



Wörterbuchstrukturen II

Claudia Kunze

kunze8@gmail.com, kunze@c1.uni-heidelberg.de

Computational Lexicography

Themen

- Was sind Wörterbuchstrukturen?
- Analyse von Wörterbuchstrukturen
- Parsing von Wörterbuchartikeln
- Kodierung von Wörterbuchartikelstrukturen
- Standardisierung von Wörterbuchartikelstrukturen

In dieser Einheit wird es vor allem um die Kodierung von WB-Artikelstrukturen gehen.

Kodierung von Wörterbuchartikelstrukturen

Auszeichnung von Dokumenten und Texten mittels so genannter MARKUP-SPRACHEN

- verschiedene Dokumente vom selben Typ können einheitlich verarbeitet werden, z.B. auch Wörterbuchartikel
- Programme können auf die Inhaltssegmente der mit markup versehenen Dokumente zugreifen
- die in lexikalischen Ressourcen enthaltenen Inhaltssegmente sind einheitlich zu kennzeichnen
- Datenfelder der Einträge einer lexikalischen Ressource sollten inhaltsbezogen und damit nachvollziehbar bezeichnet werden

Wörterbuchstrukturen - Einleitung

- Auswahl, Anordnung und Umfang der lexikalischen Elemente in WB-Artikeln variieren je nach WB-Typ und Anwendungsszenario (s. Engelberg & Lemnitzer (2001))
- häufig “pragmatische” Kriterien für Erstellung von Heuristiken, daher zweifelhafter Status der Lexikographie als Wissenschaft - eher: wissenschaftliche Praxis
- aber: Quine (1981) lobt “Logik des Wörterbuchs” gegenüber “Logik der Metasprache”
- ab den 70er Jahren: Betrachtung von Wörterbüchern als Gegenstand wissenschaftlicher Forschung und Prägung eines Fachvokabulars (Rey-Debove, Wiegand)
- Wörterbuchkonzeption für Print-WB und elektronische WB bedarf eines Lexikonmodells (Entscheidung über Auswahl und Modellierung der Elemente)

Makrostruktur

Unter der **Makrostruktur** eines Wörterbuchs verstehen wir die geordnete Menge seiner Lemmata. Das Lemma als Epiphänomen vermittelt zwischen Makro- und Mikrostruktur. (enge Definition ohne Berücksichtigung der so genannten Umtexte wie Vorwort, Benutzungshinweise, WB-Grammatik etc.)

- Lemmatisierung: Zurückführung von Textwörtern auf die kanonische Form (Grundform, Zitierform)
- Nomina: Nominativ singular, Verben: Infinitiv, Adjektive: prädikative Form
- Lemmabegriff: Darstellung mehrdeutiger Wörter als Homonyme (verschiedene Einträge) oder Polyseme (innerhalb eines Eintrags)

Makrostruktur

Probleme bei der Lemmatisierung

- fehlende Grundform bei attributiven Adjektiven wie **letzt**, unterschiedliche Varianten der Repräsentation: *letzt*, *letzte(r,s)*, *letzt-*
- ebenso bei regelmäßigen Genusvarianten wie bei Berufsbezeichnungen (*Professor*, *Professorin*), substantivierten Adjektiven (*Krank*, *Kranke*) oder Partizipien (*Verurteilter*, *Verurteilte*)
- lautliche Varianten wie *ad(e)lig*, die in einem Lemma zusammengefasst oder als zwei verschiedene Lemmata repräsentiert aufeinander bezogen werden können

Makrostruktur

Probleme bei der Lemmatisierung

- orthographische Varianten (*Potenzial, Potential* und Zahlwortvarianten (*zehnfach, 10-fach*)
- Zitierform und Umfang von Mehrwortlexemen : *einen Bärendienst erweisen, jemandem einen Bärendienst erweisen, jemandem einen (echten, richtigen, wahrhaftigen) Bärendienst erweisen*, auch die Frage, unter welchem Stichwort das MWL zu verzeichnen ist, z.B. bei *Bock schießen* unter *Bock* oder *schießen*

Makrostruktur

Die Makrostruktur stellt eine elementare Zugriffsstruktur im Wörterbuch dar. Lemnitzer und Engelberg (2001, S. 125) unterscheiden:

- monoakzessive WB mit einer äußeren Zugriffsstruktur (nur ein Wörterverzeichnis und ein Suchpfad im WB);
- monoakzessive WB mit mehreren äußeren Zugriffsstrukturen (mehrere Wörterverzeichnisse mit unterschiedlichen Stichwörtern: z.B. Allgemeinwortschatz, Eigennamenverzeichnis);
- polyakzessive WB mit mehreren äußeren Zugriffsstrukturen: zwei Verzeichnisse (Wörterverzeichnis und Register) und zwei verschiedene Suchpfade (z.B. in Thesauri);

Makrostruktur

- polyakzessive WB mit einer äußeren Zugriffsstruktur: nur ein Verzeichnis, aber Verweise zwischen Lemmata, die wir in der alphabetischen Struktur des Wörterverzeichnisses weiterverfolgen müssen.

