UBY-LMF – A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF

Developing an LMF-based model for large-scale lexical-semantic resources (LSRs)

Problem: Incompatible formats of (integrated) LSRs
Solution: Using LMF for modeling LSRs
Problem: Only few types of LSRs have been modeled in LMF so far
Solution: Creating a new, comprehensive lexicon model

Vision: A lexicon model which
1. Represents all information in various LSRs
2. Is extensible by further information types
3. Ensures structural and semantic interoperability

Contributions:
1. UBY-LMF: Lexicon model
   - ISO Standard Lexical Markup Framework (LMF)
   - Expert-built and collaboratively created LSRs
   - Fine-grained modeling of information types
   - Attributes and values refer to ISOCat data categories
2. UBY
   - 9 LSRs in English and German
   - 9 pairwise sense alignments between LSRs
   - Confirms the applicability of the model
3. UBY-API:
   - Uniform access to LSRs in UBY via JAVA API

Extensions to LMF
- SemanticLabel
  - Subclass of Sense, SemanticPredicate, SemanticArgument
  - Used for domain, style, register...
- Frequency
  - Attached to most classes
  - Used to encode frequencies derived from corpora
- SubcatFrameSetElement
  - Subclass of SubcategorizationFrameSet
  - Used for linking frames with arbitrary cardinality

New Relationships
- SemanticArgument
  - Definition
  - Allows to encode definitions of FrameNet frame elements
- MonolingualExternalRef
  - Context
  - Allows to encode annotated corpus sentences
- SenseRelation
  - FormRepresentation
  - Alleviates a special kind of ambiguity in WordNet
  - Targets of relations and translations are word forms, not senses (as required by LMF)
  - ⇒ Relation has a null target, but an attached word form

Novel Use of SenseAxis
- Used for encoding cross-lingual and monolingual alignments between LSRs

Instantiation: UBY
10 UBY Lexicons represented in UBY-LMF
Solid lines, 9 sense alignments; Dotted lines: planned sense alignments

Using UBY
UBY-API with Hibernate
- NLP application:
  - Easy swapping of resources
  - Easy combination of resources

Open licenses for data and software:
- UBY-API, LSR to UBY conversion routines: http://code.google.com/p/uby/
- UBY-LMF DTD, UBY release 1.0 (MySQL database dump): www.ukp.tudarmstadt.de/uby/

See also
UBY:
Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer
UBY – A Large-Scale Unified Lexical-Semantic Resource

Subcat-LMF
Judith Eckle-Kohler and Iryna Gurevych: Subcat-LMF – fleshing out a standardized format for subcategorization frame interoperability

This work has been supported by the European Research Area under the grant No. 798/3-1 and by the Volkswagen Foundation as part of the Leibniz Professorship Program under the grant No. 142086.