Standardizing Wordnets in the ISO Standard LMF: Wordnet-LMF for GermaNet

Verena Henrich
Universtiy of Tübingen
Department of Linguistics
verena.henrich@uni-tuebingen.de

Erhard Hinrichs
Universtiy of Tübingen
Department of Linguistics
erhard.hinrichs@uni-tuebingen.de

Abstract

It has been recognized for quite some time that sustainable data formats play an important role in the development and curation of linguistic resources. The purpose of this paper is to show how GermaNet, the German version of the Princeton WordNet, can be converted to the Lexical Markup Framework (LMF), a published ISO standard (ISO-24613) for encoding lexical resources. The conversion builds on Wordnet-LMF, which has been proposed in the context of the EU KYOTO project as an LMF format for wordnets. The present paper proposes a number of crucial modifications and a set of extensions to Wordnet-LMF that are needed for conversion of wordnets in general and for conversion of GermaNet in particular.

1 Introduction

It has been recognized for quite some time that sustainable data formats play an important role in the development and curation of linguistic resources. As witnessed by the success of the guidelines of the Text Encoding Initiative[1] (TEI) and of published standards issued by the International Standards Organization[2] (ISO), markup languages such as XML[3] (short for: Extensible Markup Language) have become lingua francas for encoding linguistic resources of different types, including phonetic transcriptions, (annotated) text corpora, and dictionaries. It is fair to say that it has become common practice among developers of new linguistic resources to consult TEI guidelines and ISO standards in order to develop standard-conformant encoding schemes that serve as an interchange format and that can be documented and validated by Document Type Definitions (DTD) and XML schemata.

However, for resources that were developed prior to or largely in parallel with the emerging acceptance of markup languages and of emerging encoding standards, the situation is far more heterogeneous. A wide variety of legacy formats exists, many of which have persisted due to existing user communities and the availability of tools that can process only such idiosyncratic formats. The development of wordnets for a large number of languages is a typical example of a type of linguistic resource, where legacy formats still persist as a de facto standard. WordNet 1.6 is encoded in the data format of lexicographer files[4] that was designed for the English Princeton WordNet (Fellbaum, 1998). It is a plain-text format for storing wordnet data and allows lexicographers to encode lexical and conceptual relations among lexical units and synsets by use of special-purpose diacritics. There exist numerous tools that can process WordNet 1.6 lexicographer files to extract relevant information or to transform the data into other special-purpose formats such as Prolog-fact databases. Even tough still widely used for the reasons just mentioned, the complexity of the format itself has a number of undesirable consequences. As Henrich and Hinrichs (2010) have pointed out,

---

[1] See http://www.tei-c.org
[2] See http://www.iso.org
[3] See http://www.w3.org/TR/REC-xml/

[4] See http://wordnet.princeton.edu/man/lexnames.5WN.html

the editing of lexicographer files is highly error-prone and time-consuming in actual lexicographic development. Moreover, format validation of the data as well as development of new tools for data visualization and data extraction become increasingly difficult since they cannot be based on generic state-of-the-art tools, that are, for example, available for XML-based encodings.

For exactly these reasons, XML-based interchange formats have been proposed in recent years also for wordnets. One of the first, if not the first, example is the XML format for GermaNet[5], a wordnet for German (Lemnitzer and Kunze, 2002; Henrich and Hinrichs, 2010). An even more recent development along these lines is the specification of Wordnet-LMF (see Soria et al., 2009), an instantiation of the Lexical Markup Framework[6] (LMF, (Francopoulo et al., 2006)) customized for wordnets.

Since LMF is an ISO standard (ISO-24613), it is a particularly attractive candidate for encoding wordnets. Everything else being equal, ISO standards have a high chance of being adopted by a wide user community and of being recognized as an interchange format.[7] Such agreed-upon interchange formats are a crucial prerequisite for interoperable linguistic resources in the context of web services and of processing pipelines for linguistic resources.

