

UBY – A Large-Scale Unified Lexical-Semantic Resource



TECHNISCHE
UNIVERSITÄT
DARMSTADT



UKP

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann,
Michael Matuschek, Christian M. Meyer, Tri-Duc Nghiem

<http://www.ukp.tu-darmstadt.de/uby/>

UBY in a Nutshell

Developing a large-scale unified lexical-semantic resource (LSR) in a standardized format for NLP

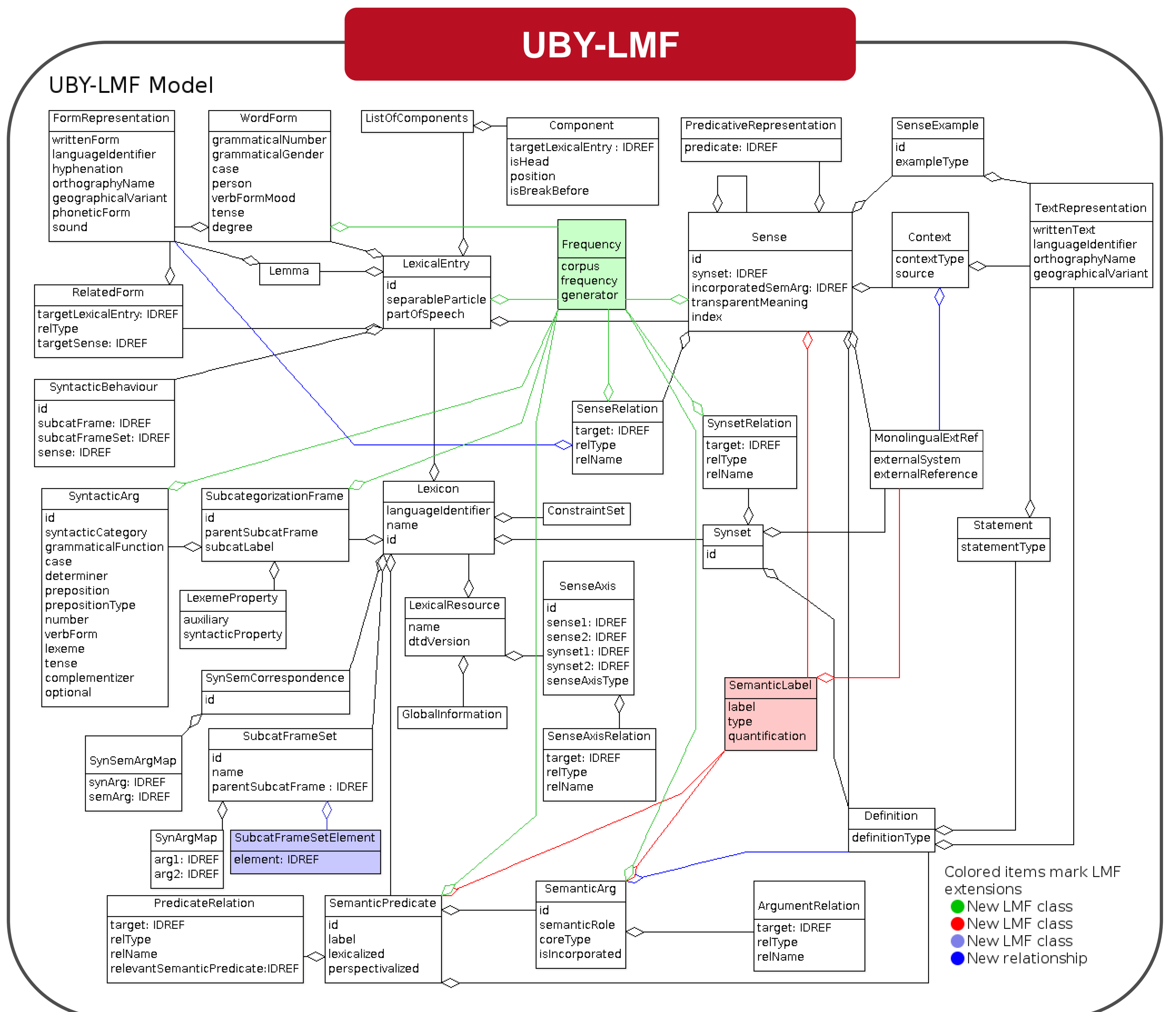
Motivation:

- Problem: Limited resource coverage in NLP
 - Solution: Integration of LSRs
- Problem: Incompatible formats of (integrated) LSRs
 - Solution: Standards for modeling LSRs

