

Tsammalex – Manual

C.Naumann & L.Sell

vs. CN, 03/02/2015 (planned changes cf. current online version in red).

1 General

This manual provides basic information about the layout and functions of the Tsammalex database online ("2 Website"), the basic data files ("3 Data files") and planned changes.

Planned structural changes

- **possibility of personalized versions** for researchers (prior to general publications) –only accessible to researchers > only if there is real need; the general policy should be that publishing data is valuable in itself
- specification of the workflow: 1) how to deal with changes for ids of species/taxa and corresponding images and words, 2) uploading process (images, which format, which metadata), 3) editing process in general (which data must be entered manually, which data are generated automatically, e.g. from EOL: EOL codes, countries of distribution, other?), versions

Possible future improvements

- possibility of "Show/Hide labels" also for map of ecoregions
- expanded export (download) function generating illustrated lists of plants and animals for individual languages (i.e. including 2-3 photos, all included lexical data, and corresponding meanings in major 2nd languages)
- information on/photos (e.g. landscapes, habitats) from ecoregions (when selecting ecoregions under 2.5.)
- export function (download) generating tables that may be used for the dialectological tool "Gabmap" (<http://www.gabmap.nl/>), i.e., tables with language varieties and their word forms for biological taxa/species (cf. <http://www.gabmap.nl/~app/doc/preparing/>)
- a function generating/showing language-specific hierarchies/taxonomies based on collected categories?

2 Website

9 main tabs: Home, Names, Languages, Taxa, Ecoregions, **Countries**, References, Images, Contribute!

2.1 Home

= introduction and general information

Introductory text (enlarged font size):

“Tsammalex is a multilingual lexical database on plants and animals including linguistic, anthropological and biological information as well as images. It has been set up as a resource for linguists, anthropologists and other researchers, language planners and speech communities interested in the conservation of their biological knowledge. In the current version, it is still focused on particular geographical regions reflecting the interests of the present contributors (Kalahari basin in Southern Africa, Dogon languages in West Africa).

Lexical and biological data can be accessed directly (tabs "Names" and "Taxa", respectively) or filtered for specific languages ("Languages") or geographical regions ("Ecoregions" and "Countries"), with varying details. The tabs "References" and "Images" include lists of sources and individual images, while "Contribute!" provides more information, especially for potential contributors."

(Smaller font size):

"Tsammalex has been developed by Robert Forkel and compiled by Christfried Naumann, Lena Sell, Noémie Jaulgey and Kathrin Heiden. Significant amounts of data were collected by Jeffrey Heath and colleagues (Dogon Languages Project) and added by Steven Moran and Robert Forkel. Peter Fröhlich, Hans-Jörg Bibiko, Jan Klom and Stefan Koch were involved in former versions of the database." [add further main contributors here]

"The project is funded by the Max Planck Insitute for Evolutionary Anthropology's Department of Linguistics, led by Bernard Comrie. Tsammalex is published as part of the Cross-Linguistic Linked Data project, led by Martin Haspelmath. The content of this web site including the downloadable database, is published under a Creative Commons Attribution 4.0 International License License (unless stated otherwise).

Contributions and comments are highly appreciated! Read more information under "Contribute!".

All images and other data included in this database should be under creative commons license. Sensitive information that may cause conflicts in communities or harm to their environment, or that may threaten biological species (such as information about specific ways of killing animals or detailed information on medical uses) must not be included here. Cultural knowledge should not be published without the consent of the communities it belongs to."

"Cite

Christfried Naumann, Steven Moran & Robert Forkel, (eds.) 2015. Tsammalex: A lexical database on plants and animals. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://tsammalex.clld.org>, Accessed on 2015-01-26.)"

2.1a Contributors

= list of editors, institutions and their edited sections

This section includes a list of all contributors, importantly also editors of biological aspects for different regions, such as "Plants of Southern Africa". Editors of lexical sections for various languages may additionally be mentioned under "Languages: Description".

Name	Section	Affiliation	Project	Homepage
<i>Brumble, Bonny</i>	<i>Plants of Southern Africa (principal contributor)</i>			
<i>Müller, Max</i>	<i>Taa</i>	<i>Humboldt University, Berlin</i>		
<i>Muster, Melanie</i>	<i>Data input and management (version October 2014)</i>	<i>MPI-EVA, Leipzig</i>		
<i>Temple, Ed</i>	<i>Taa (principal contributor)</i>	<i>MPI-EVA, Leipzig</i>	<i>Kalahari Basin Area (KBA)</i>	<i>dobes.mpi.nl/...</i>

Table layout

2.2 Names [General view]

= general tabular view on names and species and other taxa (search of names in any language > species)

Title: Names (general view)

Display the following comment: "See also 1) Languages > Individual languages > Names (linguistic, cultural and bilingual comparative views), and 2) Taxa > Individual taxa (multilingual comparative view)"

Name	Language	IPA	Grammatical info	Taxon	English name	thumbnail	References
kameeldoring	Afrikaans	kami:ldɔrɨŋ		Acacia erioloba		X	
+ sorting, search	+ sorting, search	+ sorting, search		+ sorting, search			

Table layout

2.3 Languages (map and list)

= list of all languages included in the database, cf. <http://tsammalex.clld.org/languages>

< csv table "Languages"

By clicking on individual language names, information on and from that language is accessible (2.3.1).

