

Semantikkonstruktion mit DRT

Vorbemerkungen

Die wesentlichen Aufgaben und Techniken der Konstruktion semantischer Repräsentationen im Rahmen der DRT wurden in Teil II des Semantikkapitels behandelt.

Hinweise zur Implementation der λ -DRT:

- Semantikkonstruktion mit Merkmalstrukturen:
Darstellung der DRSen durch Merkmalsstrukturen, funktionale Applikation wird durch Merkmalsunifikation realisiert — siehe oben.
Hier: *Semantische Resolution (Auflösung anaphorischer Referenzen)*
Schnittstelle zur weiteren Verarbeitung: Auswertung von Ausdrücken mit DRT-Operatoren (Junktoren, Duplexbedingung) sowie Ausführung spezieller Berechnungen (z.B. Kalenderrechnung)

- Prozedurale "direkte" Implementation;
tritt an die Stelle der Konstruktion mit Merkmalstrukturen
- Eigener logischer Kalkül für DRSen; existiert als theoretischer Entwurf (z.B. Muskens u.a.), Implementation nicht bekannt
- Implementation von DRS-Konstruktion und -Auswertung in PROLOG (Referenzauflösung nach Wada und Asher):
J. Bos, P. Blackburn: Representation and Inference for Natural Language. 2 Bde., 1999 Text und Implementation online verfügbar: www.comsem.org
- Schnittstelle Syntax–Semantik mit semantischen Konstruktionsoperatoren für verschiedene theoretische Ansätze (Montague, Situationssemantik, Davidsons Ereignislogik, DRT), implementiert in PROLOG:
M. Johnson, M. Kay: Semantic Abstraction and Anaphora. COLING 1990

Semantische Resolution in der λ -DRT: Auflösung anaphorischer Referenzen (B. Geistert)

Motivation

Auflösung anaphorischer Referenzen in der Sprachverarbeitung:

1. **Anapher:** abgekürzte linguistische Einheit, deren volle Bedeutung nur unter Bezugnahme auf den Kontext geklärt werden kann, wie z.B.
 - Personalpronomen wie *er, sie, es*
 - Reflexivpronomen wie *sich*
 - definite Nominalphrasen (NPs) beziehen sich auf Bekanntes, wie *das Buch*
2. mit dem **Antezedenten** (z.B. indefinite NPs — führen Objekte ein — oder Eigennamen im vorangehenden Kontext) steht die Anapher in Beziehung (**Koreferenz**)

Antezedent steht *vor* Anapher, sonst: Katapher

Auflösung der Referenz von Anaphern hängt von mehreren Faktoren ab:

- linguistisches Wissen (Syntax und Semantik)
- Weltwissen: Inferenzen werden notwendig

These: Wissensarten sind kooperativ, Information ist redundant

⇒ gleiche Beschränkungen bei Referenzauflösung treten in mehreren Wissensquellen auf

⇒ weitergehende (aufwendige) pragmatische Analyse ist in den meisten Fällen nicht erforderlich.

Anaphernresolution in der DRT

- Ansatz zur Behandlung von Anaphern mit prozeß- und kontextorientierter Sichtweise von Semantik
- Mächtige allgemeine Behandlung von Anaphern durch Bindung von Variablen: Kontextsensitivität über die Satzebene hinaus
- Nützliche Beschränkungen der Koreferenz, die die Nicht-Akzeptierbarkeit von Sätzen erklärt (durch DRS-Zugänglichkeitsbeschränkungen)

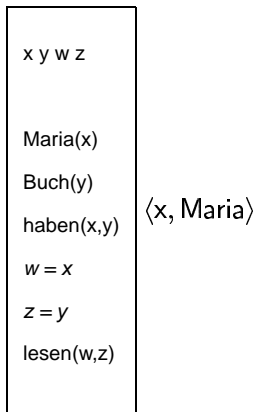
Mehrere aufeinanderfolgende Sätze: DRS-Vereinigung

$$\text{DRS-UNION } (K_1, K_2) = \langle (u(K_1) \cup u(K_2)); (con(K_1) \cup con(K_2)) \rangle$$

Für **Pronomina** wird ein neuer Diskursreferent eingeführt, der mit einem bereits vorhandenen Referenten gleichgesetzt wird.

