

Diskurs: Pronomenresolution und Koreferenzresolution

Katja Markert, einige Folien von Yannick Versley

January 28, 2020

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen
- 4 Soon et al: ML, pairwise mention-based
- 5 Evaluation

Bis jetzt: Restriktionen und Präferenzen für Pronomenresolution

- Agreement: Numerus, Genus, Person
- Syntaktische Restriktionen:
 - c-commando: “John bought him a car”
 - i-within-i: “The son of the woman who killed him was 20 years old.
- Recency preference
- Präferenz von grammatischen Rollen
- Repeated mention preference
- Parallelismus Präferenz
- Weltwissen und semantisches Wissen

C-commando (eine Definition): A c-commands B, genau dann wenn weder A B dominiert, noch B A dominiert und außerdem B vom selben Knoten dominiert wird, der A direkt dominiert

- Pronomenresolution: Der Hobbs-Algorithmus und der Lappin und Leass Algorithmus
- Koreferenzresolution: Stanford Sieve System (regelbasiert)
- Koreferenzresolution: Überwachtes maschinelles Lernen
- Evaluation von Koreferenzresolution

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen
- 4 Soon et al: ML, pairwise mention-based
- 5 Evaluation

Eingabe

- Text
- Ein Zielpronomen zum auflösen (Annahme: referentiell)
- Parse des aktuellen und der vorhergehenden Sätze (mit morphologischen Infos)

Ausgabe

Antezedent für das Pronomen

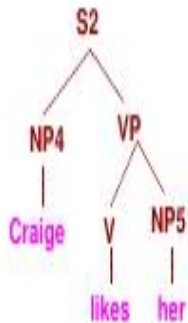
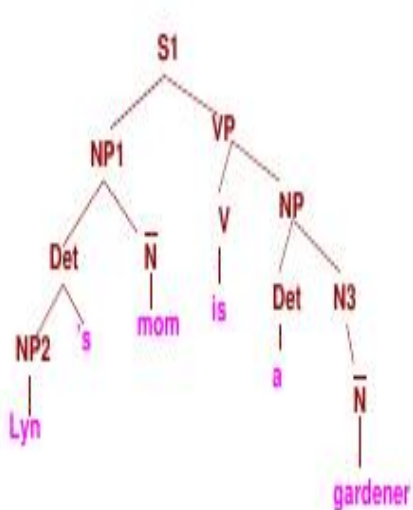
Hobbs, Jerry R., 1978, *Resolving Pronoun References*, *Lingua*, Vol. 44, pp. 311-338.

Hobbs-Algorithmus: Simplifiziert

- 1 Beginne satzintern
- 2 Laufe vom Zielpronomen den Parsebaum hinauf. Untersuche unterwegs für jeden NP oder S-Knoten, den man findet, eine **Breitensuche** von **links nach rechts** unter dessen Kindern, die links vom Zielpronomen stehen. Beachte dabei c-commando.
- 3 Sobald ein untersuchter Knoten die agreement Restriktionen erfüllt, dann haben wir den Antezedenten gefunden.
- 4 Falls die Untersuchung erfolglos war, machen wir mit dem vorherigen Satz weiter.

Salienz und andere Restriktionen/Präferenzen werden durch Suchordnung approximiert

Beispiel: korrekt



Beispiel: Fehler

- The castle in Camelot remained the residence of the king until 536 when he moved it to London

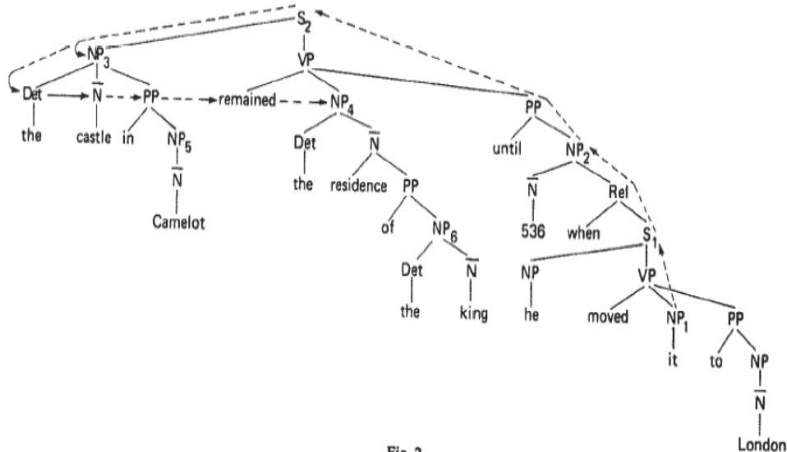


Fig. 2.