Id	Name	Glottocode	Top-level family	Lineage	2 nd languages	Region	Latitude	Longitude
default: ISO 639-3	default= Glottologue		from Glottologue	≈ genus in WALS		default: UN geoscheme		
<i>afr</i>	<i>Afrikaans</i>	<i>afri1274</i>	<i>Indo-European</i>	<i>Germanic</i>	<i>English, Xhosa</i>	<i>Southern Africa</i>	-33.6	19.4

Table layout: Headers (top row), comments and examples (base row)

2.3.1 Language X

= information on language X (cf. 2.3.2 Description) and more detailed tabular views on entries from it (2.3.3 - 2.3.5)

- 4 subordinate tabs: Description / Names (linguistic view) / Names (cultural view) / Species (bilingual comparative view)

2.3.2 Description

= information on particular languages included in the database, cf.

<http://tsammalex.clld.org/languages/afr#tdescription>

< csv tables "languages", "categories" and "habitats"

The introductory text should include the name of the editors/compiler of lexical data for the language, information on standard and dialectal varieties, orthography used under "word form", abbreviations under "Grammatical notes" and descriptions of language-specific categories (such as English translations, e.g. for German: category *Baum* "tree").

2.3.3 Names [Language X] (linguistic view)

= list of all names of plants and animals for a particular language/view on linguistic information (search of names in a particular language > species + linguistic knowledge)

cf. <http://tsammalex.clld.org/languages/afr#twords>

< csv table "Words" (Species, Word form, Basic term, ...) and csv table "Images" (thumbnails)

Name	Basic term	Categories	Taxon (English name)	Thumbnail1	IPA	Grammatical notes	Meaning	Select ... (or: Usage, from Names/Usage)
Kameldorn	Dorn	Pflanzen, Bäume, Akazien	Acacia erioloba (Camel thorn)	(picture)	kame:ldɔrn	m	camel thorn (acacia)	colloquial (construction)
+ sorting, search	+ sorting, search	+ filter, search	+ sorting, search		+ sorting, search	+ sorting, search	+ sorting, search	

Table layout

2.3.4 Names [Language X] (cultural view)

= list of all names of plants and animals for a particular language/view on cultural information (search of names in a particular language > species + cultural/encyclopedic knowledge)

cf. <http://tsammalex.clld.org/languages/afr#twords>

< csv table "Words" (Species, Word form,...) and csv table "Images" (thumbnails)

Name [preceded by basic term]	Categories	Taxon (English name)	Thumbnail1	Habitat (Language N)	Introduced?	Uses (Usage)	Importance	Select ... (or: Associations)
Dorn: Kameldorn	Pflanzen, Bäume, Akazien	Acacia erioloba (Camel thorn)	(picture)	Savanne				
+ sorting, search	+ filter, search	+ sorting, search		+ sorting, search				

Table layout

2.3.5 Taxa (bilingual comparative view)

= list of all names of plants and animals for a particular language compared to one (or two?) other language (search of names in a particular language > names in another L)

< csv table "Names" (Taxon, Word form, ...) and csv table "Images" (thumbnails)

-insert comment: "See also Taxa > Individual taxa (multilingual comparative view)"

-all taxa are listed with data from at least one of the two selected languages.

columns shown by default	Language X	Language 2 (to be selected)
--------------------------	------------	-----------------------------

Taxon	English name	Biological classification	Thumbnail 0 (general)	Name [preceded by basic term]	(Field to be selected) (or: IPA)	(Field to be selected) (or: gramm.info)	Name	(Field to be selected) (or: IPA)	(Field to be selected) (or: gramm.info)
<i>Acacia erioloba</i>	Camel thorn	Plantae --Fabales --Fabaceae --Acacia		Dorn: Kameldorn	m [grammatical info]	Dorn [basic term]	kameeldoring [here: Afrikaans]	kami:ldɔŋ [IPA]	plante, bome [categories]
+ sorting, search	+ search, sorting	+ filter, sorting		+ sorting, search	+ search, sorting	+ search, sorting	+ search, sorting	+ search, sorting	+ search, sorting

Table layout

2.3.6 Languages: Individual word forms

= all lexical data (included in the database) for an individual name in a particular language, cf.

