

Einführung in die Computerlinguistik Semantik

Robert Zangenfeind

Center for Information and Language Processing

2023-11-20

Die Grundfassung dieses Foliensatzes wurde von Dr. Annemarie Friedrich (unter Zuhilfenahme von Materialien von Prof. Dr. Manfred Pinkal und Prof. Dr. Vera Demberg) erstellt. Fehler und Mängel liegen ausschließlich in meiner Verantwortung.

Outline

- 1 Intro
- 2 Lexikalische Semantik
- 3 Textähnlichkeit
- 4 Logik
- 5 Zeit & Aspekt
- 6 Kompositionalität
- 7 Referenz

Der Begriff “Semantik”

- griechisch *sēmantikós* = bezeichnend, zu: *sēmaínein* = bezeichnen
 - griechisch *sēma* = Zeichen, Merkmal
- ⇒ Lehre von der Bedeutung sprachlicher Zeichen

Levels of Language



Kaffee machen



Motivation (1)

*Machen Sie mir
bitte
eine Tasse Kaffee!*



Aktion

Sprache/Text

Verarbeitung
Darstellung

Motivation (2)

*Machen Sie mir
bitte
eine Tasse Kaffee!*



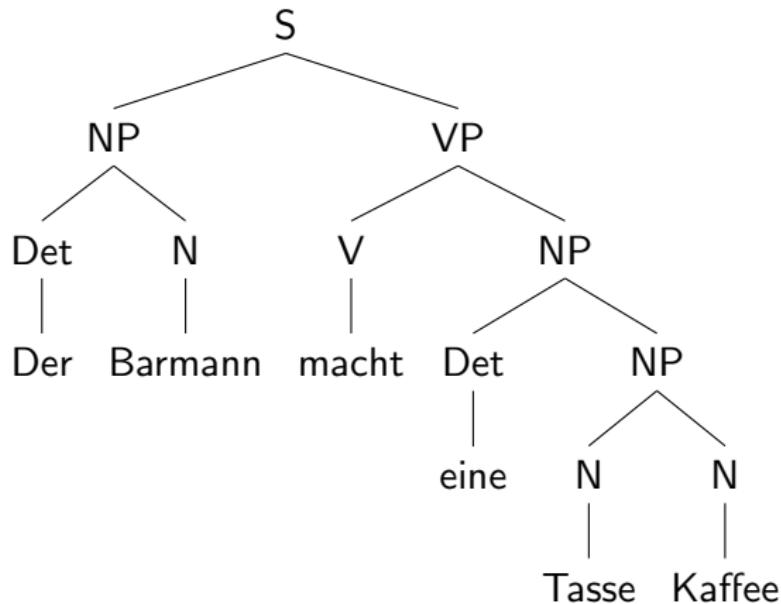
Aktion

Sprache/Text

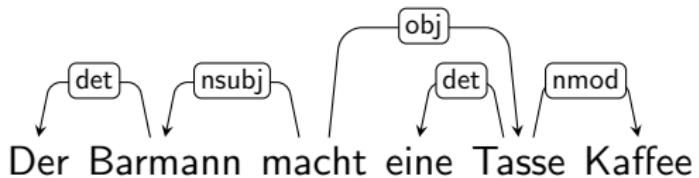
Verarbeitung
Darstellung

Verarbeitung/Darstellung,
um Aktion einleiten zu können:
Wie?

Konstituenten-Baum?



Dependenz-Baum?



Konstituenz/Dependenz:

- ⇒ nur Syntax.
- ⇒ auch Semantik wird benötigt.

Semantikrepräsentation: WordNet

WordNet: *coffee*

<noun.food>S: (n)

coffee#1,

java#2

(a beverage consisting of an infusion of ground coffee beans)

“he ordered a cup of coffee”

[...]

<https://wordnet.princeton.edu/>

PropBank

agree.01

Arg0: Agreer

Arg1: Proposition

Ex: [Arg0 The group] agreed [Arg1 it wouldn't make an offer ...]

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. Computational linguistics, 31(1), 71–106.