Beispiel: Fehler

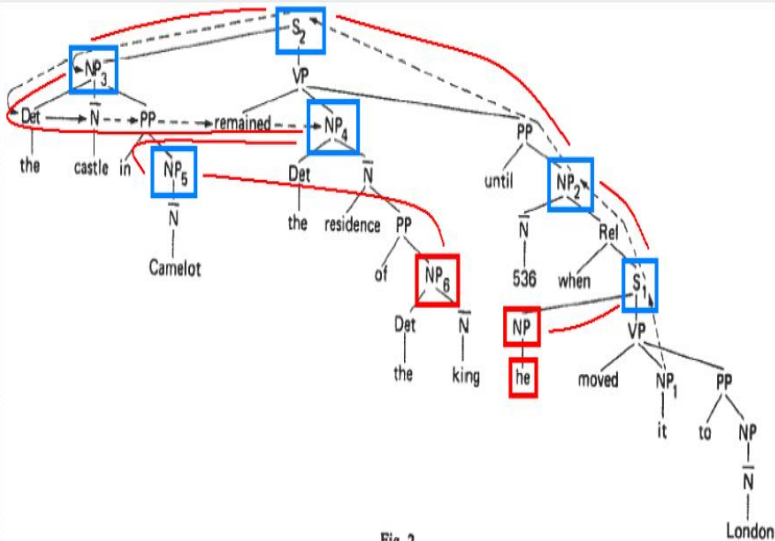


Fig. 2.

Gut für "he", nicht so gut für "it".

- ① Begin at the NP immediately dominating the pronoun.
- ② Go up tree to first NP or S encountered.
 - Call node X, and path to it, p.
 - Search left-to-right below X and to left of p, proposing any NP node which has an NP or S between it and X.
- ③ If X is highest S node in sentence, Search previous trees, in order of recency, left-to-right, breadth-first, proposing NPs encountered.
- ④ Otherwise, from X, go up to first NP or S node encountered, Call this X, and path to it p.
- ⑤ If X is an NP, and p does not pass through an N-bar that X immediately dominates, propose X. x
- ⑥ Search below X, to left of p, left-to-right, breadth-first, proposing NP encountered.
- ⑦ If X is an S, search below X to right of p, left-to-right, breadth-first, but not going through any NP or S, proposing NP encountered.
- ⑧ Go to 2.

Hobbs: Handevaluation von 100 Pronomen “he,” , “she, “they” ,
“it” . Texte:

- Buch: Early civilization in China
- Roman: Wheels
- Newsweek

72.3% korrekt unter Annahme perfekter Parses, und ohne
semantische Restriktion

Andere Evaluationen bei anderen Forschern: meist so um 82%
korrekt.

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen
- 4 Soon et al: ML, pairwise mention-based
- 5 Evaluation

Lappin & Leass (1994)

Lappin and Leass (Computational Linguistics, 1994): An algorithm for pronominal anaphora resolution

- Benutzt einfache Gewichte, um mögliche Antezedenten zu ranken
- Berücksichtigt syntaktische Präferenzen, Parallelismus, repeated mention und recency
- Gewichte sind empirisch entwickelt worde.

Wir betrachten nur den Algorithmus für nicht-reflexive Personalpronomen in dritter Person

- Wenn neuer referentieller Ausdruck per NP, ändere das DM
- assoziiere einen **Salienzwert/salience value** mit neuer referentieller NP basierend auf **Salienzfaktoren/salience factors**

1. **Salienzwert**: Summe von Salienzfaktoren

sentence recency (added to NPs in current S)	100
subject emphasis (added if NP is subject)	80
existential emphasis (added if “there is NP ...”)	70
direct object emphasis	50
indirect object emphasis	40
non-adverbial emphasis (if NP isn't in demarcated advP)	50
head noun emphasis (if NP isn't embedded in another NP)	80

John saw a car with a sunroof at the dealership. He showed it to Bob.

