

# XML-Strukturen für Worterklärungen von Verben

Vortrag bei den XML-Tagen, Berlin,  
25.09.2007

Gregor Büchel

FH Köln, Institut für Nachrichtentechnik

[gregor.buechel@fh-koeln.de](mailto:gregor.buechel@fh-koeln.de)

# Übersicht:

1. Bedeutungserklärungen in Wörterbüchern (Beispiele und Datenmodell)
2. Definitionsstruktur einer Bedeutungserklärung und XML-Repräsentation
3. XML-Strukturen von Verbfeldern und semantischen Relationen von Verben
4. Integration und Aufbau von Ontologien für Teilmengen von Verbfeldern

# 1. Bedeutungserklärungen in Wörterbüchern (Beispiele und Datenmodell)

Die Analyse knüpft an folgende computerlinguistische Ansätze zur maschinellen Repräsentation von Bedeutungserklärungen an:

- Methoden der computergestützten Lexikographie (z.B. [Heid2004])
- Untersuchungen der Feinstruktur von Bedeutungserklärungen in Wörterbüchern, speziell: Verben ([Wiegand1989], [Lenders1993])
- Wortnetze mit semantischen Relationen (WordNet 3.0 [WN30], GermaNet [GNet])
- Deutschsprachige Wörterbuchportale mit Angaben zu semantischen Relationen ([DWDS20Jh], Wortschatz Uni Leipzig [WUL]).
- Maschinelles Aufbau Ontologien für Verfahren der semantischen Ähnlichkeitssuche ([Theobald2003]).

## 1.1 Ein Formalisierungsansatz für Bedeutungserklärungen in Wörterbüchern

- Wörterbuchartikel haben durch die lexikographische Bearbeitung eine Feinstruktur.
- Wenn Stichwörter in ihrer Bedeutung mehrdeutig sind (Polysemie), trennt die Feinstruktur eines Artikels die Beschreibungen der verschiedenen Bedeutungen durch besondere Gliederungsmerkmale (z.B. römisch-arabische Gliederung, vgl. DUDEN-Bedeutungswörterbuch [DU85]).
- Ein Ziel: Analyse der lexikographischen Feinstruktur, um ein geeignetes XML-Datenmodell aufzustellen.

# 1.1 Ein Beispiel und ein Formalisierungsansatz

- **Auszug aus [W3B] zum Stichwort „streichen“:**
- **streichen** [V.158] I [o. Obj.] **1** [hat gestrichen] *mit der Hand (über etwas) gleiten, die Hand mit leichtem Druck über, auf, durch etwas bewegen; jmdm. übers Haar, über die Wange s.; ... 4 [ist gestrichen] wehen; der Wind streicht durch die Bäume 5 [ist gestrichen] ruhig fliegen; eine Schar Enten strich über den See II [mit Akk.; hat gestrichen] 1 mit Farbe bemalen; eine Tür, einen Zaun s.; die Wohnung neu s. die Wände der Wohnung neu bemalen, tünchen ... 3 mit Belag versehen; sich. jmdm. ein Brot (mit Butter) s. 4 (ein Streichinstrument) spielen; die Geige s. 5 durch Strich(e) ungütig machen, tilgen; einen Abschnitt im Text, ein Wort s. ... 8 einziehen, abnehmen; die Flagge, die Segel s. [auch übertr.] eine Sache aufgeben, den Kampf aufgeben*

- Aufbau einer Bedeutungserklärung (Lesart): Hauptkomponenten sind:**
- (1) ein identifizierendes Merkmal (hier: römisch-arabische Gliederungsnr.)**
  - (2) grammatische Angabe (z.B. [o. Obj.] )**
  - (3) Bedeutungsphrasenangabe (hier: kursiv gesetzt)**
  - (4) Beispiele der Bedeutungsverwendung**

## 1.1 Ein Beispiel und ein Formalisierungsansatz (Forts.)

- Innerhalb einer Lesart wird die Bedeutungsphrasenangabe eines Verbs durch eine Verbalphrase oder durch eine Folge von Verbalphrasen formuliert. Jede dieser Verbalphrasen endet auf den Infinitiv eines Verbs (Paraphrasen). Liest man die Form dieser Paraphrasen als „lexikographische Definition“, so könnte man sie symbolisch notieren als:

$$S := (Lx) VP(v1) \circ VP(v2) \circ \dots \circ VP(vn)$$

- Hierbei bezeichnet S das in einer Lesart Lx zu erklärende Verb (Definiendum-Verb). Jede Verbalphrase VP(vi) wird mindestens durch ein Verb vi (Definiens-Verb) gebildet. Sie kann darüber hinaus weitere Phrasen (z.B. Präpositionalphrasen, Objektphrasen) enthalten. Das Konnektor-Zeichen  $\circ$  notiert die Verknüpfung mehrerer Phrasen in einer Folge (überwiegend OR-Verknüpfung).
- Betrachtung einzelner Bedeutungserklärungen (Lesarten):
- (L.I.1) *mit der Hand (über etwas) gleiten, die Hand mit leichtem Druck über, auf, durch etwas bewegen;*  
$$\text{streichen} := (L.I.1) VP(\text{gleiten}) \circ VP(\text{bewegen})$$
- **mit** :  $VP(v1) = \textit{mit der Hand (über etwas) gleiten}$  ;  $VP(v2) = \textit{die Hand mit leichtem Druck über, auf, durch etwas bewegen}$ ;

## 1.1 Ein Beispiel und ein Formalisierungsansatz (Forts.)

- (L.II.1) *mit Farbe bemalen*

**streichen := (L.II.1) VP(bemalen)**

- **mit : VP(v1) = *mit Farbe bemalen*; [Synonym: anstreichen]**

- (L.II.5) *durch Strich(e) ungültig machen, tilgen*;

**streichen := (L.II.5) VP(machen)  $\circ$  VP(tilgen)**

- **mit : VP(v1) = *durch Strich(e) ungültig machen* ; VP(v2) = *tilgen*;  
[Synonym: durchstreichen]**

