

Robuste Anaphernresolution

Roland Stuckardt
PMG Presse-Monitor GmbH, Berlin
roland.stuckardt@presse-monitor.de

1

Phänomen: Anaphora

On a dark night in Seville, **Leporello** is keeping watch, grumbling, outside a house in which **his master Don Giovanni** is engaged in **his latest amorous pursuit**. **His target is Donna Anna**, daughter of the **Commendatore**; but **she**, discovering only that **the intruder is not her fiancée Don Ottavio**, rouses the household and hampers **Giovanni's** escape. **Her father** appears and challenges **him** to a **duel which he** reluctantly accepts. **They fight**, **the old man** is killed, and **Giovanni** makes off. **Anna** has meantime fetched **Ottavio**, and finding **her father** lying dead makes **Ottavio** swear revenge on **the unknown murderer**.

2

Phänomen: Anaphora

Definition:

Eine **Anapher** ist ein sprachlicher Ausdruck, der als inhaltliche Entität (**Diskursreferent**) zu interpretieren ist, die bereits durch (mindestens) ein weiteres sprachliches Element, das **Antezedens**, entweder direkt oder indirekt in den Diskurs eingeführt wurde.

Neben dem Elementarfall der **Koreferenz** sind auch **indirekte Relationen** (z. B. Teil-Ganzes-Beziehungen) zu beobachten.

NB: Aus theoretischer Perspektive müsste von **Kospezifikation** gesprochen werden, da betreffend die Interpretation sprachlicher Zeichen die Bezugnahme auf eine **projizierte Welt** (Habel 1986: kognitives Substrat) zu postulieren ist.

- Element der textuellen **Kohäsion**

3

Phänomen: Anaphora

Fokussierung der Betrachtungen auf

- Objekt-/NP-Anaphora
- Kospezifikation
- (schwerpunktmäßig) Dritte-Person-Pronomen

Nicht näher betrachtet werden indirekte Relationen sowie Sachverhalts- und VP-Anaphora

Kurt ging zu seinem Wagen. Die Scheiben waren vereist. Er holte den Eisschaber aus dem Handschuhfach und machte sich an die Arbeit. Hierbei hatte er große Mühe. Dies hätte er nicht erwartet.

4

Frühe Ansätze

- Frühe Arbeiten zur Anaphernresolution in den 60er und 70er Jahren
- Interpretation **nominaler** oder **pronominaler** Anaphern in **restringierten** Domänen
- Beispiele:
 - Bobrow, 1964(!): STUDENT
 - Winograd, 1972: SHRDLU / Blocksworld,
 - Woods, 1973: LUNAR
 - Wilks, 1973: Preference Semantics

5

Frühe Ansätze

Beispiel: **Bobrow, 1964, STUDENT**

„High-school algebra problem answering system“:

The number of soldiers the Russians have is half the number of guns they have. The number of guns is 7000. What is the number of soldiers they have?

Einfachstes *Pattern Matching* - nicht abgedeckt wird:

The number of soldiers the Russians have is half the number of guns they have. The number of guns is 7000. How many soldiers do they have?

Frühe Ansätze

Beispiel: **Wilks, 1973, Preference Semantics**

Pronominale Anaphernresolution als Komponente eines Ansatzes zur maschinellen Übersetzung Englisch – Französisch:

4 Schwierigkeitsstufen pronominaler Anaphernresolution – zur Behandlung der einfachsten Fälle (**Stufe A**) reicht die **Anwendung „elementarer“ selektionaler Präferenzkriterien**:

Give the bananas to the monkeys although they are not ripe, because they are very hungry.

Komplexere Fälle benötigen „analytic inference“ (**Stufe B**), „inference using real-world knowledge“ (**Stufe C**) bzw. „focus of attention rules“ (**Stufe D**).

Merkmale der frühen Ansätze

- zugeschnitten auf das jeweilige Anwendungsgebiet
- basierend auf domänenspezifischen semantischen Kriterien und entsprechender Wissensrepräsentation
- erfordern i. A. aufwändige Feinkonfiguration

→

- „skalieren“ demnach schlecht
- **nicht geeignet für die domänenübergreifende Verarbeitung von Massentexten**

8

Literatur zu den frühen Ansätzen

- Ruslan Mitkov. Anaphora Resolution. Longman, 2002. Kap. 4.