Am Ende des ersten Satzes

Referent	phrases	computation	total
John	[John]	$100 + 80 + 50 + 80$	310
car	[car]	$100 + 50 + 50 + 80$	280
sunroof	[sunroof]	$100 + 50$	150
dealership	[dealership]	$100 + 50 + 80$	230

Am Anfang des zweiten Satzes werden alle W erte halbiert
(**recency preference**)

Referent	phrases	total
John	[John]	155
car	[car]	140
sunroof	[sunroof]	75
dealership	[dealership]	115

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen
- 2 Werfe alle raus, die nicht in Numerus oder Genus kongruieren

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen
- 2 Werfe alle raus, die nicht in Numerus oder Genus kongruieren
- 3 Werfe alle raus, die binding Restriktionen nicht erfüllen (satzintern).

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen
- 2 Werfe alle raus, die nicht in Numerus oder Genus kongruieren
- 3 Werfe alle raus, die binding Restriktionen nicht erfüllen (satzintern).
- 4 Berechne den totalen Salienzwert von jedem noch übrigen möglichen Antezedenten, indem man möglichen Rollenparallelismus addiert (+35).

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen
- 2 Werfe alle raus, die nicht in Numerus oder Genus kongruieren
- 3 Werfe alle raus, die binding Restriktionen nicht erfüllen (satzintern).
- 4 Berechne den totalen Salienzwert von jedem noch übrigen möglichen Antezedenten, indem man möglichen Rollenparallelismus addiert (+35).
- 5 Wähle Antezedenten (Äquivalenzklasse/Entität) mit höchstem Salienzwert. Falls unentschieden, nimm denn nächstgelegenen.

- 1 Identifiziere alle möglichen Antezedenten aus den 4 vorherigen Sätzen
- 2 Werfe alle raus, die nicht in Numerus oder Genus kongruieren
- 3 Werfe alle raus, die binding Restriktionen nicht erfüllen (satzintern).
- 4 Berechne den totalen Salienzwert von jedem noch übrigen möglichen Antezedenten, indem man möglichen Rollenparallelismus addiert (+35).
- 5 Wähle Antezedenten (Äquivalenzklasse/Entität) mit höchstem Salienzwert. Falls unentschieden, nimm denn nächstgelegenen.
- 6 Update Salienzwerte

He kann nur nach *John* aufgelöst werden, wegen Kongruenzen (Schritt 2)

Update DM

he hat Salienzwert von $100 + 80 + 50 + 80 = 310$.

Referent	phrases	total
John	[John, he]	$310+155=465$
car	[car]	140
sunroof	[sunroof]	75
dealership	[dealership]	115

it im zweiten Satz kann nur nach *car*, *sunroof* und *dealership* aufgelöst werden.

Auflösung nach *car* erlaubt Parallelismus, damit ist der Salienzwert für *car* nun maximal $140 + 35 = 175$. Dies ist der höchste Wert und damit wird *car* als Antezedent gewählt.

it_1 hat einen Salienzwert von $100 + 50 + 50 + 80 = 280$.

Referent	phrases	total
John	[John, he]	465
car	[car, it]	$140 + 280 = 420$
sunroof	[sunroof]	75
dealership	[dealership]	115

Bob hat einen Salienzwert von $100 + 40 + 50 + 80 = 270$.

Referent	phrases	total
John	[John, he ₁]	465
car	[car, it ₁]	420
sunroof	[sunroof]	75
dealership	[dealership]	115
Bob	[Bob]	270

- Pronomen werden zu Entitäten in DM aufgelöst. Höchster Salienzwert zählt.
- Da mehrere NPs (mit verschiedenen Salienzwerten) zu der gleichen Entität gehören können, kombinieren wir alle ihre Werte in einer **Äquivalenzklasse**.
- Dadurch etabliert man auch die **repeated mention** preference
- Wenn eine Entität mehrfach innerhalb eines Satzes erwähnt wird, dann wird das Maximum gewählt. In mehreren Sätzen die Summe der Salienzwerte.