Vision: One-stop resource

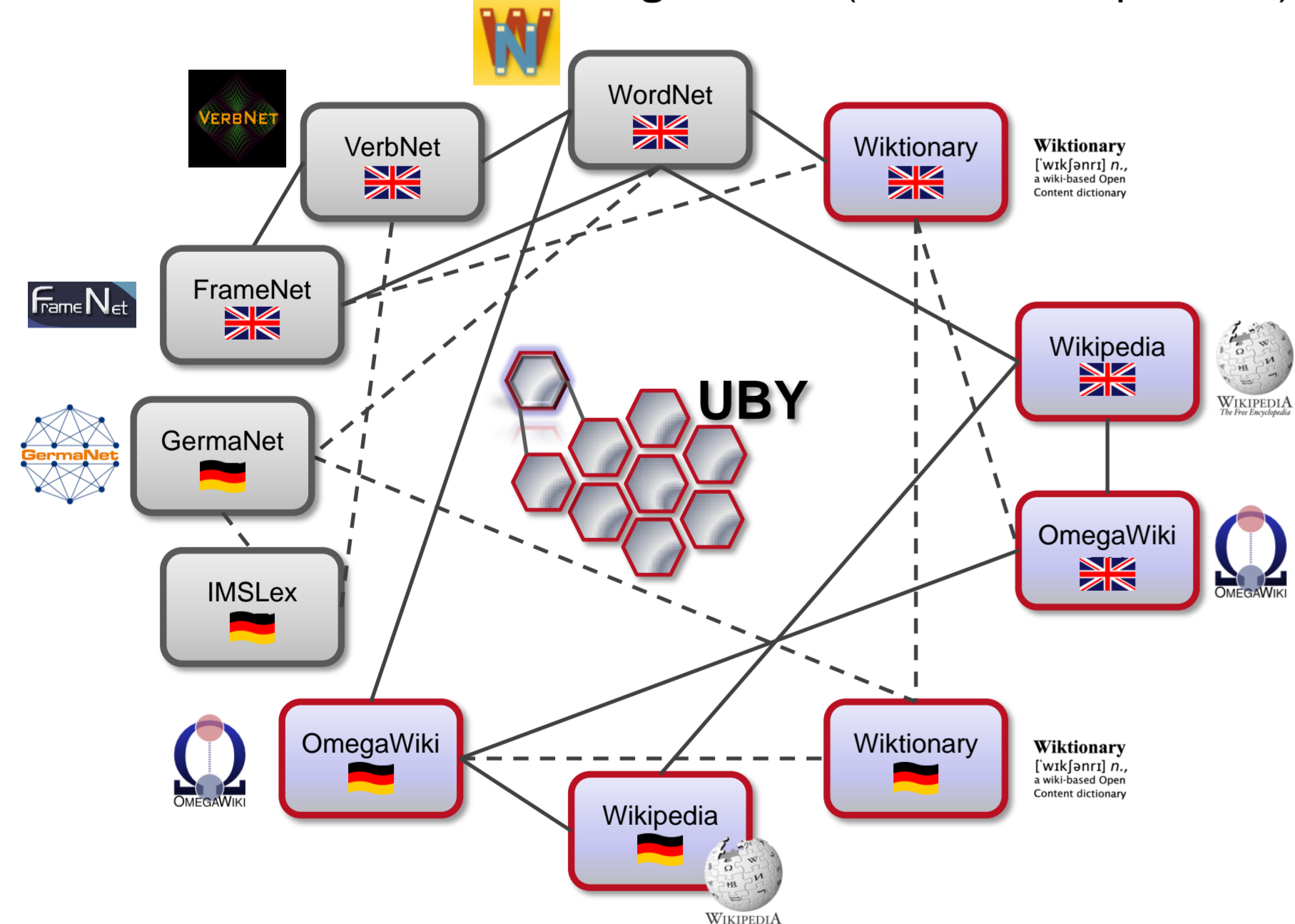
Contributions:

- 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
- UBY**
 - 11 LSRs in English and German
 - 10 pairwise sense alignments between LSRs
- UBY-API:**
 - Uniform access to LSRs in UBY via JAVA API
- UBY Web Interface**
 - Visual exploration of UBY



