Semantic Web 0 (0) 1
IOS Press

lemonUby - a large, interlinked, syntactically-rich resource for ontologies

Judith Eckle-Kohler, [a,*] John M[c]Crae [b] and Christian Chiarcos [c]
[a] *Ubiquitous Knowledge Processing Lab (UKP-TUDA), Department of Computer Science, Technische Universität Darmstadt, Germany, http://www.ukp.tu-darmstadt.de*
[b] *CITEC, Universität Bielefeld, Germany, Email: jmccrae@cit-ec.uni-bielefeld.de*
[c] *Information Sciences Institute, University of Southern California, USA, Email: chiarcos@isi.edu*

Abstract. We introduce a new lexical resource integrated in the Semantic Web called *lemonUby* which is the result of a large-scale population of the *lemon* lexicon model. This was achieved by converting a number of UBY lexica standardized according to UBY-LMF to the *lemon* format: English WordNet, FrameNet, VerbNet, Wiktionary and OmegaWiki, and German Wiktionary and OmegaWiki. *lemonUby* is significantly linked – both at the sense level within its component resources and to other lexical resources and terminology repositories in the Linguistic Linked Open Data cloud.

Keywords: Lexicon model, lemon, UBY-LMF, UBY, OLiA, ISOCat, WordNet, VerbNet, FrameNet, Wiktionary, OmegaWiki

1. Introduction

Recently, the language resource community has begun to explore the opportunities offered by the Semantic Web, lead by the formation of the Linguistic Linked Open Data (LLOD) cloud and an increasing interest in making use of Linked Open Data principles in the context of Natural Language Processing (NLP) and Linguistics [8]. The use of RDF supports data integration and offers a large body of tools for accessing this data, and furthermore, the linked data approach gives rise to novel research questions in the context of language resources and their application.

For lexical resources, data integration has been in the focus of interest for many years, resulting in numerous mappings and linkings of lexica, as well as standards for representing lexical resources, such as the ISO 24613:2008 Lexical Markup Framework (LMF) [13]. In this context, the LLOD cloud can be considered as a new data integration platform, enabling linkings not only between lexical resources, but also between lexical resources and other language resources.

Many lexical resources have already been included in the LLOD cloud, e.g., WordNet, Wikipedia (DBpedia [3]), and Wiktionary, as well as integrated resources, such as an integrated version of WordNet and Wiktionary[25], or of WordNet and Wikipedia (BabelNet, [28]). There has also been some work towards the integration of FrameNet [2] to the Semantic Web [27]. All these resources provide a substantial body of lexical knowledge, including semantic relations, multilingual information and encyclopedic knowledge.

However, what is missing in the LLOD cloud is a large-scale lexical resource rich in lexical information on verbs, including aspects such as syntactic behaviour and how semantic arguments of verbs can be realised syntactically. Such information is crucial for lexicalizing relational knowledge which is often expressed by using verbs, e.g., the relation $like(Experiencer, Theme)$ can be lexicalized syntactically as "NP likes NP".

lemon, a lexicon model for representing and sharing ontology lexica, has been proposed as a common interchange format for lexical resources on the

*Corresponding author. E-mail: eckle-kohler@ukp.informatik.tu-darmstadt.de

1570-0844/0-1900/$27.50

Semantic Web[24]. Making use of a common interchange format is important, to integrate resources such as FrameNet and WordNet, which have been characterised as complementary resources [1]. The RDF version of FrameNet currently available does not adhere to an interchange format such as *lemon*, but is specific to the underlying data model of FrameNet.

Independently from linked data principles and Semantic Web technology, the large-scale lexical-semantic resource UBY [15] has been developed[1]. UBY is based on LMF and has currently integrated 10 lexical resources in English and German. A subset of these resources is interlinked at the word sense level. The dataset presented in this paper, *lemonUby*, is the result of converting a selection of UBY lexica to the *lemon* format: it contains interoperable and interlinked versions of WordNet, FrameNet, VerbNet [18], English and German Wiktionary [2], and the English and German entries of OmegaWiki [3].

