

Das Erkennen von referentiellen Bezügen in Personal Translator

Kurt Eberle
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
&
linguatec Entwicklung & Services
Hebelstr. 14
D-69115 Heidelberg
Germany
k.eberle@linguatec-es.de

- Motivation
- morphosyntaktischer Filter
(Lappin/McCord 1990)
- pragmatisches Bewerten und Filtern
(Leass/Schwall 1991, Lappin/Leass 1994)
- semantischer Filter
- Alternativansätze
- Aufgaben

1 Motivation

Kennzeichnungen können, abhängig von ihrem Bezug, unterschiedlich übersetzt werden:

1.1 Personal- und Possessivpronomen

- (1) *Paul kennt seine Frau seit langem; er ging sogar mit ihr zur Grundschule.*

Paul has known his wife for a long time; he even went to elementary school with her.

versus

- (2) *Weil er mit ihr in der Schule erwischt wurde, wurde Pauls nagelneue Pistole beschlagnahmt.*

Because he was caught with it in school, Paul's brand new pistol was confiscated.

←Bezug findet nicht nur anaphorisch statt, sondern auch kataphorisch.

1.2 Definite Kennzeichnungen

- (3) *Tonio Kröger wurde von Herrn Mann geschrieben.
Der Schreiber wurde später sehr berühmt*

*Tonio Kröger was written by Mr Mann.
The author got very famous later.*

versus

- (4) *Ingeborg unterschrieb immer mit einem Filzstift von Staedtler.
Der Schreiber wurde später richtig berühmt.*

*Ingeborg always signed with a felt-tip of Staedtler.
The writer got quite famous later.*

Ähnlich:

- (5) *Maria hat sechs Glühbirnen gekauft. Sie will die Birnen in ihren neuen Leuchter einsetzen*

Maria bought six light bulbs. She wants to put the bulbs in her new chandelier.

versus

- (6) *Maria und Paul haben einen Birnbaum. Der Baum hat dieses Jahr viele Früchte getragen. Paul hat die Birnen alle eingemacht.*

Maria and Paul have a pear tree. The tree produced a lot of fruit this year. Paul canned all of the pears.

←schwieriger: Assoziation / 'Bridging'-Phänomene

1.3 Strukturelle Unterschiede der Referenzrelation bei der Übersetzung

- Spezialisierung

Desambiguierung des Quelltexts kann notwendig sein

- (7) *The commission informed Parliament that its budgetary power ends at the point where the financial resources are exhausted.*

Die Kommission teilte dem Parlament mit, daßihr/sein Budgetrecht da endet, wo die Finanzen erschöpft sind.

- Generalisierung

Strukturell homomorphe Pronominalisierung schafft Ambiguität im Zieltext

→stilistisch unschön - nicht zugelassen

- (8) *Diese Maschine hat einen Atommotor. Gestern ist eines seiner Räder gebrochen. Wir werden ihn zurückschicken und Ersatz verlangen.*

This machine has a nuclear engine. Yesterday one of its wheels broke. We will send it back and demand a replacement.

Beispiele: Leass/Schwall

⇒ Desambiguierungskomponente Quelltext

⇒ Resolutionskomponente Zieltext

2 Slot Grammar Bindungstheorie

(McCord/Lappin 1990)

2.1 Morphologischer Filter

(9) *Ein Mann betritt die Bühne. Sie ist nicht groß.*

A man enters the stage. It isn't big.

Sie $\left[\begin{array}{l} \text{Gen: fem} \\ \text{Num: sg} \end{array} \right]$ kann sich nicht auf *Mann* $\left[\begin{array}{l} \text{Gen: masc} \\ \text{Num: sg} \end{array} \right]$ beziehen

→

Anapher und Antezedent müssen in Person, Numerus und Genus übereinstimmen !

2.2 Syntaktischer Filter

- (10) * *Eine Frau_i spricht sie_i an.*
* *Ein Mann_i kommt auf ihn_i zu.*

(Peter kommt mit einem Verband).

* *Ein Junge_i verbindet ihn_i mit ihm_i.*

* *Peters_i Bild von ihm_i*

* *Sein_i Glaube an Peter_i*

← 'disjoint reference' (Reinhart 1983 / Lasnik 1976)

← 'c-command'-Struktur (Chomsky 81, Government & Binding)

← 'o-command' (Pollard/Sag, 94 HPSG)
(obliqueness-Hierarchie)

ähnlich Lappin/McCord 90:

Ein (Personal- oder Possessiv-)Pronomen P ist nicht koreferent mit einer NP N, wenn es lokal von N o'-kommandiert wird.