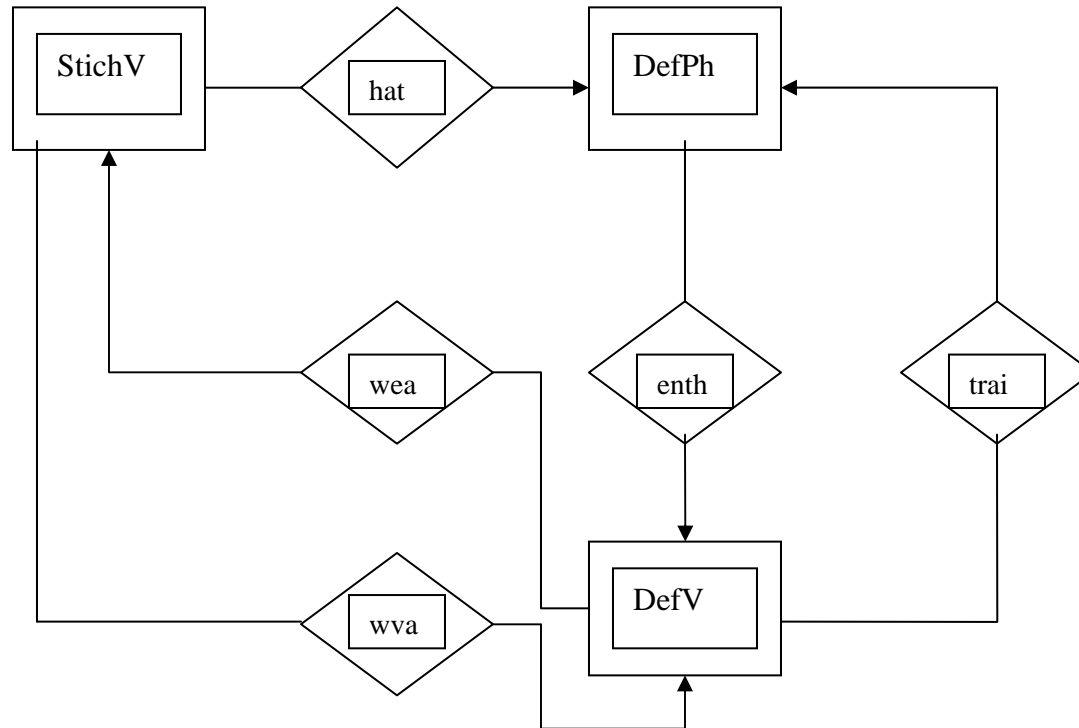
## 1.2 Ein Entity-Relationship-Modell (ERM) für Bedeutungserklärungen von Verben

In Hinsicht auf den Zusammenhang von Bedeutungserklärungen von Verben und auf ihre Feinstruktur gelten folgende Beziehungen:

- (1) Ein Verb als Stichwort wird in einem Wörterbuchartikel durch eine oder mehrere Bedeutungserklärungen erörtert.
- (2) Jede Bedeutungserklärung enthält eine Bedeutungsphrasenangabe.
- (3) Jede Bedeutungsphrasenangabe enthält ein oder mehrere Definiensverben.
- (4) Ein Definiensverb kann an anderer Stelle des Wörterbuchs als Stichwort erklärt werden. Eine Umkehrung gilt auch: Stichwort-Verben können in anderen Artikeln als Definiensverben genutzt werden.

Dieser Zusammenhang kann als ERM formuliert werden.