To summarize, our contributions are threefold: (i) an interlinked lexical resource rich in linguistic information on verbs, (ii) a mapping of the lexicon model UBY-LMF to *lemon*, and (iii) the linking of *lemonUby* to other language resources in the LLOD cloud.

2. Representing lexical-semantic resources as Linked Data: *Lemon*

There has been significant work towards integrating lexical resources using RDF and Semantic Web principles [7], and many resources have already been converted to RDF, notably the conversion of WordNet [32]. They provided a simple mapping from WordNet to RDF, and augmented it with OWL semantics so that reasoning could be applied to the structure of the resource. However, the format chosen for this resource was specific to the underlying data model of WordNet. For this reason, the *lemon* model[24] was proposed that supports publishing lexical-semantic resources as linked data on the basis of the following principles:

LMF-based : To allow easy conversion from non-linked data resources.

RDF-native : Publishing as linked data, with RDFS and OWL used to describe the semantics of the model.

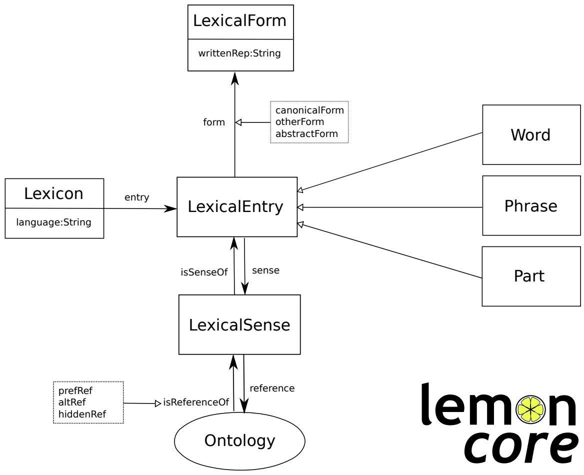

Fig. 1. The core of the *lemon* model

Modular : Separation of lexicon and ontology layers, so that *lemon* lexica can be linked to existing ontologies in the linked data cloud.

Externally defined data categories : Linking to data categories in annotation terminology repositories, rather than being limited to a specific part-of-speech tag set.

Principle of least power : The smaller the model and the less expressive the language, the wider its adoption and the higher the reusability of the data[30].

This core model is illustrated in Fig. 1, which defines the basic elements used by all lexica published as linked data. In addition to this there are a number of modules used to model linguistic description, syntax, morphology and relationships between lexica [4].

lemon has been used as a basis for integrating the data of the English Wiktionary, a (human-readable) dictionary created along ‘wiki’ principles, with the RDF version of WordNet [25]. *lemon*’s similarity to the WordNet model made this conversion straightforward, with only the need for a slight change in modelling to accommodate inflectional variants of lexical entries.

[1] `http://www.ukp.tu-darmstadt.de/uby/`
[2] `http://www.wiktionary.org`
[3] `http://www.omegawiki.org`

[4] More detail of the model and descriptions of the modules can be found at `http://lemon-model.net`

3. Large-scale integration of lexical-semantic resources: UBY and UBY-LMF

UBY is both a network of interlinked lexical-semantic resources and a project on continuous integration and linking of lexical resources for NLP applications. It is motivated by the observation that an essential requirement in NLP is the availability of a wide range of lexical resources that can be used for many different NLP tasks. In a continuous process, such resources are integrated into UBY by means of (i) making them inter-operable and (ii) linking them to other resources in UBY at the sense level.

In UBY, interoperability is achieved by standardizing lexical resources according to UBY-LMF [10,11], a lexicon model which is a full-fledged instantiation of the ISO standard LMF, specifically for NLP.

In comparison to the *lemon* lexicon model, UBY-LMF is similar in two aspects: first it is LMF-based, and second, it uses externally defined data categories from ISOCat.[5] In contrast to *lemon*, however, UBY-LMF is based on two quite different principles:

Principle of Adoption : UBY-LMF has been designed to fully cover a wide range of heterogeneous lexical resources without information loss.

