UDLex: Towards Cross-language Subcategorization Lexicons

G. Rambelli[†], A. Lenci[†], T. Poibeau[‡]

[†] Computational Linguistics Laboratory
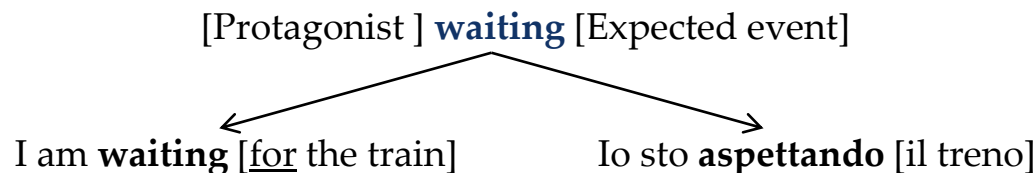*University of Pisa, Italy*

[‡] LATTICE laboratory
CNRS, ENS, U Paris 3, PSL Research U, USPC, France

International Conference on
Dependency Linguistics
Depling 2017, Pisa, September 18-20 2017

exploring dependency grammar, semantics, and the lexicon

- Create a **multilingual subcategorization lexicon** for the analysis of the syntactic realization patterns of verbs arguments across languages.

  - Synonymous predicates across languages occur with similar or different morpho-syntactic frames?

  - The same valency frame in two languages is instantiated or not by synonymous constructions?

  [Protagonist] **waiting** [Expected event]

  I am **waiting** [<u>for</u> the train]   Io sto **aspettando** [il treno]

- Goal: explore possible universals concerning the relationship between form and meaning at syntax-semantic interface.

- Develop a computational framework to automatically acquire the combinatorial properties of verbs.

- Exploiting **Universal Dependency** annotation scheme to build an universal set of rules for detecting verb dependencies in a sentence.

- *UDLex* extractor algorithm is divided into

  1. **Universal extractor** A common set of rules based on UD core argument relations (subject, object, indirect object..)

  2. **Language Specific functions** External modules for each language to undergo certain grammatical transformations, such as relativization and passivization

RESULTS AND DISCUSSION

- Resulting lexicons built from Universal Dependencies 2.0 treebanks for English, Italian, French and Finnish.

- Each verb is described by:

  - a **syntactic profile** that lists all a verb syntactic contexts (argument /SCFs)

    » *play:* SUBJ#OBJ#COMP_in, SUBJ#OBJ, SUBJ#0, SUBJ#COMP_in#COMP_with …

  - a **lexical set** with the typical lexemes occurring in each slot of a SCF.

    » *play* + OBJ: role, chess, match , song , movie..

RESULTS AND DISCUSSION

- Let's try *UDLex* to typologically different languages!

- Next step: align all languages into a single database in order to facilitate the comparison of the lexicons

See you
at poster session!

UNIVERSITY OF PISA

COMPUTATIONAL LINGUISTICS LAB

http://colinglab.humnet.unipi.it

- Create a **multilingual subcategorization lexicon** for the analysis of the syntactic realization patterns of verbs arguments across languages.

  - Synonymous predicates across languages occur with similar or different morpho-syntactic frames?

    » I am **waiting** [for the train] vs Io sto **aspettando** [il treno]

  - The same valency frame in two languages is instantiated or not by synonymous constructions?

- Goal: explore possible universals concerning the relationship between form and meaning at syntax-semantic interface.