

FreeLing: From a multilingual open-source analyzer suite to an EBMT platform

David Farwell and Lluís Padró
Centre de Recerca TALP
Universitat Politècnica de Catalunya
`{farwell, padro}@lsi.upc.edu`



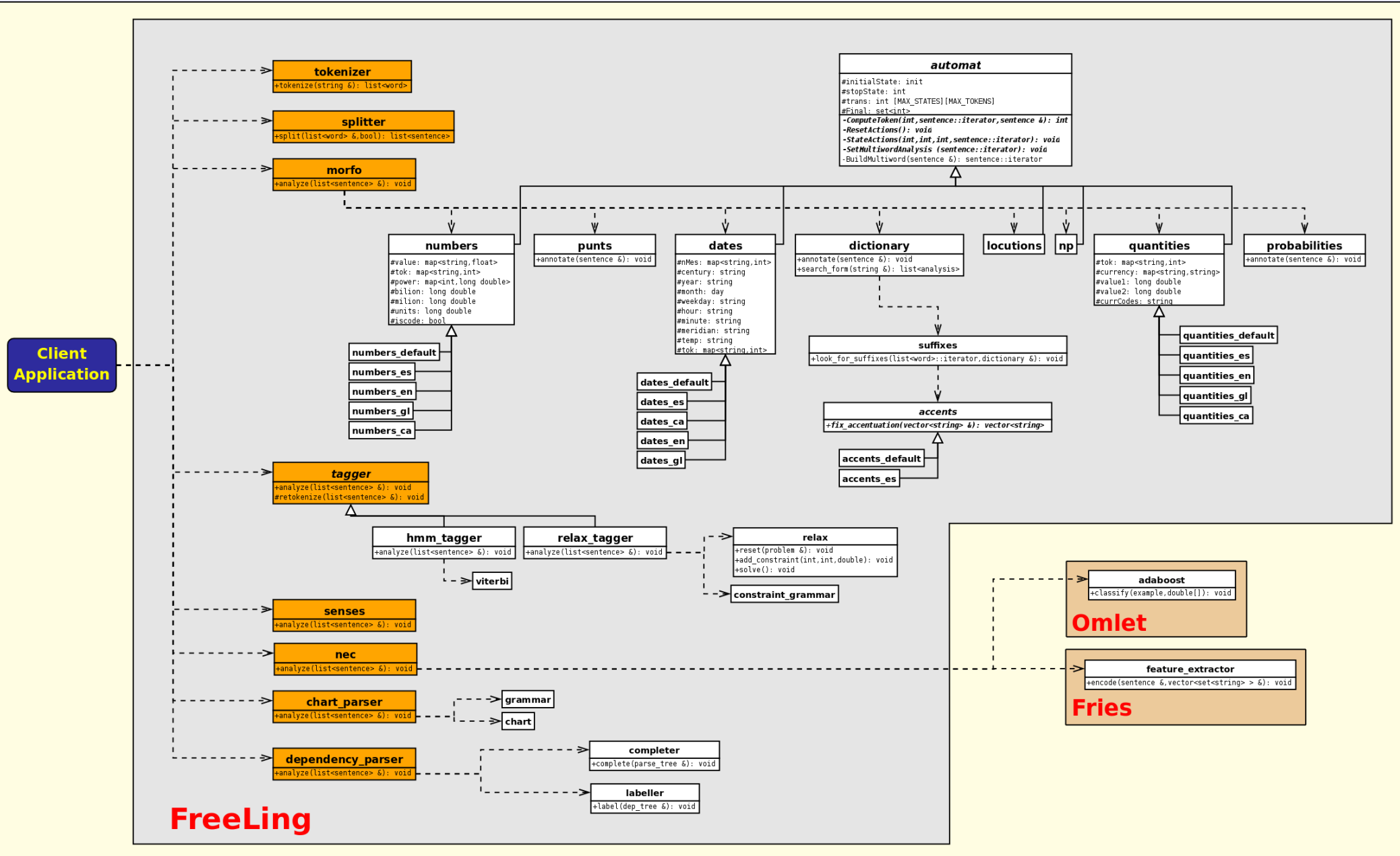
Outline

- FreeLing architecture and functionalities.
- Adapting it to new languages
- Sample applications
- Towards an EBMT open-source platform

Introduction

- What is FreeLing ?
 - A configurable and extensible linguistic analysis library, developer-oriented.
- What is FreeLing not?
 - A user-oriented off-the-shelf linguistic analyzer.
- What do people use it for?
- What has it to do with EBMT?

