

FrameNet und WordNet - Perspektiven für die Verknüpfung zweier lexikalisch-semantischer Netze

Abstract

Dieser Beitrag stellt das FrameNet-Projekt vor, das auf dem Ansatz der Frame-Semantik von Charles J. Fillmore basiert. Verglichen mit WordNet-Daten besitzen die Daten von FrameNet eine grundlegend andere Struktur. Die Verknüpfung der beiden semantischen Netze ist jedoch gerade aufgrund der komplementären Informationen, die sie bieten, lohnenswert. In Form eines Entity-Relationship-Diagramms wird eine Datenbankstruktur, die sämtliche Daten verbindet, dargestellt und diskutiert.

1 Einleitung

FrameNet ist ein lexikographisches Projekt, das seit 1997 unter der Leitung von Charles J. Fillmore am International Computer Science Institute (ICSI) in Berkeley durchgeführt wird. In diesem Projekt werden semantische und syntaktische Eigenschaften von Wort-Tokens in Korpora erfasst, in einer MySQL-Datenbank systematisiert und auf der FrameNet-Website der internationalen Forschungsgemeinschaft zur Verfügung gestellt (vgl. <http://www.icsi.berkeley.edu/~framenet/index.html>; FILLMORE, JOHNSON & PETRUCK 2003).

FrameNet basiert auf dem Ansatz der Frame-Semantik von Charles J. Fillmore und Anderen (vgl. FILLMORE 1982; FILLMORE 1985; FILLMORE & ATKINS 1992; FILLMORE & ATKINS 1994).

Im nächsten Abschnitt wird eine Einführung in die Frame-Semantik gegeben, Abschnitt 3 stellt die Umsetzung dieses Ansatzes im FrameNet-Projekt dar. In Abschnitt 4 werden in kurzer Form einige Bemerkungen zu bereits bestehenden WordNet-Projekten gemacht. Abschnitt 5 zeigt den Entwurf einer Datenbank, die Informa-

tionen von WordNet bzw. GermaNet mit FrameNet verbindet. Im Ausblick werden weitere Möglichkeiten der Erweiterung einer solchen Datenbankstruktur aufgezeigt.

2 Frame-Semantik

Im Gegensatz zu anderen Ansätzen in der Semantik (wie beschrieben bei SCHUMACHER & STEINER 2002), insbesondere der Wahrheitsbedingungen-Semantik, ist die zentrale Idee der Frame-Semantik, dass lexikalische Bedeutungen am besten vor dem Hintergrund von miteinander zusammenhängendem Wissen beschrieben werden können (vgl. FILLMORE 1985).

Lexeme wie LUKEWARM (FILLMORE 1982: 123) können auch im Bereich der Wortfeldtheorie beschrieben werden (vgl. GIPPER 1995: 337), doch Fillmore (FILLMORE 1985: 226ff.) grenzt seinen Ansatz insofern von der Wortfeldtheorie aus der Münsterschen Tradition (vgl. TRIER 1931; WEISGERBER 1962; GECKELER 1971; GIPPER 1973; GIPPER 1995; ZILLIG 1994) ab, als dass innerhalb der Frame-Semantik nicht *Wörter*, sondern *Konzepte* die Einheiten der Beschreibung sind, so dass – die aus der Sicht der Wortfeldtheorie postulierten – lexikalischen Lücken irrelevant für das zugrunde liegende Wissen sind. Konzepte sind unabhängig davon, ob es Bezeichnungen für bestimmte Wissenszusammenhänge gibt.

Zum Beispiel existiert im Deutschen kein Lexem für das Konzept „nicht mehr durstig sein“, während dies für „nicht mehr hungrig sein“ der Fall ist (*satt*) (vgl. SCHUMACHER & STEINER 2002: 185). *shore* und *coast* werden im Deutschen zwar jeweils mit *Küste* übersetzt; das erste Lexem bezeichnet allerdings die Küste, die vom

Wasser aus erreicht wird, während *coast* verwendet wird, wenn die Küste von Land aus erreicht wird (vgl. FILLMORE 1982: 121; FILLMORE 1985: 236). Um die Konzepte SHORE, COAST und KÜSTE verstehen zu können, muss nicht das gesamte Wortfeld, wohl aber die gesamte Wissensstruktur, in die sie eingebettet sind, bekannt sein. Ähnliche Unterschiede gibt es bei *land* und *ground*: Bei *land* besteht der Hintergrund in der Kontrastierung mit dem Meer, im zweiten Fall wird die Beziehung zur Luft präsupponiert (vgl. FILLMORE 1982: 121). Ein *Häretiker* präsupponiert eine Religion, von der abgewichen wird (vgl. FILLMORE 1982: 123), eine *Friedensaktivistin* die Möglichkeit oder tatsächliche Existenz eines Krieges.

Fillmore bezeichnet derartiges Wissen über Hintergründe als *Frames*.¹ *Lexikalische Einheiten* im Verlauf/Gebrauch evozieren diese Frames. So evozieren die Ausdrücke *travel* und *reisen* nicht nur eine Person, die verreist, sondern auch die Konzepte *Ausgangspunkt*, *Ziel* oder *Strecke*. Eventuell wird ein Fahrzeug oder Flugzeug benutzt – hier kommt also eine Menge Wissen ins Spiel.

3 FrameNet

Im FrameNet-Projekt werden Frames und die zu ihnen gehörenden lexikalischen Einheiten beschrieben. Die definierten Frames werden innerhalb des lexikalischen Netzwerks mit anderen Frames verbunden und eine Auswahl von Sätzen aus Korpora wird mit ihnen annotiert, wobei sämtliche Daten in einer Datenbank gespeichert werden. Annotationstools helfen bei der Bewältigung der komplexen Aufgaben. Im Folgenden werden die einzelnen Arbeitsschritte für die Gewinnung der Daten beschrieben.²

Die Definition von Frames und Frame-Elementen

Häufig wird ein neuer Frame definiert, wenn lexikalische Einheiten sich nicht bereits bestehenden Frames zuordnen lassen. Diese Defini-

tionen stammen vom Annotationsteam des Projekts oder aus dem *Concise Oxford Dictionary*. Um beim obigen Beispiel zu bleiben, wird der Frame TRAVEL folgendermaßen definiert:

In this frame a Traveler₁ goes on a journey, an activity, generally planned in advance, in which the Traveler₁ moves from a Source₂ location to a Goal₃ along a Path₄ or within an Area₅. The journey can be made with an Vehicle₆ and/or accompanied by Co-travelers₇ and Baggage₈. The Duration₉ or Distance₁₀ of the journey, both generally long, may also be described. Words in this frame emphasize the whole process of getting from one place to another, rather than profiling merely the beginning or the end of the journey.