Problem: Wahl des passenden Antezedenten

Beispiel: *Maria hat ein Buch. Sie liest es.* (Beachte Ankerliste)



Ansätze zur Anaphernresolution

- Einfache anaphorische Referenz (Allen):
 - Anapher und vorangehender Antezedent müssen in Numerus, Person und Genus übereinstimmen
 - Bindungsprinzipien: regeln Arbeitsteilung zwischen Reflexiv- und Personalpronomen
 - konfigurationelle, d.h. die syntaktische Oberflächenstruktur betreffenden Restriktionen

Problem: Komplexität der NP-vollständigen Anaphernresolution mit heuristischen Lösungsansätzen verringern

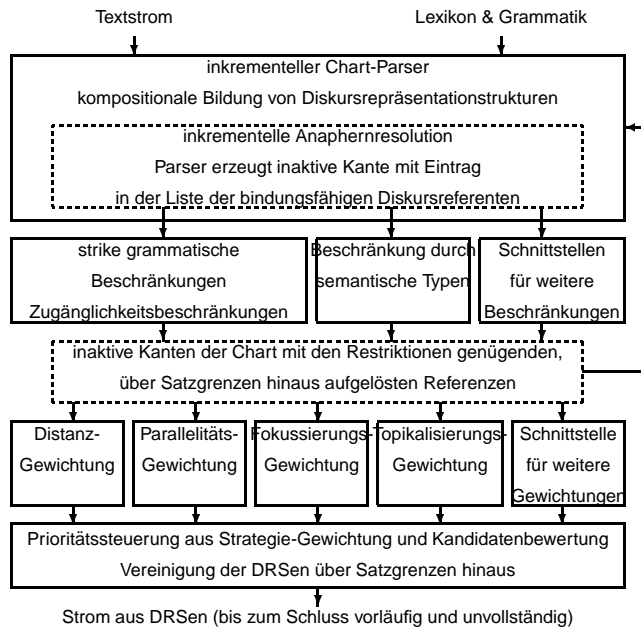
- intrasententiell: Antezedent tritt im gleichen Satz auf wie die Anapher, insbesondere Reflexivpronomina
- intersententiell: Liste von Objekten in der **history list**
- NP im Plural: Objekte zu Mengen vereinigen

Anaphernresolution in der λ -DRT

Anaphernresolution wird aufgeteilt in

- **Restriktionsmechanismen:** strikte linguistische Beschränkungen; schließen von vorneherein den Bezug zwischen Anapher und betrachtetem Antezedent aus
- **Prioritätsgewichtungen:** schwächere linguistische Einschränkungen; gewisse Lesarten werden gewichtet

- Diskursstruktur: Diskurs hierarchisch strukturiert: Stack-Modell, getriggert von Stichwörtern, Tempuswechsel, ...
- Fokussierung mit Diskursfokus (worüber der Sprecher Aussagen machen will) und Aktorfokus (Agent des letzten Satzes) nach Sidner: Aufwand für Inferenzen mit Weltwissen wird verringert

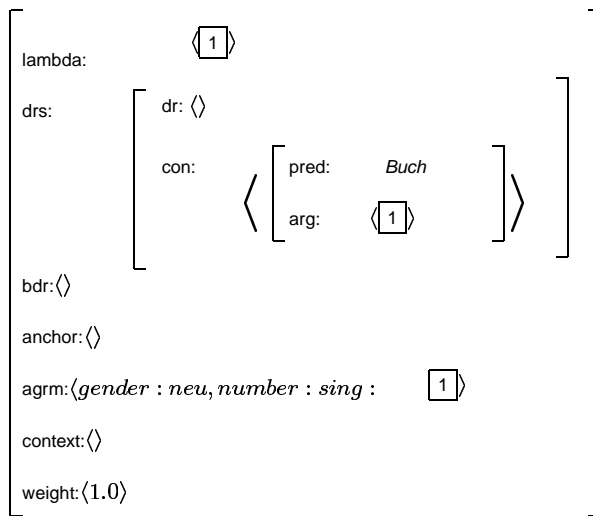


Ziel: Inkrementelle Anaphernresolution

Die Anaphernresolution soll inkrementell erfolgen. Dazu wird die λ -DRT erweitert um weitere Merkmale:

- *anchor*: Ankerlisten für Eigennamen
- *agrm*: verbindet semantische mit syntaktischen Einträgen während der Initialisierung der Chart
- *context*: Diskurskontext
- *weight*: Kantengewichtung

Beispiel:



Restriktionen

- Strikte grammatische Beschränkungen:

Kongruenz morphologischer Merkmale

- Genus (im Deutschen): m, f, n — relativ wirkungsvoller Ausschluss, z.B.
Natalia_f hat [ein Buch]_n. Sie_{f|p} liest es_n.
- Numerus: Bei Plural Gruppe von Antezedenten (ohne Genusmarkierung)
- Person: nur 3. Person; 1. und 2. Person immer deiktisch

Gleichzeitiger Vergleich der Einträge der *agrm*-Listen

Schlägt dieser fehl \Rightarrow betrachteter Antezedent wird nicht weiter bearbeitet

- Bindungsprinzipien:

Reflexivpronomina beziehen sich immer auf das Subjekt des Teilsatzes, in dem sie vorkommen: *Peter₁ mag sich₁.* ;

Personalpronomina nicht: *Peter₁ mag ihn_{*1}.*

Restriktionen (Forts.)

- **Konfigurationsbeschränkungen:** Bezüge von Nicht-Reflexivpronomen sind nicht zulässig, bei denen

- sich mehrere anaphorische Valenzstellen ein und desselben Verbs auf den gleichen Antezedenten beziehen, z.B.

*Silvia₁ und Silke₂ spielen. Sie₁ mag sie_{*1}.*

- die Anapher und der Antezedent aus einem DRS-Prädikat stammen, z.B.

*Susanne₁ mag Nicole₂ und sie_{*2}*

- sich Teile eines Summationsprädikats auf den gleichen Antezedenten beziehen, z.B.

*Andrea₁ und Maria₂ lügen. Peter₃ mag sie₁ und sie_{*1}.*

- ein Bezug von Teilen eines Summationsprädikats auf das Subjekt eines Verbs hergestellt wird, wenn das Summationsprädikat Objekt des Verbs ist, z.B.

Silvia₁ lügt. Silke₂ arbeitet. Grit₃ schläft.

*Silvia₁ mag Silke₂ und sie_{*1}.*

Prüfung der Argumente der mehrstelligen DRS-Prädikate (Verben, Summation):

Besteht Koreferenzeintrag?

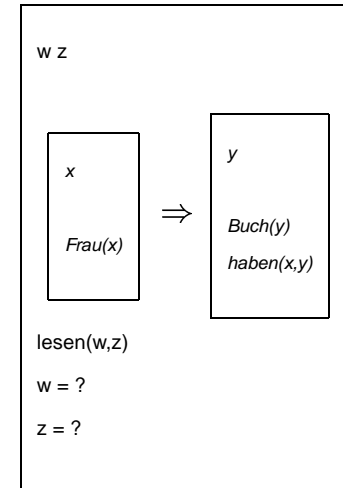
Restriktionen (Forts.)

- **DRS-Zugänglichkeitsbeschränkungen:** Struktur eines Satzes spiegelt sich in der Lage der Referenten in verschieden tief verschachtelten DRSen wider, daraus aufgestellte Zugänglichkeitsbedingungen schränken die möglichen Anaphern bei der Resolution ein.

