saharan, Niger-Kordofanian, Khoisan) + Malagasy (a late Austronesian immigration to Madagascar).
- 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?

Niger-Congo languages

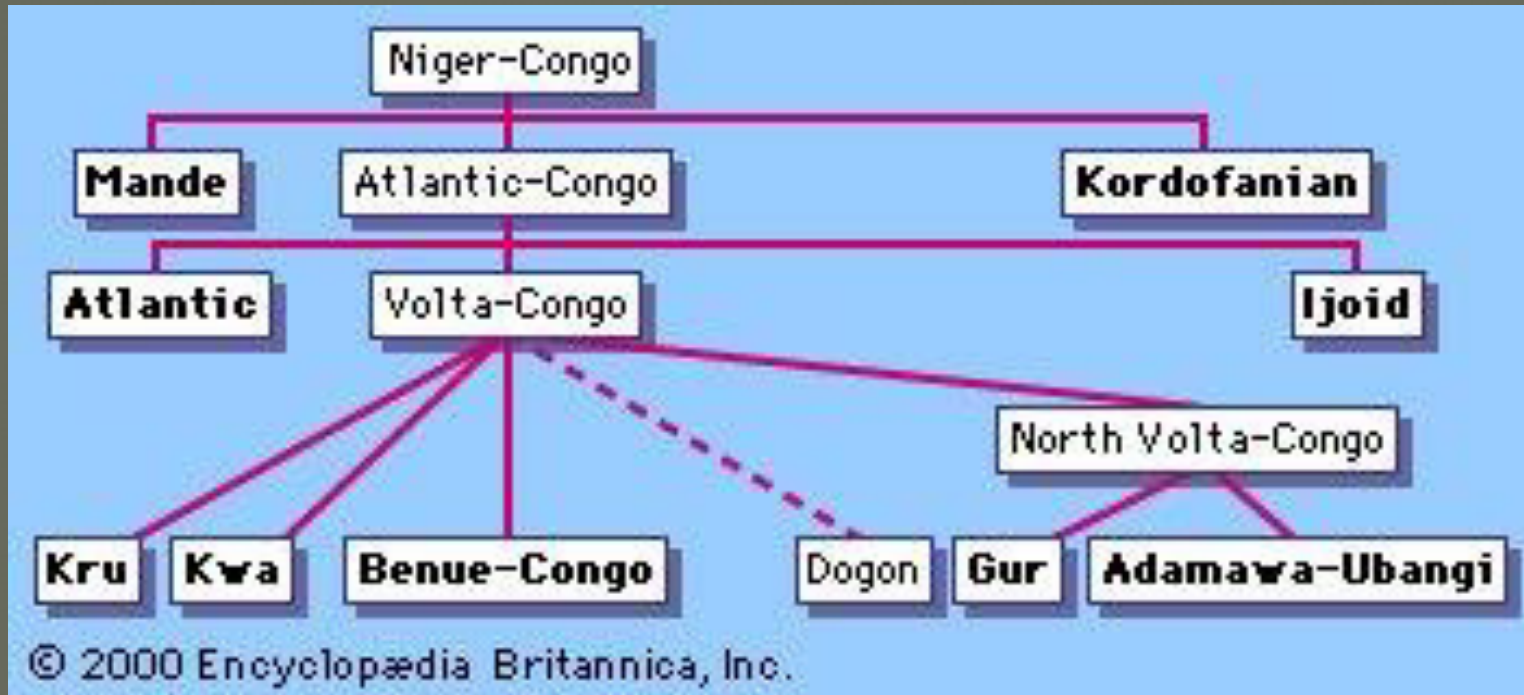


Evidence for Niger-Congo: noun classes (genders)

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

Source: Schadeberg (1989)

Sub-classification of Niger-Congo



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

	HR	Britannica	BRE	Brockhaus	Larousse	Treccani	Wiki
Proof explained	NO	NO	NO	NO	NO	NO	NO
Branches surveyed	11	9	12	No data	No data	7	9 + 5 (possible)
Lgs. surveyed	No data	1400	1500	No data	No data	No data	1540
Problems of sub-classification	YES	YES	NO	NO	NO	NO	YES
Typological characteristics	YES	YES	YES	YES	YES	NO	YES

<https://www.britannica.com/topic/Niger-Congo-languages>

[Home](#) > [Geography & Travel](#) > [Languages](#)

Niger-Congo languages

WRITTEN BY

[John T. Bendor-Samuel](#)

Executive Director, Summer Institute of Linguistics Ltd., High Wycombe, England. Editor of *The Niger-Congo Languages: A Classification and Description of Africa's Largest Language Family*.

[See Article History](#)

Alternative Title: Western Sudanic languages

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.

