

GRUNDPRINZIPIEN UND VORSCHLÄGE FÜR EINE SEMANTISCHE BESCHREIBUNG VON ARGUMENTSTRUKTUREN

Überlegungen zur EUROTRA Interface-Struktur

Erich Steiner

IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Zwischen Quell- und Zielsprache werden in der Konzeption von EUROTRA, dem europäischen Projekt zur maschinellen Übersetzung für alle EG-Sprachpaare, verschiedene linguistische Ebenen angenommen. Damit wird der gesamte Übersetzungsprozess in eine Folge von Teilübersetzungen zerlegt. Zwischen Analyse und Generierungsphase findet der Transfer auf der Ebene der *Interface-Struktur* (IS) statt. Diese IS ist keine Interlingua, sondern eine Repräsentation, die möglichst wenig Transferoperationen erfordert. Der vorliegende Beitrag diskutiert die *semantischen Relationen* (Prädikat-Argument-Beziehungen) als eine der wichtigsten Teiltheorien der IS. Für diesen Bereich gibt es noch keinen von allen EUROTRA-Gruppen akzeptierten Ansatz. Zunächst werden Anforderungen an und Funktionen von semantischen Relationen im Rahmen eines Übersetzungssystems umrissen. Anschließend wird das Konzept der EUROTRA-D-Forschungsgruppe, das sich auf die Systemic Functional Grammar abstützt, vorgestellt und begründet. Der Stand der Realisierung und die weitere Entwicklung werden skizziert.

1. Grundprinzipien der IS

Semantische Relationen (SRs) sind die semantischen Beziehungen zwischen den "governors" (govs) einer Konstruktion und den von ihnen abhängigen Gliedern, den Komplementen (comps). Semantisch interpretiert handelt es sich um die Beziehungen zwischen Prädikaten und Argumenten (preds und args). Inwieweit hierher auch die semantischen Relationen zwischen govns und modifiers gehören, ist eine Frage, mit der wir uns hier nicht beschäftigen wollen. Die verschiedenen Standardtheorien aus dem Bereich der Linguistik, die hier zugrundeliegen, sind recht gut bekannt und sollen auch hier nicht weiter diskutiert werden (Fillmore, Halliday, Longacre, Starosta, Gruber, Jackendoff, Bresnan). Der im folgenden beschriebene EUROTRA-D-Ansatz ist u.a. dokumentiert in Steiner 1985, Steiner 1986, Steiner 1987a, Steiner 1987b, Steiner et al. 1987.

3. Die Funktion von SRs in einem multilingualen MT-Projekt

Semantische Relationen haben eine Reihe von Funktionen, die sie in einem multilingualen MT-Projekt als wesentlich erscheinen lassen:

3.1. Die Repräsentation von Ambiguitäten

Innerhalb von EUROTRA ist die *Interface Struktur* definiert als Ebene minimalen Transfers zwischen einer Ausgangssprache und einer Zielsprache. Die *Interface Struktur* (IS) ist somit keine Interlingua - eine grundlegende Festlegung, die immer wieder in der Diskussion um EUROTRA übersehen wird.

Die Theorie der IS umfaßt eine Reihe von Teiltheorien: MODALITY, TIME, SEMANTIC FEATURES, SEMANTIC RELATIONS und möglicherweise weitere. Wir werden uns hier mit der Theorie der Semantischen Relationen beschäftigen.

2. Semantische Relationen in der theoretischen Linguistik

Durch semantische Relationen lassen sich grundlegende Klassen von lexikalischen Ambiguitäten darstellen, die mit syntaktischen Repräsentationen alleine nicht darstellbar sind. Da solche Fälle in der Literatur hinreichend dokumentiert sind, wollen wir uns hier das Anführen von Beispielen ersparen.