<http://tsammalex.clld.org/values/acaciaerioloba-afr-0>

< csv table "Words"

(transcription, audio, grammatical info, plural form, ... ethnobiological notes)

2.4 Taxa: List

= list of all biological species included in the database (search of species > biological/ecological information and names)

cf. <http://tsammalex.clld.org/parameters>

< csv table "Taxa" and csv table "Names" (word forms and categories)

One language can be specified, for which columns with colloquial names and language-specific categories are added.

columns shown by default								only shown if selected		
Taxon (+ sp. description)	English name	Biological classification	Thumbnail 1	Thumbnail (Select tag, e.g. seed, fruit, ...?)	Characteristics	Eco-regions	Countries	Colloquial names in ... (L1)	Categories (of L1)	Name in ... (L2)
<i>Acacia erioloba</i>	Camel thorn	Plantae --Fabales --Fabaceae --Acacia			4-12 m, yellow flowers, ... [only first lines shown in table]	(Afrotropics): Kalahari xeric savanna	(Southern Africa): Botswana, Namibia, South Africa	doring: kameeldoring	plante, bome	DELETE THIS COLUMN
+ search, sorting	+ search, sorting	+ filter, sorting; incl. Plantae,			+ search	+ filter	+ filter	+ search, sorting	+ filter	+ search, sorting

		Animalia, Fungii as filters !								
--	--	-------------------------------------	--	--	--	--	--	--	--	--

Table layout

- "Ecoregions" should be grouped according to realm/ecozone if possible, e.g.

Afrotropic (Subsaharan Africa)

Kalahari xeric savanna

....

Neotropic (Central and South America)

...

2.4.1. Taxa: Individual taxa

= all biological information (**included in the database for a particular taxon**), links, photos and names in all languages (map and list) included for individual taxa, cf. <http://tsammalex.clld.org/parameters/acaciaerioloba>

< csv tables "Taxa", "Images" and "Words"

Fields of biological information

- Scientific name (species description)
- English name (standardized international name; obligatory)
- Biological classification (kingdom, order, family, genus)
- Characteristics
- **Habitat (biological)**
- (macroregion/) countries: downloaded from EOL?
- (ecozone/) ecoregions: specify significant ecoregions by comparing maps or countries of distribution (EOL, Wikipedia, printed resources) and map of ecoregions (Olson et al. 2004)
- general_uses?
- Notes
- References
- Links (EOL, Wikipedia, others)

Map

with common names in different languages

Images

with fields for metadata (date of picture taken, place, author, permission, source (name) + source_url?)

List of common names

Language	Lineage	Name	IPA	Grammatical notes	Meaning	References

+ search, sorting	+ filter, sorting	+ search, sorting	+ search, sorting	+ search, sorting	+ search, sorting	
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	--

table layout

2.5 Ecoregions

= map and list of ecoregions of the WWF (Olson et al. 2004) with links to the WWF description and "Taxa: List" filtered for individual ecoregions (click on ecoregions > box > links "WWF" and "Taxa in Tsammalex", respectively)

2.6. Countries

= map and list of countries (linked to > species filtered for the selected country)

2.7 References

= list of sources (author, year, title, place, publisher, links)

2.8 Images

= list of all images in the database and linked taxa (file name, license, taxon, thumbnail)

- all images (here and elsewhere, e.g. under Taxa): link to original page (and title?) must be shown

2.9 Contribute!

= information for contributors, especially this manual, a template (LibreOffice Calc file?), and information of versions (latest changes, how often new versions will be uploaded)

Text on the website:

"The expansion of the Tsammalex database strongly depends on researchers and communities willing to share their knowledge or collected data. If you are interested in a contribution and have questions, don't hesitate to contact one of the editors (Christfried Naumann, Steven Moran or Robert Forkel).

The structure of the database is based on csv tables ("...csv"), i.e., a simple file format that can be opened and edited in editors such as Notepad, or calculation programs such as LibreOffice Calc or Microsoft Excel. These csv files contain all data on biological species/taxon and related linguistic or lexical data, as well as metadata of audio and image files, etc. There is a bib file containing all references cited. Images of species (licensed as Creative Commons or in Public domain) might be uploaded to websites such as Flickr or Wikimedia Commons and linked, or sent to the editors.

Here you can download the following items:

- [csv files](#) including all data of the current Tsammalex version (cf. also Home/Download/GitHub repository)
- a [sketch manual](#) of Tsammalex with details about layout, functions and data formats
- [csv table templates](#) ready for the input of lexical ("names.csv") and biological data ("taxa.csv"), etc.

Contributions of data must be committed in the form of csv files. In a Windows environment, we recommend using LibreOffice Calc. Open the csv table templates, set the font encodings to Unicode (UTF-8) and select comma separated table. Have a look into the sketch manual about obligatory vs. optional information, and start entering data. Send these files to the editors.