Independence of implementation : UBY-LMF is independent of any particular implementation. There are many ways to implement an LMF lexicon model [14], including RDF.

UBY-LMF, is currently represented in two ways: first, as a DTD, and second, as a Java Object-Relational Mapping by means of the Hibernate framework[6], which allows mapping any instance of UBY-LMF either to a SQL database or to an XML file. Both ways of representing UBY-LMF do not require the use of globally unique identifiers (URIs). However, an implementation of UBY-LMF in RDF would be possible as well.

An implementation of LMF to include URIs has already been suggested[14]. In fact, providing lexical resources, in particular interlinked resources such as UBY, as linked data is a very natural step to take and allows us to integrate UBY-LMF-based resources with other resources previously converted to RDF.

As a first step into this direction, we performed a mapping of UBY-LMF to *lemon* which allows conversion of lexical resources in UBY-LMF format to *lemon* format. Apart from the fact that the mapping of UBY-LMF to *lemon* is an interesting task per se, because *lemon* links lexical resources and ontologies, there is another reason for this mapping. The LMF standard is not an open standard (in the sense that its specification is not freely available), while *lemon* provides an interchange format which fully complies with open data and open access principles.

Although both UBY-LMF and *lemon* are based on LMF, the mapping revealed substantial differences which are mainly due to the fact that *lemon* is a model for ontology lexica where the lexicon and ontology layers are kept separate. Thus, sense representations in *lemon* primarily consist of references to the associated ontology where a rich and domain-specific sense definition is provided. The development of UBY-LMF, on the other hand, has been driven by the requirement to cover a large variety of lexical information types, which ranges from morphology and lexical syntax to lexical semantics and the mapping between syntactic and semantic arguments. Thus, the resulting lexicon model makes use of very fine-grained sense specifications which are often grounded in linguistic theories, e.g. Frame Semantics (the basis of FrameNet) or the Levin alternation classes of verbs [21] (the basis of VerbNet).

4. Converting UBY lexica to *lemon*: *lemonUby*

The actual conversion of UBY lexica to *lemon* was achieved by means of an XML style sheet transform [7]. The following UBY lexica were converted: WordNet, FrameNet, VerbNet, English and German Wiktionary, and the English and German entries of OmegaWiki. Pairs of these resources are linked at the sense level: there are monolingual links between VerbNet–FrameNet[8], VerbNet–WordNet[9], WordNet–FrameNet [31,20], as well as between WordNet–Wiktionary [26]. Moreover, *lemonUby* provides cross-lingual links between WordNet and the German OmegaWiki [15], also including the inter-language links already given in OmegaWiki.

The resulting resource *lemonUby* has been published at `http://lemon-model.net/lexica/`

[5] `http://www.isocat.org/rest/dcs/484`

[6] `http://www.hibernate.org`

[7] The XSL file can be downloaded at `https://raw.github.com/jmccrae/lemon.api/master/src/main/resources/xslt/ubylmf2lemon.xsl`

[8] `http://verbs.colorado.edu/semlink/`

[9] `http://verbs.colorado.edu/~mpalmer/projects/verbnet`

`uby` and is available under an open CC-BY-SA license. The choice of a share-alike license was due to UBY being published under the same license. Statistics for the resources are given in table 1[10].

lemonUby nicely illustrates how the two lexicon models represent senses in different ways:

- In line with previous authors [25], synsets in WordNet and OmegaWiki are considered as ontology classes. Hypernym relationships stated in the lexicon are mapped directly to subclass relationships in the ontology. In UBY-LMF, Synset is part of the lexicon.
- The semantic predicate in UBY-LMF is used to represent semantic frames from FrameNet [11]; frames consist of senses which evoke the same situation with participants. Thus, senses in the same semantic frame are semantically related, but not synonymous; e.g., the verbs *love* and *hate* are both in the same frame. In *lemon* a *broader sense* is used to capture semantic predicates.
- VerbNet verb classes and their hierarchical organization correspond to subcat[egorization] frame set in UBY-LMF. VerbNet classes group verbs that share the same syntactic subcategorization frame, semantic roles, selectional restrictions, and semantic predicate. Although the resulting verb classes are semantically coherent, the semantic relatedness of verb senses in a VerbNet class is much more distant that in FrameNet, e.g., the verbs *believe*, *swear* and *doubt* are in the same class. Verb classes in *lemon* are represented by a broader sense as well. The hierarchic organisation of VerbNet classes is thus lost, but this can be reconstructed by means of an axiomatic description of subcategorization frames in order to create a hierarchy similar as in the LexInfo linguistic ontology [23].
- Syntactic behaviour in UBY-LMF mainly provides information on subcategorization frames, which are specified by means of syntactic arguments with rich morpho-syntactic information attached [9]. SyntacticBehaviour is merely a property instead of a node in *lemon*: in contrast to LMF, there is no need for an explicit link between senses and subcategorization frames. Instead this information can be reconstructed from the *synArg* - *semArg* mapping. Lexica which do not specify semantic arguments, e.g., GermaNet [19], must provide a specific annotation to allow this information to be represeneted.
- Translations are represented differently. UBY-LMF uses Equivalent for translations that are not sense-disambiguated [22], where as *lemon* requires that this link is made at the sense level. In the case where it was not possible to infer the sense, the lexical entries a ‘stub’ lexical entry with a single sense.[11]

Resource	Triples
WordNet	5,102,744
VerbNet	570,256
FrameNet	1,110,763
OmegaWiki English	6,173,515
OmegaWiki German	5,310,551
Wiktionary German	4,766,917
Total	23,034,746

Table 1

Number of triples for each resource

To sum up, most of the differences between UBY-LMF and *lemon* derive from the fact that *lemon* keeps the lexicon and ontology layers separate. In this way, sense representations in *lemon* are more compact compared to the more distributed sense representations in UBY-LMF.

5. Linking *lemonUby* to lexical-semantic resources

As UBY is derived from existing resources the simplest links to create are those to other RDF versions of the resources that compose UBY. For WordNet these links are simply created by mapping the data of UBY to the linked data version of WordNet 3.0 [12]. Here, we provided links at both the sense level and at the lexical entry level (lexical entries are “words” in WordNet 3.0). We found that this worked apart from 7 senses that did not map, which we believe is due to a bug in the WordNet API.

In addition, we provided links to two existing resources that are also widely used, firstly to RDF WordNet 2.0 [32]. As the sense identifiers are different to the WordNet version used by UBY we only attempted

[10]At the time of submission, en.wiktionary.org resource was not available, we expect to have this ready in time for final publication

[11]UBY-LMF can also represent sense-disambiguated translations by means of a sense axis which connects senses from different lexica.

[12]`http://semanticweb.cs.vu.nl/lod/wn30/`

Uby Resource	Other Resource	Links (Percentage of Uby resource)
WordNet	WordNet 3.0	206,773 (99.9%)
WordNet	WordNet 2.0	84,416 (40.8%)
WordNet	Wiktionary	76,294 (36.9%)

Table 2

Number of external links created between lemonUby and other resources in the LLOD cloud

to link at the lexical entry level, using the assumption that the lemmas were the same. Secondly, we linked to the RDF export of Wiktionary [13]. For this resource, we first linked the WordNet data on the lexical entry level using the lemma and part-of-speech information. This was mostly effective, however some elements were initially missing due to category misalignment (Wiktionary has "Initialism" as a part-of-speech, for example for "IBM", whereas WordNet counts these as nouns); we added manual corrections to compensate for this. Statistics for all mappings are given in table 2.

6. Linking *lemonUby* with terminology repositories and corpora

In UBY-LMF, most of the linguistic terms used (e.g. values of the attribute part-of-speech) are linked to standardized descriptions of their meaning in ISOCat [14], the implementation of the ISO 12620:2009 Data Category Registry. As the mapping of UBY-LMF to *lemon* preserves this linking, *lemonUby* is linked to ISOCat as well. The content of ISOCat is also available as Linked Data [33], and therefore, provides a possible way to interconnect *lemonUby* with other LLOD resources at the level of linguistic data categories.