Die markierten Wörter sind die *Frame-Elemente*, semantische Rollen, die zum jeweiligen Frame gehören. Im Vergleich zu den semantischen Rollen der frühen Arbeiten Fillmores zur Kasusgrammatik (vgl. FILLMORE 1968 u. 1977) sind die Frame-Elemente weitaus spezifischer. Auch zu den Frame-Elementen werden Definitionen entwickelt, wie zum Beispiel:

Traveler [Trav]₁

This is the living being which travels. Normally, the Traveler₁ is expressed as an external argument.

Vehicle [Veh]₆

The Vehicle₆ is the means of conveyance on the journey.

Beim Prozess der Frame-Definition wird auch eine Liste mit lexikalischen Einheiten festgelegt, die prinzipiell erweiterbar ist. Die folgenden lexikalischen Einheiten evozieren den Frame TRAVEL:

commute.v, excursion.n, expedition.n, journey.n, journey.v,

FrameNet und WordNet

junket.n, odyssey.n, peregrination.n, pilgrimage.n, safari.n, tour.n, tour.v, travel.n, travel.v, trip.n, voyage.n, voyage.v (vgl. www.icsi.berkeley.edu/~framenet)

Homonyme evozieren unterschiedliche Frames, zum Beispiel kann *Trip* den Frame TRAVEL, andererseits aber auch den Frame INTOXICATION evozieren.

Das FrameNet-Projekt geht in seiner Konzeption nicht von Lemmata mit mehreren Bedeutungen, sondern von Frames aus, die möglichst vollständig beschrieben werden sollen. Daher sind Homonyme und polyseme Lexeme in der Datenbasis häufig nicht vollständig erfasst.³

Zwischen den Frames besteht eine *Framehierarchie*, derart, dass Frames der oberen Hierarchieebene Eigenschaften an die unteren vererben. Zum Beispiel ist der Frame TRAVEL ein Spezialfall des Frames SELF_MOTION, der wiederum Eigenschaften der Frames MOTION und INTENTIONALLY_ACT erbt. Dabei stehen die Frame-Elemente in definierten Beziehungen zueinander: Zum Beispiel entspricht das Frame-Element *Co_travelers* des Frames TRAVEL dem Frame-Element *Cotheme* bei SELF-MOTION (vgl. Abbildung 1).

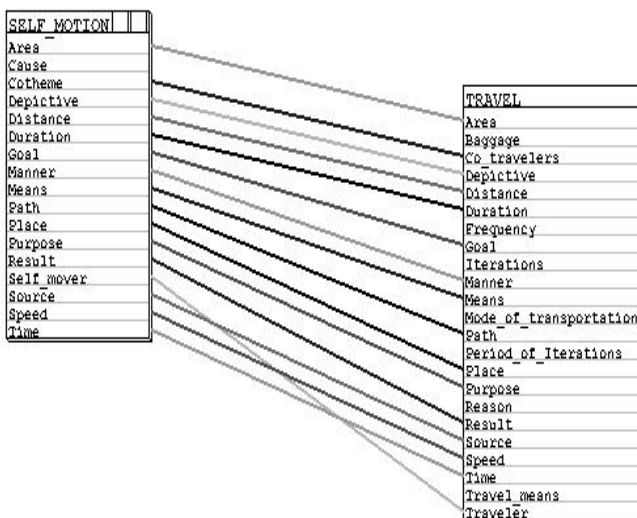


Abbildung 1: Verbindungen zwischen vererbten Frames

Wie in der Abbildung zu erkennen ist, ist der *Kind-Frame* hinsichtlich der semantischen Rollen elaborierter als der *Eltern-Frame*. Ferner können Frames die Bestandteile anderer, komplexer Frames sein. Zum Beispiel ist der Frame TRIAL ein *Subframe* von CRIMINAL_PROCESS. Ab einer gewissen Hierarchieebene werden Frames zu so genannten *Domains* zusammengefasst. CRIMINAL_PROCESS ist ein Beispiel dafür. Erfasst werden auch die Valenzmuster, also die phrasalen Kategorien und die grammatischen Funktionen (Subjekt, Objekt etc.) der Frame-Elemente.

3.2 Die Frame-Annotation in Sätzen

Nach der Definition von Frames und Frame-Elementen sowie der Festlegung der Relationen dieser Entitäten untereinander, werden Beispielsätze aus großen Textkorpora (das BNC-Korpus und das American Newswire-Korpus) extrahiert und dabei direkt unterschiedlichen Subkategorisierungsrahmen zugewiesen (so genannte *Subkorpora*⁴). Erst im nächsten Schritt werden diese Sätze annotiert.

Die annotierten Sätze sollen möglichst typische Beispiele aller möglichen Kombinationen von Frame-Elementen erfassen; sie spiegeln jedoch nicht die Verteilung dieser Muster im Text wider. Beispiele von Annotationen für den Frame TRAVEL sind:

Ellen₁ JOURNEYED₁₂ to Europe₃ with five suitcases₈.

Samantha₁ JOURNEYED₁₂ 2500 miles₂ with her family₇ by sea₄ to China₃.

The Osbournes₁ took a TRIP₁₂ from Beverly Hills₂ to London₃ on the Concorde₆.

Die annotierten Frame-Elemente werden nach der Relevanz für den jeweiligen Frame unterschieden. Bei *core frame elements* handelt es sich

um grundlegende konzeptuelle Bestandteile des Frames. Zum Beispiel sind Goal₃ und Path₄ Core-Frame-Elemente des TRAVEL-Frames. Häufig werden auch *periphere* Frame-Elemente annotiert, die Angaben zu Zeit, Ort, Zweck usw. sind. Für den TRAVEL-Frame sind Baggage₈ und Distance₁₀ Beispiele für periphere Frame-Elemente. *Extrathematische Frame-Elemente* gehören anderen Frames an, sie werden innerhalb der Sätze als solche annotiert. Zum Beispiel wird im folgenden Satz *in pursuit of suspected traffickers* dem Frame-Element *Depictive* zugewiesen. *Depictive* bezeichnet den Zustand des Reisenden.

Customs officers from each country₁ would be allowed to TRAVEL₁₂ to either country₃ in pursuit of suspected traffickers₁₁.