3.2. Vermeidung von komplexem strukturellem Transfer

Eine Repräsentation, die SRs beinhaltet, ist ein wesentliches Mittel, klassische Arten von komplexem strukturellem Transfer zu vermeiden. Es handelt sich dabei um Erscheinungen wie Thematisierung, Passivierung im besonderen, Reflexivierung, Kausativierung, verschiedene Arten von "raising". Da aus offensichtlichen Gründen komplexer struktureller Transfer in einem multilingualen MT-System sehr arbeits- und kostenintensiv ist, handelt es sich hierbei um einen wichtigen Gesichtspunkt.

3.3. Steuerung des lexikalischen Transfers

SRs können bis zu einem gewissen Grad zur Steuerung von automatischem lexikalischem Transfer eingesetzt werden. Untersuchungen zum Transfer Deutsch-Französisch im Rahmen der Zusammenarbeit zwischen EUROTRA-D Saarbrücken und der EUROTRA-D-Begleitforschung Stuttgart sind vorgelegt worden u.a. in Eckert/Heid 1987.

3.4. Behandlung anaphorischer Referenz

SRs sind im Rahmen einer entsprechenden Konzeption Bestandteil der logischen Form von Sätzen und können in diesem Rahmen auch eingesetzt werden bei der Behandlung anaphorischer Referenz.

4. Der Ansatz von EUROTRA-D

Der Ansatz von EUROTRA-D berücksichtigt zunächst einmal die in Abschnitt 2 genannten Funktionen von SRs. Des Weiteren ist davon auszugehen, daß das Ziel von EUROTRA für IS eine Ebene minimalen Transfers ist, nicht aber eine Form von Interlingua. Als linguistische Modelle kommen von daher eher funktionalistische Theorien in Frage als solche, die a priori einen starken Anspruch auf Universalität erheben. Es erscheint ebenfalls geboten, sich weitgehend auf relativ langlebige und gut ausgearbeitete Ansätze zu stützen. Kurzlebige Theorien werden rezipiert, können aber in einem Projekt von einer Größenordnung wie EUROTRA nicht die Grundlage in irgendeinem Bereich bilden. Letztendlich muß ein SR-System für ein Projekt wie EUROTRA größtmögliche Konsistenz unter einer großen Anzahl von Kodierern/-innen verschiedener Nationalität und mit verschiedenem Hintergrund gewährlei-

sten, ähnlich wie das zu entwickelnde System für die Kodierung semantischer Merkmale von lexikalischen Einheiten, das in Zelinsky-Wibbelt 1987b beschrieben ist. Aus den geschilderten Überlegungen heraus entschloß man sich in EUROTRA-D, sich in der SR-Theorie weitgehend, wenn natürlich auch nicht ausschließlich, auf die "Systemic Functional Grammar" zu stützen (vgl. Halliday 1985, Fawcett 1980, Winograd 1983, Butler 1985, Steiner 1983).

Der Ansatz arbeitet mit einer kleinen Anzahl von Prozeßtypen (4 Grundtypen, die in 17 Untertypen unterteilt sind), die durch eine sie charakterisierende semantische Rolle festgelegt sind. Diese Schlüsselrollen sind durch ein eindeutiges syntaktisches Realisierungspotential gekennzeichnet. Jeder der 4 Grundtypen ist als Eingangsbeschränkung für eine Untermenge von semantischen Rollen definiert. Die Rollensätze der 4 Grundtypen sind fast vollständig disjunkt. Bei der Lexikonkodierung wird so ein hohes Maß an Konsistenz zwischen verschiedenen Kodierern/-innen gewährleistet. Gegenwärtig sind die SRs in verschiedenen Versionen eines lauffähigen Systems implementiert, wobei systematisch bisher SRs für Verben sowie SRs für Nomen implementiert wurden.

