

Using CLDF and CLLD for language documentation and description

Florian Matter
University of Bern

2019-08-25

UPPLADOC, Uppsala

1 Introduction

2 Demo

3 Issues and open questions

The project

- **What?** A fully interactive online description of a language
- **Why?** Digital tools are available and used, but the resulting descriptive material is usually in traditional format (except dictionaries)
- **How?** Combining a bunch of existing applications with Python

“Fully interactive”

- instances (recordings) of linguistic utterances
- brought into phonemic representation and parsed into morphemes, based on linguistic analysis
- linguistic analysis written in prose form, supported by examples
- collection of morphemes in dictionary format
- **everything should be linked**

CLLD

- Forkel, Bank, et al. (2019)
- cross-linguistic **linked** linguistic data
- WALS, [Glottolog](#), [Phoible](#)...
- online interfaces for relational databases with linguist content
- common framework, easy automatization, encourages consistency and explicitness

CLDF

- externalized version of the principles underlying CLLD (Forkel, List, et al. 2017)
- combination of JSON (metadata) and CSV (content)
- CSV is simple, but encourages consistency
- files are light and software-independent (no proprietary format)
- requires little training to create and edit (?)

The workflow

1. Annotation in ELAN
2. Export to and parsing in FLEx
3. Adding metadata (CSV)
4. Conversion to CLDF and extraction of individual audio files
5. Creation and deployment of CLLD database

The application



Demo

Issues

- FLEx does not export Morpheme IDs in `.flex` text files?
 - ...nor does it export allomorphs?
- “Markdown” is hacky
- issues with providing recordings and metadata?
- CLLD is useful for representing linguistic variation, what about FLEx and CLDF?

Questions / feedback?

References

-  Forkel, Robert, Sebastian Bank, et al. (2019). “cldd: a toolkit for cross-linguistic databases”. DOI: [10.5281/zenodo.3239095](https://doi.org/10.5281/zenodo.3239095). Online: <https://doi.org/10.5281/zenodo.3239095>.
-  Forkel, Robert, Johann-Mattis List, et al. (2017). *CLDF 1.0*. DOI: [10.5281/zenodo.1117644](https://doi.org/10.5281/zenodo.1117644). URL: <https://doi.org/10.5281/zenodo.1117644>.