The purpose of this paper is threefold:

1. To compare and contrast the GermaNet XML initially proposed by Lemnitzer and Kunze (2002) with the Wordnet-LMF. This comparison is instructive since it reveals two completely different conceptions of representing semantic knowledge at the lexical level.

2. To point out a number of open issues that need to be resolved if Wordnet-LMF is to be adopted widely among

wordnets for a steadily increasing number of languages.

3. To show how these open issues can be resolved in a customized version of Wordnet-LMF suitable for GermaNet.

The remainder of this paper is structured as follows: section 2 provides a general introduction to GermaNet. Details about the adapted XML format used for GermaNet up until now are provided in section 3. Section 4 introduces the challenge of how to represent a wordnet in the Lexical Markup Framework. As one possibility, Wordnet-LMF is regarded. Issues that arise during the conversion of GermaNet into Wordnet-LMF lead to a modified version of Wordnet-LMF. Finally, section 5 concludes with a comparison of the two representation formats.

2 GermaNet

GermaNet is a lexical semantic network that is modeled after the Princeton WordNet for English. It partitions the lexical space into a set of concepts that are interlinked by semantic relations. A semantic concept is modeled by a *synset*. A synset is a set of words (called *lexical units*) where all the words are taken to have (almost) the same meaning. Thus a synset is a set-representation of the semantic relation of synonymy, which means that it consists of a list of lexical units and a paraphrase (represented as a string). The lexical units in turn have frames (which specify the syntactic valence of the lexical unit) and examples. The list of lexical units for a synset is never empty, but any of the other properties may be.

There are two types of semantic relations in GermaNet: *conceptual* and *lexical relations*. Conceptual relations hold between two semantic concepts, i.e. synsets. They include relations such as hyperonymy, part-whole relations, entailment, or causation. Lexical relations hold between two individual lexical units. Antonymy, a pair of opposites, is an example of a lexical relation.

GermaNet covers the three word categories of adjectives, nouns, and verbs, each of which is hierarchically structured in terms of the hyperonymy relation of synsets.

---

[5] See http://www.sfs.uni-tuebingen.de/GermaNet/

[6] See http://www.lexicalmarkupframework.org

[7] An anonymous reviewer raised the question why OWL is not a good candidate for encoding wordnets. On this issue, we agree with the assessment of Soria et al. (2009) who point out that *"[...] RDF and OWL are conceptual repositories representation formats that are not designed to represent polysemy and store linguistic properties of words and word meanings."*
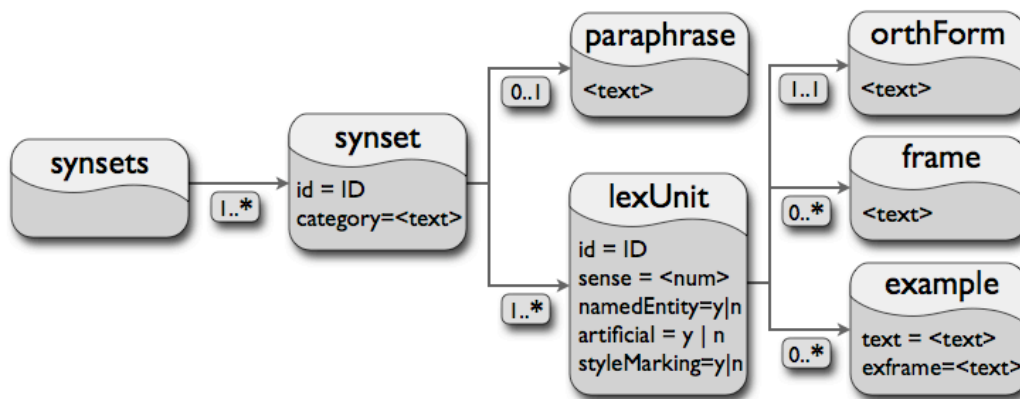
Figure 1. Structure of the XML synset files.

3   Current GermaNet XML Format

The structure of the XML files closely follows the internal structure of GermaNet, which means that the file structure mirrors the underlying relational organization of the data. There are two DTDs that jointly describe the XML-encoded GermaNet. One DTD represents all synsets with their lexical units and their attributes (see subsection 3.1). The other DTD represents all relations, both conceptual and lexical relations (see subsection 3.2).