## 1.2 Ein Entity-Relationship-Modell (ERM) für Bedeutungserklärungen von Verben (Forts.)



**Legende zum ERM:**  
**StichV := Stichwort-**  
**Verb**

**DefPH:= Definiens-**  
**Phrase**

**DefV:= Definiens-**  
**Verb**

**hat:= hat (1:n)**

**wea:= wird erklärt**  
**als (1:c)**

**wva:= wird ver-**  
**wendet als**  
**(1:c\*n)**

**tra:= tritt auf in (1:n)**

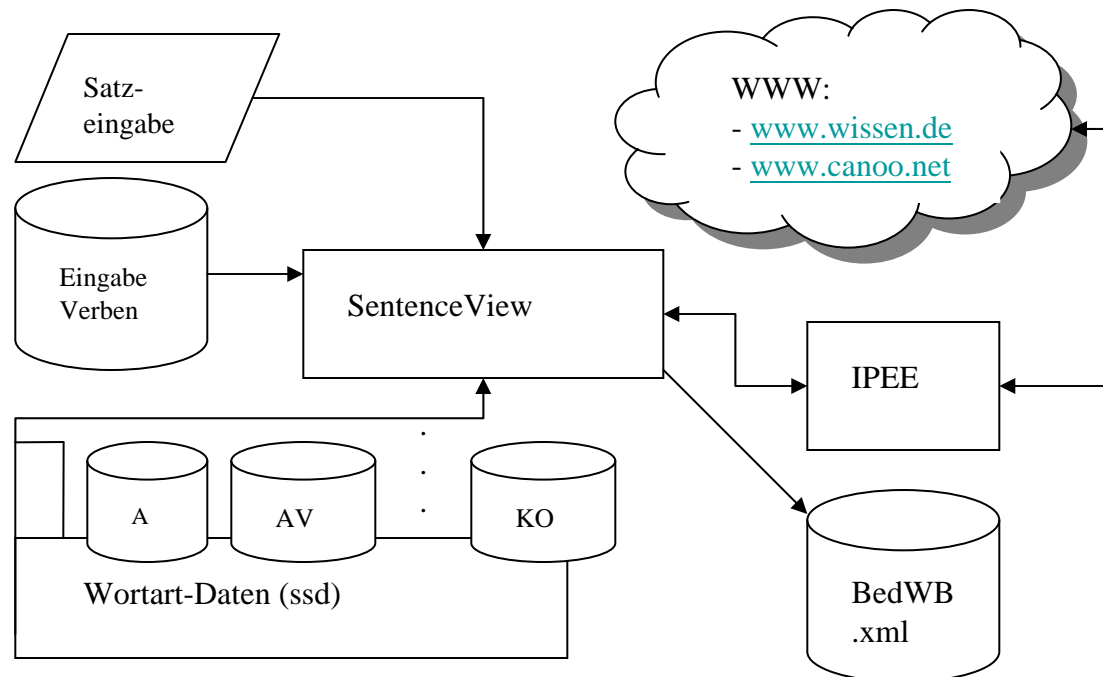
**enth:= enthält (1:n)**

## 2.1 Definitionsstruktur einer Bedeutungserklärung und XML-Repräsentation

- Im folgenden wird ein Programmsystem vorgestellt, das aus Wörterbuchartikel, die durch graphische Mittel segmentiert sind (hier: HTML-Tags), Bedeutungserklärungen von Verben in einem XML-Format extrahiert, mit dem obiges ERM implementiert ist. Das Programmsystem VERBSEMANTIK [Fr07] besteht aus folgenden Modulen:
- SentenceView: Steuermodul, Verwaltung von Benutzereingaben.
- IPEE: Internet Pattern Extract Engine.
- Modul: Wortartentagger.
- Modul: Aufbau einer XML-Datei mit Bedeutungserklärungen für Verben (BedWB.xml).

## 2.2 Definitionsstruktur einer Bedeutungserklärung und XML- Repräsentation (Forts.)

Systemübersicht: Programm **VERBSEMANTIK** :



## 2.3 XML-Repräsentation der Bedeutungserklärung von Verben (Erläuterung der DTD von BedWB.xml)

<b>Bestandteil</b>	<b>XML-Element (– Attribut)</b>
(1) ein identifizierendes Merkmal	BNr
(2) grammatische Angabe	BGramAn
(3) Bedeutungsphrasenangabe	BErkl
(4) Beispiele der Bedeutungsverwendung	BErWA
Verb-Stichwort	SchlagW
VP(vi) wird zerlegt in:	Nominalphrase / Präpositionalphrase
Definiens-Verb vi	Verb (Infinitiv)

## 2.4 Auszug aus BedWB.xml zum Stichwort „streichen“

```
<WArtikel>
<SchlagW>streichen</SchlagW>
<ArtEin>
<BFolgE>
<BNr>I.1.</BNr>
<DefNr>786</DefNr>
<BGramAn>[o. Obj.]</BGramAn>
<BErkl BErklPhrase="mit der Hand (über etwas) gleiten, ... ">
  <Praepositionalphrase>
    <Praeposition Wort="mit" Modus="Modal,Instrumental," />
    <Nominalphrase>
      <Artikel Wort="der" ... Art="bestimmt" />
      <Substantiv Wort="Hand" Genus="feminin" ... />
    </Nominalphrase>
  </Praepositionalphrase>
  <Satzzeichen Wort="(" Zeichen="rundeKlammerauf" />
  <Praeposition Wort="über" Modus="Temporal,Lokal," />
  <Zahlenwort Wort="etwas" Wert="etwas" />
  <Satzzeichen Wort=")" Zeichen="rundeKlammerzu" />
  <Verb Wort="gleiten" ... />
... </BErkl>
</BFolgE>
```

## 2.4 Auszug a. BedWB.xml zum Stichwort „streichen“ (Forts.) (Lesart L.II.1 [Synonym zu: „anstreichen“])

```
<BFolgE>
<BNr>II.1.</BNr>
<DefNr>791</DefNr>
<BGramAn>[mit Akk.</BGramAn>
<BErkl BErklPhrase="mit Farbe bemalen">
  <Praepositionalphrase>
    <Praeposition Wort="mit" Modus="Modal,Instrumental," />
    <Nominalphrase>
      <Substantiv Wort="Farbe" Genus="feminin" ... />
    </Nominalphrase>
  </Praepositionalphrase>
  <Verb Wort="bemalen" Person="1. Pers. Plu./3. Pers. Plu./Infinitiv" Tempus="Praesens"
    Praefix="no" />
</BErkl>
<BErWA>eine Tür, einen Zaun s.</BErWA>
<...> ... <...>
</BFolgE>
```

## 2.5 BedWB.xml (Statistische Übersicht)

1.	Verb-Stichwörter	:	508
2.	Bedeutungsphrasen	:	2108
3.	„Polysemiemaß“ (2./1.)	:	4,14
4.	Segmentierte Verbalphrasen (in 2.)	:	2840
5.	Verschiedene Definiensverben in 4.	:	851
6.	Verwendungshäufigkeit (von 5. in 4.; „wva“-Relation)	:	3,31

Die statistische Übersicht wurde mit einem Java-SAX-Parser-Programm, das auf BedWB.xml angewendet wurde, erstellt.

## 3.1 Strukturen von Verbfeldern

- In GermaNet sind die Verben in 15 Verbfelder angeordnet. Jedes Verbfeld ist eine XML-Datei.
- Die Einteilung der Verbfelder in GermaNet [GVC] ist in Anlehnung an Levin [Lv93] und Schumacher [Su86] ausgearbeitet worden.  
Anzahl der Verbfelder in GermaNet:

$$7 \text{ [Su86]} < 15 \text{ [GVC]} < 49 \text{ [Lv93]}$$

- Nachfolgend wird eine Übersicht über die Einteilung der Verbfelder in GermaNet und ihren quantitativen Aufbau gegeben. Die Angaben wurden durch ein eigenentwickeltes SAX-Programm ermittelt.
- GermaNet hat eine flache Hierarchie: Ein Verbfeld enthält mehrere „Synsets“. Ein Synset ist eine Menge synonymmer lexikalischer Einträge („lexical items“). Jeder lexikalische Eintrag wird durch (genau) eine orthographische Form (Wortform) repräsentiert.