Im Gegensatz zu den anderen Frame-Elementen werden extrathematische Frame-Elemente innerhalb der Frame-Hierarchie nicht notwendigerweise vererbt.

Es gibt Belege, in denen die Core-Frame-Elemente nicht im Satz auftreten, jedoch als Teil des Frames vorhanden sind. Für solche Core-Frame-Elemente wird eine spezielle Markierung verwendet. Unterschieden wird hier zwischen *Constructional Null Instantiation* (CNI), *Definite Null Instantiation* (DNI) und *Indefinite Null Instantiation* (INI). CNI umfasst zum Beispiel weggelassene Subjekte in Imperativsätzen oder ausgelassene Agens-Formen in Passivsätzen:

Days began early and ended late so that maximum distances₄ could be TRAVELLED₁₂.
(CNI₁)

Im Fall von DNI befindet sich das fehlende, obligatorische Element im situationalen oder linguistischen Kontext. Im folgenden Beispiel ergibt sich das Goal₃ aus dem Kontext.

As pilgrims the children of Israel₁ JOURNEYED₁₂, led by the guiding hand of God.
(DNI₃)

INI bezieht sich auf Frame-Elemente, bei denen die Interpretation obligatorisch ist, die Elemente jedoch nicht im Kontext zu finden sind.

Im folgenden Beispiel ist etwa das Goal₃ nicht instanziiert:

I₁ used to TRAVEL₁₂ by bus₆ a lot₁₀, so I had a season ticket.
INI₃

Ein Sonderfall der Null-Instanzierung sind *Inkorporationen*. Hier besteht eine enge morphologische Verbindung zwischen Verb und nicht instanziiertem Frame-Element, ein Beispiel hierfür ist die lexikalische Einheit *to bicycle*:

When he had leisure he₁ went BICYCLING₁₂ to Lincolnshire village churches₃. INC₆

Hier ist das Vehicle₆ nicht instanziiert.

Diese Markierungen erlauben es, Sätze mit Ellipsen von solchen mit tatsächlich nicht vorhandenen Frame-Elementen zu unterscheiden. Dies kann für die semantische Klassifizierung von Verben von Nutzen sein (vgl. SCHULTE IM WALDE 1998: 86f).

Für die Definition und Beschreibung der Frames, die Extraktion der Satzsammlungen aus den Korpora und die Annotation der Sätze werden die Tools für die FrameNet-Annotation verwendet (*FrameNet Desktop*, vgl. Fillmore u.a. 2003).

Abbildung 2 zeigt ein Beispiel für den Prozess einer Annotation. Mit Hilfe des Annotationstools werden neben den Frame-Elementen auch die Phrasenstrukturtypen und grammatische Funktionen annotiert, Fehler der Wortartenannotation können ebenfalls korrigiert werden.

3.3 Die Struktur der FrameNet-Datenbank

Die gesamten Daten werden in einer relationalen Datenbank abgelegt (vgl. BAKER, FILLMORE & CRONIN 2003), die aus zwei Teilen besteht: In der so genannten lexikalischen Datenbasis sind neben den Lemmata, Lexemen, Wortformen, Wortarten auch die Frames, Frame-Elemente und die Beziehungen zwischen Frames enthalten. Die Annotationsdatenbank enthält die mit Frames annotierten Sätze aus dem BNC-Korpus und dem American Newswire Korpus. Momentan enthält die Datenbank 456 Frames mit über 7.300 lexikalischen Einheiten. Etwa 130.000 Sätze wurden seit Beginn des Projekts annotiert.

Abbildung 3 zeigt den Teilausschnitt der lexikalischen Datenbank. *WordForm* bezeichnet hier eine grammatische Ausprägung eines Lexems, damit wird jedoch nicht nach grammatischen Funktionen unterschieden. Ein oder mehrere

WordForms sind mit einem *Lexeme*, der abstrakten lexikalischen Einheit, verknüpft. *Lexeme* umfasst lediglich lexikalische Einheiten, die keine Mehrwortlexeme, aber ggf. Bestandteile davon sind. Ein *Lemma* hingegen besteht aus einem oder mehreren Lexemen, die über *LexemeEntry* miteinander verknüpft sind. Häufig hat ein *Lemma* nur eine Verknüpfung zu genau einem *LexemeEntry*, wenn es sich nicht um einen Mehrwortausdruck wie *take off*, oder komplexe Eigennamen handelt. Für *LexemeEntry* ist vermerkt, wo es sich innerhalb des *Lemmas* befindet, ob es sich um das Head handelt und ob hier eine Trennung der Lemma-Teile vorliegen kann. Jedes *Lemma* besitzt eine Wortart (*PartOfSpeech*), ebenso wie jedes *Lexeme*. Jedes *Lemma* ist mit einem oder mehreren Einträgen in *LexUnit* verknüpft. *Lexical Units* (vgl. *lexikalische Einheiten* im Abschnitt 2) sind Verknüpfungen zwischen Lemmata und Frames.