However, ISOCat does currently not contain an explicit license statement, even though its content is available over the internet. While people involved (Menzo Windhouwer, p.c., June 2012) believe that a publication under an open license would be appreciated, it has not been clarified whether the restrictive licensing policy of the ISO that applies to the ISOcat technical specifications also extends to the ISOcat data categories. A similar problem persists with the General Ontology of Linguistic Description [12, GOLD] [15], another LLOD-linked terminology repository.

Within the LLOD cloud, the Ontologies of Linguistic Annotation (OLiA,[4,6]) represent a repository of annotation terminology that act as a central reference hub for linguistic annotations in about 70 languages, for which they provide formal definitions of annotation schemes for various linguistic phenomena as OWL/DL ontologies. Further, OLiA establishes interoperability between different annotation schemes by linking them to an overarching 'Reference Model'. Through the OLiA Reference Model, interoperability with community-maintained data category registries can be achieved, as it is linked to both GOLD and ISOCat. Unlike these *evolving* repositories that aim for definitions and categories that are generally applicable, the OLiA Reference Model only serves as an aggregation point for the annotation schemes directly linked to it. It generalizes over these schemes, interprets them against other terminology repositories, and thereby provides a stable interface between linguistic annotations and these repositories in-the-making.

The evolving character of ISOCat, for example, is evident from its appeal as a semistructured, extendable list that seems to be accumulative rather than normative: Data providers can augment the repository with their own data categories,[16] but ISOcat does not provide a formalism to define their relationship. The small set of relations between data categories it does provide are optional and not consistently applied. A relation category registry that has been announced for this purpose [29] is not yet in common usage.

An ontology-based approach allows to express complex relationships among reference categories and between reference categories and annotations. For the linking of Uby to the OLiA Reference Model, we created an ontology for the morphosyntactic concepts used in UBY-LMF, that redefined the enumeration of categories in the DTD was redefined with more elaborate hierarchical structures. This ontology is linked to the OLiA Reference Model by subClassOf relationships between its concepts and the Reference Model. As this linking is an *interpretation*, it is physically separated in an independent 'Linking Model', because different interpretations may be possible.

This interpretation may be complex: Many annotation schemes for Germanic languages postulate a category 'determinerPossessive', and so does Uby, for possessive pronouns like *his house*. However, this con-

[13] `http://wiktionary.dbpedia.org`

[14] `http://www.isocat.org/`

[15] `http://linguistics-ontology.org/`

[16] Shown for example by largely equivalent, but distinct data categories 'noun': DC-1333, DC-2704, DC-3347.

flates independent levels of analysis, syntax (determiners mark nouns as noun phrases), and semantics (pronouns "stand in for" nouns). Pronouns can have different syntactic functions, attributive pronouns can modify nouns (as in *his house*), substitutive pronouns represent independent noun phrases (as in *that's him*). In a terminology repository, however, these functions should be carefully distinguished, as there are languages where no category 'determiner' exists (e.g., Russian, which nevertheless has attributive pronouns), or where attributive pronouns are not necessarily determiners (e.g., in Italian *la donna mia*). Accordingly, the linking defines the Uby determinerPossessive as Determiner *and* AttributivePronoun *and* PossessivePronoun. Another example for a complex linking is 'ingForm', conventionally used for verb forms like *talking* in English. This category represents a language-specific merger of present participles (*he is speaking*, Old English *-inde*) and gerunds (*he finished speaking*, Old English *-inge*). The linking provides a language-neutral definition as PresentParticiple *or* Gerund, so that categories across different languages can be compared more easily.[17]

The OLiA ontologies (and the terminology repositories it is linked with) have been applied as a component of the *lemon* model before [24], and (as part of their original motivation), they can be used to represent, compare and integrate linguistic annotations in corpora on the basis of formal concepts rather than arbitrary strings [5], and in this function, they also enjoy a certain popularity in NLP applications [16]. In a Linked Data context, they explicitly allow to compare the linguistic categories used in Uby with the morphosyntactic annotations in linguistic corpora, if these are represented in RDF. An RDF version of the MASC corpus [17] has been produced, a resource that also provides FrameNet and WordNet annotations, and whose annotations can thus be directly compared with and combined with Uby resources. Information integration between Uby and corpora is one of the goals we aim to achieve with *lemonUby*.