... immer noch einen Blick wert ist:

- Graeme Hirst. Anaphora in Natural Language Understanding. A Survey. LNCS 119, Springer, 1981

9

Ziele: Skalierbarkeit, „Robustheit“

Beobachtungen:

- die genannten Verfahren setzen i. A. (zu) viel voraus
- die benötigten inhaltlichen Tiefenrepräsentationen können allenfalls für Texte der restringierten Anwendungsdomäne auf **algorithmischem** Wege erstellt werden

→

- sog. „seichte“ Ansätze – z. B. **Carter, 1987**:
 “Shallow Processing Anaphor Resolver” (SPAR)
- Jedoch: Einschränkung auf simple “story“-Domäne. Rekurs auf ein nur partiell algorithmisiertes Fokusmodell (Sidner)

10

→ ... rein „syntaktische“ Verfahren?

Hobbs, 1978: “Naïve Algorithm”

- Pronomenresolution qua Suche in der syntaktischen Oberflächenstruktur
- expliziter Test auf Kongruenz in Numerus und Genus
- implizit Algorithmisierung einer Präferenz naher Antezedenskandidaten
- implizit Algorithmisierung satzinterner syntaktisch-konfiguraler Koindexierungsbedingungen
- jedoch werden auch satzübergreifende anaphorische Bezüge in Betracht gezogen
- Basisverfahren erzielt Genauigkeit von **88.3%** auf Pronomen

Setzt Verfügbarkeit eindeutiger Syntaxbäume voraus
→ **nicht wirklich “robust”!**

11

Mehrstrategie-Paradigma

Carbonell & Brown, 1988:

- Unterscheidung zwischen **Restriktionen** und **Präferenzen**.
- **Rahmenstrategie**:
 1. Ermittle Anaphern und zugehörige Antezedenskandidaten.
 2. Für jede Anapher A: Eliminiere diejenigen Kandidaten K, die kein Antezedens sein können, da eine oder mehrere **Restriktionen** verletzt wäre(n).
 3. Für jede Anapher A: Wähle aus den verbliebenen Kandidaten denjenigen aus, der gemäß Vorhersage der **Präferenzen** zu bevorzugen ist.
- ↔ **Restriktionen** werden **so früh wie möglich** angewendet, um den Suchraum rasch zu verkleinern.

12

Restriktionen

Kongruenz in Numerus und Genus bzw. Sexus

*Die Ehefrau begrüßte den Architekten auf der Terrasse.
Er überreichte ihr einen Strauß Blumen.*

Das Mädchen trug ein neues Kleid. Es/Sie war 7 Jahre alt.

*Die Ehefrau begrüßte die Architekten auf der Terrasse.
Sie trug ein Abendkleid.*

Restriktionen

Syntaktisch-konfigurationale Bedingungen

*Behrens, rasiert sich.
* Behrens, verlangt, dass der Friseur_i sich, rasiert.*

** Behrens, rasiert ihn,
Behrens, verlangt, dass der Friseur_i ihn, rasiert.*

** Behrens, rasiert Behrens.
* Behrens, informierte Gropius_i, bevor Behrens, abreiste.
Gropius_i, informierte Behrens_i, bevor Behrens, abreiste.*

Restriktionen

Semantische Skopus-Restriktionen

*Jede Bäuerin_i, die, einen Esel_j besitzt, schlägt ihn_j.
* Er_j ist stets sehr störrisch.*

Analyse gemäß DRT (Kamp, 1981):

$\text{Forall } x.(B(x) \Rightarrow \text{Exists } y.(E(y) \ \& \ s(x,y)))$

y und auch x sind innerhalb des Skopus des Allquantors **lokal gebunden** und daher nicht von außen per anaphorischem Verweis aufgreifbar (sog. **accessibility constraint**).

Jedoch gibt es **Gegenbeispiele** – z. B. (Asher & Wada 1988):

Every cadet receives a rigorous training. First, he goes to boot camp. Then, he gets intensive flight instructions in a basic trainer. Finally, he is given 200 hours of flight time in a supersonic aircraft.