Eine definite NP D ist nicht koreferent mit einer NP N, wenn sie von N o'-kommandiert wird

Sog. lexikalische Anaphern (Reflexive und Reziproke) übernehmen die Aufgabe in solchen Kontexten Bezug herzustellen, in denen für Personal- und Possessivpronomen die disjoint reference gilt.

- (11) *Peter_h kommt mit einem Verband_j.*
Der Junge_i verbindet sich_i mit ihm_j.

Der Junge_i lässt ihn_h mit sich_{i/h} alleine.

←Slot Grammar Bindungstheorie

Problem: Es können mehrere Antezedent-Kandidaten übrigbleiben !

3 Pragmatisches Bewerten und Filtern

(Leass/Schwall 1991, Lappin/Leass 1994)

(12) a) *Plötzlich sah sie einen Mann und einen Hund.*

** Sie fürchtete sich vor ihm.*

b) *Plötzlich stand sie vor einem Mann und vor einem Hund.*

? Sie fürchtete sich vor ihm.

c) *Plötzlich sah sie einen Mann mit einem Hund.*

Sie fürchtete sich vor ihm.

d) *Plötzlich sprang ein Löwe neben einem Tiger hervor.*

Sie fürchtete sich vor ihm.

ad a)

→Bei genau demselben 'Aufmerksamkeitswert':
unglückliche/nicht akzeptable Formulierung

ad b)

→Bei fast demselben 'Aufmerksamkeitswert'
(bei funktional/struktureller Gleichheit):
relative Nähe bestimmt leichte Präferenz

→*Recency Reward*

Das gilt allgemein für Sätze:

→*Sentence Recency*

(13) *Ein Mann erscheint. Ein Hund kommt hinzu. Er schaut sich um*

Ein Hund erscheint. Ein Mann kommt hinzu. Er schaut sich um.

ad c)

Matrix-Emphase

- (14) *Plötzlich sah sie einen Hund, von einem Mann geführt.
Sie fürchtete sich vor ihm.*

ad d)

Subjektsemphase

- (15) *Plötzlich sprang ein Tiger neben einem Löwen hervor.
Sie fürchtete sich vor ihm.*

Dynamische Gewichtung

- keine gleich bleibenden Gewichte
- dynamischer Abbau des Aufmerksamkeitswerts
bei ausbleibendem Wiederaufgreifen eines Diskursreferenten
- empirisch begründetes Austarieren der Gewichte gegeneinander

Neben dynamischen 'absoluten' Gewichten

Relative Gewichte

- relative Nähe zur Anapher
- Parallelismus-Bonus

(16) *Die Katze wird von der Frau verwöhnt. Sie liebt sie.*

- Kataphora-Malus

(17) a)
(α) \emptyset versus (β) Gerda_i kennt das Problem.

Damit sie_{i/j} farbige Bilddokumente drucken kann, braucht die Vorrichtung_j einen speziellen Input.

→Aber: beschriebene Situation beeinflusst die Bewertungen
(beim Rezipienten)!

(17) b) *Wenn sie (für Peter schnell) farbige Bilddokumente drucken muss, braucht die Vorrichtung einen speziellen Input.*

c) *Wenn sie farbige Bilddokumente drucken will/soll, braucht die Vorrichtung einen speziellen Input.*

4 Zusätzlicher semantischer Filter

Nutze Selektionsbeschränkungen

- des Verbs bzgl. eines Komplements:

(18) *Die Hunde entdeckten die Kekse.*

a) *Sie fraßen sie.*

b) *Sie schmeckten lecker*

- des Nomens bzgl. eines Komplements oder des Possessivdeterminers

(19) *Die Frau, die Probleme und ihre Lösung*

Die Frau, die Probleme und ihre Idee

- Allgemeiner, nutze Präsuppositionen:

(20) *Die Hunde entdeckten die Kekse.*

a) *Sie waren nicht knusprig.*

b) *Sie waren hungrig*

- Nutze Ableitungen:

- für die Bestimmung von nichtspezifiziertem Genus

(21) *Les journalistes disent la vérité et elles mentent.*
Die Journalistinnen sagen die Wahrheit und sie lügen.

oder

- zum Erhalt von Genus/Typinformation zu unbekanntem Namen:

(22) *Aldi kommt mit superbilligen neuen Produkten. Er ist Marktführer.*

Aldi comes with super cheap new products. It is market leader.

