

Flextext interlinear import preserving morpheme glosses

Aleksandr Riaposov, Alexandre Arkhipov, Elena Lazarenko

University of Hamburg

Github repository:

<https://github.com/git-debase-verbose/FWdata-Import>

This tool can import interlinear glossed texts (IGTs) into an existing SIL FLEx project.

Read more in the [\(Riaposov et al. 2025\)](#) paper provided under /papers.

Prerequisites

Data

- One or more IGTs in **flextext XML** format

This can originate from either Toolbox, ELAN or FLEx.

- **FLEx** has a built-in export into flextext XML (`Export interlinear as "ELAN, SayMore, FLEx"`).
- **ELAN** has a built-in export into flextext (`Export as "FLEx File"`).

These files will probably need a preprocessing step to treat punctuation, unless previously exported from FLEx into ELAN.

- **Toolbox** files can be converted into flextext via ELAN: first import into ELAN with the built-in Toolbox import.

- Existing **FLEx project** (setup a new one if none)

This must exist before you start the scripts. In particular, you should:

- Know the exact **writing system codes** for (i) the analyzed text (main *Baseline* or *Vernacular language* writing system), (ii) the glosses and translations (two *Analysis languages* are supported). These codes must be provided in the launcher dialog.
- Make sure that each **part-of-speech label** used in the IGT matches one existing in the FLEx project (`Category > Abbreviation` in the first *Analysis language* writing system).

System requirements

- Any reasonably recent version of [Python 3](#)

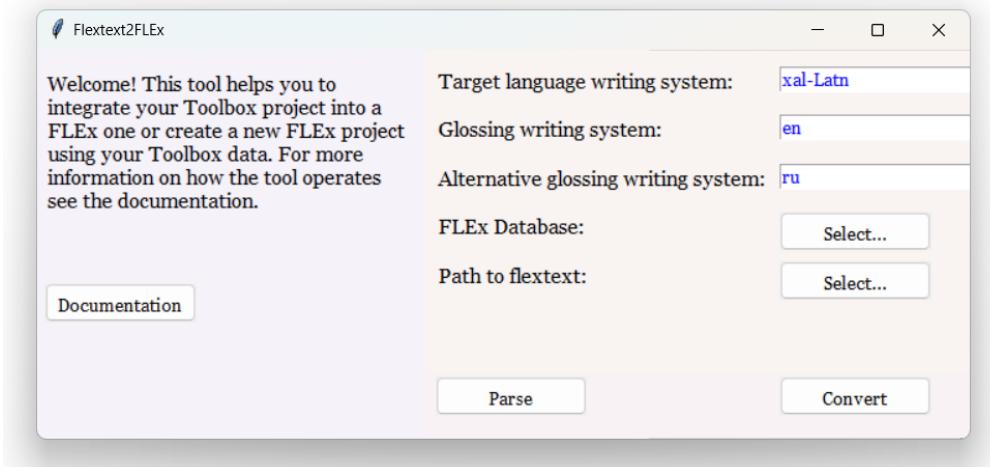
- Currently, the launcher does not yet handle the preprocessing XSLT scripts which require **XSLT 3.0/XPath 3.1** support. They can be run e.g. with the free editions of [Saxon](#) (developed with Saxon 9.9, [current version](#) is Saxon 12)
- [SIL FLEX \(FieldWorks Language Explorer\)](#) (tested with version 9.*)

FLEX does not need to be necessarily installed on the same system as the conversion scripts, but you will likely need to move some files back and forth between the two.

- [ELAN](#) Multimedia Annotator -- for texts glossed in [Toolbox](#) and/or in ELAN

How to use

- Setup/use a FLEX project (see above)
- Prepare the FLEX project file (*.fwdata*)
 - **NB! Backup the project** (`File > Project management > Backup this project`) and close the project.
 - Locate the *.fwdata* file in the FLEX Projects folder (typically `c:\ProgramData\SIL\FieldWorks\Projects\<ProjectName>\<ProjectName>.fwdata`). Alternatively, you can extract it from the *.fwbackup* file which is a regular zip archive.
- Put all your *.flextext* IGTs into one folder (nested subfolders are possible).
- Run *launcher.py*



- Enter writing system codes (Target language = *Vernacular*, Glossing & Alternative glossing = *Analysis*)
- Select the FLEX project file (*.fwdata*) (FLEX Database)
- Select the folder containing the IGTs (Path to flextext). All files in the selected folder (and subfolders) will be processed.
- Click Parse. *parserFWdata.py* analyzes the *.fwdata* file to list objects (wordforms, morphemes, analyses etc.) already present in the FLEX project.

- Click Convert.
`converter.py` creates an augmented copy of the original `.fwdata` and stores it in the folder containing the scripts as `Converted-<ProjectName>.fwdata`.
- Replace the original `.fwdata` file with the new augmented copy.
 - Rename the new copy identically to the original `.fwdata` (i.e. remove the `Converted-` prefix) and
 - either replace the original `.fwdata` file in the FLEx Projects folder and open the project in FLEx as usual
 - or
 - copy the `.fwbackup` file (to keep an unaltered version of the backup!)
 - replace the original `.fwdata` file inside the copy of `.fwbackup` file (which is a zip archive)
 - restore the project from this new backup
- Inspect the new project in FLEx and cleanup
 - It is highly recommended to run some cleanup utilities from the Tools > Utilities menu:
 - Find and Fix errors in a FieldWorks data (XML)
 This utility must be run **from another project** (close the current project, open another, run the utility and select the target project).
 - Reopen the target project and run the following utilities: Merge Duplicate Analyses, Merge Duplicate Wordforms and Write Everything.
 - Create a fresh backup.
 - Manually review changes in homonyms, allomorphs, multiple senses in the lexicon. You might need to merge or split some of them.
 - If something is still wrong, restore from the original (unaltered) backup and let us know -- perhaps we can find a solution.