### 3.1 Strukturen von Verbfeldern in GermaNet:

<b>Nr.</b>	<b>Verbfeld</b>	<b>Lex. Item</b>	<b>Syn-sets</b>	<b>Bsp. Verben</b>
<b>1</b>	<b>Allgemein</b>	<b>536</b>	<b>397</b>	<b>existieren, wohnen, neigen</b>
<b>2</b>	<b>Besitz</b>	<b>670</b>	<b>489</b>	<b>geben, nehmen, sammeln</b>
<b>3</b>	<b>Gefühl</b>	<b>502</b>	<b>341</b>	<b>spüren, hassen, jubeln</b>
<b>4</b>	<b>Gesellschaft</b>	<b>1532</b>	<b>1107</b>	<b>behandeln, versöhnen, ausbaden</b>
<b>5</b>	<b>Körperfunktion</b>	<b>692</b>	<b>469</b>	<b>erkranken, hecheln, schlafen</b>
<b>6</b>	<b>Kognition</b>	<b>1349</b>	<b>986</b>	<b>wissen, sehen, planen,</b>
<b>7</b>	<b>Kommunikation</b>	<b>1060</b>	<b>801</b>	<b>zeigen, segnen, erlauben</b>
<b>8</b>	<b>Konkurrenz</b>	<b>324</b>	<b>246</b>	<b>stören, spielen, zanken</b>

### 3.1 Strukturen von Verbfeldern in GermaNet: (Forts.)

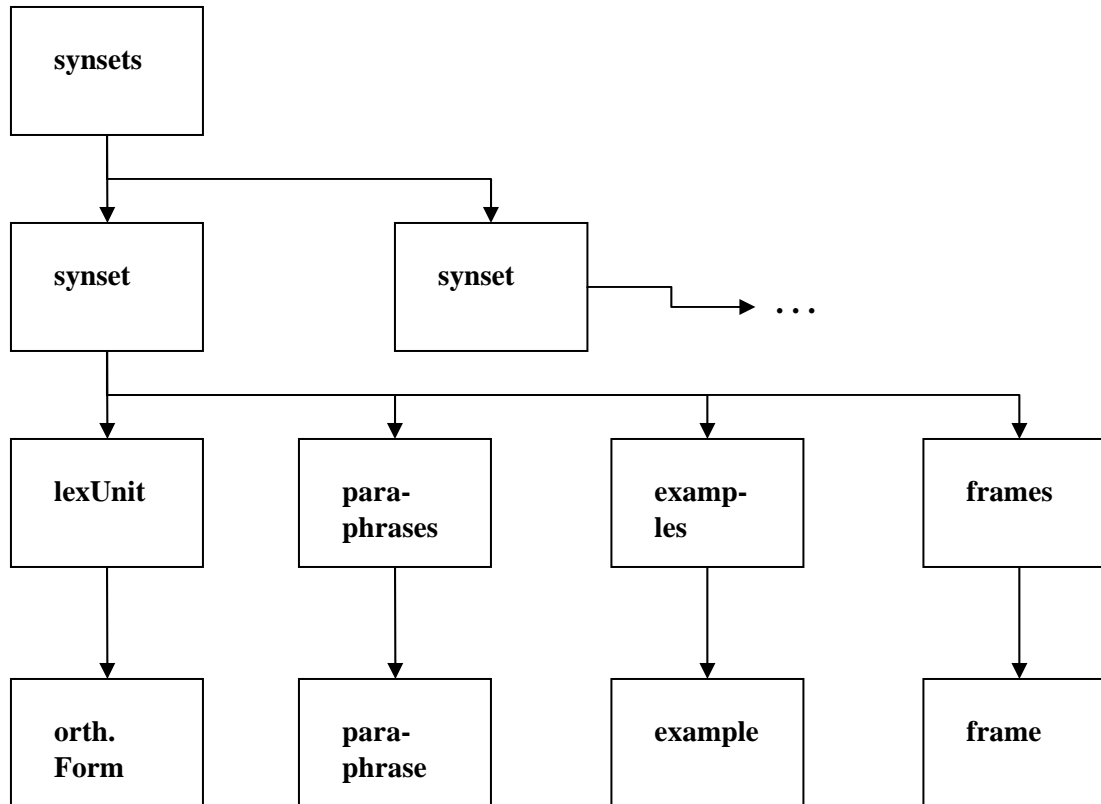
<b>Nr.</b>	<b>Verbfeld</b>	<b>Lex. Item</b>	<b>Syn- sets</b>	<b>Bsp. Verben</b>
<b>9</b>	<b>Kontakt</b>	<b>702</b>	<b>491</b>	<b>haften, abstoßen, wickeln</b>
<b>10</b>	<b>Lokation</b>	<b>1348</b>	<b>979</b>	<b>bewegen, ruhen, fliegen</b>
<b>11</b>	<b>Natürliches Phänomen</b>	<b>167</b>	<b>125</b>	<b>leuchten, regnen, fluten</b>
<b>12</b>	<b>Perzeption</b>	<b>495</b>	<b>367</b>	<b>wahrnehmen, sehen, schmecken</b>
<b>13</b>	<b>Schöpfung</b>	<b>634</b>	<b>470</b>	<b>schaffen, schneiden, komponieren</b>
<b>14</b>	<b>Veränderung</b>	<b>2251</b>	<b>1678</b>	<b>anfangen, häufen, verbessern</b>
<b>15</b>	<b>Verbrauch</b>	<b>259</b>	<b>192</b>	<b>essen, trinken, füttern</b>
<b>Σ</b>		<b>12521</b>	<b>9138</b>	

## 3.1 Verbfelder in GermaNet: Anmerkungen:

- Verben des körperlichen Handelns (z.B. handwerkliche Tätigkeiten wie anstreichen, bohren, fräsen, hobeln, lackieren, putzen, schreien, streichen) sind mit großer Streuung über verschiedene Verbfelder verteilt.
- Bestimmte Verbfelder sind Bezug auf ihre Handlungsfelder nur schwach disjunkt. Z.B. sehen, hören, beweisen, ... sind sowohl im Feld Perzeption als auch im Feld Kognition belegt.
- Die Verbfelder in GermaNet sind im wesentlichen nur Container für Synsets, in denen Synsets nach allgemeiner Zugehörigkeit zusammengefasst sind. Die Synsets sind das wesentliche Strukturierungskonzept, auf sie sind die semantischen Relationen bezogen (vgl. 4.2).

