Lexicon Formalisms - Summary

Kurt Eberle

Juli 19 2016
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Lexical information for NLP

- basic lexical information for processing of sentences (morphologic, syntactic semantic analysis)
- additional information (for human use, high quality analysis, question answering, and translation)
- representation of lexical information (standards, experience from mental lexicon)
- learning of lexical information (lemmas, morphosyntactic, semantic information, multiwords, idioms, relations (thesaurus information), word sketches)
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Basic information for NLP

- lemma
- stems, secondary stems, inflection classes
- word senses
- subcategorization frames
- syntactic and semantic typing
  (control verbs, pred_use_only, human, artefact, ...)
- selection restrictions on subcategorized roles
  (PoS, more-detailed morphosyntactic description, semantic type)
→ mostly available in the lexical components of implementations of instances of usual grammar theories (HPSG, LFG, Slot Grammar, Categorial Grammar)
→ differences with respect to partitioning the information into: lexicon, lexical rules, grammar
→ usually, information can be translated from one grammar formalism into another
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Additional information for NLP and human use

- **word sketch** information
  - collocations (following various syntactic patterns: *adj-noun*, *verb - direct obj*, . . . )
  - similarity of / difference between items with respect to collocations
  - classification of collocations (collocation types), hierarchies
  - references

- thesaurus information
  - hyponyms, hyperonyms, antonyms, meronymy, . . .
  - *(WordNet information)*

- analytical information (for high quality analysis and inferences)
Additional information for NLP and human use

- *word sketch* information
- thesaurus information
- analytical information (for high quality analysis and inferences)
  - argument structure (in terms of semantic roles: syntax-semantics interface/mapping)
  - event structure (including *Aktionsart*)
- qualia structure
  - constitutive role (material, ...)
  - formal role (color, orientation, shape, ...)
  - telic role (purpose)
  - agentive role (bringing about, causal chain, ...)
- lexical inheritance structure
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Representation of lexical information

- lexicon subformalisms of grammar formalisms (HPSG, LFG, Slot Grammar, Categorial Grammar)
- Standardized module structure for lexical information
  
  *Lexical Markup Framework*

  - lexical entry, syntax module, semantics module, ...
  - specification of core NLP information

- Standards of surface form of information
  - OLIF (Open Lexicon Interchange Format (initiative ended))
  - ISO (International Organization for Standardization)
    DIN (Deutsches Institut für Normung)
  - RDF (Resource description framework)
  - TEI (Text encoding initiative)
  - TBX (TermBase Exchange)

- experiences from mental representation
  - storage of inflected forms (less morphological operations) vs.
  - derivational forms on the fly
  - weighting of the importance of typical types of structuring:
    prototypes (less) (vs) thesaurus relations (more) etc.
Outline

Lexical information for NLP

Basic information for NLP

Additional information

Representation of lexical information

Learning of lexical information
Learning of lexical information

- resources
  - conventional dictionaries
  - monolingual, bilingual corpora with/without annotations
- type of information
  - idioms
  - lemmas
  - subcategorization frames
  - word sketches
  - classifications of collocations
  - classification/analysis of compounds
  - Aktionsart (behavior)
  - ...