The GermaNet XML format was initially developed by Kunze and Lemnitzer (2002), but modifications of the GermaNet data itself led to an adopted XML format, which is presented here.[8]

3.1   XML Synset Files

The XML files that represent all synsets and lexical units of GermaNet are organized around the three word categories currently included in GermaNet: nouns, adjectives, and verbs (altogether 54 synset files since the semantic space for each word category is divided into a number of semantic subfields).

The structure of each of these files is illustrated in Figure 1[9]. Each *synset* represents a set of lexical units (*lexUnits*) which all express the same meaning. This grouping represents the semantic relation of synonymy. Further properties of a *synset* (e.g., the word *category* or a describing *paraphrase*) and a lexical unit (e.g., a *sense* number or the orthographical form (*orthForm*)) are encoded appropriately.

Figure 1 describes the underlying XML structure. Each box in the figure stands for an element in the XML files, and the properties in each box (listed underneath the wavy line) represent the attributes of an XML element. This means, for example, that a *synset* element has the attributes of an *id* and a *category*.[10]

Figure 2 shows an example of a *synset* with two lexical units (*lexUnit* elements) and a *paraphrase*. The *lexUnit* elements in turn contain several attributes and an orthographical form (the *orthForm* element), e.g., *leuchten* (German verb for: *to shine*). The first of the two lexical units even has a *frame* and an *example*.

```
<synset id="s58377" category="verben">
  <lexUnit id="l82207"
        sense="1"
        namedEntity="no"
        artificial="no"
        styleMarking="no">
    <orthForm>leuchten</orthForm>
    <frame>NN</frame>
    <example>
      <text>
        Der Mond leuchtete in der Nacht.
      </text>
      <exframe>NN</exframe>
    </example>
  </lexUnit>
  <lexUnit id="l82208"
```

---

[8] The interested reader might compare the version at hand with (Lemnitzer and Kunze, 2002) or (Kunze and Lemnitzer, 2002), which both describe the initial GermaNet XML version.
[9] In fact, this figure is not quite complete for the reason of simplicity.

[10] Note that XML element or attribute names appear *italic* if they are referenced in the text.

```
            sense="2"
            namedEntity="no"
            artificial="no"
            styleMarking="no">
      <orthForm>strahlen</orthForm>
  </lexUnit>
  <paraphrase>
    Lichtstrahlen aussenden,
    große Helligkeit verbreiten
  </paraphrase>
</synset>
```

Figure 2. Synset file example.

3.2   XML Relation File

This type of XML file represents both kinds of relations: conceptual and lexical relations. All relations are encoded within one XML file, whose structure is illustrated in Figure 3.
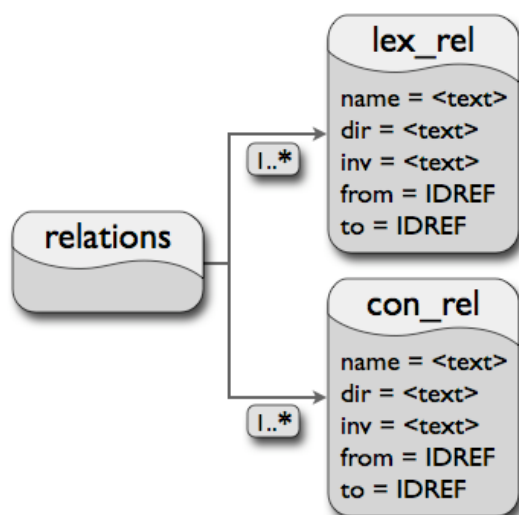

Figure 3. Structure of the XML relation file.

The boxes in Figure 3 again represent XML elements, which means that there is one *relations* element that contains all lexical relations (*lex_rel* elements) and conceptual relations (*con_rel* elements). Both relation types contain several attributes.