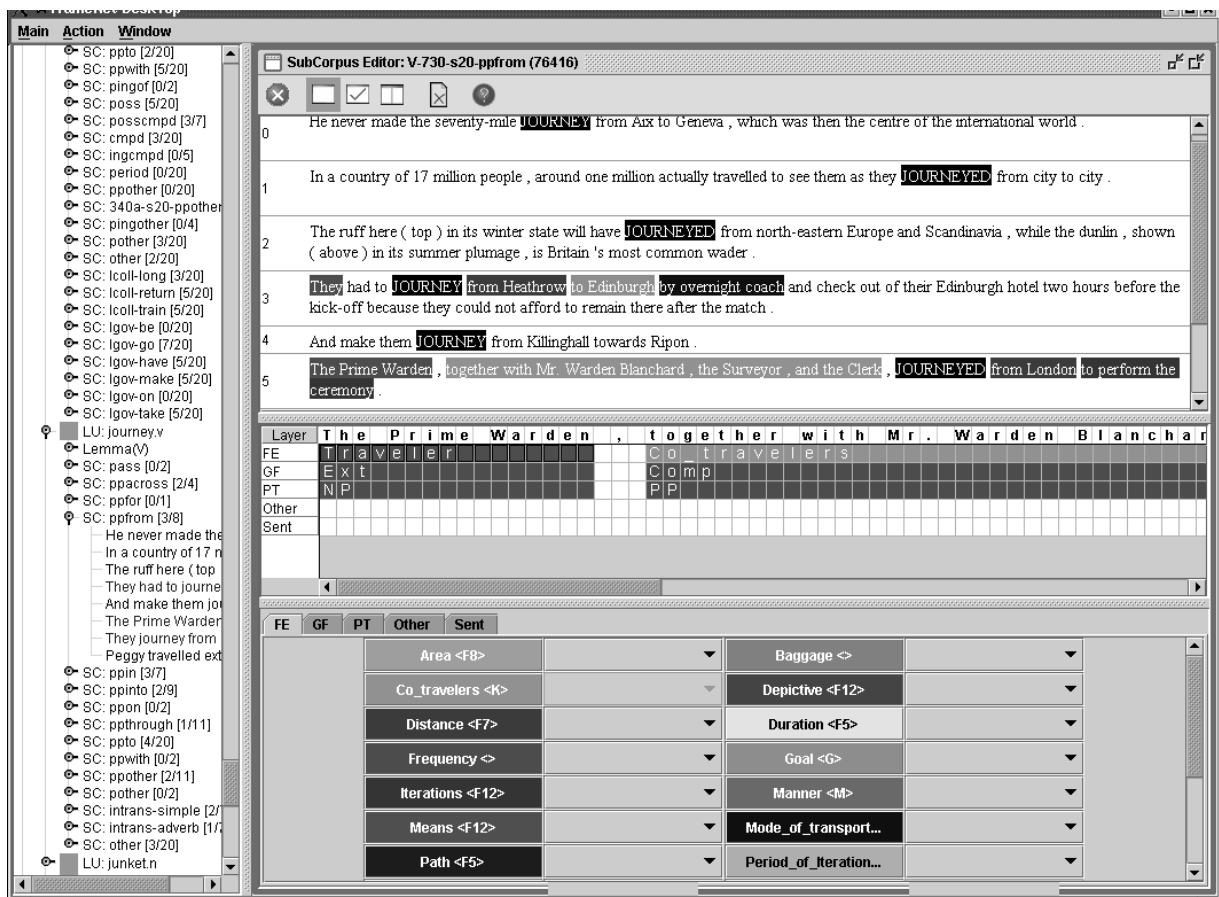


Abbildung 2: Ausschnitt aus dem FrameNet Desktop

Die Entität *Status* gibt den Bearbeitungszustand der lexikalischen Einheit an, kann aber andere Informationen enthalten, zum Beispiel, wenn die lexikalische Einheit sich analog zu anderen lexikalischen Einheiten verhält, wie es bei den Wochentagen der Fall ist. Jede lexikalische Einheit ist mit höchstens einem Frame verknüpft, umgekehrt besteht eine 1:n-Relation.

Beziehungen von Frames zu Frames und Frame-Elementen zu Frame-Elementen, wie zum Beispiel bei der Frame-Hierarchie, werden in den Tabellen *FrameRelation* und *FERelation* erfasst. Hier wird festgelegt, welche Frames bzw. Frame-Elemente in Frames in Relation zueinander stehen. Die verschiedenen Relationen befinden sich in *RelationType*.

Frame-Relationen können auch Metabeziehungen zueinander besitzen. Momentan betrifft dies die Beziehung des *Subframes* und die Metare-

lationstypen sind *Gleichzeitigkeit* und *Vorzeitigkeit*.

Die Entität *SemanticType* enthält zusätzliche, relevante Information, zum Beispiel zur Konnotation (*knickerig* vs. *sparsam* : negativ vs. positiv), ist aber im FrameNet-Projekt nicht systematisch erfasst. Zwischen diesen *SemanticTypes* können hierarchische *STInherit*-Beziehungen bestehen, zum Beispiel gehört eine negative Beurteilung wie *knickerig* generell zu affektiven Beurteilungen.

3.4 Andere FrameNet-Projekte

Neben einem japanischen und einem spanischen FrameNet-Projekt (vgl. SUBIRATS & PETRUCK 2003; <http://gemini.uab.es/SFN/>), wird ein deutsches FrameNet-Projekt (SALSA) unter der Leitung von Manfred Pinkal in Saarbrücken durchgeführt (vgl. ERK ET AL. 2003). Hier soll das 1,5 Millionen Token umfassende deutsche TIGER-Korpus mit Frames annotiert werden, wobei al-