Alphabetische (u.a.) Sortierungen:

- nischenalphabetische Sortierung (strikt alphabetisch)
 - nestalphabetische Sortierung (Durchbrechen der alphabetischen Sortierung bei Komposita), z.B.: **Apfel[...]**; Apfelbaum, Apfelschnaps, Apfeltasche, Apfeltorte, [...]
- Apfelsine**

- finalalphabetisch: z.B. arabische WB
- andere Zugriffsstrukturen nötig für Zeichensprachen (Chinesisch) und Thesauri

Mikrostruktur

Die MIKROSTRUKTUR bezeichnet die hierarchische Binnengliederung eines Wörtereintrages zu einem gegebenen Lexikonstichwort oder Lemmazeichen als konkrete Analyse dieses Eintrags, aber auch als Abstraktion über Typen lexikalischer Einträge (abstrakte Mikrostruktur).

- Begriff *Mikrostruktur* wurde von Ray-Debove (1971) geprägt
- lineare Analyse von Textsegmenten im Wörterbuch
- Begriff des *linguistischen Informationsprogramms*
- Weiterentwicklung des Mikrostruktur-Ansatzes zu einem hierarchischen Modell (durch Wiegand)

Mikrostruktur

Zugriff auf Mikrostruktur eines Lemma meist Ziel der Wörterbuchrecherche: z.B. Information über Schreibung, Lautung, morphosyntaktische oder andere grammatische Charakteristika oder über Bedeutung und/oder Verwendung eines bestimmten lexikalischen Zeichens
Beispieleintrag *Pamphlet* aus dem DUW:

Pamph—**let**, das: -[e]s, -e [frz. Pamphlet< engl. Pamphlet = Broschüre, H. u.]
(bildungsspr. abwertend): *Streit- oder Schmähschrift*: ein politisches P., ein P. gegen jmdn schreiben, verfassen

Die folgenden Angaben sind enthalten:

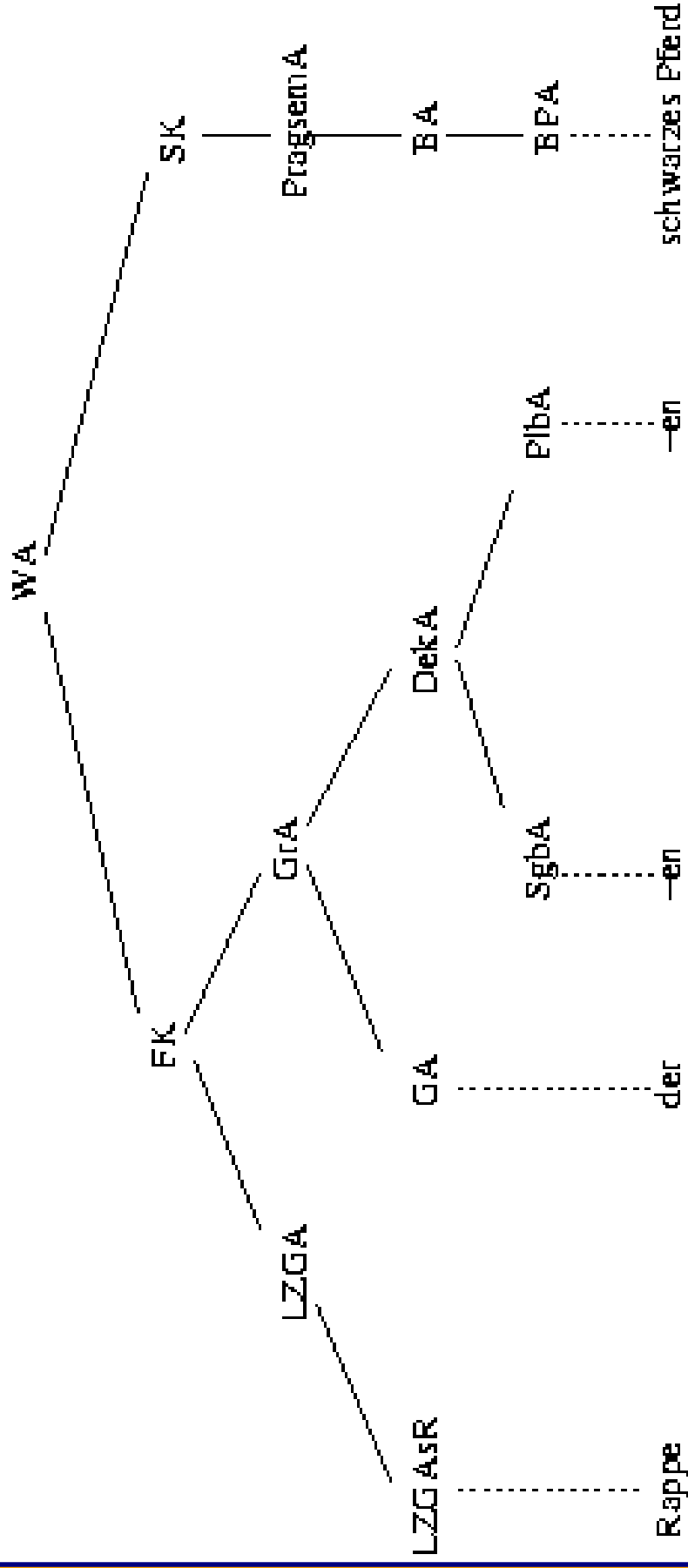
- *Pamphlet* besteht aus den Silben *Pamph* und *let*, mit Betonung auf der zweiten Silbe mit langem Vokal.
- *Pamphlet* bezeichnet ein Substantiv Neutrum Singular und kann den Genitiv Singular sowohl mit -es als auch -s sowie den Nominativ Plural mit -e bilden.
- Das Wort wurde aus dem Englischen ins Französische und dann aus dem Französischen entlehnt; die weitere Herkunft ist ungeklärt.
- Es wird im bildungssprachlichen Deutsch abwertend gebraucht und bedeutet *Kampfschrift* oder *Schmähschrift*.
- es wird wie in den folgenden Beispielphrasen verwendet: *ein politisches Pamphlet*, *ein Pamphlet gegen jemanden schreiben* und *ein Pamphlet gegen jemanden verfassen*.

Angabetypen bei Wiegand

Größere Klassen von Angabetypen: Formkommentar (FK), grammatische Angaben (GramA), semantischer Kommentar (SK), pragmatisch-semantische Angaben (PragSemA)

Konkrete Mikrostruktur:

(1) Rappe, der: -en, -en 'schwarzes Pferd'



Abstrakte Mikrostrukturen

Abstrakte Mikrostrukturen spezifizieren ein einheitliches Informationsprogramm für große Gruppen lexikalischer

Einträge:

- Angaben bestimmter Angabeklassen müssen in allen abstrakten Mikrostrukturen realisiert sein: obligatorische Angaben, z.B. die Lemmazeichengestaltangabe;
- Angaben spezifischer Angabeklassen sind nur für spezifische abstrakte Mikrostrukturen obligatorisch (z.B. Genusangaben für Substantive);
- Angaben einiger Angabeklassen sind für alle abstrakten Mikrostrukturen optional (etwa Beispielangabe oder Belegangabe);
- Angaben mancher Angabeklassen sind für einige abstrakte Mikrostrukturen optional, während sie für andere Strukturen nicht zulässig sind (Graduierungsangaben sind etwa nur bei Adjektiveinträgen möglich).

Wiegand unterscheidet zwischen einfachen integrierten Mikrostrukturen, einfachen, nicht-integrierten und erweiterten integrierten Mikrostrukturen, wobei Letztere dem Artikel nachgestellte Verweisstrukturen integrieren können.

Beschreibung der Mikrostrukturen in einem

Die vollständige Beschreibung der Mikrostrukturen in einem WB muss folgende Informationen enthalten:

1. alle funktionalen Textsegmente, die im Wörterbuch vorkommen;
2. für alle Angaben bestimmter Angabeklassen, ob sie notwendig, optional oder nicht zulässig sind;
3. alle Lemmzeichentypen;
4. Zuschreibung aller abstrakten Mikrostrukturen zu den jeweiligen Lemmzeichentypen.

Verdienstvolle Forschung zur Erfassung lexikographischer Artikelstrukturen als Wegbereiter der Computerlexikographie (s. Wiegand (1989) sowie Engelberg & Lemnitzer (2001:141ff.))