- Benutzerhandbücher für Computer
- 86% Genauigkeit (unseen test data) (82% für Hobbsalgorithmus)
- Braucht leider vollständiges Parsing
- Hat kein Weltwissen

- Pronomenresolution muss mehrere Restriktionen und Präferenzen inkludieren
- Hobbs und Lappin und Leass tun dies in verschiedener Art und Weise.
- Welche Präferenzen werden wie inkludiert??
- Übung: Wie würden Sie einen überwachten ML-Algorithmus zur Pronomenresolution entwickeln?

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen**
- 4 Soon et al: ML, pairwise mention-based
- 5 Evaluation

Beispiel

Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to 1.3 million dollar, as the 37-year-old also became the Denver-based financial-services companys president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Beispiel

Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to 1.3 million dollar, as the 37-year-old also became the Denver-based financial-services companys president. It has been ten years since she came to Megabucks from rival Lotsabucks.

- 1 Victoria Chen, Chief Financial Officer of Megabucks Banking Corp; her; the 37 year-old; the Denver-based financial services company's president; she
- 2 Megabucks Banking Corp; the Denver-based financial services company; Megabucks
- 3 2004
- 4 her pay
- 5 20%
- 6 1.3. million dollar
- 7 Denver
- 8 rival Lotsabucks

Beispiel

[Victoria Chen, Chief Financial Officer of [Megabucks Banking Corp] m_2^2] m_1^1 since [2004] m_3^3 , saw [[her] m_4^4 pay] m_5^4 jump [20%] m_6^5 to [1.3 million dollar] m_7^6 , as [the 37-year-old] m_8^1 also became the [[[Denver] m_9^7 -based financial-services company] m_{10}^2 president] m_{11}^1 . It has been ten years since [she] m_{12}^1 came to [Megabucks] m_{13}^2 from [rival Lotsabucks] m_{14}^8 .

Terminologie:

- 1 m_n^k : m -ter referentieller Ausdruck (**mention/markable**), der zur k -ten Entität (**entity**) gehört
- 2 Koreferenzketten der Länge Eins: **singletons**
- 3 Koreferenzketten der Länge l bestehen aus $l(l-1)/2$ koreferenten mention-Paaren
- 4 Koreferenz \neq Anapher!

- **Mention detection:** Finde alle referentiellen Ausdrücke.
 - Ausschluss von pleonastischem “it”
 - Ausschluss von Idiomen “raining cats and dogs”
 - Abhängig von Annotationsschemata!
 - Interessiert uns hier weniger
- **Koreferenzerkennung:** Für jeden Mention bestimme zu welcher Entität er gehört und mit welchen anderen mentions er koreferent ist (wenn überhaupt). Einstufig oder mehrstufig:
 - Singletonerkennung
 - Antezedentenerkennung für nicht-singletons

Überblick über Arten von Mentions

- **Eigennamen:** *Rafael Nadal ... Mr Nadal ... Rafa*
- **Indefinite NPs:** Seltenst koreferent zu vorherigen NPs, kann aber eine Koreferenzkette starten
 - *Ich traf einen Mann. Er ...*
 - Generisch: *Ein Vogel kann fliegen. Sie haben auch Federn.*
- **Pronomen:** Üblicherweise anaphorisch. Ausnahmen schon behandelt.