<http://propbank.github.io>

Semantikrepräsentation: FrameNet

Cooking_creation

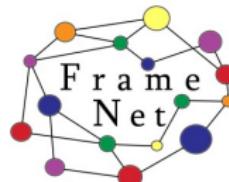
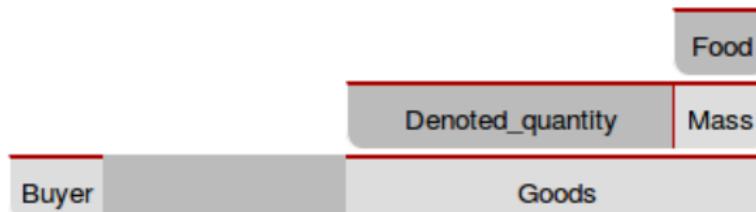
This frame describes food and meal preparation. A **Cook** creates a **Produced_food** from (raw) **Ingredients**. The **Heating_Instrument** and/or the **Container** may also be specified.

Caitlin baked some cookies from the pre-packaged dough.

<https://framenet.icsi.berkeley.edu/about>

Tina **bought** **some** **milk** .

COMMERCE_BUY RELATIONAL_QUANTITY FOOD



Semantikrepräsentation: Abstract Meaning Representations (AMR)

“Can you please make me a mug of coffee?”

```
(m / make-01 :polite + :mode imperative
  :ARG0 (y / you)
  :ARG1 (c / coffee
    :quant (v / volume-quantity
      :unit (m2 / mug)
      :quant 1))
  :ARG3 (i / i))
```

<http://amr.isi.edu>



Anforderung an Semantikrepräsentationen (1)

Verifizierbarkeit

Es soll möglich sein, eine Aussage mit einem Zustand der Welt zu verbinden, so dass überprüft werden kann, ob ein Satz wahr ist.



Das Flugzeug hat einen Propeller.



Repräsentation: part-of(Propeller, Flugzeug)



Kommt dies in der Wissensbasis (knowledge base) vor?

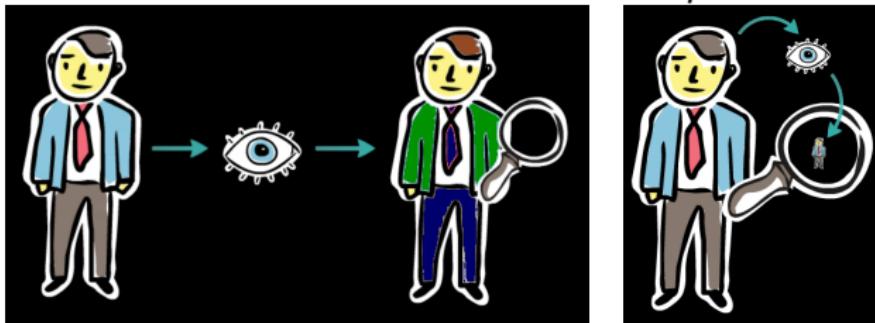


Anforderung an Semantikrepräsentationen (2)

Eindeutigkeit = Nicht ambig

Sprachliche Ausdrücke sind oft ambig – erstrebenswert und wichtig ist aber eine eindeutige Semantikrepräsentation.

Hans sah den Mann mit der Lupe.



Zum Beispiel hier (plausibel): hält(Mann, Lupe)
und nicht (wäre nicht plausibel): hält(Hans, Lupe)

Bei maschineller Analyse: schwierige Entscheidung, welche
Repräsentation gewählt werden soll

Anforderung an Semantikrepräsentationen (3)

Kanonische Form

Alle Sätze, die das Gleiche bedeuten, sollen gleich repräsentiert werden.

Gibt es im Restaurant Almaz vegane Speisen?

Bietet Almaz auch veganes Essen an?

Kann man bei Almaz vegan essen?

⇒ Egal, wie die Frage gestellt wird, es soll eine Repräsentation für die Datenbankabfrage geben.

