

LEXIKALISCHE DISAMBIGUIERUNG MIT SEMANTISCHEN MERKMALEN  
Überlegungen zur EUROTRA Interface-Struktur

Cornelia Zelinsky-Wibbelt  
IAI/EUROTRA-D  
Martin-Luther-Straße 14  
D-6600 Saarbrücken

Die Interface Struktur ist diejenige linguistische Ebene des EUROTRA-Ansatzes, auf der der Transfer von der maschinell zu übersetzenden Quellsprache in die Zielsprache stattfindet. Der vorliegende Beitrag skizziert ein Konzept zur semantischen Vereindeutigung lexikalischer Einheiten im Lexikon mittels kognitiv motivierter *semantischer Merkmale*. Der vorgestellte Ansatz basiert auf den *Cognitive Grammars* und soll ein Beitrag zu dem für maschinelle Übersetzungssysteme wesentlichen Problem der Konsistenz von Einträgen in großen Wörterbüchern sein. Die Metasprache, in die die semantischen Merkmale eingebettet sind, wird an einem Beispiel illustriert.

Auf der bereits in *Zelinsky-Wibbelt 1987a* beschriebenen Ebene der Interface-Struktur (IS) sollen Sätze möglichst "euroversell" repräsentiert werden, was bedeutet, daß die Transferschritte zwischen den Interface-Strukturen der verschiedenen Sprachen möglichst klein werden sollen. Dabei werden unter anderem verschiedene Lesarten einer lexikalischen Einheit durch "euroverselle" semantische Merkmale identifiziert. Für die semantische Vereindeutigung lexikalischer Einheiten entwickeln wir ein vom Lexikographen interaktiv zu nutzendes Generierungssystem. Mit diesem soll für jede Lesart diejenige Liste von semantischen Merkmalen erzeugt werden, die diese Lesart von anderen Lesarten abgrenzt. Diese Liste ist Teil der lexikalischen Information.

Die semantischen Merkmale basieren auf **grundlegenden kognitiven Prinzipien**, mit denen der Mensch seine Umgebung logisch im Hinblick auf allgemeine Übereinstimmungen und Unterschiede strukturiert. Diese semantischen Merkmale sind die Terme einer semantischen Metasprache. Diese verfolgt folgende Ziele im Hinblick auf eine weitestgehende Operationalisierung des semantischen Raumes, der tatsächlich ein Kontinuum ist:

- 1) Das Ziel der Objektivität
- 2) Das Ziel der psychologischen Realität
- 3) Das Ziel der Generalisierung

Die drei Ziele bedingen sich in der gegebenen Reihenfolge. Sie bilden auch das Leitmotiv der *Cognitive Grammar* (vgl. *Langacker 1982*, S. 23-25,35). Unsere Grundprinzipien, also auch die Entwicklung der Definitionen der Merkmale, basieren auf diesem Ansatz; wir gehen davon aus, daß eine Erfüllung dieser Anforderungen zu einem System führt, bei dem eine genügend hohe Konsistenz in der Kodierung lexikalischer Einträge erreicht werden kann, was bei allen existierenden Systemen als größtes Problem beim Aufbau umfangreicher Wörterbücher berichtet wird.

Das Formulieren von Generalisierungen über die Kompetenz der Sprachbenutzer impliziert die Hypothese von der "Euroversalität" dieser Metasprache.

Als Teilziel von (3) bleibt hier noch das Verfahren zu nennen, mit dem Generalisierungen ermöglicht werden. Auch hierin können wir uns auf die *Cognitive Grammar* beziehen. Für diese sind hierarchische Schemata unter anderem ein adäquates Erklärungsmittel dafür, daß Ähnlichkeit zwischen Konzepten abhängig ist von der Hierarchiestufe, auf der sie verglichen werden. Diese Erklärungskraft muß auch unsere semantische Metasprache besitzen, um die Bedeutungen lexikalischer Einheiten in analoger Weise zu vergleichen.