[17]It should be noted that such a complex linking requires the use of operators like *and* and *or* in the linking, to capture this information, OLiA ontologies employ OWL/DL. The direct mapping between annotations and reference concepts originally advocated for ISOCat and GOLD cannot represent this information.

7. Conclusion

We have presented a new linked data resource called *lemonUby*, which combines data from the standardized lexical resource UBY with the principled model, *lemon*, for representing lexical data on the web. This resource provides not only rich information about many lexical entries in two languages, it is also significantly linked both within its component resources and to other lexical resources and terminology repositories on the web.

Acknowledgments

The work on *lemon* was developed in the context of the Monnet project, which is funded by the European Union FP7 program under grant number 248458, the CITEC excellence initiative funded by the DFG (Deutsche Forschungsgemeinschaft). The work on UBY has been supported by the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant No. I/82806. The research of Christian Chiarcos has been supported by the German Academic Exchange Service (DAAD) through a PostDoc fellowship at the University of Southern California. We thank Silvana Hartmann, Michael Matuschek and Christian M. Meyer for their contributions to this work.

References

[1] Baker CF, Fellbaum C (2009) WordNet and FrameNet as complementary resources for annotation. In: Proceedings of the Third Linguistic Annotation Workshop, Suntec, Singapore, ACL-IJCNLP '09, pp 125–129
[2] Baker CF, Fillmore CJ, Lowe JB (1998) The Berkeley FrameNet project. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98), Montreal, Canada, pp 86–90
[3] Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R, Hellmann S (2009) DBpedia – A Crystallization Point for the Web of Data. Journal of Web Semantics: Science, Services and Agents on the World Wide Web 7(3):154–165
[4] Chiarcos C (2008) An ontology of linguistic annotations. LDV Forum 23(1):1–16
[5] Chiarcos C (2010) Towards robust multi-tool tagging. an owl/dl-based approach. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, pp 659–670
[6] Chiarcos C (2012) Ontologies of linguistic annotation: Survey and perspectives. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC), pp 303–310

[7] Chiarcos C, Hellmann S, Nordhoff S (2011) Towards a linguistic linked open data cloud: The open linguistics working group. Traitement automatique des langues 52(3):245–275

[8] Chiarcos C, Nordhoff S, Hellmann S (eds) (2012) Linked Data in Linguistics. Representing Language Data and Metadata. Springer, Heidelberg

[9] Eckle-Kohler J, Gurevych I (2012) 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 2012), Avignon, France, pp 550–560

[10] Eckle-Kohler J, Gurevych I, Hartmann S, Matuschek M, Meyer CM (2012) UBY-LMF — A Uniform Format for Standardizing Heterogeneous Lexical-Semantic Resources in ISO-LMF. In: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, pp 275–282

[11] Eckle-Kohler J, Gurevych I, Hartmann S, Matuschek M, Meyer CM (2013) UBY-LMF - exploring the boundaries of language-independent lexicon models. In: Francopoulo G (ed) LMF: Lexical Markup Framework, theory and practice, ISTE - Wiley, London, UK, p (to appear)

[12] Farrar S, Langendoen DT (2003) Markup and the GOLD ontology. In: EMELD Workshop on Digitizing and Annotating Text and Field Recordings, Michigan State University, pp 1–19

[13] Francopoulo G, Bel N, George M, Calzolari N, Monachini M, Pet M, Soria C (2006) Lexical Markup Framework (LMF). In: Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC), Genoa, Italy, pp 233–236

[14] Francopoulo G, Bel N, George M, Calzolari N, Monachini M, Pet M, Soria C (2007) Lexical markup framework: ISO standard for semantic information in NLP lexicons. In: Proceedings of the Workshop of the GLDV Working Group on Lexicography at the Biennial Spring Conference of the GLDV