Verweisstruktur - Mesostruktur/Mediostruktur

Durch **Verweise** im WB wird der Wörterbuchbenutzer bei der Recherche zu einem Stichwort zu einem anderen Wörterbuchartikel gelenkt, den er zur Vervollständigung der Information nachschlagen soll. Ein Verweis setzt sich zusammen aus **Verweisursprung** und **Verweisziel**, die Repräsentation des Verweisziels am Verweisursprung und das **Verweissymbol** (meist ein Pfeilsymbol).

Verschiedene Arten von Verweisen:

- GRAMMATISCHE VERWEISE: z.B. zum Ausdruck einer Derivationsbeziehung zwischen Stichwort und Grundform (z.B. *buk* → *backen*)
- INHALTLICHE VERWEISE: realisieren z.B. paradigmatische Relationen zwischen Wörtern wie Synonymie und Antonymie
- POTENTIELLE VERWEISE: erfordern weiteres Nachschlagen (*nautisch: die Nautik betreffend*)
- ZIRKULÄRE VERWEISE: *kaufen* auf *erwerben* und *erwerben* auf *kaufen*. Dies ist eher unerwünscht.
- ANDERE VERWEISE: z.B. auf ein bedeutungstragendes Wort in einem Phrasem (*jmdn. über den* → *Leisten ziehen*)

Verweisstruktur - Mesostruktur/Mediostruktur

Zu beachten sind auch die folgenden Anmerkungen:

- Konsistenz ist wichtig! Verweise dürfen nicht auf einen Eintrag verweisen, den das WB gar nicht enthält
- grammatische Verweise sind medienspezifisch und können in elektronischen Wörterbüchern entfallen
- Dafydd Gibbon (2001) schlägt lexikalische Mesostruktur vor, die paradigmatische Relationen zwischen sinnrelational aufeinander bezogenen lexikalischen Einheiten expliziert
- hilfreich für Lexikonmodelle, die mit Objekthierarchien und Defaulthierarchien arbeiten, da generalisierende Elemente in wörterbuchumspannender Verweisstruktur zusammenfasst werden

Lexikalische Informationssysteme

lexiko (Institut für deutsche Sprache Mannheim)

- ein umfangreiches lexikalisches Informationsprogramm liegt zugrunde, aus dem verschiedene abstrakte Informationsprogramme (je nach Anwendungszweck) abgeleitet werden können
- lexikalische Informationen werden von Hand auf der Grundlage von Korpusdaten erstellt
- minimale Menge von Angaben werden für alle Stichwörter ausgefüllt (ca. 300 000 Stichwörter aus den IDS-Korpora)
- zu relativ wenigen Einträgen (einige 100) wurde das vollständige Informationsprogramm erstellt

Mehrschichtige Architektur von elexiko

- Datenspeicherungsebene: Verwendung des Datenbankmanagementsystems ORACLE
- konzeptuelle Ebene: abstrakte Mikrostrukturen werden durch DTDn beschrieben, die das Format von XML-annotierten Dokumenten (hier: Einträgen) festlegen
- Präsentationsebene: aus Elementen der Mikrostruktur wird konkrete Mikrostruktur ('maßgeschneidert') aufgebaut und ggf. modifiziert (durch Einblendenden oder Ausblendenden von Angaben)
- Hypertextualisierung macht elexiko Web-fähig: Angaben zu Schreibung und Worttrennung, Morphologie; syntaktische, semantische u. pragmatische Angaben
- Artikel vollständig mit Korpusbelegen versehen



Service

Benutzerhinweise

Lesartenübergreifende Angaben

Info

Orthografie

Normgerechte Schreibung:

Service

Worttrennung:

Ser|vice

Info

Wortbildung

keine Angabe

Nationale Verteilung

Info

Dieses Stichwort ist in Texten des folgenden Sprachraumes belegt:

BRD, Österreich, Schweiz

Herkunft und Wandel

Info

Für Informationen zu Herkunft und Wandel des Stichwortes klicken Sie bitte [hier](#).

Lesartenbezogene Angaben

Info

Das Stichwort verfügt über 6 Lesart(en).

[Lesart 'das Bedienen'](#)

[Lesart 'Geschirrsel'](#)

[Lesart 'Spielechnik'](#)

[Lesart 'Dienstleistungseinrichtung'](#)

[Lesart 'Personal'](#)

[Lesart 'Bereich'](#)

Hinweis:

Klicken Sie auf eine Lesart bzw. Spezifizierung, um zu weiteren Angaben zu kommen.

