

Computerlexikographie-Tutorium 18.04.2008

- Informationen:
 - E-Buch:
 - bitte angegebene Kapitel lesen (Material im Tutorium)
 - weitere Materialien: <http://www.lemnitzer.de/lothar/CoLex>
 - Hausaufgabe: für Übungsblatt 1:
 - 10 Punkte
 - Abgabe bis Montag, gegebenenfalls noch später (gilt nur für dieses Übungsblatt!)
- Themen für heute:
 - Zeichentheorie
 - Relationale Semantik
 - Der lexikographische Prozess

Zeichentheorie

- **lexikalisches Zeichen:** jede sprachliche Einheit, der eine Bedeutung zugeordnet werden kann
 - Morphem, Wort, Mehrworteinheit (Phrasem, Kollokation)
 - Kollokation: häufig auftretende Wortverbindungen, primär semantisch motiviert
 - Phrasem = Idiom: sprachliche Wendung aus mehreren Wörtern, die Bedeutung ist nicht mehr transparent, Starrheit des Phrasems
- **lexikalisches Feld:** eine Menge lexikalischer Zeichen, deren Bedeutungen über lexikalischen-semantischen Relationen verbunden sind
 - Synonymie, Antonymie, Hyperonymie, Hyponymie
 - in Printwörterbüchern: als Verweise

Zeichentheorie

- Struktur des lexikalischen Zeichens: (E-Buch: 16–22, 22–25)
 - Formseite vs. Inhaltsseite (strukturalistisch – de Saussure)
 - langue vs. parole
 - Kompetenz vs. Performanz (generative Grammatik – Chomsky)
 - denotative vs. referentielle Bedeutung
 - paradigmatische vs. syntagmatische Beziehungen

Zeichentheorie

- <http://www.sfs.uni-tuebingen.de/~lothar/lexbuch/Seminar01/skripte/onomasio.html>
 - **Saussure** (1916): 2-er-Teilung
 - **Odgen & Richards** (1949): semiotisches Dreieck: 3-er-Teilung
 - **Ullmann** (1962): 3-er-Teilung
 - **Heger** (1976): semantisches Trapez: 4/7-er-Teilung, Sem (Bedeutungssatom, Bedeutungsmerkmal), Semem (bedeutungstragende Kombination von Semen)
 - Synonymie: Zwei sprachliche Einheiten haben verschiedene Formen, aber ein gemeinsames Semem. = Extension ist gleich
 - Antonymie: Die Sememe zweier sprachlichen Zeichen sind bis auf ein Sem gleich.
 - Hyponymie: Die Sememe zweier sprachlichen Zeichen teilen einige Seme. Das Hyponym hat zusätzliche Seme.
 - Kohyponyme: Diejenige Seme sind gleich, die mit dem gemeinsamen Hyperonym geteilt werden. Disjunkte Referenten.

Relationale Semantik

- **Lexikalisch-semantische Beziehungen** zwischen sprachlichen Zeichen:
 - paradigmatisch: zwischen Wörtern mit systematischer Bedeutungsverwandtschaft → Austauschbarkeit im gleichen Kontext
 - syntagmatisch: gemeinsames Vorkommen in sprachlichen Äußerungen

Relationale Semantik

- **Paradigmatische Beziehungen:** zwischen Wörtern mit systematischer Bedeutungsverwandtschaft → Austauschbarkeit im gleichen Kontext
 - Synonymie = echte Synonymie: gleiche denotative Bedeutung
 - Plesionymie = Quasi-Synonymie: ähnliche Bedeutung
 - Hypero- / Hyponymie: Hyponym impliziert Hyperonym
 - Holo- / Meronymie (E-Buch: 42 ff.)
 - Inkompatibilität (E-Buch: 44)
 - Antonymie (groß – klein)
 - Komplementarität (lebendig – tot)
 - Konversion (kaufen – verkaufen)

Relationale Semantik

- **Syntagmatische Beziehungen:** gemeinsames Vorkommen in sprachlichen Äußerungen
 - Kollokation (*nächste Zukunft*)
 - <http://www.dwds.de/kollokationsgraph?&sh=1&qu=Zukunft&mode=html>
 - Funktionsverbgefüge (*zur Verfügung stehen*)
 - Verb-Komplement-Beziehung (*Apfel essen*)

Relationale Semantik

- **Polysemie** = Mehrdeutigkeit: Ein lexikalisches Zeichen mit mehreren Bedeutungen, die miteinander verbunden sind (eine Grundbedeutung ist i.r.R. vorhanden). (siehe auch E-Buch: 48–54)
- Polysemie → Ambiguität → **WSD** (word sense disambiguation): Lesartenbestimmung von Textwörtern (siehe auch E-Buch: 60 ff.)
- **Lesart**: Eine Bedeutungsstelle eines Wortes, die Zuordnung der relevanten semantischen Beschreibung zu einer Wortform.
 - nach Cruse: **lexikalische Einheit** ~ Form-Bedeutungspaar
- **Reguläre Polysemie = reguläre Bedeutungsalternanz**: Anhand von Analogien regulär vorhersagbare Bedeutungserweiterung von lexikalischen Zeichen. → generalisierbar (inwiefern?) (siehe auch E-Buch: 50–54)
 - siehe „Noun“ / „Regular Polysemy in GermaNet“ unter <http://www.sfs.uni-tuebingen.de/lsd/>
- **Lexikalische Regeln**: Kodierung linguistischer Generalisierungen – Auch reguläre Polysemie kann damit erfasst werden.