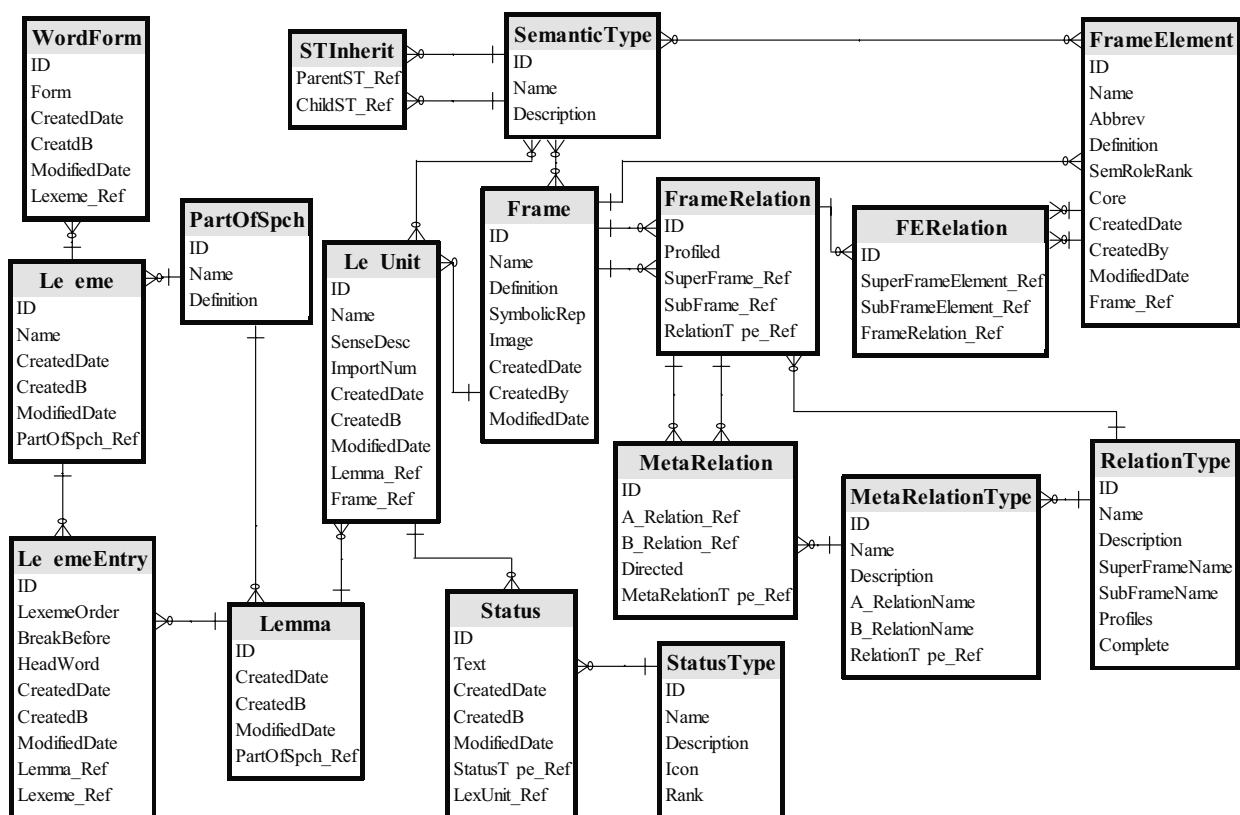


Abb. 3: Entity-Relationship-Diagramm der FrameNet-Datenbank (lexikal. Teil)

lerdings nur ein beschränktes Inventar an Frames und Frame-Elementen verwendet wird, das auf den 456 Frames des englischen FrameNet-Projekts basiert. Das bedeutet, dass die meisten Verben mit einem Tag annotiert werden, der die Existenz eines Frames und seiner Frame-Elemente lediglich markiert, diese Einheiten jedoch nicht definiert („UNKNOWN“). Informationen wie Relationen und Vererbung zwischen Frames werden nicht erfasst. Durch statistische Verfahren soll in einem späteren Schritt eine Klassifikation von Frames ermittelt werden.

4 WordNet, GermaNet, EuroWordNet

WordNet (BECKWITH ET AL. 1991; FELLBAUM 1998; MILLER 1998; <http://www.cogsci.princeton.edu/~wn/>) hat den Anspruch, eine aufgrund psycholinguistischer Evidenzen aufgebaute lexikalische Datenbank zu sein. Der Ansatz von Collins u. Quillian (COLLINS & QUILLIAN 1969), auf den mehrfach Bezug genommen wird, ist allerdings schon früh und zu Recht kritisiert worden (CONRAD 1972).

In dieser lexikalischen Datenstruktur werden hauptsächlich semantische Relationen zwischen Konzepten beschrieben, wie Synonymie, Antonymie, Hyponymie (IS-A-Relation), Hyperonymie, Meronymie und Holonymie bei Nomina. Lexikalische Einheiten mit der gleichen Bedeutung werden in so genannten *synsets* klassifiziert.

GermaNet (vgl. HAMP & FELDWEG 1997; KUNZE & LEMNITZER 2002; <http://www.sfs.uni-tuebingen.de/lsd/Intro.html>) und EuroWordNet (vgl. VOSSEN 1997) sind Parallelprojekte. EuroWordNet umfasst acht europäische Sprachen, die miteinander durch den *Inter-Lingual-Index* verknüpft sind, und integriert auch Module des GermaNet. Während EuroWordNet 1999 abgeschlossen wurde, wird das GermaNet-Projekt fortgesetzt. Hier, wie auch in EuroWordNet, sind die lexikalischen Informationen zum Teil anders strukturiert als im WordNet-Projekt. Zum Beispiel ist bei GermaNet die Struktur

der Adjektive ähnlich der Hyponymierelation der Nomina konzipiert. GermaNet enthält momentan mehr als 42.000 Synsets mit über 61.000 Lesarten.

Miller selbst (MILLER 1998) kritisiert den Mangel an Ausnahmeregeln für die Vererbung von Eigenschaften in der Begriffshierarchie, vor allem aber sind syntagmatische Zusammenhänge kaum repräsentiert; dies bezeichnet er als *tennis problem*. Ball, Schläger, Tennisplatz, Schiedsrichter etc. sind in unterschiedlichen Bereichen zu finden, es gibt keine markierten Zusammenhänge in der Datenbasis: „A framelike semantics is precluded by WordNet’s strict separation of its entries according to their syntactic category membership [...].“ (FELLBAUM 1998: 5).

GermaNet gibt für seine ca. 9.000 Verben Subkategorisierungsrahmen an (vgl. HAMP & FELDWEG 1997; KUNZE 1999; KUNZE 2003). Auch das Konzept des Entailments wird verwendet. Die Kausalrelationen werden auch dann erfasst, wenn sie zwischen unterschiedlichen Wortarten bestehen (vgl. VOSSEN 1997). Ebenso werden einige Verbalalternationen erfasst (vgl. KUNZE 1999).

Trotz der beschriebenen Erweiterungen bietet auch GermaNet keine Informationen über Verbindungen der Argumentstrukturen von Lexemen, die durch semantische Relationen verbunden sind. Die Stärke des WordNet-Ansatzes liegt hingegen auf der Seite der Nomina und ihrer Relationen und auch bei der vorhandenen Datenmenge.

5 Entwurf einer Datenbankstruktur für die Verknüpfung von WordNet- und FrameNet-Informationen

Im Folgenden wird dargestellt, wie FrameNet- und WordNet-Daten miteinander verknüpft werden können. Dabei wird besonders berücksichtigt, welche Anforderungen an eine Datenbankstruktur für deutsche Daten erfüllt werden müssen. Außerdem wird davon ausgegangen, dass

in einer solchen Datenbank nicht nur Informationen aus zwei, sondern aus mehreren Netzen oder sonstigen Datenquellen verknüpft werden können. Auch mehrere Korpora können mit den Daten zusammenhängen.

Abbildung 4 stellt die Struktur in Form eines Entity-Relationship-Diagramms dar.

Der erste Unterschied zur FrameNet-Datenbank besteht bereits darin, dass hier nicht von der Entität *WordForm* ausgegangen wird. Stattdessen wird mit *Flexform* eine Ausprägung eines *Lexems* bezeichnet, die mit grammatikalischer Information verknüpft ist, zum Beispiel gibt es drei verschiedene Einträge zu der Wortform *spielen*, die in der Tabelle *WordformTypes* enthalten ist.