Alle Diskursreferenten, die von einem Diskursreferenten x in der DRS K aus zugänglich sind, kann man dadurch finden, dass die Universen derjenigen DRSen betrachtet werden, die "links" oder über der DRS K liegen

Beispiel: *Jede Frau hat ein Buch. *Sie liest es.*

Das Pronomen *Sie* darf sich nicht auf *Frau*, und es sich nicht auf *Buch* beziehen:



Restriktionen (Forts.)

- Problem der Standard-DRS- Zugänglichkeitsbeschränkungen:

- Koordinierte Teilsätze

[Jede Frau]₁ hat [ein Buch]₂ und sie₁ liest es₂.

- Modale Subordination

[Jede Frau]₁ hat [ein Buch]₂. Normalerweise liest sie₁ es₂.

⇒ Nur **intersententielle Zugriffsbeschränkungen** modellieren: Von den DRSen des zurückliegenden Diskurses kommen nur die (atomaren) Bedingungen auf oberster DRS-Ebene als Antezedenten überhaupt in Frage (Ausnahme: Eigennamen kommen mit der Ankerliste nach "oben")

- **Semantische Restriktionen:** Prüfung auf Konsistenz einfacher semantischer Typen, sobald Valenzstelle bestimmt ist.

Antezedenten werden in *belebt* oder *unbelebt* eingeteilt, Anaphern übernehmen diesen Typ: Verbvalenzen sind in *belebt/unbelebt/unspezifiziert* eingeteilt, z.B.

Margit_{belebt} hat eine Zeitung_{unbelebt}.

Sie_{belebt} liest_{unbelebt} sie.

- Mögliche Erweiterungen mittels inkrementeller Schnittstelle:

- Semantische Beschränkungen (Konzeptlexikon für thematische Rollen)
- Pragmatische Beschränkungen (Vor- und Nachbedingungen)

Prioritätsgewichtungen

- **Distanzgewichtung** (zwischen Anapher und Antezedent): Der letzterwähnte, allen Restriktionen genügende Ausdruck ist mit hoher Wahrscheinlichkeit der richtige Antezedent.

- In ein und demselben Satz auftretende mögliche Antezedenten werden von der Distanz her immer gleich bewertet
- Gewichtungen werden bei jedem neuen verarbeiteten Satz abgewertet (durch Multiplikation mit Distanzgewichtungsfaktor $d = 1/2$)
- Wiederaufwertung bei Entscheidung für die Wiederverwendung

Im n -ten Satz gültige Distanzgewichtung d_l eines im l -ten Satz ($n > l$) letztmalig erwähnten Antezedenten:

$$d_l(n) = d^{n-l}$$

Prioritätsgewichtungen (Forts.)

- **Fokussierungsgewichtung**: Markierung einer Nominalphrase als prominenter Kandidat für den anaphorischen Bezug mit grammatischen (z.B. Demonstrativpronomina) oder prosodischen Mitteln (Betonung).

Hier zwei Fokussierungsalgorithmen:

- Mit grammatischen Markierungen: definite NP (statt Pronomen)
Aufwertung zusammen mit Wiederaufwertung der Distanzgewichtung, hier auf $(1 + d)/2$
- Vereinfachtes Modell der Diskurs- und Aktorfokus von Sidner:
Satz mit einem Verb mit zwei Valenzen erzeugt einen Diskursfokus (direktes Objekt) und einen Aktorfokus (Subjekt).
Ähnlich Parallelität, aber Diskurs- und Aktorfokus können über mehrere Sätze lang unverändert bleiben.

- **syntaktische Parallelität** koordinierter Satzteile

Peter schlägt Paul und Constanze tritt ihn mit dem Fuß.

Maria schenkt Claudia ein Buch und sie gibt ihr eine Zeitung.

Denjenigen Antezedenten des ersten Teilsatzes bevorzugen, der die gleiche Valenzstelle eines Verbs wie die betrachtete Anapher im zweiten Teilsatz ausfüllt.

Abwertung anderer Antezedenten mit Parallelitätsgewichtungsfaktor $p < 1$

- **Topikalisierungsgewichtung**: im Deutschen durch Satzgliederstellung eines Objekts, z.B. bei

Sich hat Hans gewählt.