Zum Beispiel: `bietet_an(Almaz, VeganesEssen)` → true/false?

⇒ Paraphrasen erkennen

Anforderung an Semantikrepräsentationen (4)

Inferenzen

Alle Dackel sind Hunde.

Alle Hunde mögen Wurst.

Fiffi ist ein Dackel.

⇒ *Fiffi mag Wurst.* wahr oder falsch?

Es soll nicht nur das direkt Gesagte repräsentiert werden; es sollen auch neue Aussagen / neues Wissen daraus abgeleitet werden können.

Anforderung an Semantikrepräsentationen (5)

Question Answering

Wer hat Dracula geschrieben?

Antwort: *Bram Stoker*



<https://www.youtube.com/watch?v=P18EdAKuC1U>

START-System: <http://start.csail.mit.edu/index.php>

Wortbedeutungs-Relationen (1)

Beispiel

Konzepte:



Bank



Homonym = ein Ausdruck; steht für verschiedene
Begriffe/Konzepte

Wortbedeutungs-Relationen (2)

Weitere Relationen zwischen Konzepten (\neq zwischen Wörtern)

- ① **Synonymie:** zwei Begriffe mit derselben Bedeutung
 - ② **Hyperonymie:** ein Begriff ist ein Oberbegriff eines anderen
 - ③ **Hyponymie:** ein Begriff ist ein Unterbegriff eines anderen
 - ④ **Homophonie:** zwei Begriffe sind gleichlautend, haben aber unterschiedliche Bedeutung
 - ⑤ **Meronymie:** Teil-Ganzes-Relation
 - ⑥ **Holonymie:** Umkehrung der Meronymie, d.h. Ganzes-Teil-Relation
- Verschiedene Beispiele:
- ⓐ *Dach – Haus*
 - ⓑ *Hand – Finger*
 - ⓒ *Mensch – Lebewesen*
 - ⓓ *Leere – Lehre*
 - ⓔ *Geige – Violine*
 - ⓕ *Kuchen – Nusskuchen*

WordNet (1)

- Große lexikalisch-semantische Ressource
- Netzwerk aus semantischen Relationen zwischen Konzepten; Hyponymie-Relation als Kern
- Konzepte werden als **Synsets** repräsentiert: Mengen von synonymen Wörtern, die sich gegenseitig disambiguieren
- Umschreibungen (**glosses**) für alle Synsets + Beispielsätze

Synsets für car.noun

- S: (n) car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine), *he needs a car to get to work*
- S: (n) car, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad), *three cars had jumped the rails*
- ...

WordNet (2)

- 170.000 lexikalische Einträge (Wörter)
- 120.000 Synsets
- WordNet-Versionen für etwa 45 Sprachen
- Nutzung in vielen sprach- und informationstechnologischen Anwendungen, insbesondere als Grundlage für Inferenz

Word Sense Disambiguation (WSD)

Wortbedeutung im Kontext

- (1) *Schwester Maria sprach ein Gebet.*
(2) *Meine Schwester ärgert oft meinen kleinen Bruder.*
Woher weiß man, welche "Schwester" gemeint ist?

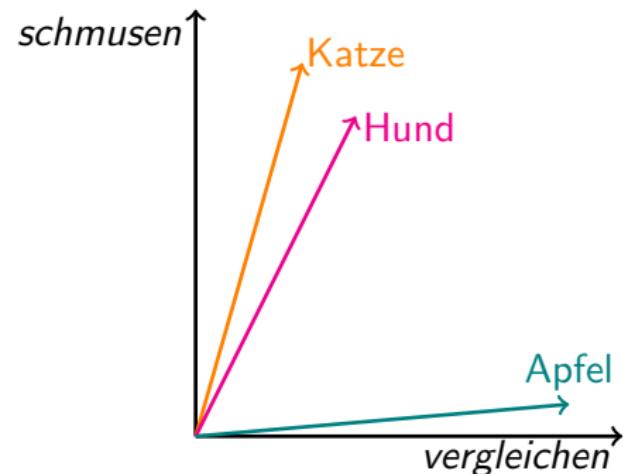
Methoden in der Computerlinguistik für WSD