Figure 4 illustrates an example for each of the two relation types. The type of the conceptual relation is *hyperonymy* (indicated by the *name* attribute), and it holds between the synset with ID *s58377* (*from* attribute) and the synset with ID *s58376* (*to* attribute). The lexical relation is of type *antonymy* (again indicated by the *name* attribute), and holds between the lexical units with the IDs *l2471* (*from* attribute) and *l12470* (*to* attribute).

```
<con_rel name="hyperonymy"
        from="s58377" to="s58376"
        dir="revert" inv="hyponymy" />
<lex_rel name="antonymy"
        from="l2471" to="l2470"
        dir="both" />
```

Figure 4. Example from relation file.

4   Wordnet-LMF

The Lexical Markup Framework (ISO-24613) is an ISO standard for encoding natural language processing lexicons and machine readable dictionaries (Francopoulo et al., 2006). The intention of LMF is to provide a common model for the creation and use of lexical resources, to manage the exchange of data between and among these resources, and to enable the merging of a large number of individual electronic resources to form extensive global electronic resources.

4.1   The Challenge

The core structure of LMF is based on the prototypical structuring of a lexicon in terms of lexical entries, each of which enumerates the different senses of the lexical item in question. This word-driven perspective contrasts the synset-driven relational structure of wordnets – the grouping of word senses (i.e., lexical units) that express the same meaning into synsets. Exactly these two radically different organizing principles (relation-based in the case of wordnets versus lexical-entry-based in the case of LMF) constitute the challenge of encoding wordnets in LMF. We take up this challenge: *How can a synset-based wordnet, e.g. GermaNet, be represented in a word-driven format like LMF?*

4.2   Apply LMF to Wordnets

The conversion of GermaNet to LMF will build on Wordnet-LMF (Soria et al., 2009; Lee et al., 2009), an existing Lexical Markup Framework subset[11]. Wordnet-LMF has been developed in the context of the EU KYOTO

---

[11] Wordnet-LMF is a proper subset of LMF since there are specifications in LMF that are not in Wordnet-LMF and since there is nothing in Wordnet-LMF which is not in LMF. Soria et al. (2009) themselves refer to Wordnet-LMF as an LMF *dialect*.
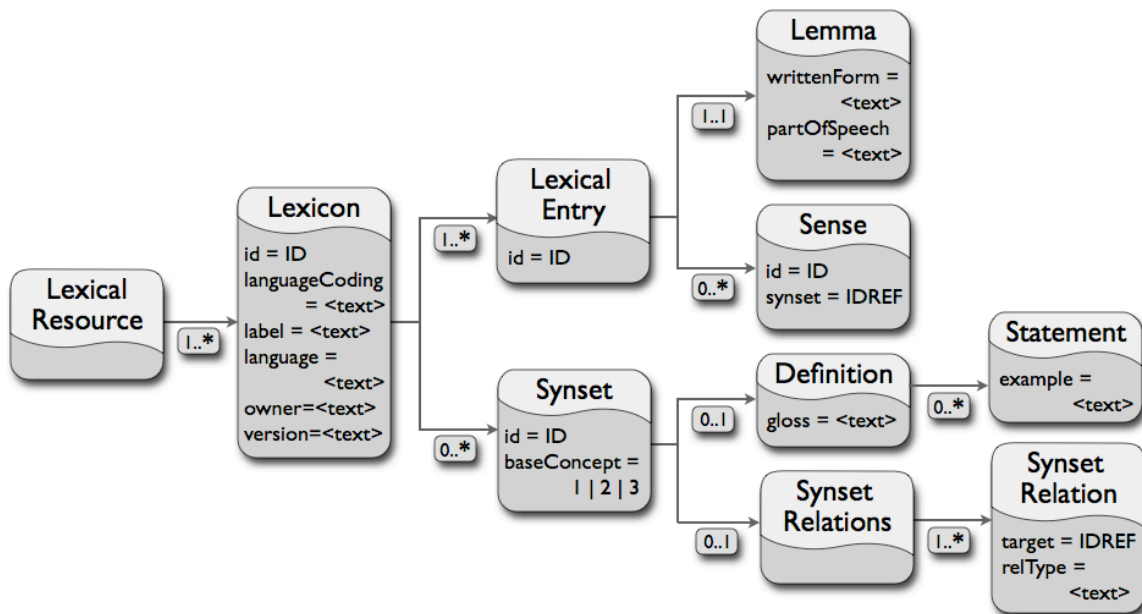
Figure 5. The Wordnet-LMF structure.

project[12] and is especially tailored to encode wordnets in the LMF standard.