Processing Classes

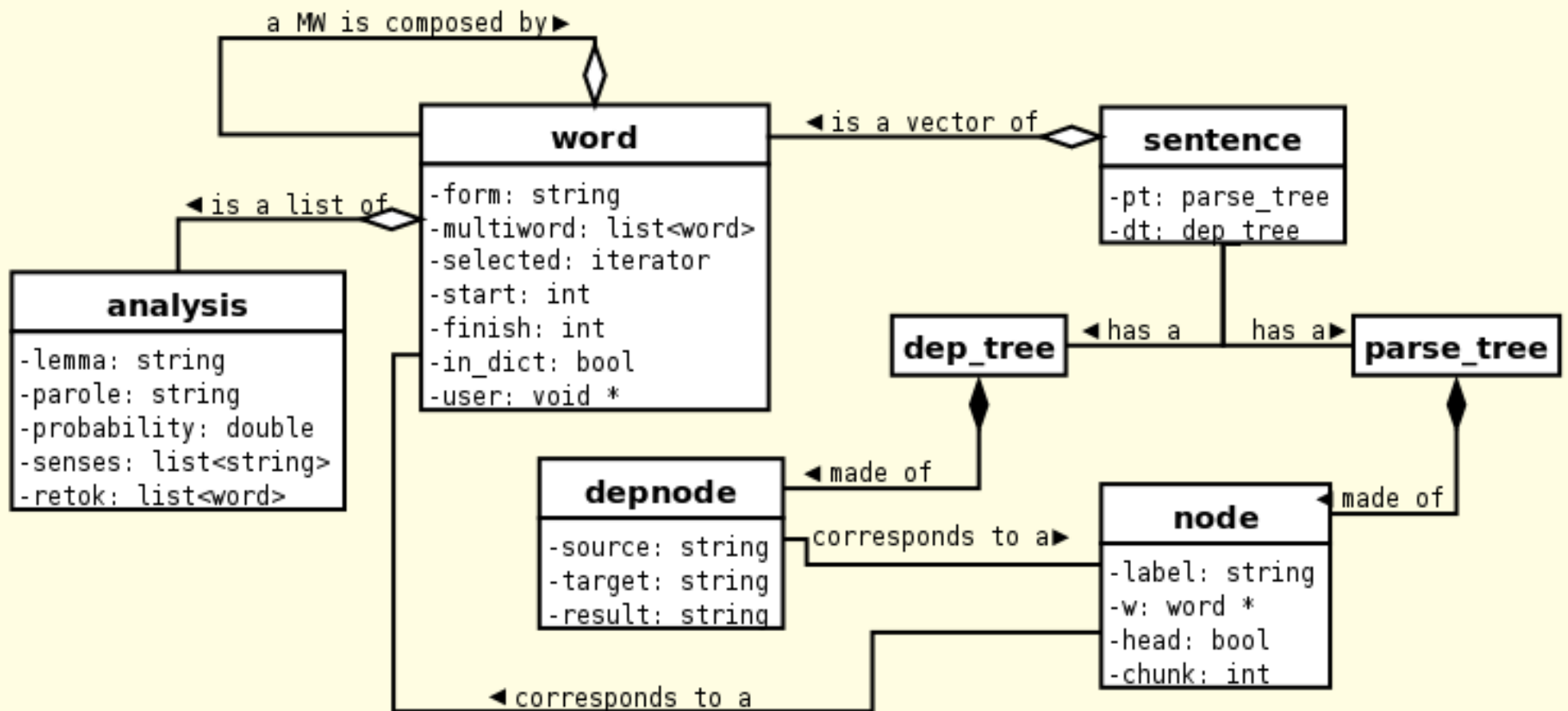


FreeLing

Omlet
adaboost

Fries
feature_extractor

Linguistic Data Classes



Including new languages (1)

- Tokenizer & Splitter:
 - Adapt config files.
- Morphological analyzer:
 - Index form dictionary
 - Adapt suffixation rules
 - Provide (if any) multiwords file
 - Provide (if any) magnitudes file
 - **Develop** (if needed) date, number, and quantities modules

Including new languages (2)

- Tagger (and probabilities module)
 - Use a tagged corpus to train taggers and compute lexical probabilities. Scripts are provided with FreeLing
- Chart parsers and Dependency parsers
 - Develop appropriate grammars (or adapt some of the existing ones to the new language)

Including new languages (3)

- NE Classifier
 - Train using annotated corpus
- Sense annotator and UKB disambiguator
 - Provide WN-like data files
- Coreference resolution
 - Use annotated training corpus with provided scripts.

Some NLP applications using FreeLing

- OpenTrad www.opentrad.org
- VKM (Cromosoma,
essukaia.lsi.upc.edu/~vkm)
- T-Incluye (Fundación CTIC
www.tincluye.org)

Towards an EBMT open source platform

- Two approaches:
 - Common architecture to integrate the EBMT system components (data interchange formats, programming, language, etc.)
 - A set of open source tools loosely glued together (as SMT does)

Towards an EBMT open source platform

- An open-source EBMT platform will require existing open-source analyzers and resources.
- Many architectures to choose from (NLTK, GATE, UIMA, ...)
- FreeLing offers homogeneous tools, and a data structure for interchange.

FreeLing: From a multilingual open-source analyzer suite to an EBMT platform

David Farwell and Lluís Padró
Centre de Recerca TALP
Universitat Politècnica de Catalunya
`{farwell, padro}@lsi.upc.edu`