- Statistische Modellierung
- **Annotation** aller Zielwort-Instanzen im Trainingskorpus mit einer Wortbedeutung
- **Annotationsschema:** Wortbedeutungen aus einem Wörterbuch / Thesaurus (Standard: WordNet-Synsets)

WSD: Distributionelle Methoden (Grundlagen)

- Wörter (\neq Konzepte) werden als Vektoren dargestellt
- Dimensionen: Anzahl Wörter des Vokabulars
- Einträge mit Vorkommenshäufigkeit von Kontextwörtern (evtl. lemmatisiert)
- Ähnlichkeit von Wörtern: Winkel zwischen den Vektoren (je kleiner der Winkel, desto ähnlicher)

	schmusen	vergleichen
Katze	35	10
Hund	30	15
Apfel	4	40
...
...
...
...



WSD: (i) Trainingskorpus, das annotiert werden soll



Bank #1:



Bank #2:

- ① *Für diejenigen, denen Komfort wichtig ist, haben wir eine Bank mit leicht schwingender Rückenlehne entwickelt. ...*
- ② *... Ich suche noch eine Bank für meinen Garten und sondiere deshalb gerade Angebote. ...*
- ③ *... Habe im März 2000 einen höheren Betrag bei einer Bank angelegt. ...*
- ④ *... Beim Test Anlageberatung der Banken löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut. ...*

WSD: (ii) Merkmalsspezifikation

Kontext: Satz (hier als Bsp.
4. Satz aus Korpus) *Beim Test Anlageberatung
der Banken löste kein Institut
die einfache Frage nach einer
sicheren Anlage wirklich gut.*

Merkmale: Kontextwörter
in diesem Satz *Anlageberatung, löste, si-
cheren, Anlage, ...*

⇒ Merkmalsvektor für
“Banken” in diesem Satz *Angebote: 0, Anlage: 1, Anla-
geberatung: 1, Garten: 0, lös-
te: 1, Rückenlehne: 0, schwin-
gender: 0, sicher: 1, ...*



Maschinelles Lernen

Algorithmus gruppiert diesen und die anderen Merkmalsvektoren je
nach ihrer Ähnlichkeit untereinander in die zwei Cluster **bank1** und
bank2.

WSD: (iii) annotiertes Korpus



Bank #1:



Bank #2:

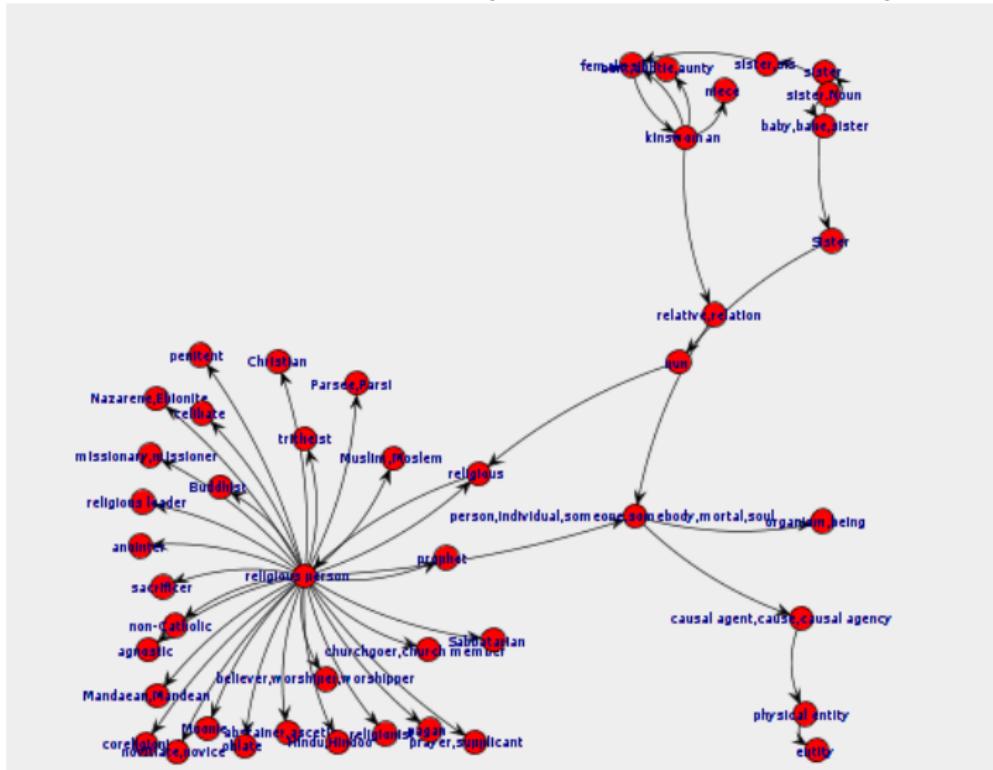
- ① *Für diejenigen, denen Komfort wichtig ist, haben wir eine Bank [bank1] mit leicht schwingender Rückenlehne entwickelt.
...*
- ② *... Ich suche noch eine Bank [bank1] für meinen Garten und sondiere deshalb gerade Angebote. ...*
- ③ *... Habe im März 2000 einen höheren Betrag bei einer Bank [bank2] angelegt. ...*
- ④ *... Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut. ...*