Wordnet-LMF is specified by a Document Type Definition (see Appendix E in (Soria and Monachini, 2008)) and fully complies with standard LMF.

The Wordnet-LMF XML structure is shown in Figure 5[13]. There is a *Lexical Resource* which contains at least one *Lexicon* (in this case a wordnet lexicon).[14] A *Lexical Entry* represents a word entry in a *Lexicon*, where the word itself is represented by the *writtenForm* attribute of the *Lemma* element. *Lexical Entries* group different *Senses* of a particular word. The *Senses* have a *synset* attribute that relates them to a *Synset* element by the corresponding ID. If two *Senses* have the same *synset* attribute, they belong to the same *Synset* and are thus synonyms.

A *Synset* can have several relations to other *Synsets*. These relations are encoded in *SynsetRelation* elements.

---

[12] See http://www.kyoto-project.eu

[13] Note that this figure does not show the whole Wordnet-LMF model. Only the monolingual part that is relevant for this paper is represented. The representation of multilingual resources (i.e., the optional *SenseAxis* element with its children) is not considered in this paper. For a complete picture, see Soria et Monachini (2008).

[14] Here, XML element or attribute names again appear *italic* if they are referenced in the text.

4.3 Apply Wordnet-LMF to GermaNet

The differences between the synset-driven structure of GermaNet (see Figures 1 and 3) and the word-driven format of Wordnet-LMF (see Figure 5) are obvious. But there is also a strong commonality: Both formats have synset elements that cluster synonymous words. In GermaNet, the words are represented by lexical units that are child elements of a synset. In Wordnet-LMF, senses, which correspond to the lexical units in GermaNet, are linked to a synset (by an attribute containing a synset ID).

The conversion of GermaNet to Wordnet-LMF proceeds as follows: Each lexical unit of GermaNet is turned into a *Sense* element in Wordnet-LMF (see Figure 5). The *synset* attribute (containing a *Synset* ID) of the *Sense* element links this *Sense* with the *Synset* that it is a member of. The different *Sense* elements are grouped by their orthographical form (the *Lemma* in Wordnet-LMF) into *Lexical Entries*.

An example of a GermaNet *LexicalEntry* in Wordnet-LMF is shown in Figure 6. This *LexicalEntry* represents the word *leuchten* (German verb for: *to shine*), as the *writtenForm* attribute of the *Lemma* element indicates. This *LexicalEntry* has two *Senses*, which belong to different *Synsets* (see the different *synset* attributes of the *Sense* elements).

460

Each *Sense* has a *MonolingualExternalRefs* element with at least one *MonolingualExternalRef* representing a reference to an external system. In this case, each *Sense* is linked to the corresponding entry in the GermaNet database[15]; the *externalReference* attribute of a *MonolingualExternalRef* specifies the database table name with a database ID.

```
<LexicalEntry id="deu-52-l4601-v">
  <Lemma writtenForm="leuchten"
                      partOfSpeech="v" />
  <Sense id="deu-52-l4601-v_1"
              synset="deu-52-s58377-v">
    <MonolingualExternalRefs>
      <MonolingualExternalRef
        externalSystem="GermaNet-Database"
        externalReference=
             "lex_uni_table#id=82207" />
    </MonolingualExternalRefs>
  </Sense>
  <Sense id="deu-52-l4601-v_2"
              synset="deu-52-s58718-v">
    <MonolingualExternalRefs>
      <MonolingualExternalRef
        externalSystem="GermaNet-Database"
        externalReference=
             "lex_uni_table#id=82677" />
    </MonolingualExternalRefs>
  </Sense>
</LexicalEntry>
```

Figure 6. Example of a *LexicalEntry*.