5. Die weitere Entwicklung

Der geschilderte Ansatz für SRs ist innerhalb von EUROTRA in der Diskussion, ist jedoch zur Zeit nicht der in allen Sprachgruppen vertretene Ansatz. Es ist die Auffassung von EUROTRA-D, daß unabhängig von wissenschaftlichen Richtungskämpfern die in Abschnitt 2 angesprochenen Grundanforderungen an jede SR-Theorie von jedem vorgeschlagenen Ansatz eingelöst werden müssen. Diese Grundanforderungen bilden eine Art von gemeinsamem Bezugspunkt für die Diskussion innerhalb des Gesamtprojektes. Weiterhin müssen in der Zukunft die verschiedenen Arten der Interaktion der SR-Theorie mit der Theorie der "Semantic Features" und den Theorien für Modality und Time geklärt werden. Damit sind unweigerlich Beiträge zu leisten zur Praxis der Implementierung solcher Systeme, aber auch zur aktuellen Diskussion innerhalb der beteiligten wissenschaftlichen Disziplinen, wobei natürlich in erster Linie die Linguistik angesprochen ist.

Literatur

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

PROJEKT SEMSYN NEW PUBLICATIONS

Walter Koll

GEOTEX Ein System zur Verbalisierung geometrischer Konstruktionen

(**GEOTEX** - a system for verbalizing geometric constructions, in German)

GEOTEX is an application of the **SEMTEX** text generator for German. The text generator is combined with a tool for interactively creating geometric constructions. The latter offers formal commands for manipulating (i.e. creating, naming and, deliberately, deleting) basic objects of Euclidean geometry. The generator is used to produce descriptive texts in German, related to the geometric construction. Descriptions of the sequence of steps describing a construction, **SEMTEX** context-handling mechanisms have been extended to support it.

Elision is no longer restricted to adjuncts. For repetitive operations verb and subject will be elided in subsequent sentences.

The distinction between known information and new one is exploited to decide on constituent ordering. The constituent referring to the known object is topicalized (i.e. put in front of the sentence).

The system allows for more ways to refer to objects introduced in the text: pronouns, textual deixis using demonstrative pronouns, names. The choice between these variants is done deliberately.

GEOTEX is implemented on ZetalISP and runs on IBM/360/370 class machines.

Cherita Russe

**The generation system of the SEMSYN project
Towards a task independent generator for German**

We report on our experiences from the implementation of the **SEMSYN** generator, a system generating German texts from semantic representations, and its application to a variety of different areas: input structures and generation tasks. In its initial version, the **SEMSYN** generator was used within a Japanese German MT project, where it produced German equivalents to Japanese titles from scientific papers. Being carefully designed in object-oriented style (and implemented with the FLAVOR system), the system proved to be easily adaptable to other semantic representations (e.g. output from MIT's Universal Parser) and extendable to other generation tasks: generating German news stories, generating descriptive texts to geometric constructions.

Cherita Russe

The automated news agency: SEMTEX - a text generator for German

As a by-product of the Japanese German machine translation project **SEMSYN** the **SEMTEX** text generator for German has been implemented (in ZetalISP for IBM/360/370 class machines). Its application has been to generate newspaper stories about jobmarket development. Starting point for the newspaper application are just the data from the monthly jobmarket reports (lots of unemployed, open jobs, ...). A rudimentary "text planner" takes these data and those of relevant previous months, checks for changes and significant developments, simulates possible argumentations of various political speakers on these developments and finally creates a representation for the intended text as an ordered list of frame descriptions. **SEMTEX** then converts this list into a newspaper story in German using an extended version of the generator of the **SEMSYN** project. The extensions for **SEMTEX** include: building up a representation for the context during the utterance of successive sentences that allows for avoiding repetitions in wording, avoiding reutterance of information until valid pronominalization and other types of references, grammatical tense is dynamically derived by checking the temporal information from the conceptual representations and relating it to the time of speech and the time period discussed by the story. When simulating arguments the text planner uses abstract rhetorical schemata, the generator is enriched with knowledge about various ways to express such rhetorical structures as German surface texts.

For copies of these papers please write to: Frau Martin, c/o Projekt SEMSYN, Institut für Informatik, Azenbergstr. 12, D-7050 Stuttgart 1. West Germany or e-mail to: semsyn@ifistf1.stg.