Konzeptfamilie und Homonymie

Info

Die Lesarten '[Dienstleistungseinrichtung](#)', '[Personal](#)' und '[Spielechnik](#)' sind Metonymisierungen der Lesart '[das Bedienen](#)'. Die Lesart '[Bereich](#)' wiederum ist eine Metonymisierung der Lesarten '[das Bedienen](#)' und '[Dienstleistungseinrichtung](#)'. Zur Lesart '[Geschirrsel](#)' kann durch die unterschiedlichen Herkunftssprachen gegenwärtig sprachlich im Deutschen kein semantischer Zusammenhang hergestellt werden.

Zusammenfassung elexiko

Wir halten insgesamt die folgenden Punkte für das lexikalische Hypertextsystem eLexiko fest:

- Aufgabe fixer Mikrostrukturen für Lemmata zugunsten eines modularen Systems zur Erzeugung dynamischer Mikrostrukturen für Lesarten, das auf adressatenunabhängiges lexikographisches Vollprogramm zugreift
- lineare Struktur eines traditionellen Wörterbuchs wird durch eine hypertextuelle, verlinkte Struktur ersetzt: mehr Eigeninitiative und Auswahl bei der Informationsrecherche
- Repräsentation der Mikrostrukturprogramme in einer Reihe von DTDn gestattet datenbankunabhängige Modellierung und damit für die Nachhaltigkeit der im Projekt aufgebauten Wissensstrukturen
- Verzicht auf eine konkrete Ausrichtung, die für traditionelle Wörterbücher unerlässlich ist, auch Aufgabe des traditionellen Term des Lemmas, Lexems oder Lexikoneintrags: elementare Einheit ist nun die *Lesart* oder *Lexical unit* bzw. *Semantic unit* (PAROLE)

Die Navigationsstrukturen als Teil des Informationssystems sind wahrscheinlich eher der Makrostruktur als der Mikrostruktur zuzurechnen, müssen noch auf Benutzerfreundlichkeit und Nützlichkeit hin evaluiert werden, z.B. die Mächtigkeit und die Funktionen der Suchmaschine, die den Zugriff auf einzelne Artikel ermöglicht.

DWDS - Digitales WB der deutschen Sprach

Digitalisierte Version des *Wörterbuchs der deutschen Gegenwartssprache* (WDG)

- Es wird kein eigenständiges Informationsprogramm entwickelt. Das Informationsprogramm des WDG wird lediglich um folgende Angabetypen erweitert:
 - Belegbeispielangaben zum Stichwort (nicht lesartenbezogen);
 - Angaben zu Kollokanten zum Stichwort;
 - Angaben zu lexikalisch-semantischen Relationen.
- Die Angaben im WDG sind z.T. veraltet. So wird die lexikalische Einheit *Service* – im Sinne von /Dienst, Bedienung/ – als „Neulexem“ bezeichnet. Neuere Wörter wie *Flyer* oder *Engine* fehlen.

DWDS-Eintrag Service aus dem WDG

Wörterbuchartikel aus dem WDG

Über das WDG

vollansicht

- **Servi'ce**¹, das; -/-s, - [.. wi. β], Pl. [.. wi. β, .. wi. βe] /*franz./* *zusammengehörendes, einen Satz bildendes Geschirr*; ein einfaches, kostbares, billiges, teures, ergänzungsfähiges S.; ein S. für zwölf Personen; das S. hat ein modernes Dekor; *umg.* von dem guten S. ist eine Tasse, ein Glas zerbrochen
- dazu Frühstücks-, ► Hotel-, ► Kaffee-, ► Likör-, ► Porzellan-, ► Saft-, ► Silber-, ► Speise-, ► Tafel-, ► Teeservice
- **Se'rvic'e**², der; -s, - /Pl. *ungebräuchl./* [sö . (r)wiβ], Pl. [sö . (r)wiβe] /*engl./* *Neuwort*
1. Kundendienst diese Auto-, Fernsehwerkstatt hat, bietet einen ausgezeichneten, vorbildlichen, erstklassigen S.; der S. an der Tankstelle läßt zu wünschen übrig; fachmännische Beratung gehört zum S. dieses Geschäfts; *Betreuung, Bedienung der Gäste*: der S. für die Hotelgäste, Urlauber wurde verbessert, erweitert; der S. des Hotels, auf dem Schiff, an Bord ist gut; diese Fluggesellschaft ist führend im S.
2. *Sitz, Einrichtung von 1*: wo ist der S. ?; im S. werden täglich 50 Autos betreut; in dem neuen S. arbeiten drei Kolleginnen
- zu 1 Zimmerservice ► zu 1 u. 2 Auto-, ► Blumen-, ► Fernseh-, ► Hotel-, ► Messe-, ► Rundfunkservice