WSD: Lesk-Algorithmus (Lesk, 1986)

- Given: (i) context of an ambiguous word
(e.g., *honey is yellow*)
- Given: (ii) lexicon definitions of the senses of the ambiguous word
- Example:
(synset1) *honey (a sweet yellow liquid produced by bees)*
(synset2) *beloved, dear, dearest, honey, love (a beloved person; used as terms of endearment)*
- Count the overlap (number of common words) of each definition with the sentence, e.g.,
synset1: overlap of one word (*yellow*)
synset2: overlap of zero words
- Select the sense with the greatest overlap

WSD: Pfad-basierte Algorithmen (WordNet-basiert)

Ähnlichkeit zwischen zwei Konzepten \propto Distanz im Graph



Textähnlichkeitsmaß: Jaccard (Vorbetrachtung)

- Grundidee des Lesk-Algorithmus: diejenige Definition eines Lexems auswählen, die zum Kontext des Lexems die **größte Ähnlichkeit** hat.
- Maß für Ähnlichkeit: **word overlap**
- Nachteile von word overlap?

Jaccard coefficient

- A commonly used measure of overlap of two sets
- Let A and B be two sets
- Jaccard coefficient:

$$\text{JACCARD}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

$(A \neq \emptyset \text{ or } B \neq \emptyset)$

- $\text{JACCARD}(A, A) = 1$
- $\text{JACCARD}(A, B) = 0$ if $A \cap B = 0$
- A and B don't have to be the same size.
- Always assigns a number between 0 and 1.

Jaccard coefficient: Example

- What is the query-document match score that the Jaccard coefficient computes for:
 - Query: *ides of March*
 - Document: *Caesar died in March*

Nachteile von Jaccard?

Warum Logik?

- ① Theorie / Formalisierung der Semantik:
Bedeutung = Wahrheitsbedingungen
- ② Inferenz / Logisches Schließen:
 - Aus bestehendem Wissen/Beobachtungen
neues Wissen ableiten.
 - Beispiel question answering:
Liegt Hamburg in Europa?
Vielleicht sagt das keine Quelle (Text, Datenbank) direkt.
Aber es folgt aus *Hamburg liegt in Deutschland* und
Deutschland liegt in Europa

Wahrheitsbedingungen-Semantik

- **Bedeutung = Wahrheitsbedingungen**
Bedeutung eines Satzes wird mit seinen Wahrheitsbedingungen identifiziert, d.h. mit Bedingungen, die die Welt erfüllen muss.
- Sätze sind synonym, wenn sie dieselben Wahrheitsbedingungen haben.
- Einen Satz verstehen, heißt, angeben können, ob er in einer gegebenen Situation wahr oder falsch ist.
- Ursprung:
 - Gottlob Frege
 - Ludwig Wittgenstein
 - Mathematische Logik: Alfred Tarski, Rudolf Carnap