[15] Gurevych I, Eckle-Kohler J, Hartmann S, Matuschek M, Meyer CM, Wirth C (2012) UBY - A Large-Scale Unified Lexical-Semantic Resource. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012), Avignon, France, pp 580–590

[16] Hellmann S, Lehmann J, Auer S, Nitzschke M (2012) Nif combinator: Combining nlp tool output. Knowledge Engineering and Knowledge Management pp 446–449

[17] Ide N, Baker C, Fellbaum C, Fillmore C, Passonneau R (2008) MASC: The Manually Annotated Sub-Corpus of American English. In: Proc. 6th Language Resources and Evaluation Conference (LREC 2008), Marrakesh, Morocco

[18] Kipper K, Korhonen A, Ryant N, Palmer M (2008) A Large-scale Classification of English Verbs. Language Resources and Evaluation 42:21–40

[19] Kunze C, Lemnitzer L (2002) GermaNet — representation, visualization, application. In: Proceedings of the Third International Conference on Language Resources and Evaluation (LREC), Las Palmas, Canary Islands, Spain, pp 1485–1491

[20] Laparra E, Rigau G (2009) Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm. In: Proceedings of the International Conference RANLP-2009, Association for Computational Linguistics, Borovets, Bulgaria, pp 208–213, URL `http://www.aclweb.org/anthology/R09-1039`

[21] Levin B (1993) English Verb Classes and Alternations. The University of Chicago Press, Chicago, IL, USA

[22] Matuschek M, Meyer CM, Gurevych I (2012) Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Computer-Aided Translation. Translation: Computation, Corpora, Cognition Special Issue on "Language Technology for a Multilingual Europe" (to appear)

[23] McCrae J, Spohr D, Cimiano P (2011) Linking lexical resources and ontologies on the semantic web with Lemon. The Semantic Web: Research and Applications pp 245–259

[24] McCrae J, Aguado-de Cea G, Buitelaar P, Cimiano P, Declerck T, Gómez-Pérez A, Gracia J, Hollink L, Montiel-Ponsoda E, Spohr D, Wunner T (2012) Interchanging lexical resources on the semantic web. Language Resources and Evaluation 46:701–719,

[25] McCrae J, Montiel-Ponsoda E, Cimiano P (2012) Integrating wordnet and wiktionary with lemon. Linked Data in Linguistics pp 25–34

[26] Meyer CM, Gurevych I (2011) What Psycholinguists Know About Chemistry: Aligning Wiktionary and WordNet for Increased Domain Coverage. In: Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), Chiang Mai, Thailand, pp 883–892

[27] Narayanan S, Baker C, Fillmore C, Petruck M (2003) Framenet meets the semantic web: Lexical semantics for the web. The Semantic Web-ISWC 2003 pp 771–787

[28] Navigli R (2012) Babelnet goes to the (multilingual) semantic web. In: Proceedings of the 3rd Workshop on the Multilingual Semantic Web, pp 6–15

[29] Schuurman I, Windhouwer M (2011) Explicit semantics for enriched documents. What do ISOcat, RELcat and SCHEMAcat have to offer? In: Proc. 2nd Supporting Digital Humanities conference (SDH 2011), Copenhagen, Denmark

[30] Shadbolt N, Hall W, Berners-Lee T (2006) The semantic web revisited. IEEE intelligent systems 21(3):96–101

[31] Tonelli S, Pighin D (2009) New features for framenet - wordnet mapping. In: Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009), Association for Computational Linguistics, Boulder, Colorado, pp 219–227, URL `http://www.aclweb.org/anthology/W09-1127`

[32] Van Assem M, Gangemi A, Schreiber G (2006) Conversion of WordNet to a standard RDF/OWL representation. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation LREC'06, Citeseer, pp 237–242

[33] Windhouwer M, Wright SE (2012) Linking to linguistic data categories in ISOcat. In: [8], pp 99–107