Restriktionen

□ **Selektionale Restriktionen**

Behrens_i gab Gropius_j einen Apfel_k. Er_{ij} aß ihn_k.

Disambiguierung der Referenz von "Er_{ij}" zugunsten der Lesart "Er_j" über sog. ...

□ **Precondition/Postcondition Constraints** (Carbonell & Brown)

□ Jedoch Vorsicht – was als "essbar" anzusehen ist, hängt stark vom Kontext ab:

Der Großmeister verspeist mit dem Turm den Freibauern.

Restriktionen

□ **Fokusbasierte Bedingungen (1)** ↔ **Nutzwert?**

*Kurt_j hat ein Saxophon. Er_i spielt Anja_j darauf vor.
Sie_j bewundert ihn_i / Kurt_j.*

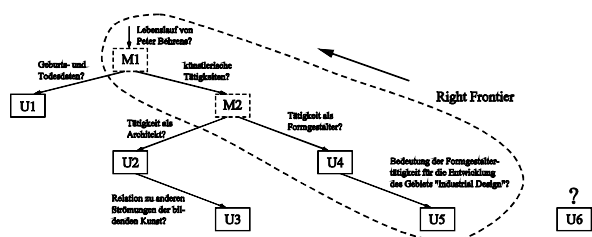
Centering-Theorie (Grosz, Joshi, Weinstein 1983/1995):
Regel R1 sagt voraus, dass Diskursreferent <Kurt> in Satz 3 pronominal zu realisieren ist:

Ist ein in den Äußerungen U_i und U_{i+1} erwähnter Diskursreferent in U_{i+1} pronominal realisiert, so ist auch Cb(U_{i+1}) pronominal realisiert.

□ **Fokusbasierte Bedingungen (2)**

Bestimmte Pronomen können nur solche Diskursreferenten wiederaufgreifen, die im offenen Teil (sog. **Right Frontier**, Webber (1988)) der Diskursstruktur vorkommen.

Right-Frontier-Bedingung nach Webber



Präferenzen

Fokusbasierte Präferenzen

*Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor.
Er_i hat ihn_j eingeladen.*

= bevorzugte Lesart gemäß **Centering-Theorie**, deren Regel **R2** besagt:

Bevorzuge CONTINUE vor RETAIN vor SHIFT-1 vor SHIFT.

Die alternative Interpretation wäre (nur) ein "RETAIN":

*Kurt_i hat ein Saxophon. Er_i spielt Peter_j darauf vor.
Er_j hat ihn_i eingeladen.*

Implementierung

Lappin & Leass: *An algorithm for pronominal anaphora resolution.* Computational Linguistics 20(4), 1994, 535-561.

- **System RAP** ("Resolution of Anaphora Procedure")
- implementiert Mehrstrategieverfahren, umfassend:
 - Kongruenzrestriktion
 - syntaktisch-konfigurationale Bedingungen
 - syntaktisch verankerte Prominenzfaktoren („salience“)
- syntaktische Voranalyse gem. Slot Grammar (McCord)
- interpretiert werden Pronomen in 3. Person inkl. Poss., Refl.

Allerdings: **Per-Hand-Korrektur** der Parses bzw. des Inputs.

Kritische Bestandsaufnahme

Grundlegende Probleme mit zahlreichen Strategien:

- Verifikation der syntaktisch-konfiguralen Bedingungen setzt Verfügbarkeit oberflächenstruktureller Beschreibungen voraus.
- Die klassischen Accessibility Constraints der DRT scheinen in Anwendungskorpora nur selten zum Tragen zu kommen; zudem wird zur Behandlung komplexerer Fälle Weltwissen benötigt – siehe etwa Asher & Wada (1988).
- Die klassische Centering Theory ist unterspezifiziert – speziell betreffend die der Definition des Vorwärtsgerichteten Centers Cf(Ui) zugrunde liegenden Prominenzordnung.
- Die fokusbasierten Restriktionen nach Webber setzen auf einem Diskursstrukturbaum auf, dessen Konstruktion auf komplexen Inferenzen sowie der Verfügbarkeit von Weltwissen aufbaut.

→ Keine geeignete Grundlagen für eine **robuste** Implementierung!