In the next conversion step, all synsets of GermaNet are listed with their relations to other synsets. The corresponding *Synset* (with the ID *deu-52-s58377-v*) of the first *Sense* in Figure 6 is illustrated in Figure 7. It has, inter alia, a describing *gloss* and two *example* sentences.

The element *SynsetRelations* encodes relations to other *Synset* instances. The relations are simply encoded with a *target* attribute that contains the ID of the referencing *Synset*. The *Synsets* in Wordnet-LMF are logically the "same" as the synsets in GermaNet XML, i.e. the concept that a synset expresses is exactly the same in both formats.

Each *Synset* has a reference to the GermaNet database. Therefore, the *MonolingualExternalRef* element links to the corresponding entry in the GermaNet database; the

externalReference attribute specifies the database table name with the synsets database ID.

```
<Synset id="deu-52-s58377-v"
                    baseConcept="1">
  <Definition gloss="Lichtstrahlen
            aussenden, große Helligkeit
            verbreiten">
    <Statement example="Der Mond leuchtete
                        in der Nacht."/>
    <Statement example="Die Lichter der
          Stadt strahlen in die Nacht."/>
  </Definition>
  <SynsetRelations>
    <SynsetRelation
            target="deu-52-s58376-v"
            relType="has_hyperonym"/>
  </SynsetRelations>
  <MonolingualExternalRefs>
    <MonolingualExternalRef
        externalSystem="GermaNet-Database"
        externalReference=
             "synset_table#id=58377"/>
  </MonolingualExternalRefs>
</Synset>
```

Figure 7. Example of a *Synset*.

These two Figures 6 and 7 represent the same example in Wordnet-LMF that was already shown in the GermaNet XML format in Figure 1.

4.4 Necessary Modifications to Wordnet-LMF

As the previous discussion has shown, Wordnet-LMF provides a very useful basis for converting GermaNet into LMF. However, a number of modifications to Wordnet-LMF are needed if this conversion is to preserve all information present in the original resource. The present section will discuss a number of modifications to Wordnet-LMF that are needed for conversion of wordnets in general. In addition, we will also discuss a set of extensions to Wordnet-LMF that are needed for conversion of GermaNet in particular.

The most glaring omission in Wordnet-LMF concerns the modeling of lexical relations which hold between lexical units (i.e., *Senses* in the terminology of Wordnet-LMF). In the current Wordnet-LMF DTD only conceptual relations (i.e., *SynsetRelations* in the terminology of Wordnet-LMF), which hold between synsets, are modeled. Thus antonymy, which is a typical example of a lexical relation (see (Fellbaum, 1998) for further details), can cur-

---

[15] For efficency reasons, GermaNet is stored in a relational database.

rently not be modeled without violating the Wordnet-LMF DTD.

Among the synset relations specified in Wordnet-LMF, the entailment relation is missing, which plays a crucial role in the modeling of verbs in the Princeton WordNet and in GermaNet alike. The list of values of attribute *relType* for *SynsetRelation* elements (see Appendix A in (Soria and Monachini, 2008)) therefore has to be amended accordingly.[16]

A third omission in the current Wordnet-LMF DTD concerns syntactic frames used in the Princeton WordNet to indicate the syntactic valence of a given word sense. Syntactic frames are also used in GermaNet, albeit using a different encoding[17]. Syntactic frames together with example sentences, which illustrate the meaning and prototypical usage of a particular word, help to distinguish among word senses.

In WordNet both syntactic frames and examples are linked to synsets. However, at least in the case of syntactic frames the linkage to synsets seems problematic since different members of the same synset may well have different valence frames. For example, the German verbs *finden* and *begegnen* both mean *meet* and thus belong to the same synset. Both are transitive verbs, but their object NPs have different cases: accusative case for *treffen* and dative case for *begegnen*. As this example shows, syntactic frames need to be associated with lexical units rather than synsets. This is exactly the design choice made in GermaNet, as shown in Figure 1.