Tools to get ready for back to school



[LEARN MORE](#)

Niger-Congo languages

Nilo-Saharan languages



Evidence for Nilo-Saharan (Bender 2004, „Les langues africaines”)

	Pl. markers	Interr. pron.	Negation	„water”	„spit”	„Moon”
Songhai				hari (Zerma)	tufa	-dar- (Gao)
Sakharan				orui „river”	tefa (Kanuri)	
Maban	-ng (Maban, pl. Pron.)	ng		are „lake”	tuf (Masalit)	
Fur	k- (pl. Dem.)			roo „river”		dul-
East Sudanic	-gu (dem. Pl., Nera)	nd, ng (Nera)	ka (pres. Nera)		tuf	
Central Sudanic	-ge (pl. Pron. Bagirmi)	nanga (Bagirmi)	-ku (Kara)	Ar-(a) ? „rainy season”	tibi	TVr
Berta	-gu (dem. Pl.)	naano (Undu dial.)		(r)ro		

Treatment of Nilo-Saharan in encyclopedias

	HR	Britannica	BRE	Brockhaus	Larousse	Treccani	Wiki
Proof explained	NO	NO	NO	NO	NO	NO	YES
Branches surveyed	6	10	11 + 4 isolates	6	NO	6	Different theories are presented
Lgs. surveyed	No data	No data	200	No data	No data	No data	15 (only languages above 1 mil. Speakers)
Problems of sub-classification	NO	YES	NO	YES	NO	NO	YES
Typological characteristic	YES	YES	NO	NO	NO	NO	YES

<https://www.britannica.com/topic/Nilo-Saharan-languages>

[Home](#) > [Geography & Travel](#) > [Languages](#)

Nilo-Saharan languages

WRITTEN BY

Morris F. Goodman | [See All Contributors](#)

Associate Professor Emeritus of Linguistics, Northwestern University, Evanston, Illinois. Author of several articles on African languages and *A Comparative Study of Creole French Dialects*.

[See Article History](#)

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 Dyer, 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" (Dyer, Matthew S. & Haspelmath, Martin 2013), unless you are citing data from more than 25 chapters simultaneously.

Glottolog

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 bibliographical 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 (in principle) all languages, i.e. all families, languages, and dialects. Any variety that a linguist works on should eventually get its own entry. The languages are organized via a genealogical classification (the [Glottolog tree](#)) that is based on available historical-comparative research (see also the [Languages information](#) section).

Bibliography

The [References](#) section ('langdoc') provides a comprehensive collection of bibliographical data for the world's lesser known languages. It provides access to 307,527 references of descriptive works such as grammars, dictionaries, word lists, texts etc. Search criteria include author, year, title, country, and genealogical affiliation. References can be downloaded as list, table, html, or with the [Open-Firefox](#) page.



New Grammars

Brenden Yoder 2020

Post Year: 2020

A grammar of Mbevi, a Lakes Plate language of Papua, Indonesia

Yamada, Atsushi 2020

Post Year: 2020

Forests ways

Nell Alexander Walker 2020

Post Year: 2020

A Grammar of Southern Pomo

New Languages

Carichans

Post Year: 2020

Mor (Bomberai Peninsula)

Post Year: 2020

Ramanon

Post Year: 2020

New Dictionaries

Aung Si 2020

Post Year: 2020

Songai-English Dictionary

Margaret Sharpe 2020

Post Year: 2020

Cangan Mbayah: Yugambel-Ngarrindjerri Ngatindjwal. A dictionary and grammar of Mbayah language varieties from the Tenet to the Logan rivers

Timotheus Adrianus Boddé 2020

Post Year: 2020

Duhumbi Dictionary

Ethnologue

The screenshot shows the Ethnologue website homepage. At the top, there is a blue navigation bar with the Ethnologue logo and tagline 'Languages of the World' on the left. On the right, there are links for 'Log in', 'Shopping Cart', and a search bar labeled 'Find a language or country...'. Below the navigation bar are links for 'Languages', 'Countries', 'Guides', 'About', and a 'Plans & Pricing' button. The main content area features a world map with regions highlighted in different colors (yellow, blue, purple, orange). To the right of the map is a purple box with the text 'Explore The World's Languages' and a description: 'Find, read about, and research the world's 7,117 known living languages.' Below this is a 'LEARN MORE' link. The bottom section contains three white boxes: 'Ethnoblog' with a 'Welcome to the 23rd edition' announcement, 'Yahadian' with a 'LANGUAGE OF THE DAY' banner and a scatter plot, and 'Coronavirus and Local Languages: How do you say, "wash your hands"?' with a world map and a 'VIEW OUR GUIDE' link.

Ethnologue
Languages of the World

Log in | Shopping Cart | Find a language or country...

Languages | Countries | Guides | About | Plans & Pricing

Explore The World's Languages

Find, read about, and research the world's 7,117 known living languages.

[LEARN MORE](#) or see the language vitality of each region.

Ethnoblog

Welcome to the 23rd edition
February 21, 2020
As our contribution to the celebration of International Mother Language Day, we are pleased to announce the

[READ MORE](#)

Yahadian

LANGUAGE OF THE DAY

A scatter plot with 'SPEAKERS' on the vertical axis and 'LANGUAGES' on the horizontal axis. A blue diagonal banner across the plot reads 'LANGUAGE OF THE DAY'. A yellow dot highlights a specific data point.