Ziel: robuster regelbasierter Ansatz

Auswahl viel versprechender Strategien mit Blick auf die Kriterien

- benötigte Ressourcen (Syntaxbäume, Weltwissen etc.)
- Algorithmisierbarkeit
- geschätzte Relevanz

→ **Kandidaten**:

- Kongruenz in Numerus und Genus/Sexu
- syntaktisch-konfigurationale Bedingungen
- fokusbasierte Präferenzen
- selektionale Präferenzkriterien à la Wilks (Stufe A)

Syntaktisch-Konfigurationale Bedingungen – ein genauerer Blick:

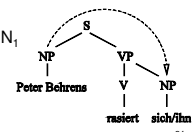
Formalisierung: (z. B.) in den **Bindungsprinzipien** der Bindungstheorie (Chomsky 1981, 1986):

- **BP A**: Reflexivpronomen sind in ihrer Bindenden Kategorie gebunden.
Behrens_i rasiert sich_i.
- **BP B**: Nichtreflexive Pronomen sind in ihrer Bindenden Kategorie frei (= nicht gebunden).
** Behrens_i rasiert ihn_j.*
- **BP C**: Nichtpronominale Anaphern („R-Ausdrücke“) sind global frei.
** Behrens_i informierte Gropius_j, bevor Behrens_i abreiste.*

→ Begriffe „Bindung“, „Bindende Kategorie“?

Unter Bezugnahme auf die „Konfiguration“ der entsprechenden NP-Knoten in der syntaktischen Oberflächenstruktur:

- Knoten N_1 ist **Bindende Kategorie** von Knoten N_2 g. d. w. N_1 der nächste Knoten ist, der N_2 dominiert und ein (näher zu definierendes) SUBJEKT N_3 enthält, das N_2 **k-beherrscht**. (*modelliert Lokalität*)
- Knoten N_1 **bindet** Knoten N_2 g. d. w. N_1 und N_2 **koindexiert** sind und N_2 von N_1 **k-beherrscht** wird. (*asymmetrische Dominanzbeziehung*)
- Knoten N_1 **k-beherrscht** Knoten N_2 g. d. w.
 - der nächste verzweigende Knoten, der N_1 dominiert, dominiert auch N_2 ;
 - weder wird N_2 durch N_1 dominiert, noch wird N_1 von N_2 dominiert, noch ist $N_1 = N_2$.



© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Syntaktisch-konfigurationale Bedingungen – ein genauerer Blick:

Insofern eindeutige syntaktische Oberflächenstruktur zur Verfügung steht, ist die Angelegenheit (vergleichsweise) einfach:

Peter Behrens rasiert sich/ihn/Peter Behrens.

Peter Behrens verlangt, dass der Friseur sich/ihn/P.B. rasiert.

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Syntaktisch-konfigurationale Bedingungen – ein genauerer Blick:

Doch die Verhältnisse – sie sind nicht so! ... der Parse ist (oft) schlecht:

Behrens beobachtet den Eigentümer des Feldstechers mit ihm.

Beobachtung: Vorhersagen der syntaktisch-konfiguralen Bedingungen hängen davon ab, wie die PP interpretiert wird.

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Syntaktisch-Konfigurationale Bedingungen – ein genauerer Blick:

Parsing ist schwierig:

- strukturelle Ambiguitäten (PP, Adverbialsatz, ...)
- Mehrdeutigkeit der syntaktischen Funktion
- leere Kategorien
- ...

→ Unter Anwendungsbedingungen ist mit **defizienten** syntaktischen Beschreibungen zu rechnen.

→ zu fordern: **Robustheit gegen defiziente Syntax**

Apropos: **Was heißt "Robustheit (gegen X)"?**

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Robuste Verarbeitung natürlicher Sprache

Definition gemäß **Menzel** (1995):

- **Autonomie:** Ein Scheitern der Verarbeitung auf einer Stufe der Analyse sollte kein Scheitern auf anderen Verarbeitungsstufen nach sich ziehen.
- **Monotonie:** Je besser (weniger defizient) der Input, desto höher die Qualität des Outputs.

Erweitert:

- **Interaktion:** Defizienz auf einer Stufe der Analyse sollte durch den Beitrag anderer Verarbeitungsstufen kompensierbar sein.