Lexikalisch-semantische Relationen zu Serv

Automatisch berechnete semantische
Relationen aus dem WDG ^{Hilfe}

Synonyme

Geschirr Kundendienst Betreuung Sitz Einrichtung

Übergeordnete Begriffe

Satz Bedienung

Untergeordnete Begriffe

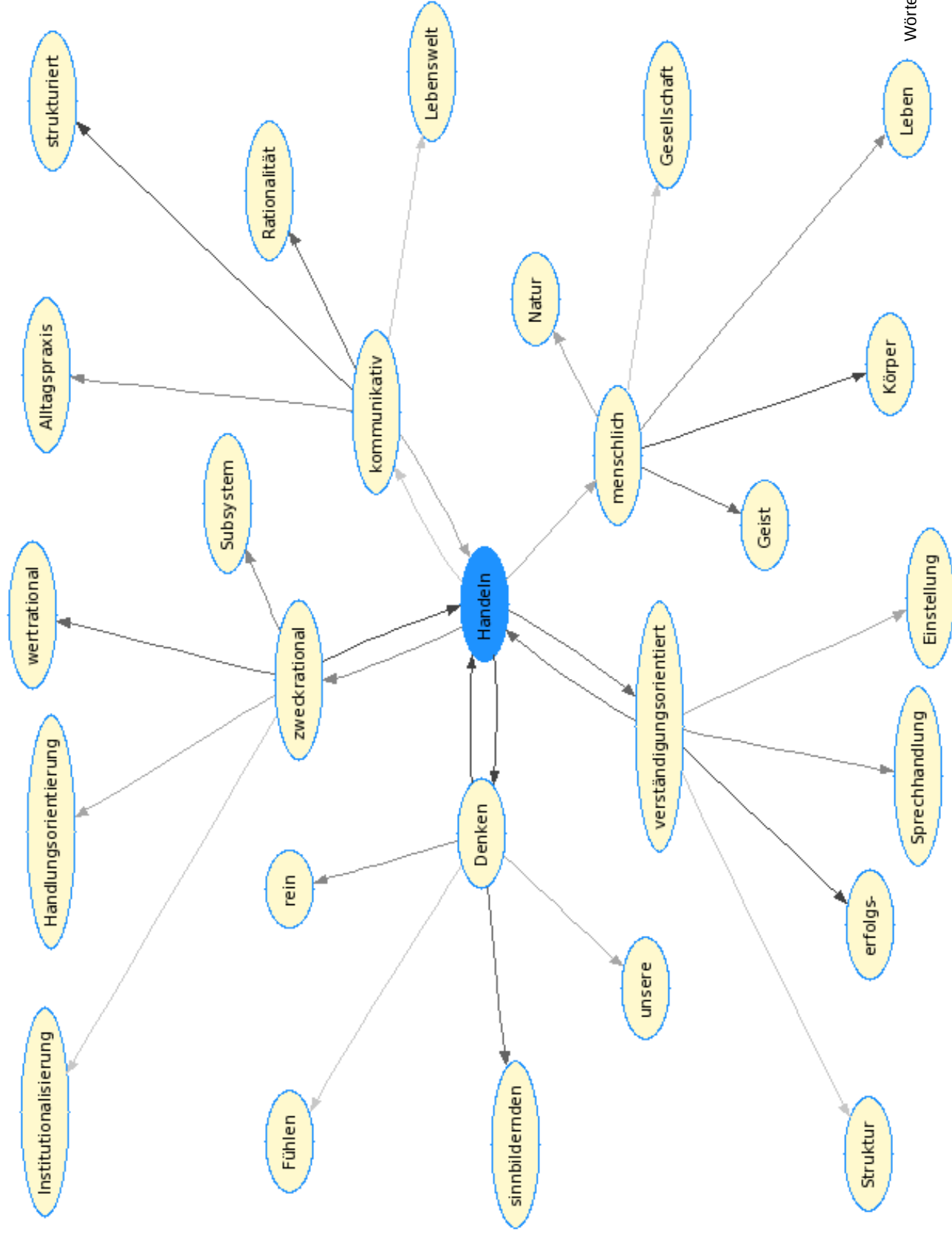
Likörservice Zimmerservice

Kollokationsgraphen zu Handeln und Tun

Automatisch berechnete Kollokationen aus dem DWDS Kerncorpus (lemmabasiert)