UBY

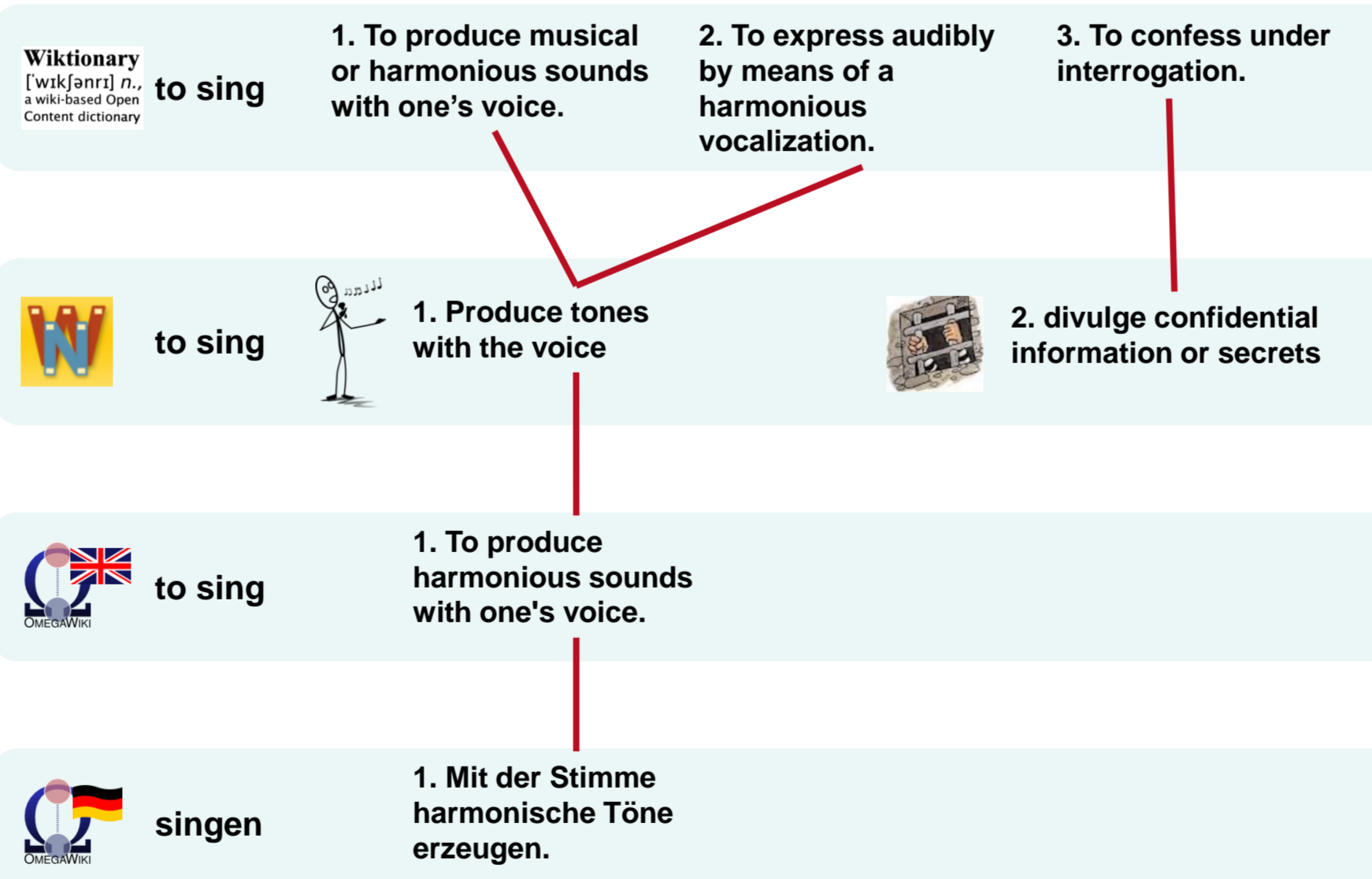
UBY Lexicons with sense alignments (dotted lines: planned)



UBY Lexicon	Lexical Entry	Sense	Sense Relation
FrameNet 1.5	9,702	11,942	-
GermaNet 7.0	87,535	99,523	350,259
IMSLex	15,342	32,396	-
OmegaWiki-DE	30,967	34,691	23,106
OmegaWiki-EN	51,715	57,921	40,348
Wikipedia-DE	789,781	838,435	571,286
Wikipedia-EN	2,709,447	2,921,455	3,364,083
Wiktionary-DE	85,575	72,752	183,684
Wiktionary-EN	335,469	420,672	23,020
WordNet 3.0	156,584	206,978	8,559
VerbNet	3,962	31,891	-
UBY	4,276,079	4,728,656	4,564,345

Sense Alignments

Enable **semantic operability** between resources in UBY:



- UBY contains:**
 - Cross-lingual and mono-lingual alignments
 - More than 760,000 sense alignments
 - Alignments modeled by **SenseAxis** class
- Generating methods:**
 - Import from existing alignments (created by experts, users, or automatically)
 - Automatic alignments
 - Created by Alignment Framework
 - Based on semantic similarity
 - Cross-lingual alignment via Machine Translation

Using UBY

UBY-API

- JAVA with Hibernate
- NLP application:
 - Easy *swapping* of resources
 - Easy *combination* of resources

UBY Web Interface

- UBY-API and Apache Wicket
- Visualization of sense alignments
- Supports *exploration* of UBY

Open licenses for data www.ukp.tu-darmstadt.de/uby/
and software <http://code.google.com/p/uby/>

See also

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth: **UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF**, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), p. 580–590, April 2012, Avignon, France.

Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek, and Christian M. Meyer: **UBY-LMF – A Uniform Model for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF**, in: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC), p. 275–282, May 2012, Istanbul, Turkey.

Judith Eckle-Kohler and Iryna Gurevych: **Subcat-LMF – Fleshing out a standardized format for subcategorization frame interoperability**, in: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), p. 550–560, April 2012, Avignon, France.