## 3.2 XML-Modellierung der Verbfelder in GermaNet

Der XML-Syntaxbaum der Verbfelder (gemäß der DTD `germanet_objects`):



## 3.2 XML-Modellierung der Verbfelder in GermaNet (Forts.)

Legende zum XML-Syntaxbaum der Verbfelder:

synsets := Menge der Synsets eines Verbfelds

synset := ein Synset

lexUnit := ein lexikalischer Eintrag (Lexem)

orth. Form := orthographische Form, die das Lexem repräsentiert

paraphrase := lexikalische Definition (optional)

frame := Valenzrahmen (optional)

## 3.3 XML-Beispiel: Synset 17 der allgemeinen Verben

```
<synset id="vAllgemein.17" wordClass="verben">
<lexUnit Eigenname="nein" artificial="nein" id="vAllgemein.17.andauern" orthVar="nein" sense="1"
  stilMarkierung="nein">
<orthForm>andauern</orthForm>
</lexUnit>
<lexUnit Eigenname="nein" artificial="nein" id="vAllgemein.17.fortbestehen" orthVar="nein" sense="1"
  stilMarkierung="nein">
<orthForm>fortbestehen</orthForm>
</lexUnit>
<lexUnit Eigenname="nein" artificial="nein" id="vAllgemein.17.fortwähren" orthVar="nein" sense="1"
  stilMarkierung="nein">
<orthForm>fortwähren</orthForm>
</lexUnit>
<lexUnit Eigenname="nein" artificial="nein" id="vAllgemein.17.fortdauern" orthVar="nein" sense="1"
  stilMarkierung="nein">
<orthForm>fortdauern</orthForm>
</lexUnit>
<frames>
<frame>NN.BT</frame>
</frames>
<examples>
<example><text>Der katastrophale Zustand in den Randbezirken dauerte Jahre an.</text>
  <frame>NN.BT</frame></example>
<example><text>Die korrupten Zustände bestanden nach den Wahlen fort.</text> <frame>NN.BT</frame></example>
<example><text>Die Monarchie währte Jahrhunderte fort.</text> <frame>NN.BT</frame></example>
<example><text>Das schöne Wetter dauerte den ganzen Urlaub fort.</text> <frame>NN.BT</frame></example>
</examples>
</synset>
```

## 3.4 Semantische Relationen in GermaNet

- GermaNet ist ein in XML repräsentiertes „semantisches Lexikon“, das „an den Strukturierungsprinzipien des Princeton WordNet“ orientiert ist [Ku04], S. 424.
- Die semantischen Relationen erzeugen die Netzstruktur in GermaNet.
- Semantische Relationen können sowohl zwischen Synonymie-  
klassen als auch zwischen lexikalischen Einträgen bestehen.
- Die semantischen Relationen, an denen Verben beteiligt sind, machen ca. 15,72 Prozent aller semantischen Relationen in GermaNet aus.
- Die nachfolgende Übersicht wurde mittels eines in Java entwickelten SAX-Programmes ermittelt.

### 3.5 Semantische Relationen, an denen Verben beteiligt sind:

<b>Relationsart</b>	<b>Anzahl Tupel <math>r(a,b)</math></b>
Antonymien	252
Pertonymien	287
Partizip_von	224
Hyperonymien	9640
Implikationsbeziehung	9
Kausation	213
Assoziation	20
$\Sigma$	10645

## 3.6 XLINK-Modellierung der semantischen Relationen

- In XLINK-Syntax [XL], die auf XML aufbaut, werden XML-Elemente, die unterschiedlichen Ressourcen (z.B. unterschiedlichen Wortfeldern in GermaNet) angehören, verknüpft. Jedes Element muß hierzu eine XML-ID besitzen, die es identifiziert (z.B. Synset-ID oder lexItem-ID). Beispiel: Antonymie zwischen den lexikalischen Einträgen „ähneln“ und „unterscheiden“:

$r_{\text{ANTONYMIE}}(\text{vAllgemein.126.ähneln}, \text{vAllgemein.150.unterscheiden})$

- xlink:href : Aufbau : DATEINAME#ELEMENT\_ID (hier: lexItem\_ID)
- Der obige Relationentupel hat damit die folgenden xlink:href-Werte:
- Resource 1:  
xlink:href="verben.Allgemein.xml#vAllgemein.150.unterscheiden"
- Resource 2:  
xlink:href="verben.Allgemein.xml#vAllgemein.126.ähneln"

## 3.6 XLINK-Modellierung (Anmerkung)

- Die Speicherung der Relationentupel erfolgt in einer flachen XLINK-Datei, die vom Persistenzkonzept einer relationalen Datenbanktabelle äquivalent ist.
- Alternative: Speicherung der semantischen Relationen in einem RDBMS.

## 4.1 Integration und Aufbau von Ontologien für Teilmengen von Verbfeldern

Ein kurzer Vergleich von XML-Elementen zur Worterklärung von Verben:

- Bei vielen Verben in GermaNet fehlen <paraphrase>-Angaben mit Bedeutungsphrasenangabe.
- In VERBSEMANTIK verfügen alle Verben über eine Bedeutungsphrasenangabe in der Form

$$\mathbf{S} := (\mathbf{Lx}) \mathbf{VP(v1)} \circ \mathbf{VP(v2)} \circ \dots \circ \mathbf{VP(vn)}$$

(Z.B.: streichen := (L.I.1) VP(gleiten)  $\circ$  VP(bewegen))

Wie kann der Verbnetz-Bestand von GermaNet um Bedeutungsphrasenangaben (Plesionymien) **maschinell** erweitert werden?