- **Definite NPs:** extrem vielfältig.
 - *Der Präsident der USA wurde kürzlich gewählt.*
 - *Der Präsident der USA ... Der Präsident*
 - *Der Präsident der USA ... Die Entscheidungen der ersten Woche*
- **Demonstrativa (inkl Demonstrativpronomen)**
 - Deiktisch: *Dieses Haus gefällt mir nicht.*
 - Anaphorisch: *Ich gehe gerne zur Schule. Das ist ungewöhnlich.*
 - Anaphorisch: *John trifft Bill. Dieser ...*

Viele ähnlich zu den Präferenzen bei Pronomen

- Kongruenzen: Numerus, Genus, Person. Wie immer gibt es Ausnahmen.
 - *Fünf Wölfe ... Das Rudel*
 - Genus nur bei Personen: **Katja ... der Professor*
 - Genus nur bei Personen: *Das Büro ... der Raum*
- Recency: am wichtigsten für Pronomen
- Subjektpräferenzen
- Wiederholungen
- ...

Ein paar neue Restriktionen und Präferenzen

- Appositionen: *Victoria Chen, Chief Financial Officer of Megabucks*
- Kopula: *Tara war die Farm von Scarlett O'Hara*
- (partieller) String Match: *Rafael . . . Rafa*
- semantische Kompatibilität: *Das Haus . . . Das Gebäude*

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen
- 4 **Soon et al: ML, pairwise mention-based**
- 5 Evaluation

Pairwise mention model

Soon et al. (CL, 2001): *A machine learning approach to coreference resolution of noun phrases*

- Überwachter ML Ansatz
- Korpora: MUC-6, Ontonotes, Conll2011. In etwas das Klammernformat mit m_n^k . (Manchmal ohne Singletons).
- Training:
 - Bilde positive und negative Beispiele von koreferenten mention Paaren
 - Stelle diese Beispiele als Merkmalsvektoren dar
 - Lerne ML-Algorithmus für binäre koreferent ja/nein Entscheidung
- Testing:
 - Bilde Mentionpaare
 - Entscheide paarweise nach Koreferenz
 - Bilde Ketten (Entitäten) durch Transitivität

Beispiel

[Victoria Chen, Chief Financial Officer of [Megabucks Banking Corp] m_2^2] m_1^1 since [2004] m_3^3 , saw [[her] m_4^1 pay] m_5^4 jump [20%] m_6^5 to [1.3 million dollar] m_7^6 , as [the 37-year-old] m_8^1 also became the [[[Denver] m_9^7 -based financial-services company] m_{10}^2 president] m_{11}^1 . It has been ten years since [she] m_{12}^1 came to [Megabucks] m_{13}^2 from [rival Lotsabucks] m_{14}^8 .

Schritt 2: Bilde positive und negative Beispiele im Training

Möglichkeiten bei freier Kombination aller mentions:

Victoria Chen ... m_1^1	Megabucks B. Company (m_2^2)	nein
Victoria Chen	2004	nein
Victoria Chen	her	ja
Victoria Chen	her pay	nein
...
Victoria Chen	the 37-year-old	ja
...
Megabucks Banking Corps	2004	nein
...
37 year-old	her	ja
...

Probleme: unbalancierte Daten. Sehr weit entfernte Paare werden auch berücksichtigt.

Schritt 2: Bilde positive und negative Trainingsbeispiele

Bei Soon et al: für ein m_i (nichtbeginnendes Mitglied einer Koreferenzkette)

- ein positives Beispiel: nächstgelegener Antezedent m_j mit $j < i$
- negative Beispiele: alle mentions zwischen m_j und m_i

reduziert negative Beispiele stark. Beschränkt sich auf kurze Distanzen.

Beispiel

[Victoria Chen, Chief Financial Officer of [Megabucks Banking Corp] m_2^2] m_1^1 since [2004] m_3^3 , saw [[her] m_4^4 pay] m_5^4 jump [20%] m_6^5 to [1.3 million dollar] m_7^6 , as [the 37-year-old] m_8^1 also became the [[[Denver] m_9^7 -based financial-services company] m_{10}^2 president] m_{11}^1 . It has been ten years since [she] m_{12}^1 came to [Megabucks] m_{13}^2 from [rival Lotsabucks] m_{14}^8 .

her (m_4)	Victoria Chen ... (m_1)	ja
her (m_4)	Megabucks Banking Company	nein
her (m_4)	2004	nein
the 37-year old (m_8)	her (m_4)	ja
the 37-year-old (m_8)	1.3 million dollar	nein
...	...	