Zur Beschreibung der vier Wortklassen erhalten wir vier Inventare semantischer Merkmale: "Situations" für Verben, "Entities" für Substantive, "Properties" für Adjektive und "Specifiers" für Adverbien. In jedem Set bilden jeweils Teilmengen wohlgeformte Ausdrücke. Ihre Wohlgeformtheit ist durch ein semantisches Regelsystem definiert, in dem die Merkmale durch die Relationen der Hierarchie, der Disjunktion und der Konjunktion verbunden sind. Zusätzliche Kookkurrenzbeschränkungen gelten für linguistisch nicht realisierte konjunktiv verbundene Merkmale (für eine genaue Erläuterung des Grammatikformalismus siehe *Zelinsky-Wibbelt 1988*).

Mit unserem semantischen Regelsystem können wir zum Beispiel für vier Lesarten der lexikalischen Einheit **Absatz** die folgenden hierarchischen Strukturen semantischer Merkmale generieren (die eckige Klammer entspricht in der Horizontalen der Relation der Hierarchie, in der Vertikalen verbindet sie Merkmale, die in konjunktiver Relation zueinander stehen).

1. Absatz als *Text* (semantischer Text)



2. Absatz als *Text* (semantischer Text)



3. Der *Text* (semantischer Absatz)



4. Der *Text* (semantischer Absatz)



Die Merkmale "individutive" und "partitive" sind Teil eines Merkmalssets, das auch für die funktionale Repräsentation eines bestimmten Typs der Determination notwendig ist, der von *Lehrer 1986* als "classifier construction" bezeichnet

wird. Oberflächenstrukturell wird dieser Typ folgendermaßen ausgedrückt:  
[ZENTRIEREN] [NP + OF + defNP]

Diesem Ausdruck entspricht das folgende semantische Schema:  
[ZENTRIEREN] [x[+]1[+]] (x[+]1[+] \_  
x[+]2[+])

Diese Merkmale, die zu den Merkmalen gehören, die "classifier constructions" strukturell identifizieren sollen, sind Teil eines umfassenderen

Merkmalsets, das zur Zeit im Rahmen eines Transfer-Experiments zur Determination entworfen und getestet wird.

#### Literatur

Die Literaturangaben finden sich für alle Fachbeiträge zum Themenschwerpunkt Maschinelle Übersetzung zusammengefaßt auf Seite 23.

#### ANZEIGE

# WordCruncher

## FEATURES OF WORDCRUNCHER \*

### Preindexing.

- With WordCruncher's *IndexETC*, a preindex is created for a specific text. It is this electronic "road map" that makes possible sophisticated search combinations of large documents with amazing speed.

- Existing electronic texts require very little preparation before preindexing with *IndexETC*. Texts not already stored in electronic form may be entered via the keyboard or an optical scanner.

- WordCruncher allows the preindexing of documents in many foreign languages. Users are able to define the proper sorting sequences of characters, such as umlauts in German, accents in French, and the "ch" in Spanish. Texts may be shown from right to left for languages such as Hebrew. Preindexed files can be linked together to create a larger file.

### Retrieving

- With WordCruncher's *ViewETC*, text material can be retrieved according to a set of user-defined parameters from very large documents.

- References can be looked up using a word or phrase, a list of words, two or more words in a defined context, a substring (such as a suffix or stem), or any logical combination of these. Supplementary dictionaries and thesauruses, supplied by the user and preindexed by WordCruncher, offer additional retrieval clout.

- Selected references can be displayed within windows of modifiable size which show the reference in its context. The Expanded Window displays an entire screen of text surrounding the reference, and you may page forward or backward through the document.

- To make it all easy, help displays are available throughout.

### Manipulating.

- Once the desired data has been retrieved from a text, it may be manipulated in any number of ways. For instance, *ViewETC* can create printable book-style indexes, as well as keyword-in-context (KWIC) concordances of words or phrases.

- Retrieved information can be printed immediately, saved as a DOS text file, or placed in a temporary file such as WordPerfect Library's clipboard function provides. WordPerfect Library allows the user to transfer text from *ViewETC* to WordPerfect without exiting either program.

### Marketing Representative

Institut für angewandte  
Kommunikations- und  
Sprachforschung e.V.

Poppelsdorfer Allee 47  
D-5300 BONN 1  
Tel.: 02 28/73-56 45 73-56 20  
EARN: DBNRHRZ1.UPK000

\*Formerly known as BYU Concordance