Relationale Semantik

- **Polysemie** ↔ **Homonymie**: (E-Buch: 48 ff.)
 - Frage: Wie viele **Einträge** im Wörterbuch?
 - Polyseme unter einem Eintrag durchnummeriert.
 - Homonyme sind verschiedene Einträge.
 - Kriterien für Homonymie:
 - formale Kriterien: Unterschiede in der Aussprache oder in den Vollformen (synchron)
 - semantische und logische Kriterien: kein Zusammenhang zwischen den Bedeutungen (synchron)
 - etymologische Kriterien: sprachgeschichtliche Angleichung der Form verschiedener Zeichen (diachron)
 - Vorgehensweisen nach Lyons:
 - Homonyme maximieren ↔ Polyseme maximieren

Der (computer)lexikographische Prozess

- E-Buch: 54–57
- Man muss die lexikographischen Daten
 - erheben ← existierende Wörterbücher, Korpora, Belege, Wortnetze, Ontologien
 - bearbeiten
 - darstellen
 - verarbeiten
- nach Kilgarriff: Für jedes Wort muss Folgendes durchgeführt werden:
 - Konkordanz abrufen (alle Vorkommen ~ Instanzen des Wortes)
 - Instanzen clustern (~ Gruppierung der Belege nach festgestellten unterschiedlichen Bedeutungen, Lesarten)
 - Cluster beschreiben, charakterisieren, ggf. neu organisieren (gefährlich)
 - Schlüsse auf die Lesarten ziehen → Bedeutungsbeschreibung

Der (computer)lexikographische Prozess

- **Lexikographische Typen** beim Clustern:
 - Zusammenfasser: grobe Unterscheidung → möglichst wenig Gruppen
 - Trenner: möglichst feine Unterscheidung → viele Gruppen
- **SFIP**: Kilgarriiffs Daumenregel: (E-Buch: 55 f.)
 - SF: *„Ein Verwendungsmuster, das zu einer Lesart abstrahiert wird, sollte hinreichend häufig belegt sein (,sufficiently frequent’ – SF) und [...]“*
 - IP: *„[...] nicht hinreichend erschließbar aus anderen Verwendungsweisen (,insufficiently predictable’ – IP)“*
- **Allgemeine Regeln**:
 - immer Ziel des Wörterbuches befolgen! (allgemeines / Fachwörterbuch, Eigenschaften der Zielgruppe, der Ressourcen, des Mediums, ...)
 - ein Cluster muss „groß genug“ sein
 - für ein Cluster müssen mehrere Belege vorhanden sein
 - reguläre Polyseme in ein Cluster

Der (computer)lexikographische Prozess

- Lexikon: polysemes Wort (siehe Vorlesung)
 - offenes System → pflegebedürftig
- Das **Lexikonmodell**
 - legt die elementaren **Einheiten** im Lexikon fest: Einträge, Lemmata
 - Morpheme
 - Wörter
 - mehrgliedrige Einheiten: Kollokationen, Phraseme
 - legt fest, wie die Einheiten **organisiert** werden:
 - formbasiert / ausdrucksseitig / alphabetisch
 - inhaltsseitig / semantisch

Der (computer)lexikographische Prozess

- **Elektronischer Zugriff** auf die für eine Aufgabe **relevante Daten**.
- **Lexikalische Ressourcen:**
 - Organisationsformen: Wortnetze, Ontologien
 - nichtdigital oder digital (lexikalische Datenbanken)

Der (computer)lexikographische Prozess

- Organisationsform lexikalischer Ressourcen:
 - **Wortnetz:**
 - WordNet, GermaNet, EuroWordNet
 - <http://wordnet.princeton.edu/perl/webwn>
 - semantisches Netz, lexikalische Zeichen nach ihren Bedeutungen (inhaltsseitig) angeordnet
 - **Ontologie:**
 - <http://de.wikipedia.org/wiki/Bild:Ontschichten.gif>
 - strengere und explizitere Formalisierung von Wissen über Objekte und Eigenschaften
 - konzeptuell ausgerichtet

Der (computer)lexikographische Prozess

- **Lexikalische Datenbanken:** digitale lexikalische Ressourcen
 - statisch oder dynamisch
 - Die Datensätze sind konsistent und formal beschrieben (konzeptuelle Schemata). (Theorie hinter der Strukturierung der Datensätze!)
 - Typen:
 - MRD (machine readable dictionary): **maschinenlesbares** Wörterbuch
 - für die Menschen → natürlichsprachlich, wie Printwörterbuch
 - LDOCE: <http://pewebdic2.cw.idm.fr/>
 - MTD (machine tractable dictionary): **maschinenverarbeitbares** Wörterbuch
 - für die Maschine → spezifiziertes, explizites Format
 - GermaNet: <http://www.sfs.uni-tuebingen.de/lsd/>, HaGenLex: <http://pi7.fernuni-hagen.de/forschung/hagenlex/hagenlex-de.html>

Der (computer)lexikographische Prozess

- **lexikalisches Informationssystem:** lexikalische Datenbank + Benutzerschnittstelle
- **Lexicon builder:** erstellen Wörterbücher automatisch aus bestehenden Wörterbüchern, Dokumenten und Korpora
- **Wissensbank:** maschinenverarbeitbar, auch außersprachliches Wissen eingebunden, z. B. Wikipedia
- **Hyperlexikon:** als Hypertexte realisierte Lexika und Lexikonsysteme
 - statisch (Indexierung) oder dynamisch (Suchfunktion)
- Links zum Selbergucken:
 - <http://www.dwds.de/>
 - http://hypermedia.ids-mannheim.de/pls/elexiko/p4_start.portal
 - <http://corpora.ids-mannheim.de/ccdb/>
 - <http://www.canoo.net/>