Vorhaben: Diese Überlegungen münden in dem Ziel einer XML-basierten und ontologisch geführten Disambiguierung von Bedeutungsphrasenangaben. Dieses Vorhaben hat einen ontologischen und einen algorithmischen Aspekt.

## 4.2 Aufbau von Ontologien für Teilmengen von Verbfeldern

- Ontologischer Aspekt: Es bedarf einer ontologischen Bearbeitung, die zu Teilmengen bisheriger Verbfelder wohlstrukturierte Hyperonymie-Bäume mit echten Verb-Lexemen (Leitverben) als Knoten bildet. Diese Bäume dienen dann als persistente Basis zur maschinellen Verarbeitung von Bedeutungsphrasenangaben.
- Beispiel: BewOT, GefOT, GesOT sind ontologisch aufgebaute Hyperonymie-Bäume mit Verb-Lexemen als Knoten, die auf Synsets in den GermaNet-Feldern vLokation, vGefuehl und vGesellschaft referenzieren. Diese drei Hyperonymie-Bäume haben als Wurzel-Knoten jeweils das Verb-Lexem „bewegen“ gemäß drei Bedeutungen als physikalische Bewegung, Gefühlsausdruck oder Handlung im gesellschaftlichen Raum. Im Baum BewOT ist das Lexem „gleiten“ als abhängiger Knoten (Hyponym) belegt.

## 4.3 Aufbau von Ontologien für Teilmengen von Verbfeldern

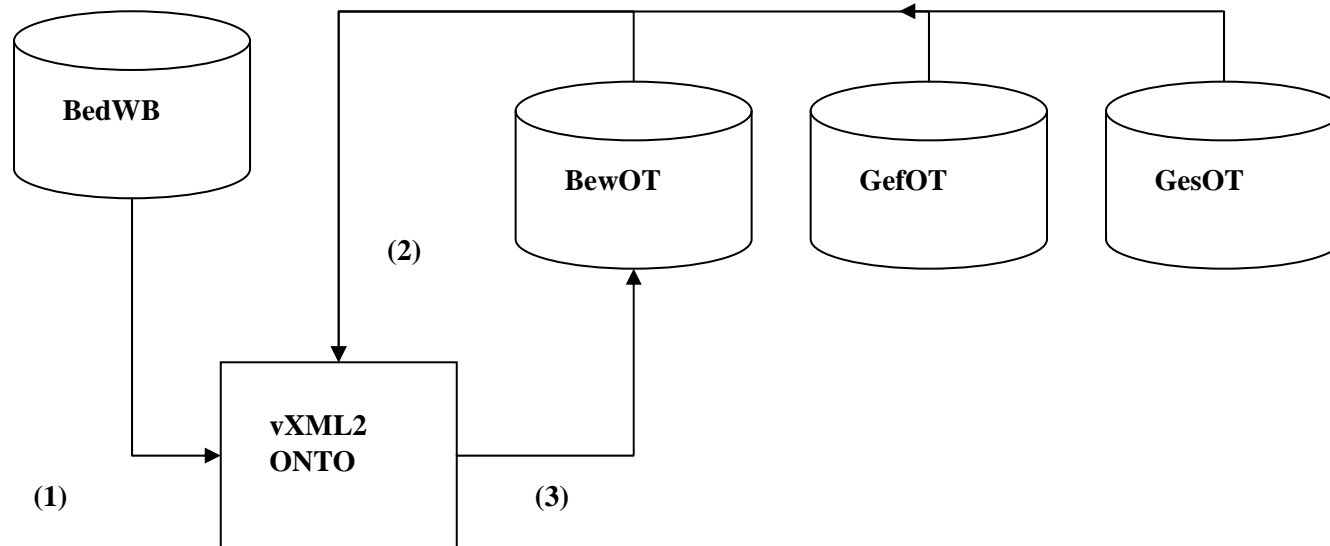
- Algorithmischer Aspekt: Gegeben ist eine Lesart eines Verbs  $s$  (z.B. streichen := (L.I.1) VP(gleiten)  $\circ$  VP(bewegen)), mit der das Verb  $s$  als Plesionym dem zutreffenden Hyperonymie-Baum angefügt werden soll. Der Algorithmus soll durch das nachfolgende Informationsflussdiagramm veranschaulicht werden. Die Informationsflüsse sind:
  - (1) {streichen -> (L.I.1) VP(gleiten); streichen -> (L.I.1) VP(bewegen)}
  - (2) {gleiten -> vBewONT559,561,726; bewegen -> vBewONT29,31,35,55; bewegen -> vGefONT238; bewegen -> vGesONT575,647,658}
  - (3) {insertUnder(gleiten, streichen, (L.I.1) VP(gleiten), vBewONT)}

## 4.4 Aufbau von Ontologien für Teilmengen von Verbfeldern

- Der Prototyp dieser INSERT-Methode soll in Pseudo-DOM-Syntax lauten:

```
public Node insertUnder(Node parent, Node newChild,  
    Paraphrase bErkl, OntologyTree bMatch)
```

- Informationsfluss des Algorithmus:



## 4.5 Ontologien für Teilmengen von Verbfeldern (Ausblick)

Kurzer Ausblick: Die Ontologie der Hyperonymiebäume ist so konzipiert, dass für darin enthaltene Verben folgende Arten von Worterklärungen abfragbar sind:

- Pfad der Hyperonyme, beginnend von der Baumwurzel
- Menge aller synonymen Verben
- Bedeutungsphrasenangabe des Verbs
- zugehörige Phrasenstruktur