Die Entität *WordformTypes* entspricht also der Entität *WordForm* bei FrameNet. Mit der Entität *Flexform* sind über *MorphDescr* zahlreiche Tabellen zur Flexion verknüpft, die hier nicht weiter behandelt werden. Über die Entität *Token* kann auf alle *Flexform*-Tokens in einem spezifischen Korpus zugegriffen werden. In *Tokeninfo* gibt es zu jedem Token im Korpus zusätzliche Informationen, wie die Angabe zur Großschreibung von Verben, Markierung in Fett- oder Kursivdruck etc. Die Verbindung zwischen *Token* und *Corpus* erfolgt mittelbar über *Sentence*, *Paragraph* und *Document*, falls Annotationen in der Datenbank enthalten sind, oder zumindest unmittelbar für andere Korpora, etwa wenn es sich lediglich um Listen von Token handelt. Für den Fall, dass in einem Korpus die morphologischen Ausprägungen der *Flexform* nicht bekannt sind, besteht zumindest eine Verbindung zwischen der Entität *WordformType* und *Sentence* bzw. *Corpus*. Da mehrere Korpora untersucht werden, besteht jeweils eine 1:n-Relation von *Corpus* zu *Token* und *WordformType*.

Ein oder mehrere Flexformen sind mit einem *Lexeme*, der abstrakten lexikalischen Einheit, verknüpft.

Die Relationen zwischen *Lexeme*, *Lemma*, *LexemEntry*, *PartofSpeech* und *LexUnit* sind identisch mit denen der FrameNet-Datenbank.

Die Tabelle *Flexinfos* ist mit weiteren Datenbanktabellen verknüpft, zum Beispiel mit Informationen zu Flexionsparadigma, Umlautung und Alternanzstamm für Nomina. In der Tabelle *alternat_orthography* sind Rechtschreibvarianten und Rechtschreibfehler vermerkt. So können Korpora, die aus der Zeit vor der und nach der Rechtschreibreform stammen, gemeinsam erfasst werden.

Die Struktur zwischen *Flexform*, *MorphemeEntry* und *Morpheme* ist analog zu der zwischen *Lexeme*, *LexemEntry* und *Lemma*. Es ist wichtig, Morpheme zu erfassen, um Frame evozierende sprachliche Einheiten, die unterhalb der Wortebene liegen, berücksichtigen zu können. Ein Beispiel hierfür ist das Kompositum *Gagenforderung*, bei dem das Determinans ein Frame-Element zum Frame *request* ist, der vom Determinandum *Forderung* evoziert wird (vgl. ERK, KOWALSKI & PINKAL 2003). Frame-Semantik auf der morphologischen Ebene kann aber auch für andere Sprachen als das Deutsche betrieben werden. Petruck und Boas (PETRUCK & BOAS 2003) zeigen anhand der hebräischen, englischen und deutschen Bezeichnungen für die Wochentage, wie unikale Morpheme anhand des Frames, vor dessen Hintergrund sie stehen, interpretiert werden können.

Die Entität *LexUnit* verknüpft in diesem Entwurf nicht nur Lemmata mit Frames, sondern auch mit anderen semantischen Informationen, unter anderem mit Synsets.

Über *LexUnit* werden Informationen aus GermaNet - im Diagramm nicht vollständig dargestellt - mit Informationen zu Frames verknüpft. Dabei kann es auch Einheiten geben, die keine Informationen zu Frames bzw. zu GermaNet besitzen.

Die lexikalischen Einheiten in *LexUnit* sind ferner mit grammatischen Funktionen, Subkat-

FrameNet und WordNet

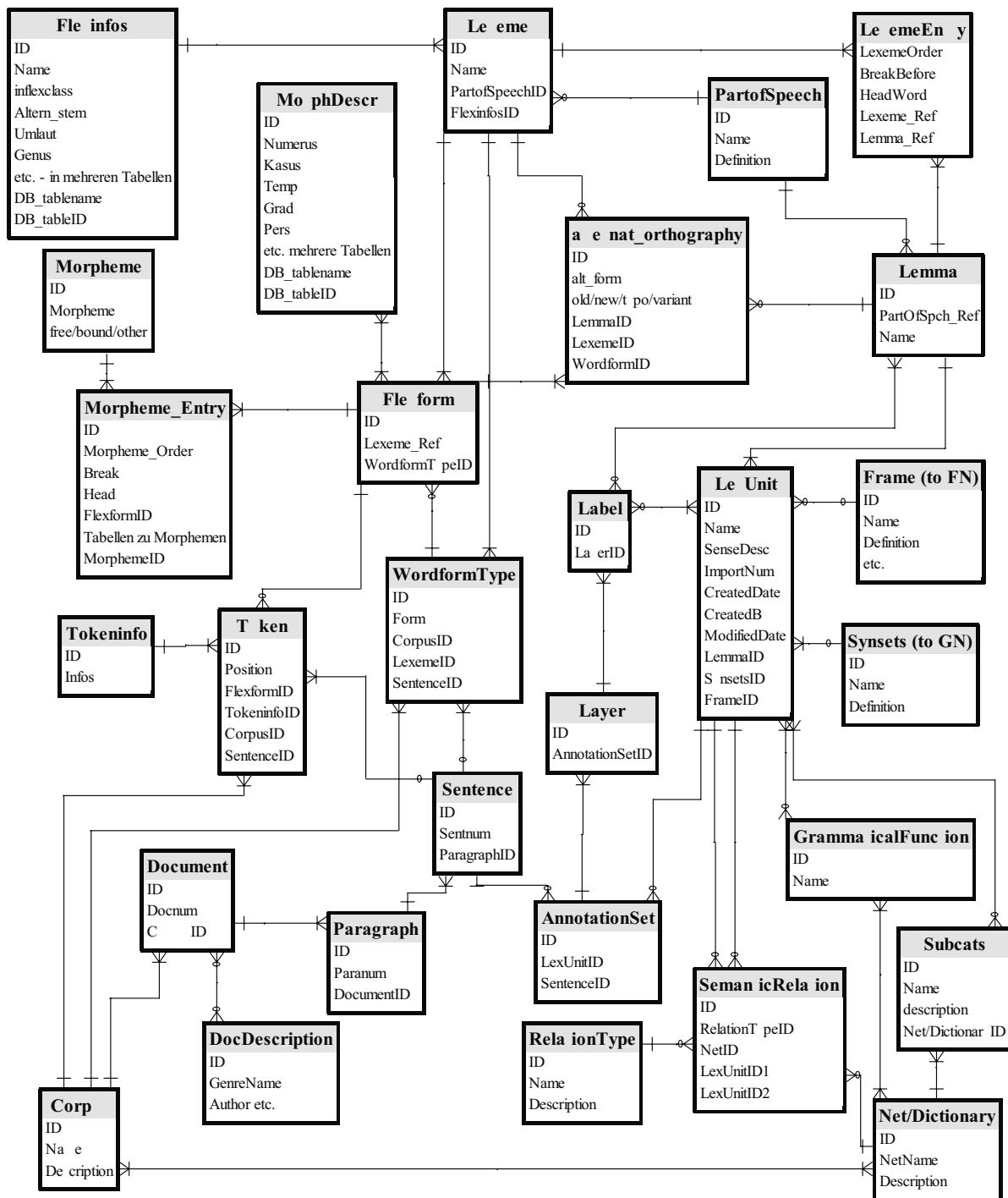


Abbildung 4: Die Verknüpfung von WordNet- und FrameNet-Daten

Informationen und semantischen Relationen verknüpft. Diese Angaben beziehen sich auf jeweils ein spezifisches Netz, Lexikon oder eine andere Datenquelle, so dass hier n:m-Beziehungen bestehen können.