Bitte vervollständigen Sie die Tabelle.

- **Merkmale, die nur einen Teil des Paares betreffen:**
 - Ist die erste mention ein Pronomen?
 - Ist die zweite mention ein Pronomen?
 - Phrasentyp
 - Grammatische Rolle
 - Semantische Klasse ...
- **Merkmale, die das Paar betreffen:**
 - String Match? Minimum edit Distance?
 - Kongruent?
 - Distanz (in Sätzen, Worten ...)
 - Hyponyme?

Beispiel

[Victoria Chen, Chief Financial Officer of [Megabucks Banking Corp] m_2^2] m_1^1 since [2004] m_3^3 , saw [[her] m_4^4 pay] m_5^4 jump [20%] m_5^5 to [1.3 million dollar] m_7^6 , as [the 37-year-old] m_8^8 also became the [[[Denver] m_9^7 -based financial-services company] m_{10}^2 president] m_{11}^1 . It has been ten years since [she] m_{12}^1 came to [Megabucks] m_{13}^2 from [rival Lotsabucks] m_{14}^8 .

Paar "her (m_4)-Victoria Chen (m_1), ja":

DIST	0
i-Pronomen	FALSE
j-Pronomen	TRUE
String Match	FALSE
j-def	FALSE
j-dem	FALSE
number-agr	TRUE
sem-agr	TRUE
gender-agr	TRUE
both-names	FALSE
Alias	FALSE
Annos	FALSE

Schritt 4: trainiere einen ML-Algorithmus

Wir haben also eine Menge Vektoren wie (0, FALSE, TRUE, FALSE ... ja) oder (3, TRUE, TRUE, FALSE ... nein). Abstrakt: $(f_1, f_2, f_3 \dots, c),,$

Beispiel NB: Berechne im Training

- 1 $p(ja), p(nein)$
- 2 $p(f_i|ja), p(f_i|nein)$

NB für Koreferenz nicht brauchbar, da zu viele Unabhängigkeitsannahmen. Kann Zusammenhänge nicht gut erkennen. Soon et al lernen einen Entscheidungsbaum.

Schritt 4: Entscheidungsbäume



Zum Lernen von Entscheidungsbäumen muss man die informativsten Merkmale bestimmen, die die Entropie am meisten reduzieren.

Schritt 4: der gelernte Entscheidungsbaum

```
STR_MATCH = +: +
STR_MATCH = -:
: ... J_PRONOUN = -:
: ... APPOSITIVE = +: +
:   APPOSITIVE = -:
:   : ... ALIAS = +: +
:     ALIAS = -: -
J_PRONOUN = +:
: ... GENDER = 0: -
:   GENDER = 2: -
:   GENDER = 1:
:     : ... I_PRONOUN = +: +
:       I_PRONOUN = -:
:       : ... DIST > 0: -
:         DIST <= 0:
:           : ... NUMBER = +: +
:             NUMBER = -: -
```

Figure 2
The decision tree classifier learned for MUC-6.

Bild aus Soon et al (2001)

Brauchen eine Suchstrategie

Soon et al: closest first

- 1 Paare eine mention m_i mit m_{i-1} (ausser sie sind ineinander verschachtelt), verwandle in Merkmale und wende Entscheidungsbaum an.
- 2 Wenn Antwort ja, beende, und gehe zu m_{i+1}
- 3 Wenn Antwort nein, paare m_i mit m_{i-2} und wende Entscheidungsbaum an
- 4 etc pp

Tafelbeispiel für Testing

```
STR_MATCH = +: +
STR_MATCH = -:
: ... J_PRONOUN = -:
: ... APPOSITIVE = +: +
:   APPOSITIVE = -:
:   : ... ALIAS = +: +
:   ALIAS = -: -
J_PRONOUN = +:
: ... GENDER = 0: -
:   GENDER = 2: -
:   GENDER = 1:
: ... I_PRONOUN = +: +
:   I_PRONOUN = -:
:   : ... DIST > 0: -
:   DIST <= 0:
:     : ... NUMBER = +: +
:     NUMBER = -: -
```

Figure 2
The decision tree classifier learned for MUC-6.