**Danke für Ihr Interesse!**

## Literaturübersicht:

- [Ca04] K.-U. Carstensen (et. al.) (Hg.): “Computerlinguistik und Sprachtechnologie“, Heidelberg (Spektrum) 2004.
- [DU85]: W. Müller (Hrsg.): „DUDEN-Bedeutungswörterbuch“, Mannheim, Wien, Zürich (Bibliographisches Institut), 1985.
- [DWDS20Jh]: Berlin-Brandenburgische Akademie der Wissenschaften: „DWDS – Das Digitale Wörterbuch der deutschen Sprache des 20. Jh.“ (<http://www.dwds.de/>)
- [Fr07]: A. Frey: „Anwendung der Konzeption semistrukturierter Daten zur semantisch orientierten Verarbeitung von Verben der deutschen Sprache“, Masterthesis, Institut für Nachrichtentechnik der FH Köln, Köln 2007.
- [GNet]: GermaNet (Homepage), <http://www.sfs.uni-tuebingen.de/lzd/>
- [GVC]: „Verbs in GermaNet“, <http://www.sfs.uni-tuebingen.de/lzd/Verbs.html>

## Literaturübersicht (Forts.):

- [Heid04]: U. Heid: „Computergestützte Lexikographie und Terminologie“, in: [Ca04], S.471-478.
- [Ku04]: C. Kunze: „4.3 Lexikalisch-semantische Wortnetze“, in: [Ca04], S.423-431.
- [Lenders93]: W. Lenders: „Strukturelle Eigenschaften von Bedeutungserklärungen deutscher Verben“, in: K. Mattheier et. al. (Hg.): „Vielfalt des Deutschen – Festschrift für Werner Besch“, Frankfurt a. M. (Lang), 1993, S.429-459.
- [Lv93]: B. Levin: „English Verb Classes and Alternations“ Chicago and London (Univ. of Chicago Press) 1993.
- [Su86]: H. Schumacher (Hrsg.): „Verben in Feldern – Valenzwörterbuch zur Syntax und Semantik deutscher Verben“, Berlin, New York (W. de Gruyter) 1986.

# Literaturübersicht (Forts.):

- [Th03]: A. Theobald: “An Ontology for Domain-oriented Semantic Similarity Search on XML Data”. In: (Weikum, G.; Schöning, H.; Rahm, E.; Hrsg.): BTW 2003 – Datenbanksysteme für Business, Technologie und Web. Tagungsband der 10. BTW-Konferenz, 16.-18. 2. 2003, Leipzig. LNI-Proceedings, Vol. P-26, GI, Bonn, 2003.
- [Wi89]: H. E. Wiegand: „Die lexikographische Definition im einsprachigen Wörterbuch. In: (Hausmann, F.-J. Hrsg.): Wörterbücher - ein internationales Handbuch zur Lexikographie. de Gruyter, Berlin, New York, 1989, S.530-587.
- [WUL]: „Wortschatz Universität Leipzig“ (<http://wortschatz.uni-leipzig.de/> )
- [W3B]: „BERTELSMANN Wörterbuch“, in: [www.wissen.de](http://www.wissen.de); (<http://www.wissen.de/wde/generator/wissen/services/suche/wbger/index.html>)
- [WN30]: „WordNet 3.0“, <http://wordnet.princeton.edu/> (letzter Besuch: 12.6.2007)
- [XL]: „XML Linking Language (XLink) Version 1.0“, W3C Recommendation 27 June 2001, <http://www.w3.org/TR/xlink/> (letzter Besuch: 12.6.2007)

# Anhang

## A.0 Auszug a. BedWB.xml zum Stichwort „streichen“ (Forts.) (Lesart L.II.1 [Synonym zu: „durchstreichen“])

```
<BFolgE>
<BNr>II.5.</BNr>
<DefNr>795</DefNr>
<BGramAn>[mit Akk.</BGramAn>
<BERkl BERklPhrase="durch Strich(e) ungültig machen, tilgen">
  <Praepositionalphrase>
    <Praeposition Wort="durch" Modus="Instrumental,Lokal,Kausal," />
    <Nominalphrase>
      <Substantiv Wort="Strich" Genus="maskulin" ... />
    </Nominalphrase>
  </Praepositionalphrase>
  <Satzzeichen Wort="(" Zeichen="rundeKlammerauf" />
  <notmatched Wort="ungültig" />
  <notmatched Wort="machen" />
  <Satzzeichen Wort="," Zeichen="Komma" />
  <Verb Wort="tilgen" Person="Infinitiv" ... />
</BERkl>
... </BFolgE>
</WArtikel>
```

## A.1 Semantische Relationen: Beispiele:

- Die Relationstupel werden nachfolgend als  $r(a,b)$  notiert.  $r$  bezeichnet die Relation;  $a$ ,  $b$  bezeichnen die beteiligten Synsets oder lexikalischen Einheiten.
- Antonymie:  $r_{\text{antonym}}(\text{ähneln}, \text{unterscheiden})$
- Pertonymie (Derivationsrelation, eigentlich morphologisch):  
 $r_{\text{pertonym}}(\text{biwakieren}, \text{Biwak})$
- Partizip\_von (eigentlich morphologisch):  $r_{\text{participleOf}}(\text{fließend}, \text{fließen})$
- Hyperonymie: Überordnung eines Oberbegriffs (Hyperonym) über zugehörige Unterbegriffe (Hyponyme). Normalerweise geht man vom Oberbegriff (z.B. „Singvogel“) aus und gelangt zum Unterbegriff (z.B. „Amsel“). GermaNet hat die Hyperonymie als eine IS-A-Beziehung modelliert: Amsel IS A Singvogel. Beispiel: Die Synonymie klasse vGes1080 ist dem Verb verarmen übergeordnet:  
 $r_{\text{hyperonym}}(\text{verarmen}, \text{vGes1080}\{\text{verschlechtern}, \text{absteigen}\})$