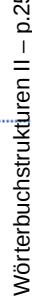
Kompaktsicht

Hilfe



Kompaktansicht

Kompaktansicht



Vergleich der Informationssysteme

- abstraktes Informationsprogramm für WB-Artikel scheint vorteilhaft (elexiko)
- Übernahme von Daten aus MRD nur nach Konsistenz- und Aktualitätscheck (s. Nachteile der direkten Abbildung der DWG-Daten in das DWDS)
- aus Korpora gewonnene Daten zu den Stichwörtern sollen für Benutzer ergänzend verfügbar sein, auch um Defizite in der Abdeckung des elexiko auszugleichen

Vorschlag zu kooperativ erstelltem Lexikonmodell

- Angaben aus anderen lexikalischen Ressourcen, vor allem MRD, übernehmen, wenn diese geprüft wurden und in das Lexikonmodell passen;
- Angaben aus Korpora automatisch generieren;
- Angaben, die nicht derart generiert werden können, manuell erstellen.

Parsing von Wörterbuchartikeln

Übersicht

- Parsing von Wörterbuchartikeln - Parser
LexParse s. separater Foliensatz
- Extraktion von semantischer Angaben aus
maschinenlesbaren Wörterbüchern

Extraktion semantischer Angaben aus dem

LINKS Projekt (Vossen et al. (1989): Entwicklung einer Datenbank semantischer Beschreibungen aus den Bedeutungsbeschreibungen des *Longman Dictionary of Contemporary English* auf der Grundlage des komponentiellen Ansatzes von Simon Dik (in welchem ohne Rückgriff auf semantische Primitiva lexikalische Bedeutungen auf ein Basisvokabular reduziert werden). Das LDOCE verwendet ein wohldefiniertes und eingeschränktes Definitionsvokabular (ca. 2000 Wörter).

- grammatische Kodierung des restringierten Definitionsvokabulars führt zum Aufbau eines grammatisch annotierten Korpus von Bedeutungsbeschreibungen
- Ermittlung syntaktischer Typologien und Entwicklung von Parsergrammatiken für die Bedeutungsbeschreibungen der unterschiedlichen Kategorien N,V,A (Segmentierung in KERNEL sowie PRÄMODIFIKATOREN und POSTMODIFIKATOREN)
- Aufbau einer zur syntaktischen Typologie parallelen semantischen Typologie
- Verknüpfung beider Typologien in einer relationalen Datenbank zur Erfassung syntagmatischer und paradigmatischer Strukturen zwischen den Wörtern
- Verknüpfung der Einträge dieser Datenbank mit pragmatischer Information aus dem LDOCE (Sachbereichsfelder, Sprachregister) für den Zugriff auf Domänenwortschatz

Beispielintrag *anaesthesist*

Spezifikation auf 5 Ebenen (LDOCE Subject field code, Box code (Stilinformation), orthographische Form, syntaktische Kategorie (POS-Code) und Bedeutungsbeschreibung)

'anaesthetist': a doctor who gives an anaesthetic to a patient before he is treated by another doctor

Dieser Eintrag wird in folgende Teile segmentiert:

- subject fields: mdon-
- semantic label:h...y.
- entry (POS): anaesthetist (n)
- marker (beginning of meaning description): ul UL
- meaning description: a doctor who gives an anaesthetic to a patient before he is treated by another doctor

Beispielintrag *anaesthesist*

Auszeichnung der Wörter aus der Bedeutungsbeschreibung mit POS-Codes:

a [D0] doctor [N0] who [P0] gives [VJ] an [D0] anaesthetic [N01A01XX] to [I0TO]
a [D0] patient [A0] before [C0I0] he [P0] is [BJ] treated [VD] by [I0] another [D0P0]
doctor [N0]

Bedeutungsbeschreibung für Nomen auf vier Ebenen:

- Wortsequenz: a man who gives
- POS Sequenz: Det Noun RelPronoun Verb
- Syntaktisches Pattern: [NP[DET a] [N (or KERNEL) man] [REL_CLAUSE w ho gives]
- Semantisches Pattern: Quantor category activity-spec. modifier

Grundstruktur einer Nominalphrase besteht aus einem (optionalen) Determiner, einer (optionalen) modifizierenden Komponente PRE-KERNEL oder POST-KERNEL) und einem obligatorischen syntaktischen KERNEL