[Ms. Washington] m_1 's candidacy] m_2 is being championed by
[several powerful lawmakers, including [[her] m_4 boss] m_5] m_3
... [She] m_5

- Merkmalsberechnungen können Fehler enthalten
- String Match, aber nicht koreferent: Energy Committee
... the committee ... Finance committee ... the committee
- Zu wenig Semantik: *Metro-Goldwyn Mayer ... the movie studio*

Prinzipielle Probleme und Weiterentwicklungen

- **Sehr einfache Merkmale:** wurden später stark weiterentwickelt und neue hinzugefügt
- **Closest first im Testing:** problematisch, da damit oft unsichere ja-Entscheidungen getroffen werden. Später best-first.
- **Mention-mention** nicht ideal, da lokale Entscheidungen, anstatt entitätsbasierte Entscheidungen
- **Fehlende Anaphorizitätsentscheidung:** vor allem für common nouns problematisch

- 1 Hobbs-Algorithmus
- 2 Lappin und Leass Algorithmus (OPTIONAL, VERALTET)
- 3 Koreferenz für alle Nominalphrasen
- 4 Soon et al: ML, pairwise mention-based
- 5 Evaluation

Tabelle aus Lee et al (CL 2013): *Deterministic Coreference Resolution based on entity-centric, precision-ranked rules*

Table 3

Corpora statistics.

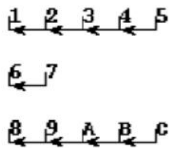
Corpora	# Documents	# Sentences	# Words	# Entities	# Mentions
OntoNotes-Dev	303	6,894	136K	3,752	14,291
OntoNotes-Test	322	8,262	142K	3,926	16,291
ACE2004-Culotta-Test	107	1,993	33K	2,576	5,455
ACE2004-nwire	128	3,594	74K	4,762	11,398
MUC6-Test	30	576	13K	496	2,136

Ein Link-basiertes Maß, das Links zählt, die sowohl in den Systemketten S als auch in den Goldketten G sind. Hierbei werden die Links so gezählt, dass nur die minimale Anzahl von Links genommen wird, die zur vollständigen Spezifikation der G/S Ketten genügen.

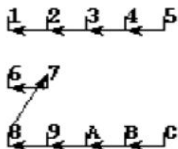
$$muc - precision = \frac{links(G \cap S)}{links(S)}$$

$$muc - recall = \frac{links(G \cap S)}{links(G)}$$

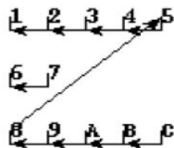
MUC-Beispiel



(a) Key



(b) Response 1



(c) Response 2

Figure 1. Shortcomings of the MUC scorer (Bagga and Baldwin, 1998)

$$P = 9/10 = 90\%$$
$$R = 9/9 = 100\%$$

$$P = 9/10 = 90\%$$
$$R = 9/9 = 100\%$$

Für jedes Mention m_i

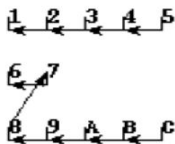
$$Precision(m_i) = \frac{|m_j|(m_i, m_j) \in S \cap G|}{|m_j|(m_i, m_j) \in S|}$$

$$Recall(m_i) = \frac{|m_j|(m_i, m_j) \in S \cap G|}{|m_j|(m_i, m_j) \in G|}$$

Bilde dann Mittelwerte über alle m_i

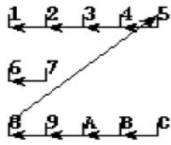
NB: Kriegt auch credit für singletons, da $m_j = m_i$ gezählt wird.