Jede semantische Relation ist mit zwei lexikalischen Einheiten und einem Eintrag in der Tabelle *RelationType* verknüpft. Diese enthält alle definierten semantischen Relationen.

Die eben aufgelisteten semantischen Informationen können aus GermaNet oder einem anderen semantischen Netz stammen.

Der Entwurf enthält jedoch noch zwei weitere Verknüpfungen zwischen GermaNet und FrameNet-Daten: In den mit Frames annotierten Sätzen sind, ggf. auf mehreren Ebenen (*Layer*), die einzelnen Realisierungen der Frame-Elemente (*Label*) enthalten. Zum Beispiel ist im Satz *Peter kauft ein Buch*. *Peter* die Instanziierung des Frame-Elements *Buyer*. Diese Label können wiederum (hier vereinfacht dargestellt) mit Lemmata verknüpft werden. So entstehen auf einfache Art und Weise Verbindungen zwischen Frame-Elementen und weiteren semantischen Informationen, wie zu semantischen Relationen aus WordNet bzw. GermaNet oder aber zu weiteren Frames. Diese Verknüpfung liefert jedoch keine Information darüber, um welche lexikalische Einheit zu dem jeweiligen (ambigen) Lemma es sich handelt. Die Zuordnung zwischen *Label* und *LexUnit* ist vielmehr der Kern der Verknüpfung zwischen den komplementären lexikalisch-semantischen Netzen.

6 Zusammenfassung und Ausblick

Der dargestellte Datenbankentwurf verknüpft Informationen aus GermaNet oder ähnlichen lexikalischen Datenbanken und FrameNet. Dabei ist auch die Hinzunahme weiterer Informationen aus anderen Netzen oder Lexika möglich. Den Spezifika deutscher Sprachdaten wird mit Entitäten zu morphologischen und graphematischen Informationen Rechnung getragen.

FrameNet und WordNet-Projekte sind in erster Linie lexikographisch orientiert, auch wenn sie Korpora für Definitionen und Beispiele verwenden. Die annotierten Sätze geben daher kein realistisches Bild der tatsächlichen sprachlichen Situation wieder. Dies könnte sich mit Projekten wie dem SALSA-Projekt (s.o., ERK ET AL. 2003) ändern. Dann wäre es sinnvoll, auch die Datenbankstruktur entsprechend zu erweitern, zum

Beispiel mit grundlegenden Informationen zu Häufigkeiten von Frames und Frame-Elementen in verschiedenen Korpora. Für den Fall, dass genügend Datenmaterial vorliegt, ist auch die Generierung von (statistischen) Zusammenhängen zwischen Frame-Elementen und Synsets denkbar.

Danksagungen

Ich danke Charles J. Fillmore und dem Team des FrameNet-Projekts, insbesondere Michael J. Ellsworth, Josef Ruppenhofer, Collin F. Baker und Miriam R. L. Petruck, für ihre Unterstützung. Claudia Kunze und Lothar Lemnitzer danke ich für hilfreiche Anregungen und stimulierende Fragen sowie für Informationen über GermaNet.

Die Arbeit wurde mit Unterstützung eines Stipendiums im Rahmen des Postdoc-Programms des Deutschen Akademischen Austauschdienstes (DAAD) ermöglicht.

Anmerkungen

- ¹ In anderen Ansätzen werden derartige zusammenhängenden Konzepte auch als Schemata, Skripte (SCHANK & ABELSON 1977) oder Szenarien bezeichnet. Frame ist ein Oberbegriff für diese Begriffe. Zum Beispiel bezeichnet Skript eine Abfolge von Ereignissen und ist damit eine spezielle Art von Frame.
- ² [*Anmerkung der Redaktion*] Die im Original des Artikels aus dem FrameNet-Desktop (vgl. Abbildung 2) hergeleiteten farbigen Hinterlegungen der Relationstypen in den Beispielen lassen sich drucktechnisch nicht wiedergeben. Statt der farbigen Hinterlegung werden die Begriffe bzw. Mehrwortgruppen durchgehend unterstrichen. Die Relationstypen werden durch tiefgestellte Indices nach folgendem Schlüssel verwendet:

Relation	Index
Traveler	1
Source	2
Goal	3
Path	4

Area	5
Vehicle	6
Co_Travelers	7
Baggage	8
Duration	9
Distance	10
Purpose	11
Motion Verb	12

- ³ ATKINS, RUNDELL & SATO (2003) allerdings verwenden FrameNet für die Analyse eines polysemen Lexems.
- ⁴ Es handelt sich hierbei eigentlich um Satzsammlungen.