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Robuste Anaphernresolution

Zwei Modelle, um Autonomie und Monotonie zu erreichen:

- **Seichte Voranalyse:** Anstelle einer vollständigen syntaktischen Beschreibung tritt eine seichte Repräsentation der Syntax, auf der die Anaphernresolutionsstrategien emuliert werden. Prominenter Vertreter: **Parserlose Anaphernresolution** n. Kennedy & Boguraev (1996)
- **Potenziell fragmentarische Voranalyse:** Adaption der Anaphernresolutionsstrategien dahingehend, dass sie auf den potenziell fragmentarischen Voranalyseergebnissen eines uneingeschränkten Parsings arbeiten. **Hier!** ↔ Idee: **möglichst viel syntaktische Information**

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Robuste Anaphernresolution über potenziell fragmentarischer Syntax

In vielen Fällen bleibt genügend Information für die nichtheuristische Verifikation der syntaktisch-konfiguralen Bedingungen erhalten:

* Kohl_i ging, bevor der Präsident kam, weil das Staatsoberhaupt sich_i verspätete.

Regelmuster: * F_d = [... K_{typ A/B/C} ...], F_e = [... bk(A)(... A_{typ A} ...)₋₃₀₇]

Robuste Anaphernresolution über potenziell fragmentarischer Syntax

Identifikation typischer Fälle syntaktischer Fragmentierung, in denen die Verifikation der syntaktisch-konfiguralen Bedingungen verlustfrei möglich ist.

→ **8 Regelmuster.**

Rahmenstrategie:

- Anapher A und Kandidat K in selbem Fragment
→ verfahren wie bei nichtfragmentarischer Syntax.
- A und K in unterschiedlichen Fragmenten:
 - Falls Regelmuster anwendbar → entscheide gemäß diesem;
 - sonst: Nimm heuristisch an, dass die Koindexierung von A und K bindungstheoretisch zulässig ist.

→ **syntaktische Information gut ausgenutzt.** (↔ **Monotonie**) 31

Robuste Implementierung der Präferenzen

Die Bedingungen der fokusbasierten Modelle lassen sich in eine Menge empirisch gleichwertiger **elementarer Präferenzbedingungen** – sog. **Salience-Faktoren** – auflösen:

- Präferenz unlängst **topikalisiert** realisierter Diskursreferenten:
Es war Gropius, der am Bauhaus zu Weimar lehrte.
- Präferenz von Diskursreferenten, die unlängst in der syntaktischen Position des **Subjekts** realisiert wurden:
- allgemein: **Präferenzhierarchie syntaktischer Funktionen**
Subjekt > direktes Objekt > indirektes Objekt > PP-Objekt
- **Oberflächen-Distanz** zur letzten Realisierung des Diskursreferenten 32

Präferenzen

Ferner: **relationale Präferenzbedingungen:**

□ Rollenträgheit:

Behrens traf Gropius. Er bat ihn um die Pläne.

Vorhersage:

Er [subje] → *Behrens* [subje]
ihn [trans] → *Gropius* [trans]

Keine Fokusbedingung im engeren Sinne!

Reproduziert jedoch in vielen Fällen die Vorhersage von R2 der Centering-Theorie.

Psycholinguistische Untersuchungen: Smith & Chambers, 1996, 33

Implementierung

Das System **ROSANA:**



Robuste syntax-basierte Anaphernresolution

- implementiert in Common Lisp
- Input:
 - potenziell fragmentarische syntaktische Beschreibungen, erzeugt mit dem Dependency Parser for **English** (Järvinen & Tapanainen, 1997)
 - morphologische Analyse (ENGTWOL)
- **Deutsche** Version, basierend auf dem CONEXOR-Parser 34

Korpusbasierte Evaluation

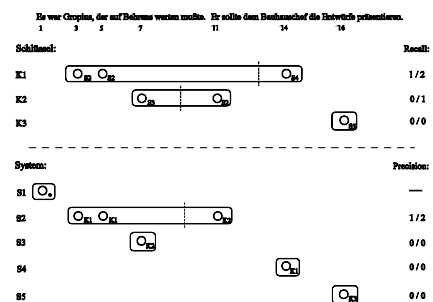
Behrens_i ist berühmt. Er_i baut Häuser_j. Seine_j Arbeit_k ...