A related question concerns the anchoring of example sentences which illustrate the meanings and prototypical usage of a particular word sense. In both the Princeton WordNet and GermaNet such examples are associated with lexical units[18]. GermaNet correlates examples additionally with particular syntactic frames and treats both examples and syntactic frames as properties of lexical units, i.e. *Senses* in the terminology of Wordnet-LMF.

The above issues lead to a modified version of the Wordnet-LMF DTD as shown in Figure 8. Compared to Figure 5, the *Sense* element is enriched by three optional subelements: *SenseRelations*, *SenseExamples*, and *SubcategorizationFrames*.

It has to be noted, though, that LMF proper contains all necessary elements. The three notions *SenseRelation*, *SenseExample*, and *SubcategorizationFrame* come from LMF proper and these elements can be used to remedy the omissions in Wordnet-LMF.

The *SenseRelation* element in Figure 8 represents relations between different *Senses* (the lexical units in GermaNet). The *SenseExamples* and *SubcategorizationFrames* elements both group several *SenseExample* or *SubcategorizationFrame* instances. A *SubcategorizationFrame* element represents the syntactic valence of a word sense. A *SenseExample* shows the prototypical usage of a word sense as an example sentence. The syntactic valence for a concrete example sentence can be specified with the optional *frame* attribute of a *SenseExample*.

5   Conclusion: Comparing GermaNet XML with Wordnet-LMF XML

We would like to conclude with a comparison between the GermaNet native XML format described in section 3 and the modified Wordnet-LMF format described in section 4.4. Since the GermaNet native XML format was particularly tailored to the structure of GermaNet, it enjoys the usual advantages of such customized solutions: it contains all and only the necessary XML elements and attributes to describe the resource. Moreover, the data are distributed over 55 different XML files, which facilitates easy data handling and efficient search by word classes and lexical fields. These properties are in fact exploited by a number of GermaNet-specific tools, including

---

[16] Piek Vossen (personal communication) has pointed out to us that Wordnet-LMF does not impose a list of relations as a standard yet.

[17] In WordNet, frames are encoded in a controlled language using paraphrases such as *Somebody ----s something* for a transitive verb with an animate subject and an inanimate object. The frames in GermaNet use complementation codes provided with the German version of the CELEX Lexical Database (Baayen et al., 2005) such as *NN.AN* for transitive verbs with accusative objects.

[18] In WordNet, the examples are placed at the synset level, but referencing to a word sense at the same time.
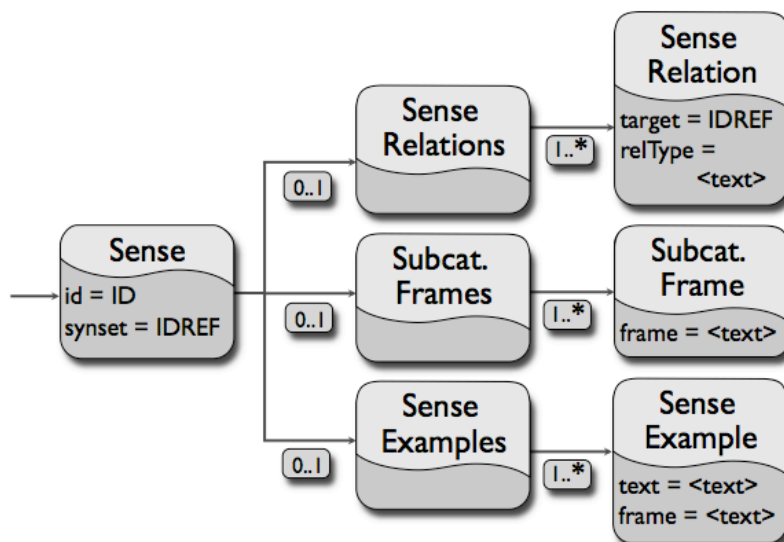
Figure 8. Revised Wordnet-LMF structure.

a GermaNet-Explorer, a tool for data exploration and retrieval, and a GermaNet Pathfinder, a tool for the calculation of semantic relatedness, similarity, and distance (Cramer and Finthammer, 2008). All of these tools utilize the Java API that has been developed for the GermaNet native XML format.

