Collaborative Dependency Annotation

Kim Gerdes
LPP (CNRS), ILPGA
Sorbonne Nouvelle

ARBORATOR

http://arborator.ilpga.fr

The Arborator software is aimed at collaboratively annotating dependency corpora.

Arborator is free software. It is realeased under the AGPL v3 Free Software Free as in Freedom licence

downloadable on launchpad

It can easily be installed on any Apache server.

relies on browser's SVG capacity
principally for use in the Firefox browser.

design choices:

- Zero setup: The tool must run on any computer without any difficult adaptation or installation procedure.
- Central storage of texts and annotations.
- Multi-audience interface: For professional annotators, it needs to include numerous keyboard short cuts for all common annotation tasks, and for starters, the annotation process has to be graphical and self-explanatory.

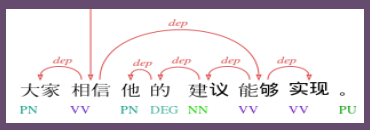
→

- web-based application
  - without plugins
- runs in a standard-compliant browser
- SVG representation
- correction of pre-annotated corpora
  - or creation of tree structures from scratch
- multi-governors + lemma and direct simple features under token + complete feature table

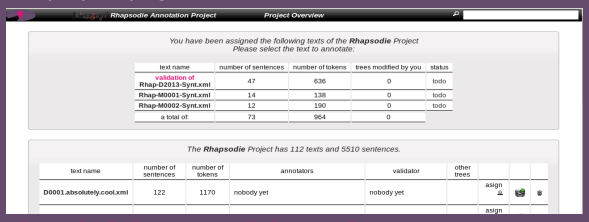
→

Technical design choices:

- Python over a Sqlite database, client-side interactions in Javascript (Jquery).
  - any Python-CGI capable Apache webserver.
- Optimized for Firefox browser.
- Multi-level user hierarchy: site administrator, corpus administrator, validator, assigned annotator, visitor.
- The appearance of the dependency structures is highly configurable in simple configuration files.
- fully Unicode capable:

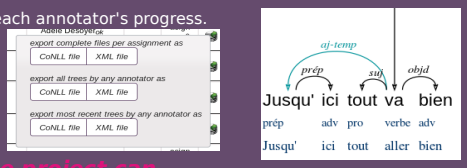


The project page



The administrator of the project can

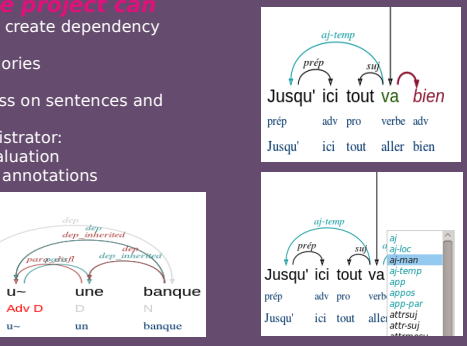
- attribute any text to a user's annotation or validation tasks,
- export the data in multiple formats and configurations
- add whole texts to the project, plain texts or pre-parsed data in CoNLL or Rhapsodie-XML format.
- Check the consistency of the annotation by obtaining tables of frequency distribution of features and 2-node connected sub-graphs of the dependency graph.
- Obtain an overview of each annotator's progress.
- Export the annotations



The editor page

The annotator of the project can

- Drag and drop words to create dependency links
- Choose or correct categories
- Change other features
- Mark annotation progress on sentences and texts
- If allowed by the administrator:
  - See numerical evaluation
  - Compare to other annotations



The experiment

carried out with a

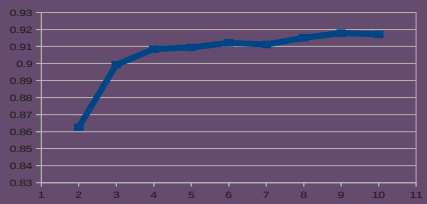
- class on corpus linguistics
- 60 3rd year linguistics majors in a French university
- about 75% French mother tongue
- Only 3 main classes and 3 tutorial session in smaller groups
- all those students had
  - other classes on syntax,
  - one class on dependency syntax with similar notational conventions as in the annotation guide
  - online annotation guide containing many concrete examples
- two sets of sentences:
  - the mini gold-standard annotated by the researcher
  - the non-annotated sentences, considered as unlimited
- Annotation task distribution: automatic, based on:
  - *t*, total number of tokens per student to annotate
  - *g*, number of sentences from the pre-annotated mini gold-standard to mix into the student's task
  - *n*, number of annotations per sentence (taken from the non-annotated sentences)
  - *p*, percentage of sentences that can be equal from one task set to the other
- each student: 48 sentences, mostly from French Wikipedia
- average length: 24.7 tokens per sentence (= real world)

The results

- annotations are kept when 80% of the words were annotated
  - only students who at least annotated 5 sentences
    → 42 student annotators.
- Results range from 64% to 90% of correct government relations (F-score), and 79% average

  How many sentences do we have to take into account in the evaluation if we want to keep similarly precise evaluation scores of the student, needed for the rover?
- the student evaluation varies very little if we base it on the first half of the corpus only (less than 1% in average), the quality of the annotation is better (80%) on the first half,

How many sentences are needed for good rovering results?



F-score in number of annotations per sentence