Evaluationsdisziplinen: Ermittlung von

- **Vorkommen:**
"Behrens", "Er", "Häuser", "Seine", "Arbeit", ...
- **Antezedenten:**
"Er" → "Behrens", "Seine" → "Er", ...
- **Klassen kospezifizierender Vorkommen:**
{ "Behrens", "Er", "Seine" }, { "Häuser" }, { "Arbeit" }, ...
- **nichtpronominalen Anknern:**
"Er" → "Behrens", "Seine" → "Behrens", ... 35

Evaluation Kospezifikationsklassen-Ermittlung

„Modelltheoretisches“ Bewertungsschema, CO-Task MUC-6, Vilain et al. (1996):



Korpusbasierte Evaluation

Korpora (englischsprachig):

- 66 **Pressemeldungen**: 31 Entwicklungsdokumente; 35 Evaluationsdokumente mit 12904 Token und 479 Pronomen.
- 3 Texte mit Beschreibungen von **Mozartopern**

Ergebnisse:

- Vorkommen: (P,R) = (0.94,0.96) bzw. (0.95,0.98)
- Antezedenten:
für 3.-Pers.-Pronomen Genauigkeit von 0.75 bzw. 0.79
- Kospez.-Klassen: (P,R) = (0.81,0.68) bzw. (0.88,0.81)
- nichtpronom. Anker: (P,R) = (0.7,0.65) bzw. (0.75,0.74)

37

Korpusbasierte Evaluation

Vergleich Ergebnisse, Disziplin "unmittelbare Antezedenten":

- Lappin & Leass: 0.86 (nicht-robust)
- Kennedy & Boguraev: 0.75 (robust, seichte Voranalyse)
- ROSANA: 0.75 bzw. 0.79 (robust, fragmentar. Voranalyse)

Qualitative Analyse der Fehlerfälle:

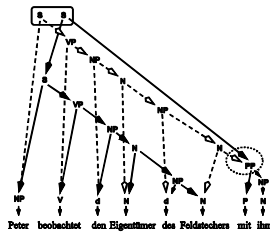
- > 30 % falsch interpretierte Pronomen aufgrund kandidaten-seitig falsch zugewiesener Genusattribute (ENGTWOL)
- weitere 30 % falsch interpretierte Pronomen außerhalb des Skopus eines rein syntaktischen Verfahrens
- nur 7 Fehler (= 2.3% von 246) bedingt durch heuristischen Anteil der syntaktisch-konfig. Filters ↔ **nahezu optimal**
- ... meist Folgefehler **falscher** (im Unterschied zu partieller) syntaktischer Beschreibungen

38

Noch größere Robustheit inkl. **Interaktion**?

Verifikation der syntaktisch-konfiguralen Bedingungen auf **Packed Shared Forests**:

- A.R.-Ergebnisse leisten **Beitrag zur strukturellen Disambiguierung**
- prinzipiell möglich (→ ACL-Workshop 1997)
- obiges Verfahren jedoch für Anwendungen ausreichend



39

Korpusbasierte Anaphernresolution?

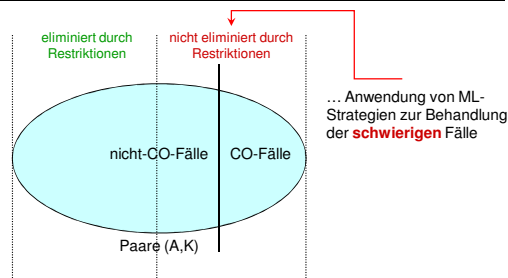
Mögliche **Anwendungsfälle statistischer bzw. ML-Techniken**:

- **Restriktionen?**
 - lassen sich mit sehr geringer Fehlerquote robust regelbasiert implementieren
 - besitzen domänen- sowie genre-übergreifende Gültigkeit
- **Präferenzen?**
 - mit Unschärfe behaftet
 - empirische Evidenz, dass Saliency-Faktoren domänen- bzw. genre-spezifisch konfiguriert werden sollten

→ **Hybridverfahren: "Lerne Präferenzen!"**

40

Hybridverfahren



41

Korpusbasierte Ansätze

Vorarbeiten (Auswahl):