Notes:

- lexical data in "names.csv" must be associated with one species or taxon (such as a genus)
- any taxon referred to under "names.csv" must be included previously under "taxa.csv""
(underlined words = links)

3. Data files (csv tables)

Abbreviations

!: obligatory tables and fields, e.g. name

(auto): data will be generated in an automatic way, based on other datasets

*: main data files

(?): necessary fields, delete?

3.0 General remarks - Editing data

- All data saved in Unicode/UTF-8!

- CSV tables should correspond to the data arrangement in the website (titles/names and order of fields)

- Don't change the layout (order of columns) of the csv tables

Only data (contents of columns B ... and rows 2 ...) may be changed. Names and order of columns may not. Data in column A ("id") may be added, but not changed.

We recommend LibreOffice Calc for editing data (free download from <http://www.libreoffice.org/>; see settings below under "Edit data in LibreOffice Calc").

Occasional contributors: Follow steps I) to III)

I) Download csv tables or templates

- <http://tsammalex.cld.org> > "Download" > "GitHub repository": list with csv files
- click csv files > click "Raw" with right mouse button and select "Save link as ..."
- or: download the csv file templates

II) Edit data in LibreOffice Calc

- open csv files with LibreOffice Calc (right mouse button > Open with > ... LibreOffice Calc)
- **Text Import settings: Character set Unicode (UTF-8), Language: Default, From row 1, Separated by Comma, Text delimiter "**
- **uncheck all AutoCorrect Options under "Tools"**
- (adjust column width, etc.)

-(for extended editing process/team work: mark your changes in colour and save as "ODF Spreadsheet" (.ods) during editing process, e.g. "words140923LS.ods"; save again as "Text CSV (.csv)" when editing process/new version is finished

III) Send your data to the Tsammalex editors

Advanced contributors: Use GitHub

- (first time: contact Robert Forkel, install GitHub and download tsammalex-data)
- update your local repository
- edit data in csv files (cf. above, "Edit data in LibreOffice Calc")
- commit your changes

3.1 contributors.csv!

= list of editors and contributors of Tsammalex (main contributors responsible for particular sections, and editors of data), for list of contributors under "Home" and citation

id!

e.g. "mj"

= abbreviation of name, by default initials of surname and first name

name!

e.g. "Miller, John"

= names of editors as they will be published on the website

sections

e.g. "Setswana", "animals of the Kalahari xeric savanna" or "plants of Southern Africa"

= sections edited by the person (individual languages [lexical and/or anthropological data > names], or biological kingdom of an ecoregion or region [biological data > taxa]). For regional subdivision, the UN geoscheme (http://en.wikipedia.org/wiki/United_Nations_geoscheme) is followed by default.

edit_ord

= order of editors

editor_sections

= edited sections

needed? (cf. "sections")

affiliation

e.g. "Max Planck Institute for Evolutionary Anthropology, Leipzig"

research_project

e.g. "Kalahari Basin Area (KBA)"

= name of the project (individual or collaborative)

homepage

= URL to institutional website

notes

= other notes concerning contributors

3.2 sources.bib!

Use BibTex format and files (.bib) for references, or send these data in a different format to the editor.

Needed: id (usually authoryear, e.g. "bleek1929"), item type (book, phdthesis, unpublished, misc), author, year, title, publisher, address

3.3 taxa.csv*!

= biological data on species and higher/lower taxa

For better editing, this table might in future be split into parts (probably by alphabetical sorting).

id!

e.g. "acaciaerioloba"

= ID of species or taxon: scientific name, all in lower case and without space (don't use "sp" for species > "anatidae" instead of "~~anatidae~~sp")

scientific_name!

e.g. "Acacia erioloba"

= scientific name as listed in 1) EOL (eol.org), 2) GBIF Backbone Taxonomy (gbif.org), 3) Catalogue of Life (catalogueoflife.org), 4) en.wikipedia, or 5) in your reference work, in this order (e.g., GBIF should be consulted only if species/taxon is not included in EOL)

- format: with spaces, first letter in upper case, other letters in lower case (except proper names of researchers), e.g. "Cephalocroton mollis Klotsch"

- abbreviations: "subsp." for "subspecies", e.g. "Acacia tortilis subsp. heteracantha"

- don't use "sp." for species, e.g. "Anatidae" instead of "~~Anatidae~~ sp."

description

e.g. "E.Mey 1843"

= species description (~ type description/Erstbestimmung)

(optional, only if relevant)

english_name!

e.g. "Camel thorn"

= international (standardized) English name, as given in 1. EOL, 2. BirdLife (birds), 3. en.wikipedia, or 4. published books or lists (in this order)

- format: singular form (also for higher taxa than species, e.g. "dung beetle" or "ant"); initial letter of 1st word in upper case, then lower case letters, except for geographic and proper names, e.g. "Southern African hedgehog" or "Muster John Henry". For several variants (dispreferred): separated by comma, 2nd variant begins with uppercase letter, e.g. "Scented thorn, Scented-pod acacia".