(23) *Aldu verführt seine Frau. Er ist ein Schlingel.*

Aldu seduces his wife. He is a rascal.

Probleme

- fehlende Spezifität / Vagheit
- Übertragungen (Metonymie)

(24) *Die Säbelzähntiger entdeckten die Faultiere.
Sie schmeckten lecker.*

→ Selektionspräferenzen

- Nutze dynamisch erstellte partielle Zugänglichkeitsrelation

- Nutze dynamisch erstellte partielle Zugänglichkeitsrelation

(25) a)

Wenn eine Funktion graue Bildsegmente zeigt, macht sie von den Defaultregeln Gebrauch.

b)

Wenn sie graue Bildsegmente zeigt, macht eine Funktion von den Defaultregeln Gebrauch.

c)

Eine Funktion macht von den Defaultregeln Gebrauch, wenn sie graue Bildsegmente zeigt.

d)

Sie macht von den Defaultregeln Gebrauch, wenn eine Funktion graue Bildsegmente zeigt.

Partielle Zugänglichkeitsrelation

- Flache semantische Darstellung
- Information zur Oberflächenordnung
- Rangordnung Skopus tragender Elemente

Indefinite können weiten Skopus haben gegenüber Wenn/dann-Einbettung
(→Präferenz bei Subjektsemphase: Indefinit vor Wenn/dann, etc.)

- wesentlich:
Skopus und Referenzbildung müssen zusammen entschieden werden!

(25.1) *Sie wird nicht nervös.*

Anschluss nur möglich, wenn 'sie' in (25) weiten Skopus hat und dh. nicht für *eine Funktion* steht.

Flache semantische Darstellung

Flache semantische Satzrepräsentation entsprechen Strukturen der *Flat underspecified discourse representation theory* (Eberle 1997).

(25.a) FUDRS:

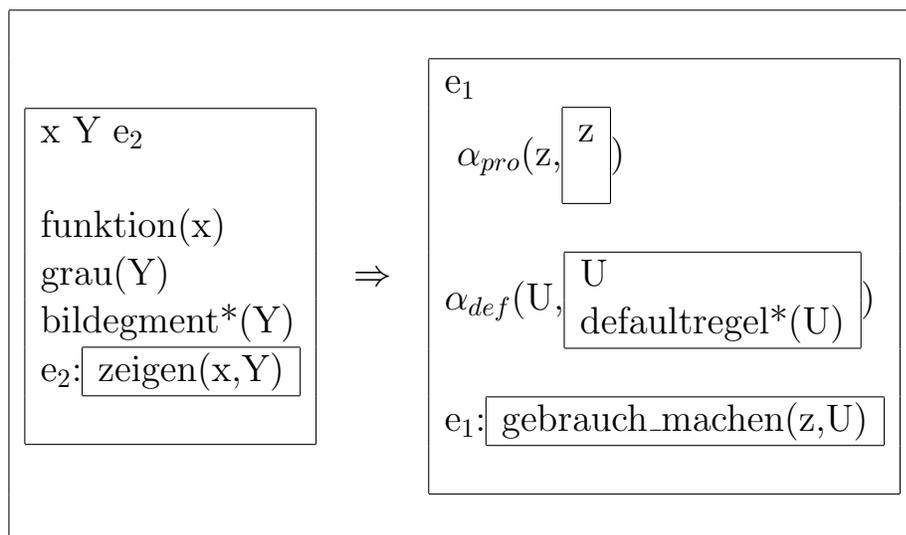
$$\left. \left. \left. \begin{array}{l} \text{vsubconj: } \underline{\text{zeigen}}_{e_2} \\ \text{subj(n): } \underline{\text{sie}}_z \\ \text{comp(p(von))}: \underline{\text{den(defaultregeln)}}_U \end{array} \right\} \right\} \&OC_1 \right\} \&OC$$

Gebrauch_machen_{e₁}

Ohne zusätzliche Annahmen sind zunächst $OC = \emptyset$ und $OC_1 = \emptyset$
wobei OC (*order constraints*) = die Menge der Angaben zum relativen Skopus der Modifikatoren von Gebrauch_machen und OC_1 die von zeigen.

Aufgrund der Oberflächenordnung und des Typs der Modifikation naheliegend: vsubconj-Operator hat Skopus über subj(n)-Operator.

Bei entsprechender Ausdifferenzierung der order constraints-Mengen, erhalten wir, nach Anwendung der Operatoren entsprechend OC und OC_1 , die folgende DRS (discourse representation structure (Kamp 1981):



⇐Der Präsuppositionstrigger (die Resolutionsaufgabe) zu z kann nach den DRT-Zugänglichkeitsbedingungen in der Restriktor-DRS aufgelöst werden, dh. z als x identifiziert werden.