Wenn die freie Wortstellung im Deutschen aufgrund von Topikalisierung in der DRS dargestellt werden soll, muss die Nominalphrase in Satzgliederstellung vom Parser als Topik markiert werden.

Aufwertung der Koreferenz auf einen topikalisierten Antezedenten mit Topikalisierungsgewichtungsfaktor t

Beispiel

- (1) *Michaela hat [einen Ball]₁.*
- (2) *[Den Ball]₁ wirft Egon₂.*
- (3) *Er ist schön.*

1. Die definite Nominalphrase in Satz (2) wird vom Parser als Topik markiert
2. Die Anapher *er* in Satz (3) erhält dann für die Referenz er_1 die Gewichtung $t \cdot (1 + d)/2 = 0,9$,
3. und für die Referenz er_2 die Gewichtung $d = 0,5$.

Inkrementelle Anaphernresolution

Vorteile eines kompositionalen und inkrementellen Ansatzes:

- Bei bereits kompositionaler Semantikkonstruktion sollen die Koreferenzen ebenso aufgelöst werden können
- Für die weitere, z.B. pragmatische, Verarbeitung der DRS stehen alle Informationen so früh wie möglich bereit
- Kognitiv adäquate Vorgehensweise, bei der Bezüge schon hergestellt werden können, bevor der *ganze* Satz verarbeitet ist.

Nachteile:

- Verarbeitung aufwendiger, da Zwischenergebnisse des Parsings betrachtet werden müssen
- Intrasententielle DRS-Zugänglichkeitsbeschränkungen können nur schwer modelliert werden

Resolution von Pronomina

Reflexivpronomina:

- Antezedent eines Reflexivpronomens ist Subjekt des Teilsatzes, in dem es auftritt
- Satzsubjekt ist Subjektstelle (1. Valenzstelle) des Verbs, in dem das Reflexivpronomen als direktes oder indirektes Objekt (2. / 3.) auftritt

Personalpronomina:

- Alle möglichen Antezedenten müssen bereits zuvor aufgetreten sein
- Intrasententiell: Suche in *antecs*-Liste
- Intersententiell: Suche in *context-drs*
- Pseudo-intrasententiell: Im Satz bereits wiederholte intersententielle Anaphern

Resolution von definiten NPs und Eigennamen

Definite NPs:

- Alle möglichen Antezedenten müssen bereits zuvor aufgetreten sein

Eigennamen:

- Treten gleiche Eigennamen in einem Text wiederholt auf, werden ihnen Koreferenzeinträge mitgegeben (Eintrag in der *bdr*-Liste)
- DRS-Zugänglichkeitsbeschränkungen sinnlos, da die Eigennamen dieses Modell mit Hilfe der Ankerliste durchbrechen

Ausblick

Ergebnisse:

- Inkrementelle Auflösung durch Restriktionen und Prioritätsgewichtungen
- Teilweise Behandlung intrasententieller Kataphern möglich
- Inkrementelle Vereinigung der Ergebnisse mit Kontext: Frühestmögliche Bereitstellung, vor Beendigung der Semantikkonstruktion
- Verwendung des besten Ergebnisses auf Satzebene zur Vermeidung kombinatorischer Explosion

Erweiterungsmöglichkeiten:

- Bisher nicht behandelte Phänomene bei Anaphern:
 - Allgemeine definite Beschreibungen
 - Allgemeine Pluralanaphern
 - Possessivpronomina, ...
 - allgemeine DRS-Zugänglichkeitsbeschränkungen

- Verwandte Phänomene:
 - Verbalphrasen- und Satzanaphern
 - Ellipsen
 - Allgemeine freie Wortstellung
 - Allgemeine DRS-Vereinigung
- Resolution mit Diskursstruktur: Diskurssegmentierung
- Problem: Fälle, in denen nicht-linguistische Plausibilitätskriterien wichtig sind, wie z.B.

Peter mag seinen Hund, obwohl er ihn manchmal beißt.