- ausschließlich ML / Statistik:
 - Soon, Lim & Ng (2001): Entscheidungsbäume
 - Aone & Bennett (1995), McCarthy & Lehnert (1995): Entscheidungsbäume
 - Connolly, Burger & Day (1994): Naive Bayes, Neuronale Netzwerke, Entscheidungsbäume
- Hybrid, ML / Statistik zum Lernen der Restriktionen (!):
 - Paul, Yamamoto & Sumita (1999): Entscheidungsbäume
- **Hybrid, ML / Statistik zum Lernen der Präferenzen**:
 - Dagan, Justenson, Lappin, Leass, Ribak (1995): RAP-Verfeinerung, selektionale Präferenzen als **Ergänzung** anderer Präferenzkriterien
 - Ge, Hale & Charniak (1998) (... allerdings nicht robust)

42

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Algorithms

Anaphor A, candidates C

- **anaphor resolution:**
 1. apply candidate filters:
 - number/gender agreement, syntactic disjoint reference, recency
 2. score and rank remaining candidates according to **ML strategy prediction** and recency
 3. select highest ranking candidate as antecedent
- **training data generation:**
 1. apply candidate filters (according to chosen data generation mode)
 2. generate feature vectors:
 - for each remaining candidate C: generate training case $fv(A,C)$
 3. classify training cases $fv(A,C)$ by consulting annotated corpus ($\rightarrow fv(A,C)::K$)
- **ML classifier learning:**
 1. **learn ML strategy** (e. g., decision tree, neural network, ...) over the classified training cases

43

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Sources of evidence

set of **robustly computable features:**

feature	examples of instances
type (O)	PER3, POS3, NAME, CN, ...
synfun (O)	subje, trans, ...
number (O)	SG, PL, SGPL
gender (O)	MA, FE, NEU, MAFE, ...
dist (A,C)	INTRA, PREV, PPREV
synpar (A,C)	YES, NO
subject (O)	YES, NO
pronoun (C)	YES, NO
theNP (C)	YES, NO
...	...

A = anaphor,
C = candidate,
O in { A, C }

\rightarrow experiments with different **signatures**

44

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

\rightarrow ML-Abkömmlinge von ROSANA

- **ROSANA-ML** (DAARC 2002):
Entscheidungsbäume vom Typ **C4.5** (Quinlan, 1993)
- **ROSANA-NN** (DAARC 2007):
Neuronale Netze vom Typ **Backpropagation**

Implementierungen nach Mitchell, 2004.

Empirische Optimierung:

- Attribut-Signaturen
- Trainingsdaten-Generierungsmodi
- *Pruning Confidence Factor* CF (C4.5)
- # innere Knoten, *learning rate*, *momentum* (Backpropagation NN)
- ...

45

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Evaluationsergebnisse im Vergleich

System	Setting	Corpus	im. antecedents: A_{1a}		non-pr. anchors: A_{na}	
			PER3	POS3	PER3	POS3
ROSANA	std.	[d ³¹ , d ₃₂ ⁵⁶]	0.71	0.76	0.68	0.66
ROSANA-NN	(a,D)	6-cv(d ₁ ⁵³)	0.64	0.74	0.61	0.64
ROSANA-ML	(1 _{nc} ^{1c} ,h)	6-cv(d ₁ ⁵⁶)	0.66	0.75	0.62	0.68
	(1 _{nc} ^{1c} ,h)	[d ³¹ , d ₃₂ ⁵⁶]	0.65	0.76	0.62	0.73

- regelbasierter Ansatz ROSANA sieht aus wie der Sieger
- ROSANA-NN und ROSANA-ML in etwa gleichauf
- ... jedoch vergleichsweise wenig Trainingsdaten
 \rightarrow vermutlich Spielraum für Verbesserung der ML-Verfahren.