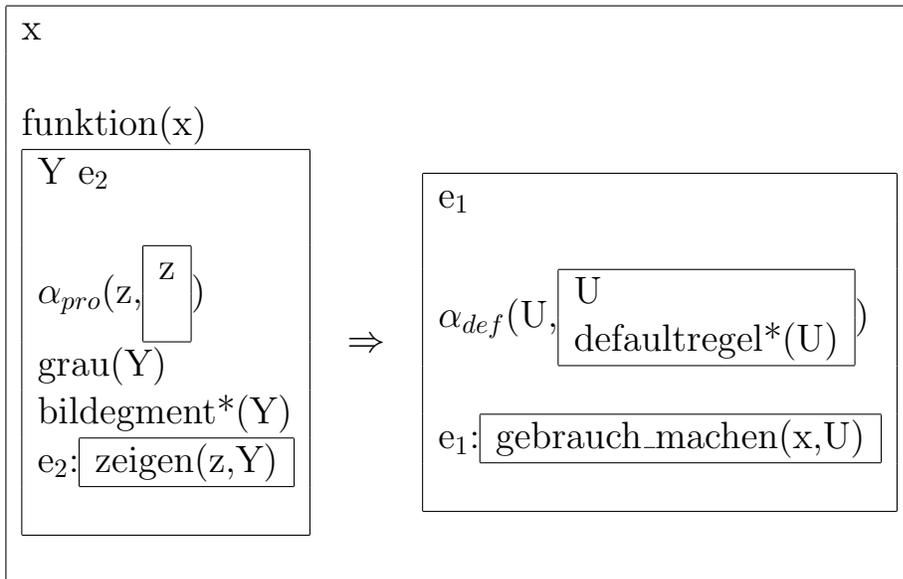
(25.b) FUDRS:

$$\left. \begin{array}{l} \left\{ \begin{array}{l} \text{vsubconj:} \quad \underline{\text{zeigen}}_{e_2} \\ \text{subj(n):} \quad \underline{\text{eine(funktion)}}_x \\ \text{comp(p(von))}: \quad \underline{\text{den(defaultregeln)}}_U \end{array} \right\} \&\text{OC}_1 \\ \underline{\text{Gebrauch_machen}}_{e_1} \end{array} \right\} \&\text{OC}$$

Wieder sind ohne zusätzliche Annahmen zunächst $\text{OC}=\emptyset$ und $\text{OC}_1 = \emptyset$

Bezug $\text{sie} \rightarrow \text{Funktion}$ ist nur möglich, wenn das indefinite Subjekt Skopus über wenn/dann erhält. Das ist durchaus naheliegend.

Bei entsprechender Ausdifferenzierung der order constraints-Mengen, erhalten wir, nach Anwendung der Operatoren entsprechend OC und OC_1 , die folgende DRS:



\Leftarrow Der Präsuppositionstrigger zu z kann nach den DRT-Zugänglichkeitsbedingungen in der Haupt-DRS aufgelöst, dh. z als x identifiziert werden.

(25.c) FUDRS:

→dieselbe Struktur, wie bei (25.b).

Nur ist an der Oberfläche das Matrix-Subjekt vor der wenn-Subordination. Damit präferieren wir hier stärker noch als bei b) die Lösung mit weitem Skopus für das indefinite Subjekt mit der nämlichen Auflösungsmöglichkeit für das Pronomen.