Typen nominaler Bedeutungsbeschreibung

LINKS: syntaktischer Kernel ist in semantischer Hinsicht ein Hyperonym des Eintragswortes und wird in der Bedeutung von den Prä- und Postmodifikatoren eingeschränkt

(1) flamingo: *a small tropical water BIRD with long thin legs*

(2) lapdog: *a small pet DOG*

SYNONYMS: keine Restriktionen oder Modifikatoren notwendig, da die gesamte Information in einem Wort, eben dem Synonym, ausgedrückt wird. Gleichzeitig impliziter Verweis auf den synonymen Eintrag (und die dort repräsentierte Information):

(3) abattoir: *slaughterhouse*

Typen nominaler Bedeutungsbeschreibung

LINKERS: relativ bedeutungsleere Kernel bezeichnet, die auf ein anderes, ebenfalls in der Bedeutungsparaphrase auftauchendes Wort verweisen, für welches an anderer Stelle im Lexikon weitere Informationen zu finden sind.

(4) breastplate: *a **PIECE** of armour worn to protect the chest*

(5) arab: *a **TYPE** of fast graceful horse*

Nicht alle nominalen Bedeutungsbeschreibungen mit einem Kernel, der ein *of*-Komplement enthält, sind aber als **LINKER** zu betrachten, z.B. formal analoge Konstruktionen, in denen der Kernel ähnliche semantische Information wie ein **LINK** trägt:

(6) **beef**: *the **MEAT** of farm cattle*

(7) **lights**: *the **LUNGS** of sheep, pigs, etc. used as food*

SHUNTER: ähnlich bedeutungsloser Kern wie bei **LINKERN**, leiten Wortklassenwechsel in der Bedeutungsbeschreibung (auf die nicht-nominale Ebene) ein, mit Relativsätzen umschreibbar

(8) **adornment**: *the **ACT** of **adorning** (an act which consists of **adorning**)*

(9) **actuality**: *the **STATE** of **being real** (a state which consists of **being real**)*

Auch Kernels mit Relativsätzen gehören zum **SHUNTER-Paradigma**:

(10) **camper**: *a **PERSON** who camps*

Typen nominaler Bedeutungsbeschreibung

Neben relativ klaren Fällen gibt es auch Probleme der Klassifizierung, weil manche Kernels zwischen Links und Shuntern stehen. Wegen der gleichwertigen Verteilung des semantischen Gehalts werden sie aber als Linker charakterisiert.

(11) **acting**: *the ART of representing a character. esp. on stage or for a film*

(12) **admixture**: *a SUBSTANCE that is added to another in a mixture*

Außerdem gibt es noch Kernel, die Shuntern ähneln, bei denen aber das „Shunten“ auf den Kernel selbst erfolgt:

(13) **advent**: *the COMING of Christ to the world*

(14) **aspiration**: *the PRONUNCIATION of the letter h*

Bedeutungsparaphrasen im LDOCE

Beispiel der Analyse von Bedeutungsparaphrasen eines Printwörterbuchs zeigt, dass

- sich Bedeutungsparaphrasen in syntaktische Strukturmuster klassifizieren lassen, wenn diese linguistisch annotiert und diese Annotationen analysiert werden;
- sich aus diesen syntaktischen Strukturmustern unterschiedliche Beziehungen zwischen der paraphrasierten lexikalischen Einheit, dem Kern der Paraphrase sowie weiteren Wörtern der Paraphrase ableiten lassen;
- sich aus diesen Relationen sinnrelationale Beziehungen zwischen lexikalischen Einheiten ableiten lassen;
- dass diese Informationen in eine lexikalische Datenbank überführbar sind, wenn das zugrunde liegende Lexikonmodell die Kodierung von Sinnrelationen vorsieht.

Es wird aber im Falle jedes einzelnen Wörterbuchs bzw. Analyseprojekts genau zu untersuchen sein, mit welcher Präzision das Stichwort mit den lexikalischen Einheiten der Bedeutungsparaphrase verbunden werden kann. Zu viele Fehlzusammenhänge wirken sich negativ auf das Kosten-Nutzen-Verhältnis einer solchen Analyse aus. Eventuell kann es danach günstiger sein, sinnrelationale Strukturen im Wörterbuchüberwiegend händisch zu erstellen. Ein Verfahren wie das hier beschriebene kommt sowieso nur für solche Wörterbücher in Frage, in denen das Verfahren der Bedeutungsparaphrasierung bereits stark formalisiert ist.