B-Cubed Beispiel



(b) Response 1

$$\frac{1}{12} * \left(\frac{5}{5} * 5 + \frac{2}{7} * 2 + \frac{5}{7} * 5 \right) = \frac{16}{21} (76\%)$$



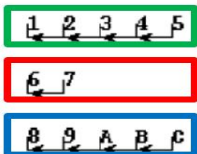
(c) Response 2

$$\frac{1}{12} * \left(\frac{5}{10} * 5 + \frac{2}{2} * 2 + \frac{5}{10} * 5 \right) = \frac{7}{12} (58\%)$$

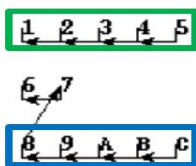
Recall for both responses is 100%

Source: Amit Bagga

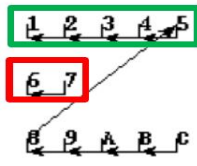
- Berechnet 1-zu-1 Alignierung der System und Goldketten, nach Anzahl der überlappenden mentions
- CEAF-score: Anzahl der mentions, die passend aligniert wurden



(a) Key



(b) Response 1



(c) Response 2

$$\text{CEAF} = 10/12$$

$$\text{CEAF} = 7/12$$

- MUC-6 Korpus
- MUC-Recall 58.6
- MUC-Precision 67.3
- MUC-F-measure 62.6

Berkeley Coreference Resolver:

Durrett and Klein (EMNLP 2013): *Easy victories and uphill battles in coreference resolution*

- überwachtes ML (log-linear)
- mention ranking, nicht mention-mention (globale Optimierung)
- eine sehr große Menge von Oberflächenmerkmalen, die frei kombiniert werden können
- Beispiel: Antecedent-Head = Voters und Anapher = they

Conll 2011 testset (viel schwieriger als MUC-6). Mit system mentions.

- MUC-F1 = 66.43
- B-Cubed F1 = 66.16
- CEAF F1 = 47.79

Stanford Sieve System:

Lee et al (CL 2013): *Deterministic Coreference Resolution based on entity-centric, precision-ranked rules*

- Regelbasiertes System
- Trick: verwendet Regeln, bei denen man sich sicher ist, zuerst
- Suchstrategie für Antezedenten: breadth-first, left-to-right. Nahe Sätze zuerst.
- Verwendet dann die ganze Entität, die man bisher zusammen gestellt hat, in den zukünftigen Regeln. (**entity-centric**)

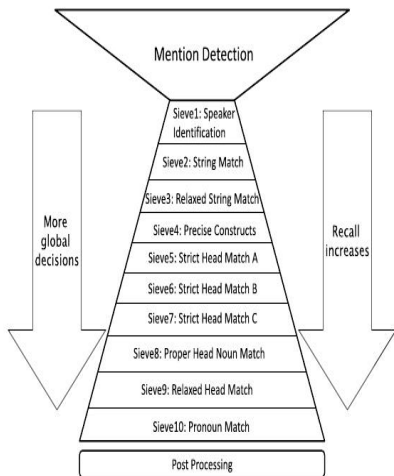


Figure 1
The architecture of our coreference system.

Auf Conll 2011. Mit System mentions und boundaries:

- MUC-F: 61
- B-Cubed: 68.9
- CEAF: 45

Mit gold boundaries für NPs

- MUC-F: 63.9
- B-Cubed: 70
- CEAF: 48.3

- Inkludierung semantischen Wissens, besonders für common nouns: the house — the building
- Dies scheitert meistens daran, dass es zu vielen false positives führt
- Weitere Entwicklung Richtung word embeddings und neuronalen Netzen
- Problem: lexical memorization?
- Ist abhängig von allen anderen bisher angeschauten Problemen: POS, parsing etc.

- ① Koreferenz nicht gelöst. Nicht einmal die Evaluierung ist vollkommen gelöst...
- ② Präferenzen und Restriktionen vielfältiger als bei Pronomen
- ③ Überwachtes ML: mention-pair (lokal) vs. entity-based (global)
- ④ Probleme: common nouns, viele singletons, Semantik

- Soon et al. (CL, 2001): *A machine learning approach to coreference resolution of noun phrases*
- Lee et al (CL 2013): *Deterministic Coreference Resolution based on entity-centric, precision-ranked rules*
- Durrett and Klein (EMNLP 2013): *Easy victories and uphill battles in coreference resolution*
- Systeme: Berkeley <http://nlp.cs.berkeley.edu>
- Systeme: Stanford Core NLP
stanfordnlp.github.io/CoreNLP/ .