A language of Indonesia. Population: 500 (1,000 SIL). Status: 04- (Threatened). Ig 1/nc
Some young people, all adults. Ig Development:

[READ MORE](#)

Coronavirus and Local Languages: How do you say, "wash your hands"?

"Wash your hands" in local languages

A world map with blue dots indicating the locations of various local languages. A small inset map shows a zoomed-in view of a region.

All of us are at risk. And for the millions who speak a lesser-known language, that risk is compounded by the lack of health materials in a language they understand.

[VIEW OUR GUIDE](#)

ISO / Glottolog codes

Fula language

From Wikipedia, the free encyclopedia

Fula (fulɑ̀ⁱ, also known as **Fulani** fʊlɑ̀ⁱ or **Fulah**^[ˈfʊlɑːh] (Fula: *Fulfulde* *awccooɗ*, *Fulaar* *uttooɗ*, *Fulaɗ* *accɗɗ*, French: *Foul*), is a Senegambian 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: *Fulbe*) 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 Kirdi of northern Cameroon and northeastern Nigeria.

Contents [hide]

- Nomenclature
- Morphology
 - Noun classes
 - Noun
 - Consonant mutation
 - Prefixes
- Varieties
- Status
- Writing systems and phonology
 - Latin alphabet
 - Sample Fula alphabets
 - Phonology
 - Consonants
 - Vowels
 - Arabic script
 - Adlam script
- ISO codes
- References
 - Works
 - Notes
- External links

Nomenclature [edit]

Several names are applied to the language, just as to the Fula people. They call their language *Fulaar* or *Fular* in the western dialects and *Fulfulde* in the central and eastern dialects. *Fula*, *Fulah* and *Fulani* in English come originally from *Manding* (esp. *Mandinka*, but also *Malinke* and *Rambana*) and *Hausa*, respectively. *Foul* in French, also occasionally found in literature in English, comes from *Wolof*.

Morphology [edit]

Fula	
Fulani, Fulah, Foul	
 <div>Fulfulde <i>awccooɗ</i><div>Fulaar <i>uttooɗ</i></div>Fulaɗ <i>accɗɗ</i></div>	
Native to	Western Africa
Region	The Sahel
Ethnicity	Fula
Language family	<div>Niger–Congo <ul style="list-style-type: none"> Atlantic–Congo<ul style="list-style-type: none">Atlantic<ul style="list-style-type: none">Senegambian<ul style="list-style-type: none">Fula–Serer<ul style="list-style-type: none">Fula</div>
Writing system	Latin (Fulɓi) <p>Adlam</p>
Language codes	
ISO 639-1	ff – Fulaɗ
ISO 639-2	ful – Fulah
ISO 639-3	ful – inclusive code – Fulah <div> <div>Individual codes:</div> <div> <ul style="list-style-type: none">ful – Fulah (Senegambia, Mauritania) ffl – Fulah (Guinea, Sierra Leone) ffn – Fulaɗ (Fulfulde) (Sen., Ghana) ful – Borgu/Fulfulde (Sen., Niger) ful – Western Niger (Guinea, Niger) ful – Central-Eastern Niger (Niger) ful – Nigerian Fulfulde (Niger) ful – Adamawa Fulfulde (Cameroon, Chad, Sudan) ful – Bagirmi Fulfulde (con.) </div> </div>
Glottolog	FuLad129a:g [1]

Wikipedia Fula/Fulani/Peul

	Fula
	Fulani, Fulah, Peul
	<i>Fulfulde</i> 𞤄𞤵𞤲𞤴𞤸𞤲𞤴𞤸 <i>Pulaar</i> 𞤄𞤵𞤲𞤴𞤸 <i>Pular</i> 𞤄𞤵𞤲𞤴𞤸
Native to	Western Africa
Region	The Sahel
Ethnicity	Fula
Language family	Niger–Congo <ul style="list-style-type: none"> Atlantic–Congo <ul style="list-style-type: none"> Atlantic <ul style="list-style-type: none"> Senegambian <ul style="list-style-type: none"> Fula–Serer <ul style="list-style-type: none"> Fula
Writing system	Latin Arabic Adlam

Language codes	
ISO 639-1	ff – Fulah
ISO 639-2	ful – Fulah
ISO 639-3	ful – inclusive code – Fulah
	Individual codes:
	foo – Pulaar (Senegambia, Mauritania)
	fuf – Pular (Guinea, Sierra Leone)
	ffm – Maasina Fulfulde (Mali, Ghana)
	foe – Borgu Fulfulde (Benin, Togo)
	fuh – Western Niger (Burkina Faso, Niger)
	foq – Central–Eastern Niger (Niger)
	fov – Nigerian Fulfulde (Nigeria)
	fub – Adamawa Fulfulde (Cameroon, Chad, Sudan)
	fui – Bagirmi Fulfulde (CAR)
Glottolog	fula1264 ^[1]

Conclusions

- 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

References

- R. Matasović, *Jezična raznolikost svijeta*, Algoritam, Zagreb, 2011 (2. izdanje)
- R. Matasović, *Uvod u poredbenu lingvistiku*, MH, Zagreb 2001.

Thanks!

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