Towards an integrated OWL model for domain-specific and general language wordnets

Harald Lüngen[i], Claudia Kunze[ii], Angelika Storrer[iii], Lothar Lemnitzer[ii]

[i] Justus-Liebig-Universität Gießen, [ii] University of Tübingen,
[iii] University of Dortmund
luengen@uni-giessen.de, kunze@sfs.uni-tuebingen.de
angelika.storrer@uni-dortmund.de, lothar@sfs.uni-tuebingen.de

Abstract. This paper presents an approach to integrate the general language wordnet GermaNet with TermNet, a German domain-specific ontology. Both resources are represented in the Web Ontology Language OWL. For GermaNet, we adopted the OWL model suggested by van Assem et al. 2006 for the Princeton WordNet, for TermNet we developed a slightly different model better suited to terminologies. We will show how both resources can be inter-related using the idea of plug-in relations (as proposed by Magnini and Speranza 2002). In contrast to earlier plug-in approaches, our method of connecting general language wordnets with domain-specific terminology does not impose changes on the structure of these two types of lexical representations. We therefore consider our proposal to be a step towards the interoperability of lexical-semantic resources.

Keywords. *wordnets; GermaNet; OWL; terminology*

1. Introduction

Wordnets (like the Princeton WordNet, cf. Fellbaum 1998) have been used in various applications of text processing, information retrieval, and information extraction. When these applications process documents dealing with a specific domain, one needs to combine knowledge about the domain-specific vocabulary represented in domain ontologies with lexical repositories representing general language vocabulary. In this context, it is useful to represent and inter-relate the entities and relations in both types of resources using a common representation language. In this paper we discuss an integrated representation model for domain-specific and general language resources using the Web Ontology Language OWL. The model was tested by relating entities of the German wordnet GermaNet to corresponding entities of the German domain ontology TermNet (Kunze et al. 2007).

In Section 3, the main characteristics of these two resources are described. We built on the W3C approach to convert Princeton WordNet in RDF/OWL (van Assem et al. 2006) and adapted them to GermaNet. For the domain ontology TermNet a different model was developed. The main classes and properties of both models are discussed in Section 4. The focus of this paper is on the question how the entities of the two

OWL models — the model of the general language wordnet GermaNet and the model of the domain ontology TermNet — can be linked in a principled fashion. For this purpose, we defined OWL-properties that relate entities of the two lexical resources, following the basic idea of the so-called plug-in approach by Magnini & Speranza (2002) for linking general language with domain-specific wordnets. Section 5 discusses the plug-in approach and our adaptation of it with reference to appropriate examples from GermaNet and TermNet. With our work, we aim at contributing to the emergent issue of interoperability between language resources.

2. Related Work

The work presented in this paper is inspired by the plug-in approach, which was developed in the context of ItalWordNet (Roventini et al. 2003) and was originally proposed by Magnini and Speranza (2002). However, rather than focusing on the processing aspects of the original method, in the present study we propose a declarative model of interlinking general language with domain-specific wordnets from the perspective of explicitly defined plug-in relations, which differ slightly from the ones proposed by Magnini and Speranza. These relations allow for connecting specific terms with appropriate concepts, but do not modify the original resources and concepts.

Subsequent applications of the plug-in approach, like ArchiWordNet (Bentevogli et al. 2004) or Jur-WordNet (Bertagna et al. 2004), implement plug-in relations for extending generic resources with domain terms from a processing perspective. The procedures lead to merged concepts and additional features being integrated into or added to the original databases.

De Luca and Nürnberger (2007) describe an approach that relates an OWL representation of EuroWordNet to an OWL representation of domain terms. In their approach, terms are directly mapped onto synsets without any reference to intermediate relations. By defining distinct OWL plug-in properties our model aims to capture, in addition, different types of semantic correspondence between general language and domain-specific concepts

In Vossen's (2001) approach, WordNet is adapted to the field and the needs of a specific organisation by extending it to include domain-specific vocabulary and removing concepts (and thus word senses) that are irrelevant for the organisation. In contrast, both the plug-in approach and the approach introduced in this paper are neutral with respect to the question whether a global ontology is extended by a specialised ontology or the other way around. Furthermore, the plug-in approach and the present approach do not address the question of how to automatically derive a domain ontology from a text collection; they are applicable to both automatically derived ontologies and hand-crafted ones. Moreover, Vossen's (2001) approach is procedural, meaning that its focus is on the specification of an extraction and integration algorithm, whereas the aim of the present paper is to declaratively model and specify the relational structure of the interface between a general and a domain-specific ontology in a formal language, i.e. the Semantic Web Ontology Language OWL.

3. Lexical and Terminological Resources

Our approach was developed and tested using an OWL model for a representative subset of the German wordnet GermaNet and an OWL model for the German terminological wordnet TermNet. In this section we outline the main characteristics of GermaNet and TermNet.

3.1 Characteristics of GermaNet

GermaNet is a lexical-semantic wordnet for German which has been developed along the lines of the Princeton WordNet (Fellbaum 1998), covering the most important and frequent general language concepts and lexical-semantic relations holding between the concepts and lexical units represented, like hyponymy, meronymy and antonymy (Kunze 2001). As is typical of wordnets, the central unit of representation is the synset, which comprises all synonyms or lexical units of a given concept. GermaNet presently covers more than 53 000 synsets with some 76 000 lexical units, among them nouns, verbs and adjectives. A basic subset of GermaNet (15 000 concepts) has been integrated into the polylingual EuroWordNet database (Vossen 1999). The following features distinguish GermaNet from the data model of the Princeton WordNet, version 2.0:

1. The use of so-called artificial, non-lexicalised concepts, in order to achieve well-formed taxonomic hierarchies. For example, the artificial concept *Schultyplehrer* ('school type teacher') has been introduced to act as a hyper(o)nym of the lexicalised concepts *Grundschullehrer* ('primary school teacher'), *Realschullehrer* ('secondary school teacher'), *Berufsschullehrer* ('vocational school teacher') etc.;
2. Named entities are explicitly marked. Proper names in GermaNet primarily occur in the geographic domain;[1]
3. In GermaNet, the taxonomic approach is also applied to the representation of adjectives, as opposed to WordNet's satellite approach (based upon the notion of similarity with regard to different adjective clusters);
4. Meronymy is deemed a generic relation in GermaNet;
5. GermaNet verbs are provided with an exhaustive list of sub-categorisation frames and example phrases.

The data model of GermaNet is depicted in Fig. 1 as an entity-relationship graph. This model guided the conversion process of GermaNet objects and relations into XML elements and attributes.

1 In version 2.1 of WordNet, however, over 7,600 synsets were manually classified as instances and tagged as such (cf. Miller/Hristea 2006, p. 3).

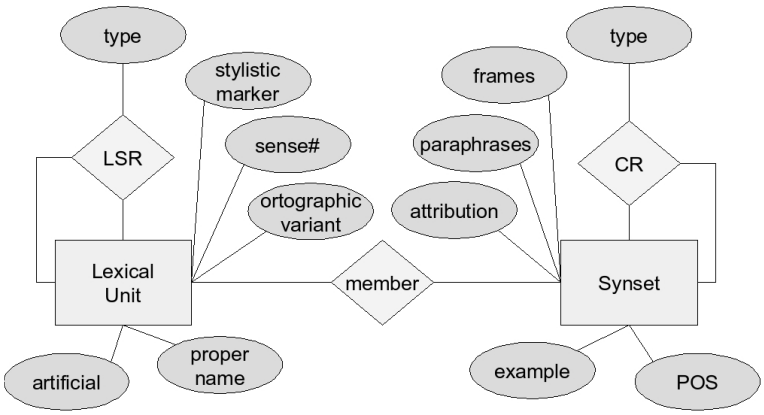

Fig. 1. Entity-relationship diagram for GermaNet

3.2 Characteristics of TermNet

TermNet is a lexical resource that was developed in a project on automated text-to-hypertext conversion (cf. Lenz/Storrer 2006). TermNet represents more than 400 German technical terms occurring in a corpus with documents in the domains "text-technology" and "hypertext research." Most terms are noun terms, including multiword terms composed of a noun and an adjective modifier such as *bidirektionaler Link* (engl. 'bidirectional link'). The entities and relations introduced for the Princeton WordNet (Fellbaum 1998) are fundamental for the structure of TermNet. The two basic entities of the TermNet model are *terms* (the analogue to word/lexical unit in the WordNet model) and *termsets* (the analogue to synsets in the WordNet model). Terms in TermNet are lexical units for which the technical meaning is explicitly defined in the documents of our corpus. Termsets contain technical terms that denote the same or a quite similar topic in different approaches to a given domain (cf. Beißwenger et. al 2003). Terms are related by lexical relations, e.g. *isAbbreviationOf*, and termsets are related by conceptual relations, e.g. is*HyponymO*f, *isMeronymOf*. The data model of TermNet is illustrated by the ER-diagram in Fig. 2.

For automated hyperlinking, and probably for other applications, it is useful to know that term A occurring in document X denotes a category similar to the one denoted by term B occurring in document Y. Unlike other standards and proposals for representing thesauri (e.g. ISO 1986, ANSI/NISO 2003, SKOS 2005), TermNet focuses on the representation of semantic correspondences between terms defined in different taxonomies or in competing scientific schools. Since competing taxonomies or schools may all have their benefits, we do not want to decide which terminology is to be preferred. Thus, the current TermNet model deliberately does not label terms as "preferred term."

Since the entity type *TermSet* is crucial for the purpose of representing semantic correspondences between technical terms defined in competing schools, we want to explain the idea behind it using an example from German hypertext terminology: Kuhlen (1991) and Tochtermann (1995) both introduced a terminology for hypertext concepts that influenced the usage of technical terms in German publications on hypertext research. Both authors provide definitions for the concept *hyperlink* and specify a taxonomy of subclasses (*external link*, *bidirectional link* etc.). But Kuhlen uses

the term *Verknüpfung* in his taxonomy (*extratextuelle Verknüpfung*, *bidirektionale Verknüpfung*) while in Tochtermann's taxonomy the term *Verweis* is used (with subclasses like *externer Verweis*, *bidirektionaler Verweis*). The definitions of the concepts and subconcepts given by these authors are slightly different, and the two taxonomies are not isomorphic. As a consequence, in a scientific document on the subject domain, a term from the Kuhlen taxonomy cannot be replaced by the corresponding term from the Tochtermann taxonomy. After all, the purpose of defining terms is exactly to bind their word forms to the semantics specified in the definition. The usage of technical terms in documents may then serve to indicate the theoretical framework or scientific school to which the paper belongs. In our OWL model of TermNet, on the one hand we represent relations between terms of the same taxonomy, on the other hand we capture categorial correspondences between terms of competing taxonomies by assigning similar terms to the same termset.

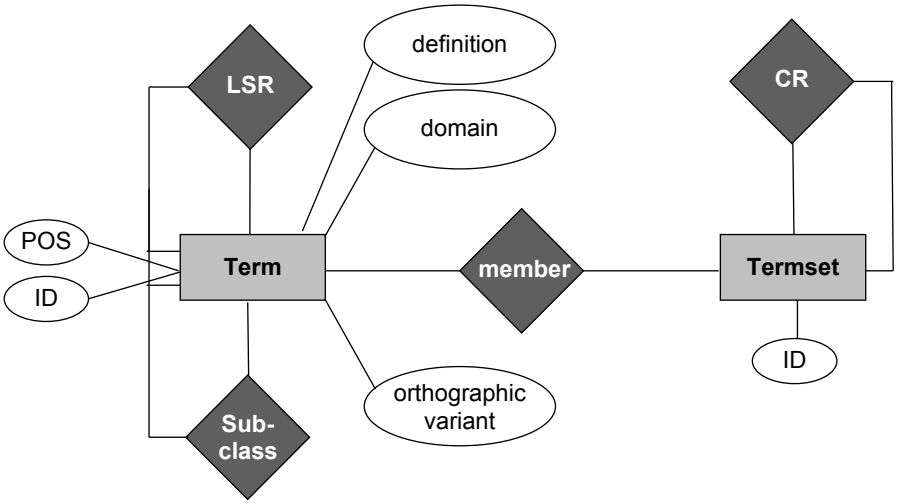

Fig. 2. Entity-relationship diagram for TermNet

4. OWL Models of GermaNet and TermNet

The Web Ontology Language OWL was created by the W3C Ontology Working Group as a standard for the specification of ontologies in the context of the Semantic Web. OWL comprises the three sublanguages OWL Light, OWL DL, and OWL Full, which differ in their expressivity. An ontology in the sublanguage OWL DL can be interpreted according to description logics (DL), and DL-based reasoning software (e.g.

RacerPro[2] or Pellet[3]) can be applied to check its consistency or to draw inferences from it. To take advantage of this, our OWL models of GermaNet, TermNet and the plug-in structure all remain within the OWL DL dialect.

Several approaches to convert PWN to OWL and to make it available for Semantic Web applications exist (e.g. Ciorăscu et al. 2003, van Assem et al. 2004, and van Assem et al. 2006). In all these, the individual synsets and lexical units are rendered as instances of the OWL ontology. Although alternative modelling options have been discussed (cf. Lüngen/Storrer 2007), in the present project we adhere to an instance model as proposed by van Assem et al. 2006.

4.1 GermaNet OWL Model

In our OWL model, sets of GN concepts are represented as classes (<owl:class>), while the properties of and relations between concepts are represented as OWL properties (<owl:ObjectProperty> or <owl:DatatypeProperty>) of these classes. For the two basic objects in the E-R-model of GN (Fig. 1), the classes *Synset* and *LexicalUnit* are introduced. Following the W3C model for PWN (van Assem et al. 2006), we introduce *NounSynset, VerbSynset, AdjectiveSynset,* and *AdverbSynset* as immediate subclasses of *Synset*, as well as *NounUnit*, *VerbUnit*, *AdjectiveUnit*, and *AdverbUnit* as immediate subclasses of *LexicalUnit*.

Table 1. Features of OWL object properties for GermaNet

Property	**Domain**	**Range**	**Characteristics**	**Inverse Property**	**Local Restrictions**
hasMember	Synset	LexicalUnit	inverse-functional	memberOf	*POS-based*
memberOf	LexicalUnit	Synset	functional	hasMember	*POS-based*
hasExample	Synset	Example	-	-	-
Conceptual Relations (CR)					
isHyperonymOf	Synset	Synset	transitive	isHyponymOf	*POS-based*
isHyponymOf	Synset	Synset	transitive	isHyperonymOf	*POS-based*
isHolonymOf	NounSynset	NounSynset	-	-	-
isMeronymOf	NounSynset	NounSynset	-	-	-
IsAssociated-With	Synset	Synset	-	-	-
entails	VerbSynset	VerbSynset	-	-	-
causes	VerbSynset	VerbSynset ∪ AdjectiveSynset	-	-	-
Lexical-semantic relations (LSR)					
hasAntonym	LexicalUnit	LexicalUnit	symmetric	hasAntonym	*POS-based*
hasPertainym	LexicalUnit	LexicalUnit	-	-	-
isParticipleOf	VerbUnit	VerbUnit	-	-	-

For modelling the lexicalisation relation between synsets and lexical units, an OWL *Object Property* called *hasMember* with domain *Synset* and range *LexicalUnit*

2 cf. http://www.racer-systems.com
3 cf. http://pellet.owldl.com

is introduced. For each POS-based subclass of *Synset* (e.g. *NounSynset*), a restriction of the range of *hasMember* to the corresponding subclass of *LexicalUnit* (e.g. *NounUnit*) is encoded using <owl:allValuesFrom>.

OWL is particularly well-suited to model the two basic relation types CR and LSR. Both types hold between internally defined classes and thus correspond to object properties in OWL. Like classes, properties can be arranged in a hierarchy in OWL

```
<owl:TransitiveProperty rdf:about="#isHyperonymOf">
   <rdfs:subPropertyOf rdf:resource="#conceptualRelation"/>
   <rdfs:domain rdf:resource="#Synset"/>
   <rdf:type rdf:resource=
             "http://www.w3.org/2002/07/owl#ObjectProperty"/>
   <owl:inverseOf>
      <owl:TransitiveProperty rdf:about="#isHyponymOf"/>
   </owl:inverseOf>
</owl:TransitiveProperty>
```

Listing 1: OWL Code for the introduction of hypernymy

using the <rdfs:subPropertyOf> construct. Our model thus contains two top-level object properties, *conceptualRelation* (with domain and range = *Synset*) and *lexicalSemanticRelation* (with domain and range = *LexicalUnit*). The OWL characteristics of their respective subproperties are shown in Table 1. Hypernymy, for example, is encoded as an <owl:TransitiveProperty> called *isHyperonymOf* with domain and range = *Synset*, as an immediate subproperty of *conceptualRelation*, and as the inverse property of hyponymy, cf. Listing 1.

Similar to *hasMember*, for each POS-based subclass of *Synset*, the range of *isHyperonymOf* is restricted to synsets of the same subclass. Relations that do not hold between internally defined classes, but in which a range in the form of an XML Schema data type like string or boolean is assigned to an internal class, are modelled as OWL datatype properties. In the case of GN, they are obviously the ones that are represented as ellipses in the E-R model of GN (Fig. 1). Table 2 contains a survey of datatype properties in the OWL model of GN with their respective domain, range and function status.

Table 2. Features of OWL datatype properties for GermaNet

Property	**Domain**	**Range**	**Functional**
POS	Synset	“N”|”V”|”A”|”ADV”	yes
hasParaphrase	Synset	xs:string	no
isArtificial	Synset ∪ LexicalUnit	xs:boolean	(yes)
isProperName	NounSynset ∪ NounUnit	xs:boolean	(yes)
hasOrthographicForm	LexicalUnit	xs:string	yes
hasSenseInt	LexicalUnit	xs:positiveInteger	yes
isStylisticallyMarked	LexicalUnit	xs:boolean	(yes)
hasFrame	VerbUnit ∪ Example	xs:string	no
hasText	Example	xs:string	yes

A subset of GermaNet (54 synset and 104 lexical unit instances including all conceptual and lexical-semantic relations holding between them) has been encoded in OWL according to the model presented above, using the Protégé ontology editor[4]. The GermaNet subset contains most of the candidate synsets for plugging in TermNet terms. Furthermore, this exemplary subset contains at least one instance of each conceptual and each lexical-semantic relation. We employed the reasoning software RacerPro to ensure its consistency within OWL DL. An automatic conversion of the complete GermaNet 5.0 is under way.

4.2 TermNet OWL Model

The complete TermNet in its OWL representation contains 425 technical terms and 206 termsets. In the OWL model we define all terms as classes, the instances of which are those objects in the real world that are denoted by the respective terms (e.g., an instance of the term *externer Verweis* is a concrete hyperlink in a hyperdocument compliant with Tochtermann's definition of this term). Since we only account for nominal terms, all terms are subclasses of the superclass *NounTerm*. We use the <rdf:subClassOf> property to relate narrower terms to broader terms within the same taxonomy (e.g., we define Kuhlen's term *extratextuelle Verknüpfung* as a subclass of his broader

```
<owl:Class rdf:ID="Verweis">
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty>
            <owl:ObjectProperty rdf:ID="isMemberOf"/>
         </owl:onProperty>
         <owl:allValuesFrom>
            <owl:Class rdf:ID="TermSet_Link"/>
         </owl:allValuesFrom>
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
```

Listing 2: OWL code for the assignment of terms to termset

term *Verknüpfung*). By modelling terms as classes we benefit from the mechanism of feature inheritance related to the predefined <rdf:subClassOf> property. In addition, we are able to represent disjointness between classes using the OWL <owl:disjointWith> construct. By defining that the sets of instances denoted by the terms *externer Verweis* and *interner Verweis* are disjoint, we make sure that a link object in a document can only be assigned to one of these classes. In other words, a link object can either be an instance of the class *externer Verweis* or an instance of the class *interner Verweis* (although it may quite well be an instance of both *externer Verweis* and *bidirektionaler Verweis*). Terms of competing taxonomies that represent similar categories (like *externer Verweis* and *extratextuelle Verknüpfung* from the example in Sect. 3.2) are assigned to the same termset. For this purpose termsets are defined as subclasses of the superclass *NounTermSet*. Terms are assigned to termsets using the object property *tn:isMemberOf* (with *NounTerm* as domain and *NounTermSet* as range). The inverse property is *tn:hasMember*. Since termsets and terms are modelled as

4 http://protege.stanford.edu

classes, we cannot simply adopt the definition of the *gn:MemberOf* object property specified in the GermaNet OWL model (cf. Sect. 4.1.). Instead, we had to use the <owl:allValuesFrom> restriction to assign all instances of a term class to the respective termset class. Listing 2 illustrates how the term *Verweis* is assigned to the termset *Termset_Link* (which comprises other terms like *Verknüpfung*, *Link*, *Hyperlink*, *Kante* etc.).

In addition to the taxonomic relations specified between terms of the same taxonomy by means of the <rdf:subClassOf> property, we also represent hierarchical relations between termsets, e.g. we want to account for the fact that all terms assigned to the termset *TermSet_Link* have a broader meaning than the terms assigned to the

```
<owl:Class rdf:ID="tn:TermSet_MonodirektionalerLink">
   <rdfs:subClassOf rdf:resource="#tn:NounTermSet"/>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:onProperty rdf:resource="#tn:isHyponymOf"/>
         <owl:allValuesFrom rdf:resource="#tn:TermSet_Link"/>
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
```

Listing 3: OWL code for a hyponym relation between Termsets

termset *TermSet_Monodirektionaler_Link*. For this purpose, we defined the *tn:isHypernymOf*–Property, which relates termsets containing broader terms to termsets containing more specific terms. Its inverse property is *isHyponymOf*. Listing 3 demonstrates how the more specific termset *TermSet_Monodirektionaler_Link* is defined to be a hyponym of the broader termset *TermSet_Link* by means of the property *isHyponymOf* and the <owl:allValuesFrom> restriction.

Table 3. Features of OWL object properties for TermNet

Property	**Domain**	**Range**	**Characteristics**	**Inverse Property**
hasMember	NounTermSet	NounTerm	inverse-functional	isMemberOf
isMemberOf	NounTerm	NounTermSet	functional	hasMember
Relations between termsets				
isHypernymOf	NounTermSet	NounTermSet	transitive	isHyponymOf
isHyponymOf	NounTermSet	NounTermSet	transitive	isHypernymOf
isHolonymOf	NounTermSet	NounTermSet		
isMeronymOf	NounTermSet	NounTermSet		
Relations between terms				
IsAbbreviationOf	NounTerm	NounTerm		isExpansionOf
IsExpansionOf	NounTerm	NounTerm		isAbbreviationOf

The object properties *isMeryonymOf* and *isHolonymOf* were introduced to account for part-whole-relations between objects denoted by the terms of two termsets. The property *isAbbreviationOf* relates short terms to their expanded forms within the same taxonomy. Table 3 provides an overview of the properties defined in the OWL Term-

Net model. The <rdf:subClassOf> property between terms of the same taxonomy is not included in this overview because its semantics is predefined.

5. Representing Plug-in Relations in OWL

Providing domain-specific extensions for general language resources in order to capture and exploit the respective advantages of both resource types in natural language processing and semantic web applications has been discussed in the approaches by Vossen (2001) and Magnini and Speranza (2002).

Vossen (2001) describes a procedure to extract a hierarchy of terms (called "topics") from a document collection, e.g. the set of all documents used in a specific organisation, and to subsequently combine it with WordNet. This is achieved by merging topics from the extracted hierarchy with matching WN concepts. The kind of matching criterion used is not specified; from the examples given one can assume that simple string matching is applied. Similar to the plug-in approach, one of the features of the resulting hierarchy is that the lower levels of the WN hierarchy and the possible upper levels of the terminological hierarchy are discarded. Vossen's procedure only identifies plug-in synonymy and plug-in near synonymy, which are not differentiated in the new hierarchy.

The resulting hierarchy is subsequently trimmed by automatically removing those concepts that are irrelevant in the domain of the document collection, i.e. removing unwanted sense ambiguities that were introduced by the merger of the two resources. Finally, a procedure to fuse the compositional hierarchy with a so-called "private" or "personal" ontology, which apparently is a more domain-specific upper level ontology designed for the organisation and its document collection, is presented. For the fusing procedure, an "interface level" with matching concepts or topics from the source and target hierarchies seems to be externally defined, i.e. criteria other than string matching could potentially be applied. In this step, subtrees of the combined hierarchy are placed under the interface nodes of the private ontology; thus, it can be regarded as another instance of merging a global with a specialised ontology.

Whereas Vossen first builds ad-hoc terminologies from large document collections using information retrieval and term extraction methods and then links the resulting terms to WordNet synsets, Magnini and Speranza have proposed the plug-in approach which serves to link two (independently) existing resources of different types, namely the general-language ItalWordNet (IWN) and the specialised ontology ECOWN from the economic domain. Plug-in is a special instance of *ontology merging*, which is normally concerned with aligning resources of the *same* type.

Various kinds of plug-in relations serve to combine the relevant synsets of both resources. The plug-in approach yields a common hierarchy in which the top concepts of the specialised ontology are "eclipsed" while the subordinate concepts, the terms, are imported into the general language ontology. A relatively small number of instances of plug-in relations (269) suffices to integrate 4662 ECOWN concepts into ItalWordNet (cf. Magnini and Speranza 2002).

ECOWN synsets are linked to a small domain ontology; 100 basic terms dominating relevant subhierarchies have been selected by experts due to relevance and frequency of use. The following scenarios of correspondences between IWN synsets and ECOWN terms are discussed:

1. Overlapping concepts: generic terms from the economic domain which also play a role in general language;
2. Overdifferentiation: a given ECOWN synset corresponds to more than one IWN concept, or an IWN synset corresponds to more than one ECOWN concept—these phenomena can be traced to different sense distinctions made by lexicographers vs. terminologists;
3. Gaps: for terms which have no general language counterpart, a suitable hypernym in the generic resource is selected.

The first scenario is captured by *plug-in synonymy* for overlapping synsets in IWN and ECOWN. A new plug-synset is created which replaces the corresponding IWN and ECOWN synsets in the integrated resource. This plug-synset takes its synonyms and hyponyms from the terminological resource and its hypernym from the generic resource. As a consequence, the terminological hypernym and the general language hyponyms are eclipsed.

The case of overdifferentiation is dealt with by *plug-in near-synonymy*. A new plug-synset is being created which also takes its hypernym from IWN and its synonyms and hyponyms from ECOWN.

In order to bridge the gap between IWN synset and ECOWN synset in the third scenario, *plug-in hyponomy* is applied. Two new plug-synsets are derived: one for the superordinate IWN synset (Plug-IWN) and one for the subordinate ECOWN synset (Plug-ECOWN). Plug-IWN takes its synonyms and hyponyms from IWN, and its hyponyms also include the Plug-ECOWN node. Plug-ECOWN relates to synonyms and hyponyms from ECOWN. Plug-ECOWN is assigned a new hypernym: Plug-IWN replaces the former hypernym from ECOWN.

The integration process is realised in four steps that centre around the plug-in relations. Thus, plug-in can be seen as a dynamic device with regard to merging two resources. The procedure yields new concepts, the plug-in concepts. The status of these merged plug-in concepts remains unclear—whether they constitute new lexical items, new terms or artificial concepts.

The plug-in approach has also been used and enhanced for Jur-WordNet (Bertagna et al. 2004) and ArchiWordNet (Bentevogli et al. 2004), two domain-specific wordnet extensions. Jur-WordNet addresses theoretical considerations regarding common language versus expert language, and emphasises the citizens' perspective on law terms, applying more or less the original plug-in relations. For ArchiWordNet, several plug-in procedures (substitutive, integrative, hyponymic and inverse plug-ins) are developed to replace or rearrange MultiWordNet hierarchies and integrate them with ArchiWordNet hierarchies. Furthermore, synsets may be enriched with terminological features, synonyms may be added or deleted from synsets, and relations may be added or deleted for specific synsets. Within this merging process, a lot of manual work specific to the resources in question had to be done which might possibly not be representative for any other pair of resources.

The plug-in approach offers an attractive model for linking TermNet to GermaNet, as both resources are also wordnet-based and of different coverage and specificity with a significant number of overlapping concepts. We primarily focus on modelling the relationships between general language and domain-specific concepts, and we use the plug-in metaphor for the relational model, less for the integration process. Thus, from our linking procedure, no new plug-in concepts evolve as the outcome of merging general language synsets with terms. The original databases, GermaNet and

TermNet, remain unchanged, but are supplemented with the relational structure provided by the established plug-in links.

As described in Section 4, in our OWL models, TermNet terms are modelled as classes and GermaNet synsets as individuals. Within OWL DL, a meta-class of term classes cannot be built, i.e. OWL classes cannot be declared to be OWL individuals without resorting to OWL Full. Thus, within OWL DL, the alignment can only be realised by restricting the range of a plug-in property to the individual that represents the corresponding GN synset. We distinguish three different linking scenarios between TermNet terms and GermaNet synsets:

```
<owl:Class rdf:ID="tn:Term_Link">
   <rdfs:subClassOf rdf:resource="#tn:NounTerm"/>
   <rdfs:subClassOf>
      <owl:Restriction>
         <owl:hasValue rdf:resource="#gn:Link"/>
         <owl:onProperty>
            <owl:ObjectProperty rdf:ID="plg:attachedToNearSynonym"/>
         </owl:onProperty>
      </owl:Restriction>
   </rdfs:subClassOf>
</owl:Class>
```

Listing 4: OWL code for a relation instance of *attachedToNearSynonym*

1. Correspondence between a given TermNet term and a GermaNet synset, for example between the term *tn:term_Link* and the GermaNet noun synset *gn:Link*. The corresponding object property *plg:attachedToNearSynonym* has *tn:NounTerm* as domain and *gn:Synset* as range. By using an <*owl:Restriction*> over *plg:attachedToNearSynonym*, every individual of the class *tn:Term_Link* is assigned the individual *gn:Link* (see Listing 4). Since we do not assume pure synonymy for a corresponding term-synset pair, no synonymy link is established for plugging general language with domain language; the closest sense-relation being near-synonymy.
2. A TermNet term cannot be assigned a corresponding GermaNet synset but is a subclass of another TermNet term which in turn is linked to a GermaNet synset by *plg:attachedToNearSynonym*. For instance, the term *tn:Term_MonodirektionalerLink* stands in a subclass relation with the term *tn:Link*, which itself is linked to the GermaNet synset *gn:Link* by *plg:attachedToNearSynonym*. The property *plg:attachedToGeneralConcept* relates a term class like *tn:Term_MonodirektionalerLink* with a GermaNet synset which stands in a *plg:attachedToNearSynonym* relation with a superordinate term. Thus, a relation between indirectly linked concepts is made explicit and also serves to reduce the path length between semantically similar concepts for applications in which semantic distance measures are calculated. In this respect, we go beyond the scope of the plug-in approach which does not account for indirect links.
3. A TermNet term cannot be assigned a corresponding GermaNet synset, and, furthermore, no suitable hypernym for linking the term is available in the GermaNet data. But the term can be linked to a holonym concept in GermaNet, via the plug-in relation *plg:attachedToHolonym*. For example, the TermNet term *tn:Term_Anker* (meaning 'anchor', i.e. a part of a link in the domain of hypertext research) has no

semantic counterpart in GermaNet, but can be linked to the superordinate holonym in GermaNet, the synset *gn:Link*, by a *plg:attachedToHolonym* relation. This plug-in relation is unique to our approach and has not been derived from the original model.

Using the Protégé ontology editor and the reasoners RacerPro and Pellet, we encoded 150 OWL restrictions representing plug-in relations for plugging terms into the Synsets of the representative subset of GermaNet, 27 of which are *plg:attachedToNearSynonym*, 103 *plg:attachedToGeneralConcept*, and 20 *plg:attachedToHolonym*. In the actual integration of resources, the OWL construct <owl:imports> is applied to import both GermaNet and TermNet into the OWL file containing the plug-in specifications, while the original GermaNet and TermNet OWL ontologies remain unchanged and reside in their separate files. The integrated ontology is within OWL DL, and the reasoning software confirms its consistency.

For identifying the necessary plug-in relation instances, we adapted the *basic concepts identification* and *alignment* steps specified for the integration procedure in Magnini and Speranza (2002), using a correspondence list of GermaNet synset and TermNet term pairs which was derived on the basis of string matching. The remaining 377 TermNet terms will be linked when the complete GermaNet is available in OWL.

Applying this approach to integrating the residual TermNet terms or even further terminologies, we might possibly encounter terms without any corresponding hypernymic or holonymic concept in GermaNet. A complete alignment of both resources will yield the relevant number of instances regarding different plug-in relations and the number of concepts that cannot be linked by one of the three relations. The outcome will show whether the introduction of further types of plug-in relations is required.

Since we decided to model TermNet terms as OWL classes and GermaNet synsets as OWL individuals, the inverse relations of the plug-in relations cannot be defined within OWL DL, i.e. with *Synset* as their domain and the meta-class of all terms as range. This would however be a desirable feature of the model, even if drawing inferences is possible without it.

6. Conclusions and Future Work

Recently, the discussion about interoperability of language resources, including lexical resources of all kinds, has gained momentum. Interoperability issues are, for example, the focus of the newly-launched EU-project CLARIN (*Common Language Resources and Technology Infrastructure*, cf. www.clarin.eu). Interoperability issues include the development of standards for various kinds of resources. For wordnets and similar resources, the *Lexical Markup Framework* (LMF, Francopoulo 2006) is of utmost importance. True interoperability, however, is more than imposing format standards on resources. It should pave the way to merging and combining resources in the context of an application, even if they do not adhere to a common format standard, a requirement which often cannot be met. The plug-in approach, as we present it here, shows how lexical resources can be merged by a set of relations, while the resources themselves are left untouched. We will demonstrate that our approach can be applied to other terminological resources and wordnets.

The next steps in our research are the automatic conversion of the complete GermaNet into the OWL model presented above, and a completion of the definition of plug-in relation instances needed to connect TermNet to it. We will also implement and process a test suite of queries to the integrated ontology that are typical of text-technological applications such as thematic chaining and discourse parsing (cf. Lüngen and Storrer 2007Cramer and Finthammer 2008, in this volume), i.e. determining (transitive) hypernyms and calculating path lengths and semantic distances between synsets or units. In our approach, the merging of plug-in configurations and the pruning of the upper level of the specialised ontology as well as the lower level of the general ontology are deliberately shunned. Thus, if the effect of *eclipse* as described in Magnini and Speranza (2002) is desired, it will have to be produced by the query resolution procedure. However, we believe that this is the right place for it to go.

Another aspect of our work is worth mentioning. The aforementioned conversions of Princeton WordNet into an OWL format (Ciorăscu et al. 2003, van Assem et al. 2004, 2006) convert synsets into OWL individuals. This is surprising both from a lexicographical and a terminological point of view. Synsets are assumed to represent concepts that are lexicalised by the lexical units which a synset contains. The conversion of a synset into an OWL class seems therefore more natural. For instance, the concept *dog* represents a class of animals, of which e.g. *Fido* is an instance. Arguably a conversion of synsets into instances is due to restrictions of the OWL-DL formalism and in particular of the tools which process OWL-encoded data. Sanfilipo et al. (2005), for instance, deem the modelling of a larger amount of synsets as classes "impractical for a real-world application." Elsewhere we have reported about experiments with an alternative modelling of GermaNet, in which synsets as well as lexical units have been modelled as OWL classes (cf. Lüngen and Storrer 2007 for details).

We will therefore investigate how and with which consequences OWL classes such as the *Synset* class can be modelled as meta-classes, with the individual synsets being instances of this meta-class. Schreiber (2002) pointed out the growing need for such an extension in the Semantic Web. Thus far, the definition of meta-classes was only possible within the dialect of OWL Full. Pan et al. (2005) introduce a variant of OWL, called *OWL FA*, which provides a well-defined meta-modelling extension to OWL DL, preserving decidability. Still, the success of such an extension of OWL DL hinges on the availability of processing tools for this dialect of OWL. From our point of view, though, such an extension will facilitate linguistically more adequate representations of lexical-semantic and terminological resources. We will continue to investigate and to tap the potential of upcoming modelling standards.

References

ANSI/NISO: Guidelines for the construction, format and management of monolingual thesauri. ANSI/NISO z39.19-2003 (2003).

Beißwenger, M., Storrer, A., Runte, M.: Modellierung eines Terminologienetzes für das automatische Linking auf der Grundlage von WordNet . In: *LDV-Forum*, 19 (1/2) (Special issue on GermaNet applications, edited by Claudia Kunze, Lothar Lemnitzer, Andreas Wagner), pp. 113--125 (2003)

Bertagna, F., Sagri, M.T., Tiscornia, D.: Jur-WordNet. In: Sojka, P. et al. (eds.): Proceedings of the Global WordNet Conference 2004, pp. 305--310 (2004)

Bentevogli, L., Bocco, A., Pianta, E.: ArchiWordNet: Integrating WordNet with Domain-Specific Knowledge. In: Sojka, P. et al. (eds.): Proceedings of the Global WordNet Conference 2004, pp. 39--46 (2004)
Cramer, I.M., Finthammer, M.: An Evaluation Procedure for Word Net Based Lexical Chaining: Methods and Issues. In this volume (2008)
Ciorăscu, I., Ciorăscu, C., Stoffel, K.: Scalable Ontology Implementation Based on knOWLer. In: Proceedings of the 2nd International Semantic Web Conference (ISWC2003), Workshop on Practical and Scalable Semantic Systems. Sanibel Island, Florida (2003)
DeLuca, E.W., Nürnberger, A.: Converting EuroWordNet in OWL and extending it with domain ontologies. In: Proceedings of the GLDV-Workshop on Lexical-semantic and ontological resources. pp. 39--48 (2007)
Farrar, S.: Using Ontolinguistics for language description. In: Schalley, A. and Zaeferer, D. (eds.): Ontolinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts. Mouton de Gruyter, Berlin (2007)
Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. The MIT Press, Cambridge, MA (1998)
Francopoulo, G., Bel, N., George, M., Calzolaria, N., Monachini, M., Pet, M., Soria, C.: Lexical Markup Framework (LMF) for NLP Multilingual Resources. In: Proceedings. of the Workshop on Multilingual Language Resources and Interoperability. pp 1--8. Sidney (2006)
ISO 1986: International Organisation for Standardization. Documentation – Guidelines for the establishement and development of monolingual thesauri. ISO 2788-1986 (1986)
Kuhlen, R.: Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Springer, Berlin (1998)
Kunze, C.: Lexikalisch-semantische Wortnetze. In: Carstensen, K.-U. et al. (eds.): Computerlinguistik und Sprachtechnologie: Eine Einführung, pp. 386--393. Spektrum, Heidelberg (2001)
Kunze, C., Lemnitzer, L., Lüngen, H., Storrer, A.: Repräsentation und Verknüpfung allgemeinsprachlicher und terminologischer Wortnetze in OWL. In: Zeitschrift für Sprachwissenschaft 26 (2) (2007)
Lenz, E.A., Storrer, A.: Generating hypertext views to support selective reading. In: Proceedings of Digital Humanities, pp. 320--323. Paris (2006)
Lüngen, H., Storrer, A.: Domain ontologies and wordnets in OWL: Modelling options. In: *OTT'06. Ontologies in Text Technology: Approaches to Extract Semantic Knowledge from Structured Information.* In: Publications of the Institute of Cognitive Science (PICS), vol. 1. University of Osnabrück, (2007)
Magnini, B., Speranza, M.: Merging Global and Specialized Linguistic Ontologies. In: Proceedings of Ontolex 2002, pp. 43--48. Las Palmas de Gran Canaria, Spain (2002)
Miles, A., Brickley, D. (eds.): SKOS Core Guide. W3C Working draft 2, November 2005. Online: http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102 (2005)
Miller, G.A., Hristea, F.: WordNet Nouns: Classes and Instances. In: Computational Linguistics 32(1) (2006)
Pan, J.Z., Horrocks, I., Schreiber, G.: OWL FA: A Metamodeling Extension of OWL DL. In: Proceedings of the Workshop OWL: Experiences and directions. Galway/Ireland (2005)
Roventini, A., Alonge, A., Bertagna, F., Calzolari, N., Cancila, J., Girardi, C., Magnini, B., Marinelli, R., Speranza, M., Zampolli, A.: ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of the Italian Language. In: Zampolli, A., Calzolari, N. and Cignoni, L. (eds.): Computational Linguistics in Pisa, Special Issue of Linguistica Computazionale, Vol. XVIII-XIX. Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma (2003)
Sanfilippo, A., Tratz, S., Gregory, M., Chappell, A., Whitney, P., Posse, C., Paulson, P., Baddeley, B., Hohimer, R., White, A.: Automating Ontological Annotation with WordNet. In: Proceedings of the 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot2005) located at the 4th Semantic Web Conference. Galway/Ireland (2005)

Schreiber, G.: The Web is not Well-formed. In: IEEE Intelligent Systems, vol. 17-2, pp. 79--80 (2002)
Staab, S. Studer, R. (eds.): Handbook on Ontologies. International Handbooks on Information Systems. Springer, Heidelberg (2004)
Tochtermann, K.: Ein Modell für Hypermedia: Beschreibung und integrierte Formalisierung wesentlicher Hypermediakonzepte. Shaker, Aachen (1995)
van Assem, M., Menken, M.R., Schreiber, G., Wielemaker, J., Wielinga, B.: A Method for Converting Thesauri to RDF/OWL. In: Proceedings of the 3rd International Semantic Web Conference (ISWC 2004), Lecture Notes in Computer Science 3298 (2004)
van Assem, M., Gangemi, A., Schreiber, G.: RDF/OWL Representation of WordNet. W3C Public Working Draft of 19 June 2006 of the Semantic Web Best Practices and Deployment Working Group. Online: http://www.w3.org/TR/wordnet-rdf/ (2006)
Vossen, P.: EuroWordNet: A multilingual database with lexical-semantic networks. Kluwer Academic Publishers, Dordrecht (1999)
Vossen, P.: Extending, trimming and fusing WordNet for technical documents. In: Proceedings of NAACL-2001 Workshop on WordNet and other Lexical Resources Applications. Pittsburgh, USA (2001)