## A.2 Semantische Relationen: Beispiele (Forts.):

- Implikationsbeziehung (entailment) wird von GermaNet restriktiv verwendet, z.B.:  $r_{\text{entailment}}(\text{schnarchen}, \text{schlafen})$ . D.h. der Zustand „schnarchen“ impliziert den Zustand „schlafen“.
- Kausation: Der Zustand a verursacht in der Zeitfolge den Zustand b. Z.B. a = Synonymklasse vAllgemein311 = {halten3, behalten} (Z.B.: „Die Firma will sie behalten“), b = Synonymklasse vAllgemein290 = {bleiben}:  $r_{\text{kausation}}(\{\text{halten3}, \text{behalten}\}, \{\text{bleiben}\})$
- Assoziation: wird von den GermaNet-Verfassern als vages Konzept betrachtet, von dem nur selten Gebrauch gemacht wird ([GermaNetSP]): Z.B. a = Synonymklasse vBesitz461 = {haushalten, wirtschaften} , b = Synonymklasse vBesitz411 = {sparen} :  $r_{\text{assoziation}}(\{\text{haushalten}, \text{wirtschaften}\}, \{\text{sparen}\})$

### A.3 XLINK-Quelltext (Auszug zu $r_{\text{ANTONYMIE}}$ (ähneln, unterscheiden) ) :

```
<lex_rel name="antonymy" dir="both" sense="1"
  xmlns:xlink="http://www.w3.org/1999/xlink" xlink:type="extended">
  <locator xlink:type="locator"
    xlink:href="verben.Allgemein.xml#vAllgemein.150.unterscheiden"
    xlink:label="vAllgemein.150.unterscheiden"/>
  <locator xlink:type="locator"
    xlink:href="verben.Allgemein.xml#vAllgemein.126.ähneln"
    xlink:label="vAllgemein.126.ähneln"/>
  <arc xlink:type="arc" xlink:from="vAllgemein.126.ähneln"
    xlink:to="vAllgemein.150.unterscheiden" xlink:actuate="onRequest"
    xlink:show="other"/>
  <arc xlink:type="arc" xlink:from="vAllgemein.150.unterscheiden"
    xlink:to="vAllgemein.126.ähneln" xlink:actuate="onRequest"
    xlink:show="other"/>
</lex_rel>
```

## A.4 Maschinelle Verarbeitung von Plesionymien

- Die Plesionymie ist eine semantische Relation, die in GermaNet als Relationstyp nicht belegt ist.
- $r_{\text{PLESIONYM}}(a,b) : \Leftrightarrow$  a hat eine zu b benachbarte Bedeutung, aber a ist nicht synonym zu b ( $\pi\lambda\eta\sigma\iota\omicron\varsigma$  : nahe, benachbart). Man könnte auch sagen, a ist partiell synonym zu b.
- Beispiel ([Klenner2003]): b=Kies, a=Geld. b=Kies verfügt über eine Lesart, in der es als synonym zu a=Geld angesehen wird: Aus WAHRIG Rechtschreibung ([W3Wa]):  
**Kies** *m.* 1 nur Sg. **1.** kleine, zerbröckelte, glatt geschliffene Gesteinstrümmer, viele Kieselsteine; **2.** *ugs.:* Geld
- In Hinsicht auf eine vereinfachte Form der Bedeutungserklärung von Verben (  $s := (Lx) VP(v1)$  ) gilt damit:  $r_{\text{PLESIONYM}}(s,v1) : \Leftrightarrow$  es gibt eine Lesart Lx von s, so dass  $r_{\text{SYNONYM}}(s, (Lx) VP(v1))$  gültig ist.

## A.4 Maschinelle Verarbeitung von Plesionymien (Forts.)

- (1) Im XML-File BedWB.xml (vgl. Kap.2) werden Lesarten, die Plesionymien für Verben darstellen, verwaltet.
- (2) In Hinsicht auf eine vereinfachte Form der Bedeutungserklärung von Verben ( $s := (Lx) VP(v1)$ ) gilt damit:  $rPLESIONYM(s, v1) : \Leftrightarrow$  es gibt eine Lesart  $Lx$  von  $s$ , so dass  $rSYNONYM(s, (Lx) VP(v1))$  gültig ist.
- (3) In GermaNet sind bei einer Reihe von Synsets (aber bei vielen auch nicht) im XML-Element `<paraphrase>` (vgl. Kap.2) Plesionymie-Kandidaten hinterlegt. Von der XML-Struktur her sind die `<paraphrase>`-Elemente leider nicht an die lexikalischen Einheiten gekoppelt.
- (4) Es gibt zwar Ansätze zu einer ontologischen Bearbeitung von GermaNet mit OWL ([KuLeLüST06]), die aber noch in der bisherigen Verbklassenstruktur befangen bleiben, die u.a. Hyperonymie-Knoten enthält, die auf keine **echten** lexikalischen Einträge bezogen sind (z.B. `vLokation.261. Manner_spez`).

## A.5 Literaturübersicht (Forts.):

- [Klenner2003]: M. Klenner: „Lexikonaufbau: Semantische Netze“ (Skript zur Vorlesung, 27.5.2003, Universität Zürich, Institut für Computerlinguistik),  
<http://www.ifi.unizh.ch/cl/klenner/lehre/ss03/morph/wordnet.2.pdf>  
(letzter Besuch: 12.6.2007)
- [KuLeLüSt06]: C. Kunze, L. Lemnitzer, H. Lungen, A. Storrer: „Modellierung und Integration von Wortnetzen und Domänenontologien in OWL am Beispiel von GermanNet und TermNet“, in: M. Butt (Ed.): „Proceedings of KONVENS 2006“, Universität Konstanz, KOPS: <http://www.ub.uni-konstanz.de/kops/volltexte/2006/2013/> , S.91-96.
- [W3Wa]: „Wahrig Rechtschreibung“, in: [www.wissen.de](http://www.wissen.de);  
(<http://www.wissen.de/wde/generator/wissen/services/suche/wbger/index.html>)