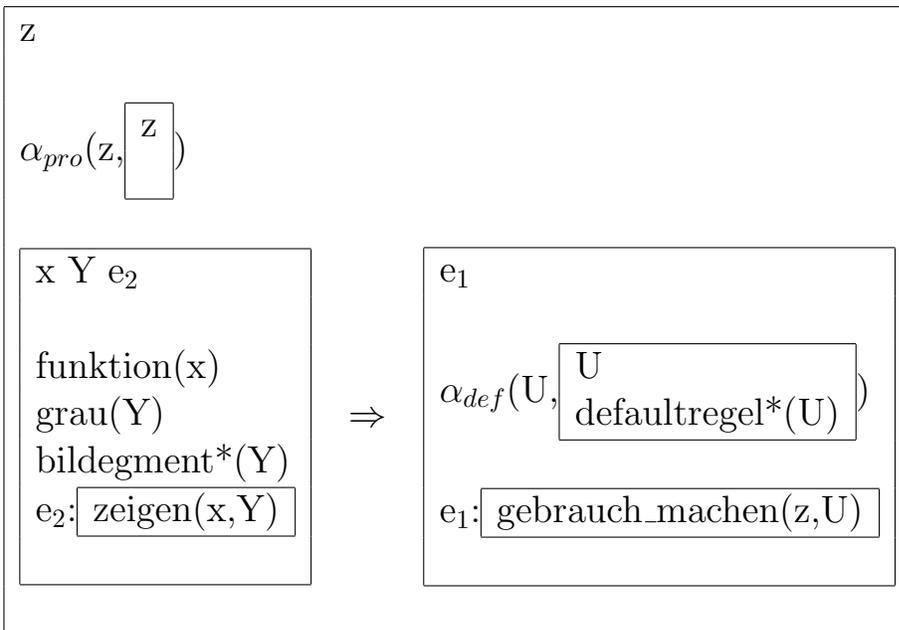
(25.d) FUDRS:

$$\underbrace{\left\{ \begin{array}{l} \text{subj}(n): \quad \underline{sie}_z \\ \text{comp}(p(\text{von})): \quad \underline{\text{den}}(\underline{\text{defaultregeln}})_U \\ \text{vsubconj}: \quad \underline{\text{zeigen}}_{e_2} \end{array} \right\}}_{\text{Gebrauch_machen}_{e_1}} \left\{ \begin{array}{l} \text{subj}(n): \quad \underline{\text{eine}}(\underline{\text{funktion}})_x \\ \text{obj}(n): \quad \underline{\text{graue}}(\underline{\text{bildsegmente}})_Y \end{array} \right\} \&\text{OC}_1 \&\text{OC}$$

Wieder, ohne zusätzliche Annahmen, zunächst $\text{OC}=\emptyset$ und $\text{OC}_1 = \emptyset$

sie in dieser Situation auf *Funktion* beziehen wollen, hieße dem Matrix-Subjekt engen Skopus gegenüber dem wenn-Operator zu geben, obwohl es an der Oberfläche vor diesem Operator an der ersten Stelle, also an der kanonischen Stelle für das schon bekannte Thema steht (das als solches natürlich weiten Skopus haben muss).

Deshalb stark präferiert oder überhaupt nur möglich
Ausdifferenzierung zur folgenden DRS:



\Leftarrow Der Präsuppositionstrigger zu z kann nach den DRT-Zugänglichkeitsbedingungen **nur** der Haupt-DRS aufgelöst, dh. **nicht** mit z identifiziert werden.

ein Test

<http://www.spiegel.de/kultur/kino/0,1518,229936,00.html>

ohne jegliche Vorbereitung

24 Sätze, im Schnitt 17 Wörter

3 männliche, eine weibliche Person, Dinge, Pluraldiskursreferenten

- Personalpronomen
19 17 ok / 2 f 89.5 %
- Possessivpronomen
13 10 ok / 3 f 76 %
- Reflexivpronomen
8 4 ok / 4 f 50 %

5 Probleme

- syntaktischer Filter:

falsche/fehlende Analyse

→56 % beim Beispieltext

typische Auswirkung:

Subj-Emphase wird konterkariert durch Analyse mit falscher subj/obj-
Interpretation

- semantischer Filter:
entsprechend + fehlende/falsche Typangaben, 'Umwertungen'

Frage: Alternativverfahren ?

6 Alternativverfahren

- statistische Verfahren
(Dagan Itai 1990)
(Ge, Hale, Charniak 1998)
- robuste Verfahren
in der Regel gemischt:
'etwas Analyse (getaggte Texte, kurze NPs) '/'etwas Statistik'
(Azzam et al 1998, Bontcheva et al 2002, Baldwin 1997, Kennedy & Boguraev, Mitkov 1997)

→ Ergebnisse:

Wenn besser, dann sehr eingeschränkte Texte oder Aufgabenstellungen.

Im General-Purpose-Bereich in der Regel viel schlechter.

Typisch: 90 % Präzision, aber 60 % Recall.

→in der Regel zu eingeschränkt in der Anwendung

→Analysen liegen in PT ohnehin vor und werden laufend optimiert

7 Resumee

- Verfahren mit sparsamem Einsatz von Semantik
 - Verzahnt mit dynamischer Spezifizierung von Lesarten
 - Besser als robuste (sparsame) Alternativverfahren im GP-Bereich
- Berücksichtige unvollständige Analysen
- Fehlerkorrektur
- Vervollständige Lexika