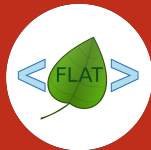


FLAT

FoLiA Linguistic Annotation Tool



LANGUAGE MACHINES

**CLST | Centre for Language and
Speech Technology**
Radboud University



Introduction (1/2)

What is FLAT?

- **FoLiA Linguistic Annotation Tool**
- A web-based tool for linguistic annotation of language resources
- Many different annotation types
- Flat is anything but “flat”, it deals with deeply nested data and retains document structure.
- Aims to be a fully-featured annotation environment for FoLiA documents (currently in an early beta stage!)



Introduction (2/2)

What is FoLiA?

- Format for **L**inguistic **A**nnotation
- An XML-based file format for storing and exchanging language resources
- Generic paradigm independent of language or label set
- Single format for all
- Implemented in many tools (*ucto, frog, ucto, pynlpl, ticcl, folia-tools*)
- Used in many projects (*SoNaR500, Nederlab, DutchSemCor, Basilex, OpenSoNaR*)
- Funded by CLARIN-NL



Notable Features

- Modern **web-based** interface
- **Multi-user** environment
- Download and upload FoLiA XML documents
- **Concurrency** (multiple users may edit the same document simultaneously)
- Storage model with [git] **versioning control** (unlimited undo operations)
- Full **annotator information**: who annotated what and when?
- Layered, retains **document structure**
- Annotations may pertain to single tokens or span over multiple (span annotation)
- Support for **corrections** of either text or linguistic annotations. Corrections are never mere substitutes, originals are always preserved!
- Support for alternative annotations.
- Introduces FoLiA Query Language for automated and batch editing



Visualisation of annotations on hover

Stemma

Stemma is een ander woord voor **stamboom**. In de historische wetenschap wordt zo'n stamboom, onder de naam stemma verwantschap tussen handschriften weer te geven.

Werkwijze

Hiervoor worden de handschriften voor het originele handschrift. De andere handschriften waarvan men toch vermoedt dat ze ook van de oorspronkelijke afstamming afstammen, worden ook in het stemma opgenomen en worden weergegeven.

Tenslotte gaat de verwantschap tussen de handschriften duidt op een verwantschap, terwijl een andere handschrift duidt op een onzekere verwantschap.

- Eerste testitem
- Tweede testitem

Een stamboom

Text

undefined

stamboom

Part-of-Speech

cgn-combinedtags

N(soort, ev, basis, zijd, stan)

Lemma

lemmas-nl

stamboom

Syntax

syntax-set

pp

voor stamboom

Dependency

alpine-set

obj1

voor stamboom

Chunk

chunk-set

undefined

voor stamboom

timesegment

event-set

utterance

voor stamboom

Token • WR-P-EJ-0000000001.p.1.s.1.w.7 • undefined

Global visualisation of classes

Legend • Named Entity

(Hide)

- N(soort,ev,basis,zijd,stan)
- VZ(init)
- LET()
- LID(bep,stan,rest)
- BW()
- WW(vd,vrij,zonder)
- ADJ(prenom,basis,met-e,stan)

Stemma

Stemma is een ander woord voor stamboom. In de historische
verwantschap tussen handschriften weer te geven.

Werkwijze

Hiervoor woorden de handschriften genummerd en gedateerd
voor het originele handschrift. De andere handschriften krijgen
waarvan men toch vermoedt dat ze ooit bestaan hebben worden
haken geplaatst.

Tenslotte gaat de verwantschap tussen de mooie handschriften
volle lijn
duidt op een verwantschap, terwijl een stippellijn op een onzeker

Editor

De FoLiA developers zijn:

Naam	Universiteit
Maarten van Gompel	Radboud University Nijmegen
Ko van der Sloot	Tilburg University

Annotation Editor x

example.table.1.w.5

Text
undefined

Named Entity
<http://raw.githubusercontent.com/proycon/folia/master/setdefinitions/namedentities.foliaset.xml>

Maarten van Gompel

Person

Location

Miscellaneous

Organisation

Person

Select Multiple

Ok

In the field

Currently used in two projects

1. A project with *College van Examens*: Spelling correction on texts of children and on automatic corrections by Valkuil.net .
2. A project concerning the annotation of named entities in historical dutch texts from World War II, with *Susan Evans*, *Marten Düring* and *Antal van den Bosch*



Demo

- Demo: <http://flat.science.ru.nl>
- Source: <https://github.com/proycon/flat>
- FoLiA website: <http://proycon.github.io/folia>

