

Conlang Database Project

New proposal for fields and options

Written 9-10, revised 23-24 Dec 2020

Main planning document:

<https://docs.google.com/document/d/1y5QhCZpn3XcaosbP3wAEtoNGluLHHLzvateNZNAI6Ek/>

We need to finalise the choice of fields for the database before we start work on creating it. We can potentially make changes to these settings once we've started work compiling entries, but it would add a lot of work to do so. It's better to try to get the best starting point we can.

I created a particular set of fields, with particular names and requirements, in the earliest conlang spreadsheet, and those were copied in the editable spreadsheet. Those people who have added conlang details have added them within that scheme of fields and names. However they were not meant to be final, and from the start I felt that there should be a discussion amongst all those interested in creating the database as to how they could be improved. Also that time should be spent thinking about it rather than rushing into a possibly inadequate solution.

Once I opened up the question of what should be changed there were lots of suggestions, requests and opinions about how it should be. They are of course not all in agreement, but I've taken on board a lot of ideas and have tried to cover all reasonable issues in the new scheme below. I'm also mindful that it's important not to put people off from entering data due to there being too many fields and it looking like too much work. It might be a good idea if we make it clear with the proper database that filling in all fields is not compulsory, guiding contributors to concentrate on the most important/relevant factors.

Name(s) of Language (in English/interlingual)

Should be multi-line to allow for more than one name. Can include former names.

Name(s) of Language (in itself)

Should be multi-line to allow for more than one name. Should also have a facility to allow for upload of an image file showing the name in the language's conscript.

Conlang Code Registry code

Gives the code (up to 8 letters long) the CLCR has applied to the language, if it has.

Name(s) of Creator(s)

Should be multi-line to allow for more than one name per creator, and also for more than one creator. Neither real names nor pseudonyms should be required, it should be up to the

creator which or what they prefer. Real names should be entered such that they can be ordered alphabetically by family name. In the case of languages created by a group, the name of the group should be entered.

Online Links

Should be multi-line to allow for more than one link. Links should be clickable. The linked resources could be an online PDF, Google Doc, full website, individual page, forum post, wiki page, YouTube video etc. In the case of some pre-internet languages there will not be an 'official' web page or presence but if there is some page that presents a reasonable amount of information about it that could be used. We may have to consider the necessity of using a book reference instead of a weblink if there is not.

Start Year

The year the creator(s) started creating the language. When the start point is not known, enter the earliest year the language is known to have been in existence.

Physical Mode(s)

What physical media the language is expressed in, such as speech, writing, signing, other rarer media. Options:

1. Speech and writing
2. Speech only
3. Writing only
4. Sign
5. Other

Note: "Speech and writing" should only be used if there is both a 'native' pronunciation and writing system. If the language is only spoken in-world, and a Romanisation or IPA transcription is only for our benefit, that can be assigned to "Speech only".

Scripts

Should be multi-line to allow for more than one entry. Lists the scripts the language uses.

Group

For if the language is in some sort of language group, such as a genealogical one (either a natlang one in the case of altlangs, or a constructed one in a conworld situation) or else a 'suite' in the case of interrelated experimental languages. Can be left blank if the language is in no group.

Type(s)

Classification of the conlang in various types, such as the author's purpose for its creation, its setting, the uses to which it's put etc. As conlangs can be classified in several different types at once this field will allow for multiple choices.

However, the usual top-level categories such as *auxiliary*, *artistic*, *engineered* will not be used; instead the second-order types like *fantasy*, *alternate history*, *logical*, *IAL* will be the options. There should be guidance on which to pick, and it will be up to database searchers, rather than us designers or contributors/compiler, which second-order types belong to any of the top-level ones. For instance if someone wants to search all artlangs, and considers "personal languages" and jokelang to be artlangs, they can include those types in the search, along with *fantasy*, *alternate history* etc. If they want to search all artlangs but *don't*

consider personal languages and jokelangs to be artlangs, they can exclude them from the search. We just leave aside from the database design the question of what counts as what major categories of conlang.