46

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Anwendungen

- **Koreferenz-basierte Text Summarization (TS):** z. B. Baldwin & Morton (1998), Azzam, Humphreys, Gaizauskas (1999)
- **Koreferenz-basiertes Question Answering (QA):**
z. B. Breck et al. (1999), Morton (1999)

47

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Koreferenzinformation für TS und QA

Gemeinsame Strategie:

- Schritt 1: Inbeziehungsetzung der sprachlichen Ausdrücke einer Anfrage mit den Ausdrücken im Text (QA, user-focused TS)
- Schritt 2: Rückgriff auf Koreferenzinformation:
 - QA: Koreferenz**klassen** nutzen, um relevante Information zu finden.
 - TS: Traversierung von Koreferenz**ketten** und Auswahl einer Teilsequenz von Sätzen.
 - TS & QA: Identifikation informativer Antezedenten für Anaphern (\rightarrow nichtpronominale Anker)

48

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Schritt 1: Zugriff auf anfragerrelevante Koreferenzklassen

- Zugriff über **informative** Vorkommen
- → Pronomen und informative(re)n Vorkommen kommt unterschiedliche Bedeutung zu

49

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Bewertung inkorrektter Koreferenz- bzw. Antezedens-Entscheidungen

50

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Bewertung inkorrektter Koreferenz- bzw. Antezedens-Entscheidungen

Beobachtung:
Nach dem modelltheoretischen Koreferenz-Bewertungsschema (MUC-6, Vilain et al, 1996) zählen folgende Fehler gleich viel:

- (1) Leporello ← he ← him ← his
- (2) Leporello ← he ← him ← his

Jedoch hat Fehler (1) i. d. R. erheblich größere negative Auswirkungen auf die Qualität eines nachgeschalteten TS bzw. QA.

Zudem ist die **erstmalige** pronominale Realisierung eines Diskursreferenten vergleichsweise schwieriger zu resolviere.

51

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

→ Perspektivwechsel: **Anaphern**resolution

... genauer:
Entscheidend für die Qualität einer "Koreferenz-basierten" TS bzw. QA ist eine möglichst hohe Genauigkeit betreffend die Ermittlung von **nichtpronominalen Anker**.

... empirische Experimente bestätigen:
Precision-Fehler (= falsch resolvierte Anaphern) sind problematischer als Recall-Fehler (= unresolvierte Anaphern) (Paper ARQAS 2003)

→ **high precision anaphor resolution**

52

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Hochpräzisionsverfahren:

Weiterentwicklungen:

- **ROSANA-θ**: („θ“ = Saliency-Schwellenwert)
- **ROSANA-ML-θ**: („θ“ = Schwellenwert betreffend die C4.5-Entscheidungskonfidenz)

→ **unterschiedliche** (P,R)-Tradeoffs erzielbar.

Erzielte Ergebnisse, Disziplin „nichtpronominale Anker“: (P_{na}, R_{na}) für

- Nichtpossessiva: (0.77,0.56) (ROSANA-θ)
- Possessiva: (0.83,0.54) (ROSANA-ML-θ)

... besser als CognIAC (Baldwin, 1997)

53

© Roland Stuckardt Computerlinguistisches Kolloquium, Jonas Kuhn / Manfred Stede, Universität Potsdam, 23. Januar 2008

Ausblick

Zentrale **Forschungsthemen**:

- umfangreichere referentiell annotierte **Korpora**
→ Fortschritt bei statistischen und ML-Verfahren
- sequentielle, „kompetenz-orientierte“ **Architekturen**
Lappin (2005): *A Sequenced Model of Anaphora and Ellipsis Resolution*.
- **empirische Verankerung theoretischer Modelle**
Poesio, Stevenson, Di Eugenio, Hitzeman (2004): *Centering: A Parametric Theory and Its Instantiations*. CL 30(3).
- **Resolution nichtpronominaler Anaphern**
defNP, VP Anaphora
- **Anwendungen**

54

DAARC

Discourse Anaphora and Anaphor Resolution Colloquium (DAARC):

- 1996: Lancaster
- 1998: Lancaster
- 2000: Lancaster
- 2002: Lissabon
- 2004: Furnas, São Miguel, Azoren
- 2007: Lagos, Algave
- **António Branco**, Univ. Lisbon, Tony McEnery, Univ. Lancaster
- Ruslan Mitkov, Univ. Wolverhampton, Fátima Silva, Univ. Oporto
- 2009: Madeira(?)

55

Vielen Dank!

Weitere Informationen unter:

www.stuckardt.de



www.presse-monitor.de

PMG Berlin

PMG-Press-Monitor®

56