- if no common English can be found, indicate Genus + "sp.", e.g. "Aristida meridionalis" > "Aristida sp.".

kingdom (auto)

e.g. "plantae"

order (auto)

e.g. "Fabales"

- from 1) GBIF, 2) Catalogue of Life, 3) en.wikipedia

family (auto)

e.g. "Fabaceae"

- from 1) GBIF, 2) Catalogue of Life, 3) en.wikipedia

genus (auto)

e.g. "Acacia"

- from 1) GBIF, 2) Catalogue of Life, 3) en.wikipedia

characteristics

= characteristic/diagnostic properties of species/taxon (in short, "by which criteria did you recognize the species?" and characteristics typically to be searched for, e.g. "white flower", "thorns 3-4 cm" or "hairy legs"

- optional, but recommended (if possible, short!)

Try to give brief standardized descriptions, e.g. size, single vs. multi-stem, flowering colour, floral formula, size and form of leaves and fruit for plants, or size, colour, size and form of special body parts and distinctive features compared to closely related species for animals.

biotope

e.g. "grassland"

= biological biotope/habitat of the taxon (free field)

- optional (only if information is available, short)

countries_ids (auto)

e.g. "BW,NA,ZA", automatically downloaded from EOL

= countries of (natural) distribution (cf. EOL/Data) in ISO 3166 code (cf. Wikipedia country profiles) separated by comma and in alphabetical order, e.g. "BW,NA,ZA" for a distribution in Botswana, Namibia, and South Africa

- information has not to be complete (countries of evidence)

general_uses

= description of uses (as given in biological guides, across cultures)

- optional

notes

= other biological or ecological notes and comments

refs_ids!

e.g. "vanrooyen2001[30]" for Van Rooyen et al., 2001, Flowering plants of the Kalahari dunes, p.30

= sources/references consulted for biological information (other than EOL and en.wikipedia)

links

e.g.

= specific links you would like to add, e.g. The Plant List (for plants) or xeno-canto.org or birdlife.org (birds)

3.4 images.csv*

= links, taxa illustrated and metadata of images.

- images are in principle optional for taxa (not always available), but highly recommended

- add only images (or links) with one of the following licenses: Public domain, CC BY, or CC BY-SA (cf. links below)

id! (auto)

e.g. "d551860e4ac2af444299bd9a53144c2c"

= md5sum of an image file, used as ID

taxa_id!

e.g. "acaciaerioloba"

= ID of the associated taxon (cf. taxa.csv/id)

tags!

e.g. "thumbnail1", "general", "flower", "seed", "bark", "leaves" for plants, or "general", "head", "face", "feather", "footprint", "dropping", "female", "male", "adult" or "young" for animals

= characterization of contents in keywords (+ number); choose the best general view as "general1" (basic thumbnail)

mime_type (auto)

e.g. "image/jpeg"

src!

e.g. "http://commons.wikimedia.org/wiki/File:Cheetah_Botswana.jpg "

= URL/link to the original source, e.g. Flickr or Wikimedia

- download images always from the original/primary source (e.g. Flickr, not EOL)

title!

e.g. "Standing Cheetah (Okavango, Botswana, 2002)"

= title as given in the original source

creator!

e.g. " Paul Maritz (paulmaz)" or "unknown"

date!

e.g. "02/01/2008" or "unknown"

= date of photo taken, in DD/MM/YYYY format.

place!

e.g. "Brikama, Western, The Gambia" or "unknown"

gps

= coordinates of the place where the picture was taken (if known)

permission!

= specification of permission and link to respective URL on Creative Commons (in the format exemplified!)

- select one of the following links (or similar):

http://en.wikipedia.org/wiki/Public_domain

<http://creativecommons.org/licenses/by/4.0/>

<http://creativecommons.org/licenses/by-sa/4.0/>

source

e.g. "Wikimedia", "Flickr", "Hans Meyer"

- obligatory for images from other sites

source_URL

e.g. " http://upload.wikimedia.org/wikipedia/commons/c/c7/Cheetah_Botswana.jpg" (first input)

e.g. "<http://edmond.mpd.lmpg.de/imeji/file/d2J...>" (after data upload to Tsammalex)

first data submission: = URL of the original source (high resolution image!) > specify the URL from which the image file must be downloaded to the Tsammalex repository

after data processing: = URL of the image in Tsammalex (Max Planck Digital Library "mpdl")

comments

e.g. "correctly identified?"

= comments on metadata, status or problematic identification of species/taxon

3.5 distribution (auto)

= list of taxa and ecoregions where they have been observed

(This file is generated automatically - you don't need to edit it.)

taxon_id!

e.g. "acaciaerioloba"

= ID of the associated species (cf. species.csv/id)

ecoregions

e.g. "AT0111;AT0722"

= list of ecoregions

3.6 names.csv*!

For better editing, this table may be split into parts in future (possibly by language or region, such as names-eng.csv or names-Southern_Africa.csv).

id!

e.g. "acaciaerioloba-afr-0"

= ID of name/word form in the format [ID of taxon-ID of language-Number 0-999]

if possible, try to choose the preferred, standardized or most frequent name in a language as number "0"

name!

e.g. "kameeldoring"

= name (word form) used in a particular language to refer to the biological species or taxon, preferably in the (standard) orthographic transcription (Unicode, UTF-8) (the chosen orthographic convention should be mentioned in language.csv/description)

languages_id!

e.g. "afr"

= ID for language of the name (cf. language.csv)

taxa_id!

e.g. "acaciaerioloba"

= species or taxon to which word form is associated (prefer higher taxon if unclear, e.g. "acacia" if exact species is unknown)

Linguistic information

ipa

e.g. "k^ha'miəldɔ:ɾɪŋ" or "kami:ldoriŋ"

= (broad) phonological or (narrow) phonetic transcription in IPA [International Phonetic Alphabet]

- optional, but strongly recommended especially for names/languages in non-Latin or unfamiliar orthographies

audio

= URL of audio file (recording of word form)

- entirely optional

grammatical_info

e.g. "masculine", "plural" or "NCL1"

= grammatical category of the name (mainly gender), according to reference grammar (explain abbreviations under language description!)

- optional (highly recommended for languages with gender/noun class systems)

plural_form

e.g. "kameeldorings"

-optional

stem

- optional

root

- optional

basic_term

e.g. "doring" (for Afrikaans name "kameeldoring"), "thorn" (for English name "Camel thorn")

Can we allow multiple entries? (e.g. "Kurzhaardackel" > basic terms "Dackel" (morphologically) and "Hund" (semantically)?

= base/generic term serving as **headword for the alphabetical listing** (> "thorn: Camel thorn"). It is equivalent to the headword of indexes, and to names of plants and animals in general dictionaries (= designations by non-experts). Unless specifically elicited, it should be obtained as the (typically non-derived, mono-stem) head of the noun phrase (typically referring to genera, e.g. "Common agama" > agama [Agama agama]). **Names of subspecies, for example, may have more than one basic term, e.g. "Kurzhaardackel" > "Dackel" (morphological base) plus "Hund" (colloquial genus).** Alternative names may have different basic terms, e.g. for the alternative names "Water thorn" and "Water acacia" > "thorn" and "acacia". The generic term typically corresponds to basic-level term of Cognitive linguistics.

Biologically (semantically) wrong basic terms must not be selected, such as "*baby" for "bushbaby", "*claw" for "devil's claw".

This field should be used consistently for a particular language.

- optional (recommended for languages with modifier - head structure, e.g. English "elephant" > "African elephant")

- use consistently for a particular language!

- should be based on speakers' judgements (or: good documentations, e.g. indexes of common names in nature guides)

meaning(!)

e.g. "scorpion" or "any tree with thorns"

= denotation (exact meaning/reference/extension: what the name/word form given stands for) of word form in English

- necessary for cases where name is not precisely associated with the given taxon, but with a higher taxon or colloquial category. Example: German speakers usually refer to *Coccinella septempunctata* [Siebenpunkt-Marienkäfer] by "Marienkäfer" ("ladybird, ladybug") > name "Marienkäfer", *taxa_id* "coccinellaseptempunctata" (as well as "coccinella" [genus] and even "coccinellidae" [family]), meaning/extension in each case: "ladybird, ladybug (Coccinellidae)".

literal_translation

e.g. "bird of dawn" or "black bird"

= literal meaning of the name

usage

e.g. "taboo", "avoidance term", "literary language"

= notes on linguistic usage (sociolects, register, style)

source_language

e.g. "Latin", "Proto-Bantu"

= immediate or closest known source language

source_form

e.g. "acacia", "*mti"

= form in source language or reconstructed form

linguistic_notes

e.g. "reduplicated form", "ultimate source language probably Arabic"

= other linguistic notes.

related_lexemes

e.g. "snawel 'beak', klou 'claw'"

= field for related lexical notes in the format [vernacular_term "meaning"], such as for parts of plants, or terms for female vs. male animals or particular motion verbs.

(example_sentence, example_glosses, example_translation, example_reference ?)

Ethnobiological information

categories_ids

e.g. "afr-bome;afr-plante"

= vernacular categories to which the taxon is attributed to (hypernyms, superordinate taxa in folk taxonomy), coded with language ID ("afr-"), as listed in "categories.csv"

- optional, but consistent usage for a particular language recommended

habitats_id

e.g. "afr-woestyn", "eng-desert"

= vernacular (language-specific) categories of habitats where the taxon occurs (requires new csv table "habitats"), listed in habitats.csv (cf. 3.11)

introduced

e.g. "yes", "19th century", "by Austronesian settlers"

= free field: has the taxon been introduced to the speech area?

uses_ids

e.g. "food & drinks", "recreation"

= uses by the speech community, to be selected (**no detailed or sensitive information!**):

- based on dedicated research!

- further comments and suggestions welcome

- “animal feeding”
- “arts & crafts”
- “boats” (e.g. material used for building canoes, paddles, rafts, ...)
- “clothing & textiles” (e.g. woven clothes, furs, leather, shoes, hats, ...)
- “construction” (e.g. house building, roof thatching, barns, stables, ...)
- “cosmetics & hygiene”
- “dye” (e.g. plants used for dyeing)
- “fences” (all kinds of material used for fences and enclosings)
- “fibre” (e.g. plant fibres, wool, etc., used for ropes and strings)
- “food & drinks” (including all kinds of human food and drinks)
- “fuel” (e.g. firewood)
- “furniture” (materials used for the manufacturing of tables, seats, beds, shelves, ...)—(> "household items"?)
- “games & recreation”
- “household items” (e.g. baskets, pots, dishes, ...)
- “hunting & fishing” (e.g. weapons, traps, poisons, nets, ...)
- “jewellery & adornment”
- “medicine & healing” (all kinds of things used against illnesses)
- “musical instruments”
- “narcotics”????????
- “other” (all uses that are not included in the list)
- “religious usage”
- "shade" (e.g. trees)
- “tools” (materials used for the manufacturing of tools)
- “trade” (things that are produced or collected for selling or exchanging them)
- “transportation” (e.g. mules, carts, ...)
- “washing & cleaning” (e.g. materials used as soap or detergent)

importance

e.g. "basic subsistence crop" or "unknown by most speakers"

= cultural importance; "Smith's saliency" index or other quantification (1 ... 10)?

associations

= cultural associations (e.g. danger, messenger of death, ...)

ethnobiological_notes

= other ethnobiological notes

Metadata

comment

e.g. "unclear", "check"

= comments concerning status of information

source

e.g. "http://www.plantzafrica.com/plantab/acaciaataxa.htm"

= links to online sources on name

refs_ids

e.g. "vanrooyen2001[26]"

= references consulted for lexical or ethnobiological information (publications or databases)

original_source

= original source of information, e.g. speakers of which variety/village, specialists or non-specialist speakers, where/when elicited

3.7 languages.csv!

id!

e.g. "afr" (for Afrikaans)

= ID of language – use ISO 639-3 code by default (Ethnologue, listed also in en.wikipedia entries).

For dialects of a language, use the language name plus hyphen plus initial of dialect, e.g. "nmn-w" for the West !Xoon dialect of Taa (nmn).

name!

= name of the language – recommended: use English name of the language (cf. <http://glottolog.org/glottolog/language> or as given in first line (in bold) of data boxes in en.wikipedia entries). Try to avoid tonal and IPA symbols not common in English.

glottocode!

e.g. "afri1274" (for Afrikaans)

= glottocode (cf. <http://glottolog.org/glottolog/language>)

description!

= short text about the language, such as on genealogical classification, speech area, number of speakers, status (endangered vs. minor vs. major vs. national), dialects (and abbreviations listed under "id"), orthographic representation chosen, editor and grammatical abbreviations used in csv table "names/words".

-optional, but strongly recommended

- Template: "LANGUAGE (VARIANTS) is a FAMILY language spoken by about X million native speakers mainly in X. It is represented here in the standard orthography/a linguistic transcription. Grammatical abbreviations: f 'feminin', ..."

lineages_id!

e.g. "tuu", "bantu"

= (reconstructed or highly trusted) family/lineage of the language on a level preferred by the editors of a region. Prefer widely used, lower level families where linguistic similarities are evident to non-linguists, e.g. "Germanic" instead of "Indo-European". Compare Dryer's concept "genus" in WALS (cf. the textbox in <http://wals.info/linguoid/genealogy>).

latitude

e.g. "-33.6"

= geographical latitude of "place" where language is spoken. The point (latitude x longitude) might represent the approximate centre of the speech area, the area with the highest number of speakers, or the historical origin. The geographical coordinates may, for instance, be obtained in Google Earth.

- optional (necessary for mapping); may be copied from Glottolog or other CLLD databases

longitude

e.g. "19.4"

- optional (necessary for mapping); may be copied from Glottolog or other CLLD databases

region

e.g. "Southern Africa" for Xhosa

= region of subdivision in Tsammax (default: UN geoscheme, cf. en.wikipedia.org/wiki/United_Nations_geoscheme)

2nd languages

e.g. "eng, afr, sot, zul" (English, Afrikaans, Zulu) for Xhosa

= important 2nd or contact (neighbouring or vehicular) languages for which translations are to be exported in the language-specific inventory of names for plants and animals

3.8 lineages.csv

id

e.g. "bantu" or "khoekwadi"

= one-word terms for linguistic lineages (cf. WALS "genus"), used in "languages.csv"/lineages_id

name

e.g. "Bantu" or "Khoe-Kwadi"

= name of the lineage (family)

description

glottocode

e.g. "bant1294"

family

e.g. "Atlantic-Congo"

= top-level family (\approx phylum) as listed in Glottolog

family_glottocode

e.g. "atla1278"

color (Robert Forkel)

e.g. "ffff00"

= colour code for the representation in the Tsammalex maps

3.9 audio.csv

= links to URL with audio recording of names in particular languages.

- audio recordings are optional for names

- cf. 3.4 Images

3.10 categories.csv

= language-specific categories, such as for growth forms of plants or groups of animals (cf. "trees" or "fish" in English)

The use of categories is optional for individual languages. Categorization may be simple [e.g. "plants" vs. "animals"] or complex, but should be based exclusively on speaker judgements - to be elicited!

id! (obligatory if categories are to be included for a given language)

e.g. "afr-diere"

= ID of category (used in csv table "names/words"), consisting of language ID + word form

name!

e.g. "diere"

= word form expressing a category in the particular language (example: Afrikaans)

meaning!

e.g. "animals"

= meaning of the word form in English (one word or a description)

languages_id!

e.g. "afr" for Afrikaans

= ID of language (according to ISO 639-3)

notes

= other notes

3.11 habitats.csv

(The use of habitats is optional for individual languages.)

= language-specific names of habitats, such as "woodland", "freshwater", "coast" or "desert" for English

- should be elicited for any particular language

id!

e.g. "afr-woestyn"

= ID of habitat (used in csv table "names/words"), consisting of language ID + habitat

name!

e.g. "woestyn"

= word form of the habitat in the particular language, here: Afrikaans

meaning!

e.g. "desert"

= meaning (one word or description) of the category in English or metalanguage

languages_id!

e.g. "afr" for Afrikaans

= ID of language (according to ISO 639-3)

notes

= other notes

3.12 uses.csv

= a list of uses (in English) that may be selected under 3.6 names.csv/uses (part of culture-specific information)

- should be elicited for any particular language/culture group (general, cross-cultural uses can be indicated separately under 3.3 taxa.csv/general_uses)

id

e.g. "artscrafts"

= one-word term

name

e.g. "arts & crafts"

description

= description or other comments

4. Export functions

4.1 List Taxa - names

- sorted by kingdom/order/family, then alphabetically by scientific name, lexical subentries alphabetically)
- for "big" languages (e.g. English): constrained by (macro-)region, e.g. "English names of plants and animals in Southern Africa" or (hypothetically) "Swahili names of plants and animals in Eastern Africa"

Structure

Scientific name (2nd_language1 *name1* [IPA], *name2* [IPA], ...; 2nd_language2 *name1* [IPA], *name2* [IPA]..) Characteristics: characteristics. Biotope: biotope.

- **name** [IPA]. *Grammatical info*. "meaning", lit. "literal_translation". [usage]. Plural *plural_form*. From *source_language* *source_form*. *Linguistic notes*. Related words: *related_lexemes*. Categories: *categories*. Habitats: *habitats*. introduced. Uses: *uses*. Importance. *associations*. *ethnobiological_notes*.
- **name2 related to same species ...**

thumbnail1 thumbnail2 thumbnail3

Example (Afrikaans)

Plants

...

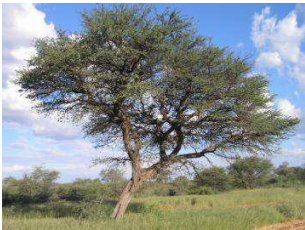
Fabales

Fabaceae

...

Acacia erioloba (Eng *Camel thorn*, *Giraffe thorn*; Tsn *mogôtlhô*, *mokala*; Kho *ǃgánàb*, *ǃgánàs*). Characteristics: 4 - 12 m tall, yellow flowers, woody pods 130 mm long, grey-green leaves 100 mm, leaflets 13 mm, straight thorns 60 mm, rough fissured bark. Biotope: arid savanna, grassland, dry river courses.

- **kameeldoring** [k^ha'miəldɔ:ŋ] *n*¹. "Camel thorn (tree)", lit. "camel thorn". [Literary language]¹. Plural *kameeldoringe*. From Dutch *kameeldoring*¹. Related words: *kameel* 'camel', *giraffe*¹. Categories: *plante*, *bome*. Habitats: *bosveld*¹. Native species¹. Uses: construction, fuel¹. Typical species of the Kalahari.¹
- **kameeldoringboom ...**



Creators: NN, MM, OO

¹ Hypothetical illustration of "Grammatical info" (this entry does not exist in the actual database).