Literatur

- ATKINS, B.T.S.; RUNDELL, M.; SATO, H. (2003). "The Contribution of FrameNet to Practical Lexicography." In: *International Journal of Lexicography* 16(3) (2003), 333-357.
- BAKER, C. F.; FILLMORE, C. F. & CRONIN, B. (2003). "The Structure of the FrameNet Database." In: *International Journal of Lexicography* 16(3) (2003), 281-296.
- BECKWITH, R. ET AL. (1991). "WordNet: A Lexical Database Organized on Psycholinguistic Principles." In: ZERNIK, U. (Hrsg.) (1991). *Lexical acquisition. Exploiting On-line Resources to Build a Lexicon*. Hillsdale, NJ: Lawrence Erlbaum, 211-232.
- COLLINS, A. M.; QUILLIAN, M. R. (1969). "Retrieval Time from Semantic Memory." In: *Journal of Verbal Learning and Verbal Behavior* 8 (1969), 240-247.
- CONRAD, C. (1972). "Cognitive Economy in Semantic Memory." In: *Journal of Experimental Psychology* 92(2) (1972), 149-154.
- ERK, K.; KOWALSKI, A.; PINKAL, M. (2003). "A Corpus Resource for Lexical Semantics." In: *Proceedings of The 5th International Workshop on Computational Semantics (IWCS5)*. Tilburg, The Netherlands, January 2003, <http://www.coli.uni-sb.de/~erk/OnlinePapers/LexProj.ps> [accessed April 2004].
- ERK, K. ET AL. (2003). "Towards a Resource for Lexical Semantics: A Large German Corpus with Extensive Semantic Annotation." In: HINRICHS, E. ROTH, D. (eds.) (2003). *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 537-544.
- FELLBAUM, CH. (1998). "A Semantic Network of English: The Mother of all WordNets." In: *Computers and the Humanities* 32 (1998), 209-220.
- FILLMORE, CH. J. (1968). "The Case for Case." In: BACH, E.; HARMS, R. T. (Hrsg.) (1968). *Universals in Linguistic Theory*. New York: Holt, Rinehard, and Winston, 1-88.
- FILLMORE, CH. J. (1977). "The Case for Case Reopened." In: COLE, P.; SADOCK, J. M. (Hrsg.) (1977). *Syntax and semantics 8: Grammatical relations*. New York: Academic Press, 59-81.
- FILLMORE, CH. J. (1982). "Frame Semantics." In: *The Linguistic Society of Korea (Hrsg.) (1982). Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Company, 111-137.
- FILLMORE, CH. J. (1985). "Frames and the Semantics of Understanding." In: *Quaderni di Semantica* 6(2) (1985), 222-254.
- FILLMORE, CH. J.; ATKINS, B.T.S. (1992). "Towards a Frame-based Organization of the Lexicon: The Semantics of RISK and its Neighbors." In: LEHRER, A.; KITTAI, E. (Hrsg.) (1992). *Frames, Fields and Contrast: New Essays in Semantics and Lexical Organization*. Hillsdale, NJ: Lawrence Erlbaum, 75-102.
- FILLMORE, CH. J.; ATKINS, B.T.S. (1994). "Starting where the Dictionaries Stop: The Challenge for Computational Lexicography." In: ATKINS, B.T.S.; ZAMPOLLI, A. (Hrsg.) (1994). *Computational Approaches to the Lexicon*. Oxford: Oxford University Press, 350-393.
- FILLMORE, CH. J.; JOHNSON, C. R.; PETRUCK, M.R.L. (2003). "Background to FrameNet." In: *International Journal of Lexicography* 16(3) (2003), 235-250.
- FILLMORE, CH. J. ET AL. (2003). "FrameNet in Action: The Case of Attaching." In: *International Journal of Lexicography* 16(3) (2003), 297-332.

- GECKELER, H. (1971). Strukturelle Semantik und Wortfeldtheorie. München: Fink.
- GIPPER, H. (1973). „Das sprachliche Feld.“ In: GREBE, P. ET AL. (1973³). Duden – Grammatik der deutschen Gegenwartssprache. Mannheim: Bibliographisches Institut Dudenverlag [= Der große Duden, Bd. 4].
- GIPPER, H. (1995). „Jost Trier und das sprachliche Feld. Was bleibt?“ In: Zeitschrift für germanistische Linguistik 23(3) (1995), 326-341.
- HAMP, B.; FELDWEG, H. (1997). „GermaNet - a Lexical-Semantic Net for German.“ In: VOSSEN, P. ET AL. (Hrsg.) (1997). Proceedings of the ACL / EACL-97 Workshop on Automatic Information Extraction and Building of Lexical-Semantic Resources for NLP Applications, 9-15.
- KUNZE, C. (1999). „Semantics of Verbs within GermaNet and EuroWordNet.“ In: KORDONI, V. (ed.) Proceedings of the ESSLI-99 Workshop on Lexical Semantics and Linking in Constraint-Based Theories, 189-200, <http://www.folli.uva.nl/CD/1999/library/coursematerial/Kordonikunze.pdf>, <http://citeseer.nj.nec.com/kunze99semantics.html> [accessed April 2004].
- KUNZE, C. (2003). „Verbsemantik in GermaNet – eine Exploration.“ In: CYRUS, L. ET AL. (Hrsg.) Sprache zwischen Theorie und Technologie. Festschrift für Wolf Paprotté zum 60. Geburtstag. Language between theory and technology. Studies in honour of Wolf Paprotté on occasion of his 60th birthday. Wiesbaden: Deutscher Universitätsverlag, 113-122.
- KUNZE, C.; LEMNITZER, L. (2002a). „GermaNet - Representation, Visualization, Application.“ In: Proceedings LREC 2002. 3rd International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain, May/June 2002, 1485-1491.
- MILLER, G. A. (1998). „Nouns in WordNet.“ In: FELLBAUM, CH. (ed.) (1998). WordNet: An Electronic Lexical Database. Cambridge, MA / London: MIT Press, 23-46.
- PETRUCK, M.R.L.; BOAS, H. (2003). „All in a Day's Week.“ Presentation at Workshop on Frame Semantics, International Congress of Linguistics. Prague, July 2003, <http://framenet.icsi.berkeley.edu/~framenet/papers/weekday.pdf> [accessed April 2004].
- SCHANK, R. C.; ABELSON, R. R. (1977). Scripts, Plans, Goals and Understanding. Hillsdale, NJ: Lawrence Erlbaum.
- SCHULTE IM WALDE, S. (1998). Automatic Semantic Classification of Verbs According to their Alternation Behaviour. Diplomarbeit. Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung (IMS), <http://citeseer.ist.psu.edu/im98automatic.html> [Zugriff April 2004].
- SCHUMACHER, F.; STEINER, P. (2002). „Aspekte der Bedeutung: Semantik.“ In: MÜLLER, H. M. (Hrsg.) (2002). Arbeitsbuch Linguistik. Paderborn: Schöningh [= UTB 2169], 170-198.
- SUBIRATS, C.; PETRUCK, M.R.L. (2003). „Surprise: Spanish FrameNet!“ In: Proceedings of XVII International Congress of Linguistics, Prague, July 2003, 24-29.
- TRIER, J. (1931). Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes. Heidelberg: Winter.
- WEISGERBER, J. L. (1962). Grundzüge der inhaltsbezogenen Grammatik. Düsseldorf: Schwann.
- ZILLIG, W. (HRSG.) (1994). Jost Trier. Leben – Werk – Wirkung. Münster: Aa Verlag.