At the same time the GermaNet native XML format is a proprietary data format that was developed at a time when the only de facto encoding standard for wordnets consisted of the lexicographer files, originally developed for the Princeton WordNet. As such GermaNet XML was never developed with the goal of providing an XML standard for modeling wordnets in general. With Wordnet-LMF a candidate standard has now been proposed that is compliant with the LMF ISO standard for lexical resources and that strives to provide a general encoding standard of wordnets for different languages. As the discussion in section 4.4 has shown, the current Wordnet-LMF DTD still needs to be amended to account for the full range of wordnet relations, frames, and examples (see Figure 8). These elements are not in Wordnet-LMF because Wordnet-LMF is a subset, but these elements are defined in the ISO document 24613 where LMF proper is defined. However, Wordnet-LMF appears to be suitably mature to serve as an interchange format for wordnets of different languages as well as for linking wordnets of different languages with one another[19].

Acknowledgements

References

Baayen, R. H., R. Piepenbrock, and L. Gulikers. 2005. *The CELEX Lexical Database (Release 2) CD-ROM.* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania (Distributor).

Cramer, Irene, and Marc Finthammer. 2008. Tools for Exploring GermaNet in the Context of CL-Teaching. In: *Angelika Storrer, Alexander Geyken, Alexander Siebert, and Kay-Michael Würzner, (Eds.): Text Resources and Lexical Knowledge*. Selected Papers from the 9th Conference on Natural Language Processing KONVENS 2008. Berlin/New York: Mouton de Gruyter, 195-208.

---

[19] For example, the Interlingual Index, based on the Princeton WordNet, can be used to link different wordnets with one another.

Kunze, Claudia, and Lothar Lemnitzer. 2002. GermaNet – representation, visualization, application. *Proceedings of LREC 2002*, main conference, Vol V. pp. 1485-1491.

Fellbaum, Christiane (eds.). 1998. *WordNet – An Electronic Lexical Database*. The MIT Press.

Francopoulo, Gil, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet, and Claudia Soria. 2006. Lexical markup framework (LMF). *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*. Genoa, Italy.

Henrich, Verena, and Erhard Hinrichs. 2010. GernEdiT – The GermaNet Editing Tool. *Proceedings of LREC 2010*, main conference. Valletta, Malta.

Lee, Lung-Hao, Shu-Kai Hsieh, and Chu-Ren Huang. 2009. CWN-LMF: Chinese WordNet in the Lexical Markup Framework. *Proceedings of the 7th Workshop on Asian Resources*. Suntec, Singapore, August 06 - 07, 2009, pp. 123-130

Lemnitzer, Lothar, and Claudia Kunze. 2002. Adapting GermaNet for the Web. *Proceedings of the First Global Wordnet Conference*. Central Institute of Indian Languages, Mysore, India, 21.-25.01.2002, pp. 174-181

Soria, Claudia, Monica Monachini, and Piek Vossen. 2009. Wordnet-LMF: Fleshing out a Standardized Format for Wordnet Interoperability. *Proceedings of ACM Workshop on Intercultural Collaboration*.

Soria, Claudia, and Monica Monachini. 2008. *Kyoto-LMF – Wordnet representation format*. KYOTO Working paper: WP02_TR002_V04_Kyoto_LMF.

Vossen, Piek, Eneko Agirre, Nicoletta Calzolari, Christiane Fellbaum, Shu-kai Hsieh, Chu-Ren Huang, Hitoshi Isahara, Kyoko Kanzaki, Andrea Marchetti, Monica Monachini, Federico Neri, Remo Raffaelli, German Rigau, Maurizio Tescon, and Joop VanGent. 2008. KYOTO: A system for mining, structuring and distributing knowledge across languages and cultures. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco.