

Repräsentation und Verknüpfung allgemeinsprachlicher und terminologischer Wortnetze in OWL

CLAUDIA KUNZE, LOTHAR LEMNITZER, HARALD LÜNGEN
& ANGELIKA STORRER

Abstract

This paper describes an approach to modelling a general-language wordnet, GermaNet, and a domain-specific wordnet, TermNet, in the web ontology language OWL. While the modelling process for GermaNet adopts relevant recommendations with respect to the English Princeton WordNet, for TermNet an alternative modelling concept is developed that considers the special characteristics of domain-specific terminologies. We present a proposal for linking a general-language wordnet and a terminological wordnet within the framework of OWL and on this basis discuss problems and alternative modelling approaches.

Keywords: wordnets, ontologies, terminologies, text technology

1. Einführung

Wortnetze haben vielfältige Anwendungsfelder in der Sprach- und Texttechnologie sowie im Information Retrieval, vgl. z. B. Fellbaum (1998), Kunze et al. (2003). In aktuellen Forschungen werden zunehmend Vorschläge entwickelt, das englische Princeton WordNet in RDF oder OWL zu konvertieren und damit für entsprechende Verfahren und Technologien des „Semantic Web“ nutzbar zu machen (Ciorăscu et al. 2003, van Assem et al. 2004, Graves & Gutierrez 2005).¹ Die *Web Ontology Language* OWL ist die W3C *recommendation* eines XML-basierten Standards zur Repräsentation von Ontologien im Semantic Web. Inzwischen liegt ein *Public Working Draft* einer W3C Arbeitsgruppe über ein standardisiertes Modell für die Repräsentation von Wortnetzen auf der Grundlage der genannten Arbeiten vor (van Assem et al. 2006b).

1. Ein Vorschlag zur Standardisierung des Formats lexikalischer Ressourcen, der auf eine Vereinheitlichung der in Wortnetzen verwendeten Kategorien und Strukturen abzielt, ist in ISO/TC-37/SC4 (2006) vorgelegt worden.

In unserem Beitrag beschreiben wir, wie sich allgemeinsprachliche und domänenspezifische Wortnetze in OWL modellieren und miteinander verknüpfen lassen. Dementsprechend wurden repräsentative Teilausschnitte des an der Universität Tübingen entwickelten deutschen allgemeinsprachlichen Wortnetzes GermaNet und des an der Universität Dortmund entwickelten deutschen domänenspezifischen Wortnetzes TermNet modelliert und miteinander verlinkt. Langfristig sollen beide Netze vollständig in OWL konvertiert werden.

Unser Modellierungsvorschlag orientiert sich vor allem am Ziel der gemeinsamen Nutzung von Wortnetzen und Domänenontologien in text-technologischen Anwendungen im Rahmen der DFG-Forschergruppe *Texttechnologische Informationsmodellierung*.

Zunächst beschreiben wir in Abschnitt 2. wesentliche Charakteristika von Ontologien und der Web-Ontologiesprache OWL. In Abschnitt 3. erörtern wir unseren Modellierungsansatz für das allgemeinsprachliche Wortnetz GermaNet, welcher weitgehend die W3C-Standardisierungsvorschläge für das Princeton WordNet aufgreift und diese an das Datenmodell von GermaNet adaptiert. Abschnitt 4. beschreibt die Modellierung für das terminologische Wortnetz TermNet, das semantische Bezüge zwischen Fachtermini aus den Domänen „Hypertextforschung“ und „Texttechnologie“ repräsentiert, die im Anwendungskontext der automatischen Hypertextualisierung genutzt werden. In Abschnitt 5. wird gezeigt, wie sich die beiden Ressourcen im Rahmen von OWL verknüpfen lassen. Wir orientieren uns dabei an dem sogenannten „Plug-in-Ansatz“ (Magnini & Speranza 2002).

In Abschnitt 6. diskutieren wir Modellierungsalternativen für GermaNet, in denen Synsets und lexikalische Einheiten nicht als Instanzen (wie im W3C-Vorschlag), sondern als Klassen bzw. als beides modelliert werden. Wir argumentieren, dass diese nicht nur die Nachbildung des Plug-in-Ansatzes im gegebenen Beschreibungsrahmen erleichtern, sondern auch semantische Intuitionen zum Verhältnis von Konzepten zu Instanzen adäquater widerspiegeln würden.

2. Ontologien und Wortnetze in der Web-Ontologiesprache OWL

In der Philosophie gilt die Ontologie als eine Disziplin, die sich mit dem Sein und den fundamentalen Typen von Entitäten beschäftigt. Der Begriff wurde in die Informatik übertragen und bezeichnet dort die Repräsentation eines Wissensausschnitts in einer computerverarbeitbaren Form; Staab & Studer (2004: VII) charakterisieren eine Ontologie nach Gruber (1993) als „a formal explicit specification of a shared conceptualisation for a domain of interest“. Ob es sich bei Wortnetzen um Ontologien handelt, wird kontrovers diskutiert; z. B. anhand der von Erd-

mann (2001: 73) eingeführten Differenzierung zwischen *light-weight ontologies*, das sind „Ontologien, die primär aus einem Schema (Begriffstaxonomie, Attribut- und Relationsdefinitionen) bestehen“ und *heavy-weight ontologies*, verstanden als „Ontologien mit (beliebigen oder eingeschränkten) prädikatenlogischen Formeln“ (Erdmann 2001: 72). In diesem Sinne lassen sich Wortnetze als *light weight ontologies* bestimmen. In ähnlicher Weise differenziert John F. Sowa zwischen den in logikbasierter Sprache repräsentierten *axiomatized ontologies* (Sowa 2000: 493), und den nicht vollständig formalisierten *terminological ontologies* (Sowa 2000: 497); als Beispiel für eine solche terminologische Ontologie nennt Sowa das englische WordNet.

Mit der Web Ontology Language OWL hat die W3C Web Ontology Working Group einen Standard zur Spezifikation von Ontologien geschaffen, der aus drei unterschiedlich expressiven Subsprachen besteht: OWL Full, OWL DL und OWL Lite. Wenn die in der Ontologie verwendeten Konstrukte aus dem als OWL DL bezeichneten Subset stammen, kann deren Semantik deskriptionslogisch interpretiert werden – dies war ein explizites Ziel beim Entwurf des OWL Standards (vgl. Baader et al. 2004). In OWL DL erstellte Ontologien können mit deskriptionslogischer Reasoning-Software (z. B. Racer oder FACT) verarbeitet werden, um die Konsistenz der Beschreibungen zu prüfen, neue Fakten und Relationen zwischen Klassen zu erschließen oder um bestehende Ontologien automatisch zu erweitern (vgl. Antoniou & van Harmelen 2004).

Um diese Möglichkeiten zu nutzen und das englische Wortnetz für Anwendungen im Kontext des Semantic Web zu erschließen, gibt es bereits mehrere Ansätze, diese Ressource in OWL zu überführen (vgl. Ciorăscu et al. 2003, van Assem et al. 2004, 2006a). Wird die deskriptionslogische Fundierung von OWL DL auch wirklich genutzt, so wird dabei eine Light-weight- in eine Heavy-weight-Ontologie transformiert, bei der sich verschiedene Modellierungsalternativen eröffnen. Zur Diskussion dieser Modellierungsalternativen wollen wir kurz die wichtigsten Konstrukte erläutern, und dabei insbesondere auf die Unterscheidung von Klassen und Individuen eingehen, und auf die zentrale Frage, ob Synsets und lexikalische Einheiten, die Repräsentationseinheiten von Wortnetzen, als Klassen (*classes*) oder als Individuen (*individuals*) modelliert werden.

Nach der OWL-Terminologie sind die drei wichtigsten Typen von Komponenten in OWL-Ontologien Klassen, Properties und Individuen. (In vielen Ontologiebeschreibungen wie z. B. Erdmann 2001 ist alternativ von Konzepten, Relationen und Instanzen die Rede, und aus unserer Sicht sind die Paare Instanz:Individuum, Klasse:Konzept und Property:Relation jeweils Quasi-Synonyme). Individuen repräsentieren die Objekte einer Domäne, und Klassen sind Mengen von Individuen, die be-

stimmte Merkmale gemeinsam haben. Klassen werden in OWL durch das `<owl:class>`-Konstrukt definiert und sind mit einer formalen Beschreibung der Merkmale, die für die in ihnen enthaltenen Individuen gelten, assoziiert (durch `<owl:Restriction>`-Ausdrücke). In einer Ontologie könnte z. B. eine Klasse *Hund* existieren, deren Mitglieder individuelle Hunde wie ‚Fido‘ oder ‚Enno‘ sind. Individuen werden in OWL mit der Angabe der Klasse, zu der sie gehören, eingeführt, dies kann durch die Verwendung des Elements `<rdf:type>` erfolgen. Eine Besonderheit von OWL ist, dass die Vergabe unterschiedlicher Namen für Individuen nicht bedeutet, dass es sich um unterschiedliche (nicht-identische) Individuen handelt, dies muss, falls erwünscht, explizit spezifiziert werden.

Erdmann (2001) zählt die Menge der Individuen einer Domäne nicht zur eigentlichen Ontologie, vielmehr bilden sie seiner Auffassung nach eine *Wissensbasis* zur Ontologie. Einer der Vorteile von OWL als Ontologiebeschreibungssprache des Semantic Web ist, dass die Teile einer Ontologie sowie auch die Wissensbasis auf unterschiedlichen Servern gespeichert sein und durch die Verwendung entsprechender URIs dennoch aufeinander bezogen werden können.

Klassen können in einer Subklassen-Hierarchie (Taxonomie) angeordnet sein, was in OWL durch Konstrukte wie `<rdfs:subClassOf>` ausgedrückt wird. In einer Ontologie könnte z. B. die Klasse *Hund* eine Subklasse der Klasse *Tier* sein, was gleichbedeutend damit ist, dass *Tier* eine Superklasse von *Hund* ist und ausdrückt, dass alle Individuen der Klasse *Hund* auch Individuen der Klasse *Tier* sind (vgl. Abb. 1). Es ist

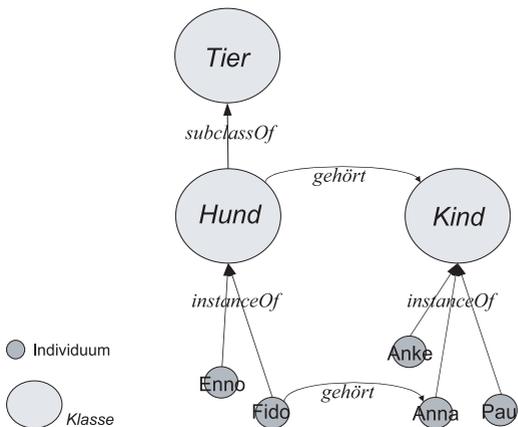


Abb. 1: Klassen, Individuen, Properties

wichtig, die Subklassen-Relation von der Instanz-Relation zu unterscheiden, was sich manchmal als schwierig erweist, weil etwa Eigenschaften wie *isA* oder *hyponym* an anderer Stelle für beides verwendet werden. Es gilt, dass *subclassOf* der mathematischen Relation ‚ist-Teilmenge-von‘, *instanceOf* hingegen ‚ist-Element-von‘ entspricht. Die Entscheidung über die Instanzebene einer Ontologie, z. B. die Frage, ob vielleicht *Fido* nicht doch eine Klasse sei und alle Vorkommen der Wortform **Fido** in Texten ihre Instanzen sein sollten, obliegt letztlich dem Ontologie-Designer und ist abhängig vom Anwendungszweck der Ontologie.

Eine Property entspricht in OWL einer binären Relation zwischen genau zwei Klassen. Beispielsweise könnte die Property *gehört* zwischen den Klassen *Hund* und *Kind* gelten. Die entsprechende Verbindung zwischen zwei Individuen wird hingegen als Property-Instanz bezeichnet, z. B. wäre *gehört* (*Fido*, *Anna*) eine Instanz der Property *gehört*, vgl. Horridge et al. (2004).

Sollten in einer OWL-Ontologie die Mengen der Klassen, Individuen und Properties nicht disjunkt sein, z. B. wenn bestimmte Objekte gleichzeitig Klassen und Individuen sind, so befindet man sich im Bereich der Subsprache OWL Full und außerhalb von OWL DL, (vgl. Smith et al. 2004).

3. OWL-Modell für GermaNet

3.1. GermaNet – Überblick und Datenstruktur

Mit GermaNet (Kunze 2001) ist ein lexikalisch-semantisches Wortnetz im Stil des Princeton WordNet entwickelt worden, das die wichtigsten und häufigsten Konzepte des deutschen Grundwortschatzes abbildet und die grundlegenden semantischen Relationen zwischen Konzepten und lexikalischen Einheiten wie Hyperonymie, Meronymie und Antonymie modelliert. Die zentrale Organisationseinheit in Wortnetzen, das sogenannte *Synset*, fasst die Synonymenmenge zu einem gegebenen Konzept zusammen, z. B. {*Streichholz*, *Zündholz*}, {*fleißig*, *eifrig*, *emsig*, *tüchtig*} und {*vergeben*, *verzeihen*}. Ein Synset wie {*öffnen*, *aufmachen*} wird im Wortnetz als Konzeptknoten mit seinen semantischen Verknüpfungen repräsentiert: mit seinem Hyperonym-Synset {*wandeln*, *verändern*}, und mit einigen Hyponym-Synsets, z. B. {*aufstoßen*}. Außerdem besteht eine kausative Beziehung mit dem Synset {*öffnen*, *aufgehen*}. Die lexikalische Einheit *öffnen* ist mit seinem Antonym *schließen* relationiert, vgl. Abb. 2.

Durch den Synset-Ansatz eignen sich Wortnetze für zahlreiche computerlinguistische Anwendungen, die eine Lesartendisambiguierung voraussetzen. Im Wortnetz sind semantische Relationen zwischen den Konzepten (Synsets) oder einzelnen Varianten (Synonymen aus den Synsets,

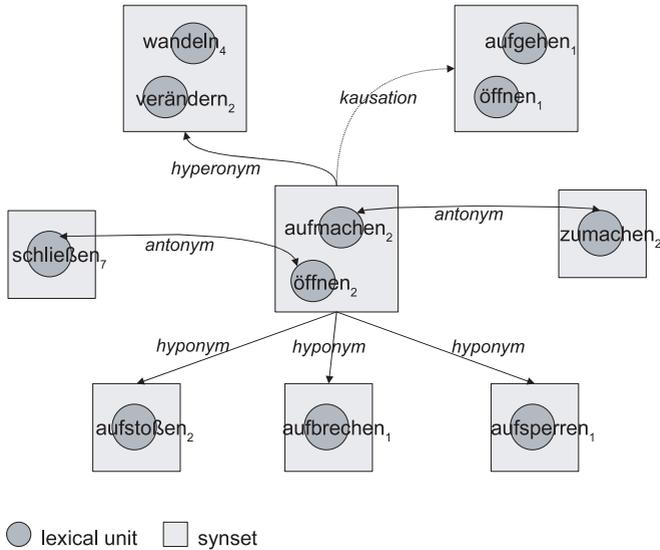


Abb. 2: Semantisches Netz für öffnen

die als *Lexical Units* bezeichnet werden) kodiert. GermaNet enthält ca. 53 000 Synsets mit mehr als 76 000 Lexical Units, darunter Nomina, Verben, Adjektive und Adverbien, überwiegend aus dem allgemeinsprachlichen Bereich. Es sind nur wenige Mehrwortlexeme wie *gesprochene Sprache* berücksichtigt. Ca. ein Drittel des Datenbestandes aus GermaNet ist im Rahmen eines europäischen Projektes in die polylinguale Ressource EuroWordNet (Vossen 1999) eingebunden worden. Einige Besonderheiten der GermaNet-Modellierung gegenüber WordNet betreffen:

- den expliziten Gebrauch nicht-lexikalisierten sogenannter künstlicher Konzepte zur Bildung wohlgeformter Hierarchien wie z. B. *Schultyplehrer* als artifizieller Oberbegriff zu (den lexikalisierten) Konzepten *Grundschullehrer*, *Realschullehrer*, *Berufsschullehrer* etc.
- die Auszeichnung von Eigennamen. In GermaNet wurden insbesondere geografische Namen aufgenommen
- den taxonomischen Gliederungsansatz zur Modellierung der Adjektivhierarchien gegenüber der Gruppierung in Ähnlichkeitsverbänden wie in WordNet
- ein umfangreiches Inventar an Subkategorisierungsrahmen zur Beschreibung des syntaktischen Komplementierungsverhaltens der GermaNet-Verben
- die Modellierung der Meronymiebeziehung als generische Relation

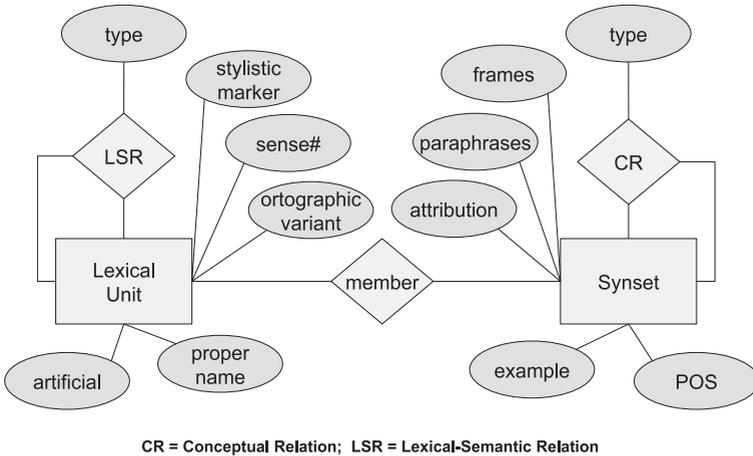


Abb. 3: Entity-Relationship-Graf GermaNet

Das Datenmodell des GermaNet ist in Abb. 3 als ein Entity-Relationship-Graf wiedergegeben, welcher als Vorlage für eine XML-Modellierung des GermaNet (Lemnitzer & Kunze 2002) diente. Im Diagramm sind die Objekte, Synsets und Lexical Units,² als Rechtecke repräsentiert. Attribute, welche die lexikalischen Objekte charakterisieren, sind als Ellipsen gekennzeichnet. Die Relationen zwischen den Objekten werden als Rauten repräsentiert: konzeptuelle Relationen (CR) zwischen Synset Objekten wie z. B. Hyperonymie und lexikalisch-semantische Relationen (LSR), die zwischen Lexical Units gelten, wie z. B. die Antonymie.

3.2. Spezifikation der OWL-Modellierung des GermaNet

Wie in Kapitel 2. erläutert, entspricht die OWL-Subsprache OWL-DL in ihrer Expressivität und Entscheidbarkeit der Deskriptionslogik, und zahlreiche *Reasoning-* und *Classifier-Systeme* für OWL sind explizit auf die Subsprache OWL DL zugeschnitten. Es ist fragwürdig, ob jemals Systeme existieren werden, die die gesamte Ausdrucksmächtigkeit von OWL Full verarbeiten können. Aus diesem Grund entwerfen wir das OWL-Modell für GermaNet ebenfalls im Rahmen der Subsprache OWL DL.

In einer Vorstudie untersuchten wir die Merkmale dreier existierender Ansätze zu Repräsentation des Princeton WordNet in OWL: des *W3C-*

2. Eine *Lexical Unit* ist nach Cruse (1986) als ein Paar mit einer lexikalischen Form und einer einzigen Bedeutung charakterisiert. Ein Synset ist eine Menge mit mindestens einer *Lexical Unit*. Im selben Synset enthaltene *Lexical Units* sind synonym.

Ansatzes (van Assem et al. 2006a, b), des *Neuchâtel-Ansatzes* (Ciorăscu et al. 2003) und des *Amsterdam-Ansatzes* (van Assem et al. 2004). Der hier vorgestellte Ansatz orientiert sich hauptsächlich am W3C-Ansatz, berücksichtigt aber die Besonderheiten von GermaNet gegenüber dem Princeton WordNet.

Für die OWL-Modellierung gilt es, die Konzepte von GermaNet als Klassen (OWL-Konstrukt `<owl:class>`) sowie die Eigenschaften der Konzepte und die Relationen zwischen den Konzepten als Properties (`<owl:DatatypeProperty>` oder `<owl:ObjectProperty>`) darzustellen. Die o. g. grundlegenden Objekte des Entity-Relationship-Modells (im Folgenden: E-R-Modell) führen unmittelbar zu den beiden Klassen *Synset* und *LexicalUnit* im OWL-Modell. Eine solche Modellierung entspricht auch den drei oben genannten Ansätzen zu OWL-Konvertierungen des Princeton WordNet. Wir behalten die bisher in GermaNet üblichen Bezeichnungen *Synset* und *LexicalUnit* bei, da sie mittlerweile in der Wortnetz-Forschung etabliert sind. Wir verzichten auf eine Entsprechung der Klasse *Word* aus dem W3C-Ansatz (van Assem et al. 2006a) (etwa: ein rein formal bestimmtes Lemma). Sie spielt in GermaNet nur eine indirekte Rolle und ist z. B. auch im Ansatz von Ciorăscu et al. (2003) (dort als *StemObject*) nicht Teil der WordNet-Ontologie.

3.2.1. Die Klassenhierarchie

Analog zum W3C-Vorschlag legen wir *NounSynset*, *VerbSynset*, *AdjectiveSynset* und *AdverbSynset* als unmittelbare Subklassen von *Synset* fest, sowie *NounUnit*, *VerbUnit*, *AdjectiveUnit* und *AdverbUnit* als unmittelbare, wortartbedingte Subklassen von *LexicalUnit*. In Ciorăscu et al. (2003) hat die der *LexicalUnit* entsprechende Klasse keine Subklassen, sie ergeben sich hier aber dadurch, dass es verschiedene konzeptuelle sowie lexikalisch-semantische Relationen gibt, die sich mit ihrem Definitionsbereich (*Domain*) und Wertebereich (*Range*) auf sie beziehen. So hat etwa die Property *isParticipleOf* als Domain die Klasse *AdjectiveUnit* und als Range die Klasse *VerbUnit*. *NounSynset* hat eine weitere Subklasse *ProperName*, deren Instanzen Eigennamen sind. Eine weitere ausgezeichnete Untermenge der Synsets sind die künstlichen Konzepte (vgl. Abschnitt 3.1.). Eine eigene Subklasse für diese einzuführen hätte, um die Property-Vererbung in OWL nutzen zu können, bedeutet auch deren Komplement, die Subklasse der nicht-künstlichen Synsets, zu kreieren und jeder Synset-Instanz zwei Klassen zuzuordnen (beispielsweise *NounSynset* und *NaturalSynset*). Da eine Einführung der zwei komplementären Klassen den Ausnahmecharakter der künstlichen Konzepte verschleiert und da die Nutzung von multipler Vererbung in diesem Fall

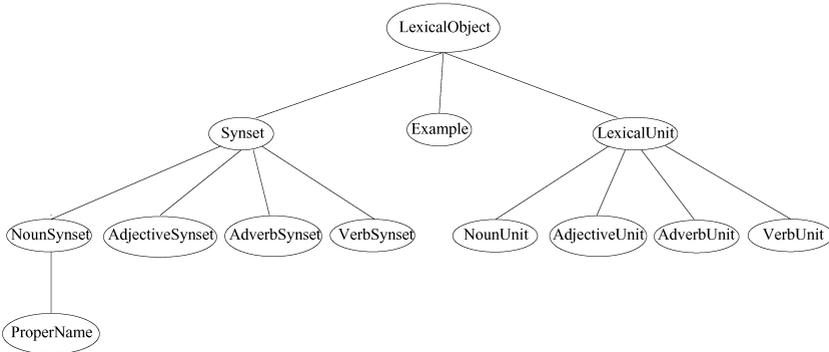


Abb. 4: Klassenhierarchie für die OWL-Modellierung von GermaNet

kein Gewinn gegenüber der alleinigen Verwendung der Property-Spezifikation `<isArtificial>true</isArtificial>` ist, haben wir darauf verzichtet, *ArtificialSynset* als weitere Subklasse von *Synset* anzusetzen. Eine grafische Darstellung der Klassenhierarchie der OWL-Modellierung von GermaNet findet sich in Abb. 4.

Wie in den von uns untersuchten Ansätzen zur OWL-Modellierung des Princeton WordNet sind in der vorliegenden GermaNet-Modellierung die einzelnen Synsets und Lexical Units Instanzen der Klassen dieser Hierarchie, sie stellen also die Population der Ontologie dar. Außer Synsets und LUs gibt es eine weitere Klasse *Example*, deren Instanzen Textbeispiele für Synset-Instanzen sind, vgl. Abschnitt 3.2.4.

3.2.2. Object Properties

Die Lexikalisierungsrelation

Die Verbindung zwischen Synsets und Lexical Units wird durch die Lexikalisierungsrelation hergestellt, wobei ein Synset durch mehrere Lexical Units lexikalisiert werden kann (die Synonyme). In OWL wird dies durch eine *Object Property* namens *hasMember* mit der Domain *Synset* und dem Range *LexicalUnit* modelliert. Die potenzielle Multiwertigkeit dieser Property ist durch das Fehlen einer `<owl:maxCardinality>`-Beschränkung realisiert. Für jede wortartbedingte Subklasse von *Synset* existiert eine Einschränkung des Wertebereichs von *hasMember* auf diejenige Subklasse von *LexicalUnit*, die der gleichen Wortart entspricht. Die Einschränkung wird, wie in Listing 1 zu sehen, durch das `<owl:all-ValuesFrom>`-Konstrukt ausgedrückt.

Listing 1: OWL-Kode für eine Einschränkung der Lexikalisierungsrelation

```

    gn:hasMember
<owl:Class rdf:ID="NounSynset">
  <rdfs:subClassOf >
    <owl:Restriction>
      <owl:allValuesFrom>
        <owl:Class rdf:about="#NounUnit"/>
      </owl:allValuesFrom>
      <owl:onProperty>
        <owl:InverseFunctionalProperty rdf:ID="hasMember"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf >
  ...
</owl:Class >

```

Die Hyperonymie-Relation

OWL eignet sich vorzüglich für eine Modellierung der beiden Relationstypen (der konzeptuellen und lexikalisch-semantischen Relationen), bei denen die Relationen jeweils zwischen intern definierten Klassen bestehen und somit Object Properties in OWL entsprechen. Sowohl der Typ der konzeptuellen Relation als auch der lexikalisch-semantischen Relation bestehen zwischen intern definierten Klassen und entsprechen somit Object Properties in OWL. Object Properties können wie Klassen hierarchisch angeordnet sein, dies wird durch das Konstrukt `<rdfs:subPropertyOf>` erreicht.

Unsere Modellierung beinhaltet demnach zwei top-level Object Properties, *conceptualRelation* (Domain *Synset* und Range *Synset*) und *lexicalSemanticRelation* (Domain *LexicalUnit* und Range *LexicalUnit*). Die jeweiligen Subproperties sind der Hierarchie der Object Properties in Abb. 5 zu entnehmen.

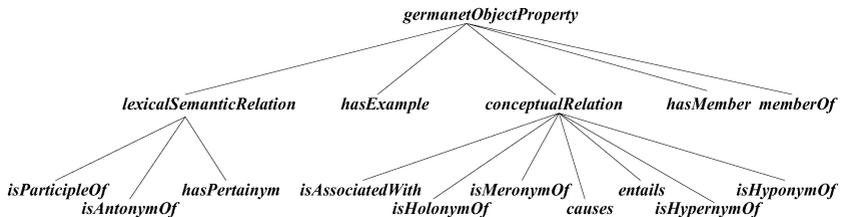


Abb. 5: Hierarchie der Object Properties für GermaNet

Gleichgeordnet mit *conceptualRelation* und *lexicalSemanticRelation* befinden sich auf der oberen Ebene der Property-Hierarchie auch die bereits vorgestellten Relationen *hasMember* und *memberOf* sowie die Property *hasExample* zur Einführung von Beispielen für Synsets (vgl. Abschnitt 3.2.4.).

Die zentrale, eine Hierarchie aufspannende Relation in lexikalisch-semantischen Wortnetzen ist die Hyper(o)nymie-Relation zwischen Synsets. Ihre inverse Relation ist die Hyponymie. In einer klassenbasierten OWL-Modellierung stellte sich die Frage, ob Hyponymie nicht mit der Subclass-Relation gleichzusetzen sei. In der Vergangenheit vorgebrachte Gegenargumente sind, dass die Hyperonymie-Relation in Wortnetzen nicht nur Paarungen der Art Klasse-Subklasse beschreibt, sondern ebenso die ontologisch bedeutsamen Unterscheidungen Klasse-Rolle (Beispiel aus GermaNet: *See-Badesee*) und Klasse-Instanz (Beispiel aus GermaNet: *Ozean-Pazifik*) bezeichnen kann, vgl. hierzu Gangemi et al. (2002). Da wir in unserer Modellierung die einzelnen Synsets als Instanzen repräsentieren, besteht die Möglichkeit, Hyponymie als `<rdfs:subClassOf>` auszudrücken ohne den Dialekt OWL DL zu verlassen, ohnehin nicht. Hyperonymie wird daher als eine `<owl:TransitiveProperty>` mit der Domain und dem Range *Synset* eingeführt, vgl. Listing 2. Gleichzeitig wird die Hyponymie-Relation als zugehörige inverse Property deklariert.

Ähnlich wie im Fall der Property *hasMember* gibt es für jede wortartbedingte Subklasse von *Synset* eine Einschränkung des Wertebereichs von *isHyponymOf* auf Synsets der gleichen Subklasse mittels des `<owl:allValuesFrom>`-Konstrukts.

Weitere Object Properties Tabelle 1 bietet einen Überblick über alle für GermaNet relevanten Object Properties und ihrer in OWL modellier-

Listing 2: OWL-Kode für die Einführung der Hyperonymie-Relation
gn:isHypernymOf

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

Table 1: *Charakteristika der Object Properties für GermaNet*

Property	Domain	Range	Charakteristika	Inverse Property	Lokale Restriktionen
hasMember	Synset	LexicalUnit	invers-funktional	memberOf	<i>wortartbezogen</i>
memberOf	LexicalUnit	Synset	funktional	hasMember	<i>wortartbezogen</i>
<i>Konzeptuelle Relationen (CR)</i>					
isHypernymOf	Synset	Synset	transitiv	isHyponymOf	<i>wortartbezogen</i>
isHyponymOf	Synset	Synset	transitiv	isHypernymOf	<i>wortartbezogen</i>
isHolonymOf	NounSynset	NounSynset	–	–	–
isMeronymOf	NounSynset	NounSynset	–	–	–
isAssociatedWith	Synset	Synset	–	–	–
entails	VerbSynset	VerbSynset	–	–	–
causes	VerbSynset	VerbSynset ⊔ AdjectiveSynset	–	–	–
<i>Lexikalisch-semantische Relationen (LSR)</i>					
hasAntonym	LexicalUnit	LexicalUnit	symmetrisch	hasAntonym	<i>wortartbezogen</i>
hasPertainym	LexicalUnit	LexicalUnit	–	–	–
isParticipleOf	AdjectiveUnit	VerbUnit	–	–	–
hasExample	Synset	Example	–	–	–

baren Charakteristika und Einschränkungen. Gegenüber dem E-R-Modell in Abb. 3 können in OWL einige präzisere Spezifikationen vorgenommen werden, so sind zum Beispiel Domain und Range von *isHolonymOf* und *isMeronymOf* auf die Klasse *NounSynset* eingeschränkt. Wie aus der Tabelle hervorgeht, sind Holonymie und Meronymie in GermaNet keine transitiven Relationen und auch nicht invers zueinander.³ Die Bezeichner für Klassen und Relationen wurden im Wesentlichen beibehalten, aber mit Affixen wie *has*, *is* oder *of* versehen, um z. B. die Richtung einer Relation zu verdeutlichen. Für die Properties *hasMember*, *isHypernymOf* und *hasAntonym* sind inverse Properties definiert. Diese bieten keine zusätzliche Information, können aber die Verarbeitung beschleunigen. Außerdem können wir in OWL die wortartbezogenen Einschränkungen von Relationen ausdrücken, z. B. dass ein Antonym einer LU der Wortart *x* wieder auch wieder eine LU der Wortart *x* sein muss, vgl. auch die Einschränkung für *hasMember* in Listing 1.

3.2.3. Datatype Properties

Relationen, die nicht zwischen intern definierten Klassen bestehen, sondern vielmehr einer intern definierten Klasse einen Wertebereich in Form eines XML Schema-Datentyps wie `xs:string` oder `xs:boolean` zuzuordnen, werden als OWL Datatype Properties modelliert. Für GermaNet sind dies offensichtlich diejenigen Charakteristika von lexikalischen Einheiten und Synsets, die im E-R-Modell in Abb. 3 als Ellipsen dargestellt sind. (Eine Ausnahme bildet *example*, da diese Eigenschaft aus zwei Komponenten besteht, vgl. Abschnitt 3.2.4.)

Tabelle 2 bietet einen Überblick über die Datatype Properties in der OWL-Modellierung von GermaNet mit Angabe von Domain, Range und Funktionseigenschaft. Wiederum konnten im Rahmen von OWL gegenüber dem Modell in Abb. 3 genauere Spezifikationen vorgenommen werden, so z. B. über Einschränkungen der Domains sowie überhaupt die Angabe von Wertebereichen.

3. Hier besteht ein Unterschied zur Konzeption der Holonymie-Relationen im Princeton WordNet. Diese ist in drei spezifischere Relationen untergliedert und jede dieser spezifischeren Relationen wird als transitiv modelliert. Bei einer generischen Meronymie/Holonymie-Relation ist dies nicht möglich. So ist zum einen *Finger* als Teil von *Hand*, und *Hand* als Teil von *Arm* charakterisierbar, aber *Finger* nicht als Teil von *Arm*. Zum anderen enthält ein *Tomatensalat* als Ingredienz *Tomaten*, aber eine *Tomate* fungiert nicht notwendigerweise als Bestandteil eines *Tomatensalats*.

Table 2: Charakteristika der Datatype Properties für GermaNet

Property	Domain	Range	funktional
POS	Synset	„N“ „V“ „A“ „ADV“	ja
hasParaphrase	Synset	xs:string	nein
isArtificial	Synset \sqcup LexicalUnit	xs:boolean	(ja)
isProperName	NounSynset \sqcup NounUnit	xs:boolean	(ja)
hasOrthographicForm	LexicalUnit	xs:string	ja
hasSense	LexicalUnit	xs:positiveInteger	ja
isStylisticallyMarked	LexicalUnit	xs:boolean	(ja)
hasFrame	VerbUnit \sqcup Example	xs:string	nein
hasText	Example	xs:string	ja

3.2.4. Beispiele

Einem Synset können in GermaNet mehrere Beispiele zugeordnet sein; ein Beispiel hat ein oder zwei Komponenten: einen Beispieltext oder, im Falle von Verben, ein Paar aus einem Verbrahmen und einem Beispieltext, in dem der Verbrahmen realisiert ist. Im OWL-Modell wird daher eine weitere Klasse *Example* angenommen, der die beiden Komponenten als die Datatype Properties *hasText* und *hasFrame* zugeordnet sind. Die Object Property *hasExample* ordnet einer Instanz von *Synset* eine oder mehrere Instanzen von *Example* zu.

3.2.5. OWL-Kodierung eines repräsentativen Ausschnitts

Um OWL-Kode für das vorgestellte Modell zu generieren und es auf Konsistenz zu testen, wurde ein repräsentativer Ausschnitt von 37 Synsets und 83 LexicalUnits aus GermaNet sowie Relationen zwischen ihnen als Ontologie in OWL kodiert.⁴ Bei der Auswahl wurde darauf geachtet, dass jede (Sub-)Klasse und jede in den Tabellen 1 und 2 aufgeführte Property mindestens einmal instanziiert ist. Somit kann nun der entstandene OWL-Kode als Spezifikation für die Konvertierungsroutinen verwendet werden, die erforderlich sind, um schließlich das gesamte GermaNet nach dem oben vorgestellten Modell in OWL zu konvertieren. Als Ausgangsformat für die Konvertierung bieten sich entweder die Prolog-Version oder die XML-Version von GermaNet (Lemnitzer & Kunze 2002) an.

4. Als Ontologie-Editor wurde Protégé 3.1.1 (<http://protege.stanford.edu>) in Verbindung mit der Reasoner/Classifier-Software RacerPro 1.9.1 (<http://www.racer-systems.com>) verwendet.

4. OWL-Modell für das terminologische Wortnetz TermNet

TermNet ist ein terminologisches Wortnetz, das mit dem Ziel entwickelt wurde, Ähnlichkeiten und Unterschiede zwischen Termini aus unterschiedlichen Terminologien zur selben Fachdomäne zu repräsentieren. Es wurde im Rahmen eines Projekts zur automatischen Hypertextualisierung aufgebaut und diente in der ersten Projektphase zur Generierung von Glossareinträgen in einem Fachtextkorpus (Beisswenger et al. 2003, Lenz et al. 2003). Das in der ersten Projektphase aufgebaute terminologische Netz enthält ca. 200 Termini aus den Fachdomänen „Texttechnologie“ und „Hypertextforschung“. Dabei wurden alle Termini aufgenommen, die im Korpus explizit definiert sind.

In der zweiten Phase des Projekts soll TermNet darüber hinaus für Strategien der automatischen Verlinkung nach thematischen Kriterien genutzt werden. Green (1999) hat für derartige Verlinkungsstrategien für das Englische lexikalische Ketten (*lexical chains*) vorgeschlagen. Für die Übertragung dieses Ansatzes auf das Deutsche ist es notwendig, das terminologische Wortnetz mit einem großen allgemeinsprachlichen Wortnetz zu verbinden. Zu diesem Zweck wäre es wünschenswert, das terminologische Netz in OWL zu remodellieren und mit einer OWL-Repräsentation des GermaNet zu verknüpfen. Der folgende Modellierungsansatz wurde vor diesem Hintergrund entwickelt.

Bei der Übertragung des Wortnetzansatzes auf Fachwortschatz möchten wir folgenden Unterschieden zwischen Fach- und Allgemeinwortschatz Rechnung tragen:

1. In terminologischen Taxonomien können vertikale Relationen der Über- und Unterordnung meist als Klasseninklusion dargestellt werden. Wir nutzen deshalb die in OWL vordefinierte Subclass-Property, um die Unterordnung eines spezifischen Terminus, z. B. *Monodirektionale Verknüpfung* unter einen generelleren Terminus, in diesem Fall *Verknüpfung*, zu repräsentieren.
2. Typisch für solche Taxonomien ist es, dass sich die einem generellen Terminus untergeordneten Termini nach dem relevanten klassifikatorischen Merkmal in wechselseitig disjunkte Subklassen gruppieren lassen. Zum Beispiel gibt es zum Terminus *Verknüpfung* die spezifischeren Termini *Monodirektionale Verknüpfung*, *Bidirektionale Verknüpfung* (Merkmal: Direktionalität) und *Intratextuelle Verknüpfung*, *Intertextuelle Verknüpfung*, *Extratextuelle Verknüpfung* (Merkmal: Position des Linkziels). Die Termini, die nach demselben Merkmal klassifiziert sind, denotieren jeweils disjunkte Teilklassen. Diese Disjunkтивität kann man, wenn man Termini als Klassen modelliert, durch entsprechende `<owl:disjointWith>`-Konstrukte festlegen.

3. Für viele Fachdomänen ist es charakteristisch, dass derselbe Untersuchungsbereich theorie- und schulenabhängig in unterschiedlicher Weise terminologisiert wird. Für den Terminus *Verknüpfung*, wie er im Standardwerk von Rainer Kuhlen (Kuhlen 1991) definiert ist, finden sich in Dokumenten, die auf anderen Ansätzen basieren, entsprechende Termini (z. B. *Link*, *Verweis*). Im Gegensatz zu lexikalischen Synonymen als Mitglieder desselben Synsets können formverschiedene Termini aus verschiedenen Taxonomien in den Fachtexten nicht wechselseitig ersetzt werden, auch wenn die denotierten Kategorien sehr ähnlich sind. Schließlich ist es ja gerade das Ziel von Terminologisierungsprozessen, die Formseite des Terminus eindeutig an eine fachsprachliche Kategorie zu binden. Andererseits bestehen zwischen den Kategorien konkurrierender Systeme durchaus Entsprechungsbeziehungen. Zum Beispiel haben die durch die Termini *Verknüpfung* (i. S. von Kuhlen 1991) und *Verweis* (i. S. von Tochtermann 1995) denotierten Instanzen einen großen Schnittbereich. Um derartige Entsprechungen von Kategorien zu repräsentieren, nutzt TermNet, als fachsprachliche Entsprechung zu den Synsets der allgemeinsprachlichen Wortnetze, das Modellierungskonstrukt *TermSet*. Ein *TermSet* enthält die Termini einer Fachdomäne, die aus verschiedenen terminologischen Systemen stammen, aber eine sehr hohe Schnittmenge in den denotierten Instanzen haben.

Im OWL-Modell von TermNet werden TermSets als Klassen modelliert. Für die Darstellung hierarchischer Bezüge zwischen TermSets, also zwischen einem TermSet, das *Verknüpfung*, *Verweis* enthält, und einem spezifischeren TermSet, das *MonodirektionaleVerknüpfung*, *MonodirektionalerVerweis* enthält, möchten wir aber bewusst nicht den vordefinierten `<rdfs:subClassOf>`-Konstruktor verwenden. Schließlich besteht die damit implizierte Klasseninklusion nur innerhalb desselben terminologischen Systems. Wir nutzen statt dessen hierfür die „weichere“ hierarchische Relation der Hyponymie und ihre Inverse, die Hyperonymie. Im Gegensatz zur Modellierung dieser Relation im OWL-Modell von GermaNet (s. Abschnitt 3.1.) besteht die Hyponymie-Beziehung aber nicht zwischen Instanzen, sondern zwischen Klassen (TermSets). Sie muss also als Property definiert werden, die paarweise einzelne TermSets relationiert. Hierzu beschränken wir den Wertebereich unserer *tn:isHyponymOf*-Property durch das `<owl:allValuesFrom>`-Konstrukt, durch das die Menge aller Instanzen der betreffenden Klasse denotiert wird. Das OWL-Statement in Listing 3 drückt entsprechend aus, dass das TermSet *MonodirektionalerLink* ein Hyponym des TermSets *Link* ist.

Listing 3: OWL-Kode zur Einführung der Klasse *TermSet_MonodirektionalerLink*

```

<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>

```

Für die Zuordnung von Termini zu TermSets kann man nicht auf die *gn:hasMember*-Property zurückgreifen, wie sie im oben beschriebenen Modellierungsvorschlag für GermaNet definiert wurde. Denn dort sind ja – entsprechend dem Standardisierungsvorschlag für das Princeton WordNet – sowohl Lexical Units als auch Synsets als Instanzen modelliert. In TermNet hingegen ist jeder Terminus und jedes TermSet als Klasse definiert. Die TermNet-Entsprechung zur *gn:hasMember-Property* wird deshalb – ähnlich wie bei der Hyperonymie – als eine paarweise Zuordnung von Klassen der Superklasse *NounTermSet* (als Domain) und Klassen der Superklasse *NounTerm* (als Range) spezifiziert.

Ein Beispiel für eine Restriktion von *isMemberOf* ist das OWL-Statement in Listing 4, in dem das TermSet *TermSet_Link* durch die Property *tn:isMemberOf* mit dem Terminus *Term_Verweis* verknüpft wird. Hiermit wird in der Modellierung festgehalten, dass der Terminus *Verweis* zu ei-

Listing 4: OWL-Kode für die *isMemberOf-Property* in TermNet

```

<owl:Class rdf:ID="Term_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>

```

Table 3: Charakteristika der Object Properties für TermNet

Property	Domain	Range	Charakteristika	Inverse Property
hasMember	NounTermSet	NounTerm	invers-funktional	isMemberOf
isMemberOf	NounTerm	NounTermSet	funktional	hasMember
<i>Konzeptuelle Relationen</i>				
isHypernymOf	NounTermSet	NounTermSet	transitiv	isHyponymOf
isHyponymOf	NounTermSet	NounTermSet	transitiv	isHypernymOf
isHolonymOf	NounTermSet	NounTermSet	–	–
isMeronymOf	NounTermSet	NounTermSet	–	–
<i>Lexikalisch-semantische Relationen</i>				
isAbbreviationOf	NounTerm	NounTerm	–	isExpansionOf
isExpansionOf	NounTerm	NounTerm	–	isAbbreviationOf

nem TermSet gehört, der alle Termini zur Bezeichnung des Link-Konzepts (*Link*, *Hyperlink*, *Kante*, *Verweis*, *Verknüpfung* etc.) enthält. Der Vorteil dieser Modellierung besteht darin, dass die Suche nach Termini in Fachtextkorpora auf verschiedenen Granularitätsstufen erfolgen kann: Je nach Nutzerinteresse und Anwendungskontext kann nach einem Terminus (z. B. *Verweis*) oder aber nach einem Konzept der Fachdomäne (also nach allen Termini des TermSets *Link*) gesucht werden.

Da für die geplanten Anwendungen vor allem nominale Termini relevant sind, wurden in der Modellierung bislang nur nominale TermSets und nominale Termini berücksichtigt. Zwischen diesen wurden, neben der Hyperonymie und Hyponymie, die Relationen Meronymie und Holonymie sowie die Relation zwischen der expandierten Form eines Terminus (z. B. *Hypertext Markup Language*) und seiner abgekürzten Variante (*HTML*) modelliert; hierzu nutzen wir die Property *isExpansionOf* und die inverse Property *isAbbreviationOf*.

Tabelle 3 gibt eine Übersicht über die im Modell definierten Object Properties. Die OWL-Statements für die Properties *isHolonymOf*, *isMeronymOf*, *isMeronymOf* und *isExpansionOf* lauten analog zu den oben gezeigten Kodebeispielen für *isHyponymOf* und *isMemberOf*. Die Konsistenz des Modells wurde an einem Ausschnitt von 51 TermSets und 141 Termini sowie den entsprechenden Relationen zwischen ihnen getestet.⁵

5. Auch hierfür wurde Protégé 3.1.1 (<http://protege.stanford.edu>) in Verbindung mit der Reasoner/Classifier-Software RacerPro 1.9.1 (<http://www.racer-systems.com>) verwendet.

5. Schnittstellen zwischen Wortnetzen und Terminologien

Die Verknüpfung allgemeinsprachlicher mit domänenspezifischen Wortnetzen zur Ausnutzung der Vorteile beider Ressourcentypen in Sprachtechnologie und Semantic Web wird u. a. in den Ansätzen von Vossen (2001) und Magnini & Speranza (2002) diskutiert. Während Vossen aus großen Dokumentensammlungen mit Verfahren der Informationerschließung und der Termextraktion ad-hoc Terminologien aufbaut, deren Termini anschließend an WordNet-Synsets gelinkt werden, entwickeln Magnini und Speranza einen Ansatz zur Verknüpfung zweier bestehender Ressourcen, des ItalWordNet mit einer spezialisierten Ontologie aus der Domäne der Wirtschaft (ECOWN), den so genannten *Plug-in-Ansatz*. Verschiedene Plug-in-Relationen (Synonymie, Quasi-Synonymie, Hyponymie) dienen dazu, die jeweilig relevanten Synsets aus beiden Ressourcen zu vereinigen. Mittels relativ weniger Plug-ins können große Wortschatzmengen aus beiden Ontologien korreliert werden (269 Plug-in-Relationen für die Anbindung von 4662 ECOWN Konzepten an ItalWordNet). Das vorgestellte Verfahren führt zu einer Gesamthierarchie, welche die Topkonzepte der spezialisierten Ontologie ausblendet und ihre Unterkonzepte, die Termini, in die allgemeinsprachliche Ontologie integriert.

Der Plug-in-Ansatz bietet eine attraktive Modellierungsvorlage für die Anbindung des TermNet an GermaNet, da beide Ressourcen ebenfalls wortnetzbasierend und von unterschiedlicher Größe bzw. Spezifität sind und sich in Bezug auf eine signifikante Anzahl von Konzepten überschneiden. Aufgrund der verschiedenartigen Modellierung von Termini als Klassen und von Synsets als Instanzen in unserem Ansatz kann die Verknüpfung nur über die Einschränkung einer Property über die Klasse der Termini *NounTerm* als Domain erfolgen, da in OWL DL nicht die Klasse der Termini-Klassen als Meta-Klasse gebildet werden kann, d. h. Klassen nicht als Individuen betrachtet werden können. Wir unterscheiden die folgenden drei Fälle der Verlinkung von TermNet-Termini und GermaNet-Synsets:

1. Entsprechung von TermNet-Termini und GermaNet-Synsets: z. B. für den Terminus *tn:Term_Link* und das NounSynset *gn:Link*. Die Object Property *plg:attachedToNearSynonym* hat die Domain *tn:NounTerm* und den Range *gn:Synset*. Mittels einer `<owl:Restriction>` über *plg:attachedToNearSynonym* wird jedem Individuum der Klasse *tn:Term_Link* das Individuum *gn:Link* zugeordnet, vgl. Listing 5.

Listing 5: OWL-Kode für eine Relationsinstanz von *attachedToNearSynonym*

```

<owl:Class rdf:ID = " tn:Term_Link">
  <rdfs:subClassOf rdf: resource = "#tn:NounTerm"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:has Value rdf:resource = "#gn:Link"/>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID = "plg:attachedToNearSyno-
          nym"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf >
</owl:Class>

```

2. Ein TermNet-Terminus hat kein semantisches Synset-Pendant in GermaNet, aber der Terminus steht in einer Subclass-Relation zu einem Terminus, der über *plg:attachedToNearSynonym* an ein GermaNet-Synset angebunden ist. Zum Beispiel ist der Terminus *tn:Term_MonodirektionalerLink* über die Subclass-Property mit dem Terminus *tn:Term_Link* verbunden, welcher wiederum über *plg:attachedToNearSynonym* an das NounSynset *gn:Link* in GermaNet geknüpft ist. Die Eigenschaft *plg:attachedToGeneralConcept* verbindet eine Terminusklasse wie *tn:Term_MonodirektionalerLink* mit einem GermaNet-Synset, das mit einem übergeordneten Terminus in einer *plg:attachedToNearSynonym-Relation* steht. Die Relation zwischen indirekt verknüpften Konzepten kann somit auch direkt spezifiziert werden und dient der Verkürzung von Pfadlängen zwischen semantisch ähnlichen Konzepten, z. B. bei der Berechnung semantischer Distanzmaße. Hier gehen wir über den Plug-in-Ansatz hinaus, der diese Relation nicht vorsieht.
3. Für einen Terminus aus TermNet gibt es kein semantisch entsprechendes Synset in GermaNet und auch kein passendes hyperonymisches GermaNet-Synset, an das der Terminus adäquat gelinkt werden könnte. Es existiert aber ein passendes holonymisches GermaNet-Synset, mit dem der TermNet-Terminus verknüpft werden kann, und zwar mittels der Plug-in-Relation *plg:attachedToHolonym*. Ein Beispiel ist der Terminus *tn:Term_Anker*, der in seiner terminologischen Bedeutung in der Hypertextforschung (Anker = Teil eines Links) keine semantische Entsprechung in GermaNet hat, statt dessen aber über *plg:attachedToHolonym* dem meronymisch übergeordneten GermaNet-Synset *gn:Link* zugeordnet werden kann.

In den von uns bearbeiteten Ausschnitten konnten alle TermNet-Termini über eine der drei Plug-in-Properties an GermaNet angeschlossen werden. Bei der Anknüpfung weiterer Terminologien ist aber natürlich nicht auszuschließen, dass Termini weder eine Entsprechung noch ein passendes hyperonymisches oder holonymisches übergeordnetes Synset in GermaNet haben. Nach einer möglichst vollständigen Verlinkung beider Ressourcen lässt sich ermitteln, wie häufig die unterschiedlichen Plug-in-Relationstypen jeweils instanziiert werden. Außerdem sollte sich abzeichnen, wie viele Konzepte mit den drei Plug-in-Relationen nicht anbindbar sind und ob sich die Einführung weiterer Relationstypen empfiehlt.

Durch die Modellierung der Termini als Klassen in TermNet und der Synsets als Instanzen in GermaNet sind die von uns vorgeschlagenen Schnittstellen-Relationen in OWL DL nicht aus der Sicht von GermaNet formulierbar, d. h. mit *Synset* als Domain. Dies ist für Inferenzen zwar nicht unbedingt erforderlich, aber dennoch wünschenswert.

6. Alternative Modellierungen und Ausblick

Die von uns untersuchten OWL-Konvertierungen des Princeton WordNet (Ciorăscu et al. 2003, van Assem et al. 2004, 2006a) modellieren die einzelnen Synsets sowie die einzelnen Lexical Units als Instanzen. Aus linguistisch-lexikologischer Sicht ist dies verblüffend, da Synsets gemeinhin als Konzepte aufgefasst werden, die durch im Synset enthaltene Lexical Units lexikalisiert werden. Man hätte also erwartet, dass ein Synset für das Konzept *Hund* als Klasse modelliert ist, der Hunde-Individuen wie „Fido“ und „Waldi“ als Instanzen zugeordnet werden können, und dass konzeptuelle Relationen zwischen solchen Klassen definiert werden. Die Entscheidung, Synsets und Lexical Units als Instanzen zu modellieren, scheint verarbeitungstechnisch und nicht linguistisch motiviert zu sein: So weisen z. B. Sanfilippo et al. (2005) daraufhin, dass die Modellierung einer großen Zahl von Lexical Units und Synsets als Klassen „impractical for a real-world application“ sei. Wir haben uns hier mit dem in Abschnitt 3. vorgestellten OWL-Modell für GermaNet dieser Perspektive angeschlossen.

In Lünen & Storrer (2007) wird – wiederum anhand eines repräsentativen Ausschnitts von GermaNet – mit der alternativen Möglichkeit experimentiert, alle Synsets und LUs als OWL-Klassen zu modellieren. Eine Relation wie der Hyponymie-Link zwischen den Synsets *öffnen*₂ und *aufbrechen*₁ wird dabei in OWL als Restriktion über der Domain und dem Range der allgemeinen Hyperonymie-Relation auf der Synset-Klasse ausgedrückt (analog zur hier vorgestellten Modellierung der Hyperonymie-Relation in TermNet mit `<owl:allValuesFrom>`). Eine

solche alternative Modellierung ist also im Rahmen von OWL DL auch möglich und bietet einige Vorteile: Synsets, LUs, Termini und TermSets haben den gleichen Status (nämlich Klassenstatus), und es lassen sich auch die in Abschnitt 5. angesprochenen inversen Plug-in-Relationen auf diese Art definieren. Die Hyperonymie-Relation könnte mit der Subclass-Relation gleichgesetzt werden (analog zu dem Vorschlag von van Assem et al. 2006b), ohne dass der Dialekt OWL DL verlassen werden müsste.⁶ Falls zudem in GermaNet – wie in der Version 2.1 des Princeton WordNet – neben Synset-Klassen und LUs auch Instanzen eingeführt werden sollten (vgl. Miller & Hristea 2006), könnte ein klassenbasiertes Modell von GermaNet auf einfache Weise angepasst werden. Es ist außerdem, wie oben argumentiert, linguistisch adäquater. Es stellte sich allerdings heraus, dass, um Anfragen wie „Liefere alle Termini, die mit dem GN-Synset WWW synonymisch verbunden sind“ in der Query-Sprache nRQL (Haarslev et al. 2004) über dem GermaNet-Ausschnitt in OWL zu formulieren, doch wieder pro Klasse mindestens ein Individuum vorhanden sein musste, um ein Ergebnis zu erhalten.

Unser nächstes Ziel ist daher zu prüfen, inwieweit eine Klasse wie *Synset* als Meta-Klasse (vgl. Pan et al. 2005) modelliert werden kann, deren Instanzen dann die Klassen der Einzel-Synsets sind. Es würden also in einer solchen Modellierung Objekte existieren, die gleichzeitig als Klassen wie als Individuen fungieren, und es ist bekannt, dass eine solche Konzeption nicht im Rahmen von OWL DL realisiert werden kann. Pan et al. (2005) weisen aber darauf hin, dass die von ihnen definierte Sprachvariante *OWL FA* eine derartige, wohldefinierte, sogenannte *Metamodelling-Extension* von OWL DL beinhaltet und im Gegensatz zu OWL Full immer noch entscheidbar ist. Es wird argumentiert, dass aus Sicht der Nutzer des Semantic Web ein derart hoher Bedarf an einer solchen Erweiterung besteht (Schreiber 2002), dass zu erwarten ist, dass sie in absehbarer Zeit in den Standard mit aufgenommen wird. Eine Konvertierung von GermaNet in ein solches Modell würde aus unserer Sicht voraussetzen, dass auch Verarbeitungssoftware für die entsprechende Sprachvariante von OWL verfügbar ist. Neben der Entwicklung und Evaluierung linguistisch adäquaterer Modellierungsalternativen werden wir den Ansatz zur Verknüpfung von GermaNet mit terminologischen Wortnetzen detaillierter ausarbeiten und den Ansatz auch für andere Domänenontologien testen. Die entstehenden Ressourcen werden in texttechnologischen Anwendungen evaluiert.

Eingereicht: 2. Februar 2007

Überarbeitete Fassung eingereicht: 23. April 2007

Claudia Kunze, Tübingen
Lothar Lemnitzer, Tübingen
Harald Lingen, Gießen
Angelika Storrer, Dortmund

6. Folgte man dem Vorschlag von van Assem et al. (2006b), so erhielte man eine Ontologie in OWL Full.

Literatur

- Antoniou, Grigoris & Frank van Harmelen (2004). Web Ontology Language. In *Handbook on Ontologies*, Steffen Staab & Rudi Studer (Hgg.), 67–92. Heidelberg: Springer.
- Baader, Franz, Ian Horrocks & Ulrike Sattler (2004). Description Logics. In *Handbook on Ontologies*, Steffen Staab & Rudi Studer (Hgg.), 3–28. Heidelberg: Springer.
- Beisswenger, Michael, Angelika Storrer & Maren Runte (2003). Modellierung eines Terminologienetzes für das automatische Linking auf der Grundlage von WordNet. In *Proceedings „Anwendungen des deutschen Wortnetzes in Theorie und Praxis“*, 113–126. Tübingen.
- Ciorăscu, Iulian, Claudia Ciorăscu & Kilian Stoffel (2003). Scalable Ontology Implementation Based on knOWLer. In *Proceedings of the ISWC 2003, Workshop on Practical and Scalable Semantic Systems*. Sanibel Island, Florida.
- Cruse, Alan D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Erdmann, Michael (2001). *Ontologien zur konzeptuellen Modellierung der Semantik von XML*. Karlsruhe: Books on Demand.
- Fellbaum, Christiane (Hg.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Gangemi, Aldo, Nicola Guarino, Claudio Masolo & Alessandro Oltramari (2002). Restructuring WordNet's top-level: The OntoClean Approach. In *Workshop Proceedings of OntoLex'2, Ontologies and Lexical Knowledge Bases, LREC2002*. Las Palmas.
- Graves, Alvaro & Claudio Gutierrez (2005). Data Representation for WordNet: A case for RDF. In *Proceedings of the Third International WordNet conference, Jeju Island*.
- Green, Stephen J. (1999). Lexical Semantics and Automatic Hypertext Construction. *ACM Computing Surveys* 31 (4).
- Gruber, Thomas R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5 (2): 199–220.
- Haarslev, Volker, Ralf Möller & Michael Wessel (2004). Querying the Semantic Web with Racer + nRQL. In *Proceedings of the KI-2004 International Workshop on Applications of Description Logics (ADL'04)*, Ulm. Ulm.
- Horrige, Matthew, Holger Knublauch, Alan Rector, Robert Stevens & Chris Wroe (2004). A Practical Guide To Building OWL Ontologies using the Protégé Plugin and CO-ODE Tools. Edition 1.0. Unveröffentlichtes Manuskript, University of Manchester.
- ISO/TC-37/SC4 (2006). Language Resource Management – Lexical Markup Framework. <http://www.tagmatica.fr/doc/ISO24613cdRev9.pdf>.
- Kuhlen, Rainer (1991). *Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank*. Berlin: Springer.
- Kunze, Claudia (2001). Lexikalisch-semantische Wortnetze. In *Computerlinguistik und Sprachtechnologie: Eine Einführung*, Kai-Uwe Carstensen et al., (Hg.), 386–393. Heidelberg: Spektrum Verlag.
- Kunze, Claudia, Lothar Lemnitzer & Andreas Wagner (2003). *Anwendungen des deutschen Wortnetzes in Theorie und Praxis. Sonderheft der Zeitschrift für Computerlinguistik und Sprachtechnologie*. Bonn.
- Lemnitzer, Lothar & Claudia Kunze (2002). Adapting GermaNet for the Web. In *Proceedings of the first Global WordNetConference, Central Institute of Indian Languages*, 174–181. Mysore.
- Lenz, Eva Anna, Benjamin Birkenhake & Jan Frederik Maas (2003). Von der Erstellung bis zur Nutzung: Wortnetze als XML Topic Maps. In *Anwendungen des deutschen Wortnetzes in Theorie und Praxis*, 127–136. Tübingen.
- Lüngen, Harald & Angelika Storrer (2007). Domain ontologies and wordnets in OWL: Modelling options. In *OTT'06. Ontologies in Text Technology: Approaches to Extract Semantic Knowledge from Structured Information*, Uwe Mönnich & Kai-Uwe Kühnber-

- ger (Hgg.), Band 1 von *Publications of the Institute of Cognitive Science (PICS)*. Osnabrück: Universität Osnabrück.
- Magnini, Bernardo & Manuela Speranza (2002). Merging Global and Specialized Linguistic Ontologies. In *Proceedings of Ontolex 2002*, 43–48. Las Palmas de Gran Canaria.
- Miller, George A. & Florentina Hristea (2006). Word Net Nouns: Classes and Instances. *Computational Linguistics* 32 (1): 1–3.
- Pan, Jeff Z., Ian Horrocks & Guus Schreiber (2005). OWL FA: A Metamodeling Extension of OWL DL. In *Proceedings of the Workshop OWL: Experiences and directions*. Galway.
- Sanfilippo, Antonio, Stephen Tratz, Michelle Gregory, Alan Chappell, Paul Whitney, Christian Posse, Patrick Paulson, Bob Baddeley, Ryan Hohimer & Amanda White (2005). 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.
- Schreiber, Guus (2002). The Web is not Well-formed. *IEEE Intelligent Systems* 17 (2): 79–80.
- Smith, Michael K., Chris Welty & Deborah L. McGuinness (Hgg.) (2004). OWL Web Ontology Language – Guide. Technischer Bericht, W3C (World Wide Web) Consortium. <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.
- Sowa, John F. (2000). *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Pacific Grove: Brooks/Cole.
- Steffen Staab & Rudi Studer (Hgg.) (2004). *Handbook on Ontologies*. Heidelberg: Springer.
- Tochtermann, Klaus (1995). *Ein Modell für Hypermedia: Beschreibung und integrierte Formalisierung wesentlicher Hypermediakonzepte*. Aachen: Shaker.
- van Assem, Mark, Aldo Gangemi & Guus Schreiber (2006a). Conversion of WordNet to a standard RDF/OWL representation. In *Proceedings of the LREC 2006*. Genoa.
- van Assem, Mark, Aldo Gangemi & Guus Schreiber (2006b). RDF/OWL Representation of WordNet. *W3C Public Working Draft of 19 June 2006 of the Semantic Web Best Practices and Deployment Working Group*. <http://www.w3.org/TR/wordnet-rdf/>.
- van Assem, Mark, Maarten R. Menken, Guus Schreiber, Jan Wielemaker & Bob Wielinga (2004). A Method for Converting Thesauri to RDF/OWL. In *Proceedings of the ISWC 2004*, Nummer 3298 in Lecture Notes in Computer Science. Hiroshima.
- Vossen, Piek (1999). *EuroWordNet: a Multilingual Database with Lexical-semantic Networks*. Dordrecht: Kluwer.
- Vossen, Piek (2001). Extending, Trimming and Fusing WordNet for Technical Documents. In *Proceedings of NAACL-2001 Workshop on WordNet and Other Lexical Resources Applications*. Pittsburgh, PA.