Options:

- "Personal" (created only for the creator's own enjoyment)
- "Jokelang" (created purely for amusement)
- "Story-based" (created to feature in a formal narrative)
- "Conworld" (created as part of an imaginary (constructed) world)
- "Future" (imagined as arising in the future relative to the time of construction)
- "Geofictional" (imagined as being in a fictional part of our world)
- "Alternate History" (imagined as part of an alternative timeline)
- "Lostlang" ([imagined as an undiscovered part of our timeline](#))
- "Exo-/Xeno-lang" (imagined as being used by alien/ET beings)
- "Pseudo-Auxlang" (created to mimic auxlangs but without an intention of auxiliary purpose)
- "Global Auxiliary" (intended for auxiliary use by the whole world)
- "Zonal Auxiliary" (intended for auxiliary use by a limited area of the world)
- "Other Auxiliary" (intended for some other sort of auxiliary use)
- "Ideal" (created to attempt an optimum representation of ideas)
- "Philosophical" (expressing a philosophical viewpoint)
- "Logical" (aiming to express meaning without syntactic ambiguity)
- "Experimental" (testing or demonstrating linguistic possibilities)
- "Conpidgin" (forming a new language by collective evolution)
- "Spiritual/Mystical/Ritual" (inspired mystically or used for ritual purposes)
- "Secret" (stealthlangs, for clandestine communication)
- "Other"

Vocab Source

Classification of the conlang by the origin of its vocabulary. Based on extensive discussion, there should be the following options:

1. 'A priori'/original vocab/made up from scratch/'ex nihilo' - including where the language has been derived from another one the same creator has created in the same constructed family.
2. 'A posteriori'/vocab derived from natlang(s)
3. 'A posteriori'/vocab derived from conlang(s), except where the creator has merely 'derived' one language from another in a family they've created. The conlang copied from should ideally be someone else's creation.
4. A mixture of several of the above
5. Other/Unknown

This issue can be confusing but I don't think we can omit it because it has been a key question in analysis of some kinds of conlang such as auxlangs. It also has some relevance to alllangs. However it would get too confusing to make further distinctions and we'll have to put aside the status of other aspects than vocabulary, such as phonology.

Development Level

How developed the language is, from sketchlang at one end to native speakers at the other. I've added some finer distinctions around the 'developed' area. I'm reluctant to include options like "finished" or "abandoned" because these are not levels of development of the language but current decisions by its creator. Options:

1. "Unknown"
2. "Sketch/Minimal" (either a mere plan or a superficial fictional language)
3. "Some Development" (moving beyond a sketch)
4. "Considerable Development" (halfway to usability)
5. "Well-developed" (the language has been built to a usable level)
6. "Learners" (the language has attracted people to learn it besides its creator)
7. "Active Community" (users are conversing and developing the language)
8. "Fluent Users" (some learners have achieved fluency)
9. "Native Users" (some children have learned the language as a mother tongue)
10. "Other"

Notes

For contributors to add noteworthy information about the conlang that doesn't fit elsewhere. Can contain a "feature summary" or "features of interest". This field should take the form of free text entry.

Of the above, four fields will have a menu of limited options. This will group conlangs of the same kind together, which will help people find them by searching for that kind. The alternative of allowing free text is possible, and would be preferred by some, but would have the negative effect of reducing the accuracy of statistics of each type and returning fewer hits for type searches. There will always be disagreement about how conlangs should be classified, or what terms people use, and the evidence from the spreadsheet shows many creators or other people adding data can be quite individual in how they answer these questions. The rights and feelings of creators should always be considered, but since the purpose of the database is to be accurate and useful we will also have to think in terms of regularising the entries for these fields. However each of them should have an 'Other' option to let contributors avoid having to apply types they don't agree with (or understand). And it's entirely possible that good arguments could lead to us adding new types in some of these fields after the database is started.

What has changed?

The following have been removed:

- The **Lexico-Semantics** field, for classification of conlangs according to the organisation of vocabulary in relation to areas of meaning, e.g. taxonomic vs. naturalistic. This was primarily intended to catch taxonomic languages, but they are a small minority and this is the sort of information that can just be put in **Notes**.
- The original **Purpose Type** field, with top-level options of artistic, auxiliary, engineered etc.

The following have been added:

- The **Conlang Code Registry code**, **Scripts**, **Group** and **Notes** fields.
- More options in the **Purpose Type** (now **Type(s)**) and **Development Level** fields.

The following have been changed:

- The three '**Names**' fields are now multi-line to allow for multiple names.
- Some of the field names have changed, e.g. "**Page**" is now "**Online Links**".
- Some options have been changed for the fields with a limited list, e.g. **Physical Modes** and **Vocab Source**.

There are now 13 fields where there were 10 in the original spreadsheet (plus **Notes** added more recently).

Overall summary of changes:

This new scheme adds a few more fields, but not many more, and a couple have been removed which makes that easier.

Suggestions for fields for e.g. previous names or pseudonyms have been accommodated by making the name fields multiple entry rather than single entry. This is neater and will look less complicated on the eventual website data input form.

More options have been added, and others changed or renamed, in response to suggestions and reminders. At least one multi-option field has been made multiple choice.

A few have been removed as involving too much complication.

Field names have been changed to become clearer.

All those involved in setting up the database are welcome to critique this new proposal including by making comments or suggestions on this document or by discussing in our various discussion sites.

Matthew McVeagh