Aussagenlogik (AL)

Aussagesatz (**Proposition**) → Wahrheitswert

Beispiele

Schnee ist weiß. = T (true / wahr)

Ein Junggeselle ist verheiratet. = F (false / falsch)

Syntax der Aussagenlogik

- Satzbuchstaben: Ausdrücke, die ganzen Sätzen entsprechen
- Beispiele: p, q, r
- 5 Satzoperatoren (Junktoren, logische Zeichen von AL):
 - Negation (nicht) \neg
 - Konjunktion (und) \wedge
 - Disjunktion (nicht ausschließendes oder) \vee
 - Implikation (wenn, dann) \rightarrow
 - Äquivalenz (genau dann, wenn) \leftrightarrow
- Hilfszeichen: ()
- A ist ein Satz von AL, wenn eine der folgenden Bedingungen erfüllt ist:
 - (i) A ist ein Satzbuchstabe
 - (ii) B und C sind Sätze von AL und A ist:
 $\neg B, (B \wedge C), (B \vee C), (B \rightarrow C)$ oder $(B \leftrightarrow C)$

Semantik der Aussagenlogik

- Die Semantik einer Aussage ist ihr Wahrheitswert.
- Die Semantik einer komplexen Aussage wird aus der Semantik der Junktoren und der Teilaussagen berechnet.
- Semantik der Junktoren: [Wahrheitstabellen](#)

Konjunktion: "und"

Tim mag Kaffee.

p	T	T	F	F
-----	---	---	---	---

Anna mag Tee.

q	T	F	T	F
-----	---	---	---	---

Tim mag Kaffee und Anna mag Tee.

$p \wedge q$	T	F	F	F
--------------	---	---	---	---

T = wahr / true F = falsch / false

Negation: “nicht”, “kein”

Tim mag Kaffee. p T F

Tim mag keinen Kaffee. $\neg p$ F T

bzw.

Tim ist in Italien. p T F

Tim ist nicht in Italien. $\neg p$ F T

Disjunktion: "oder"

Tim mag Kaffee.

p	T	T	F	F
-----	---	---	---	---

Tim mag Tee.

q	T	F	T	F
-----	---	---	---	---

Tim mag Kaffee oder Tim mag Tee.

$p \vee q$	T	T	T	F
------------	---	---	---	---

nicht ausschließendes 'oder' (vs. ausschließendes 'oder')

Implikation: (kein perfekt passender deutscher Ausdruck)

<i>Anna trinkt Kaffee.</i>	<i>p</i>	T	T	F	F
<i>Tim kocht.</i>	<i>q</i>	T	F	T	F
<i>Wenn Anna Kaffee trinkt, dann kocht Tim.</i>	$p \rightarrow q$	T	F	T	T

Wenn p, dann muss q auch eintreten, damit Gesamtaussage wahr ist.

Aber q kann auch ohne p eintreten → d.h. kein Rückschluss möglich.

Äquivalenz: "genau dann wenn"

Tim trinkt Kaffee.

p

T

T

F

F

Anna kocht Tee.

q

T

F

T

F

*Anna kocht Tee **genau dann,***
wenn Tim Kaffee trinkt.

$p \leftrightarrow q$

T

F

F

T

Kontravalenz / Exclusive Or: "entweder . . . oder"

Tim rennt.

p T T F F

Anna kocht.

q T F T F

Entweder rennt Tim oder Anna kocht.

$\neg(p \leftrightarrow q)$ F T T F

- (Blick zurück zur Aussagenlogik:
elementare Aussagen werden als nicht analysierbare Einheiten behandelt.)
- Prädikatenlogik (PL):
Darstellung auch der internen Struktur von Aussagen.

Individuenkonstanten

Hans, Maria, die Sonne, 'Dracula' usw.

Prädikate und Argumente

Hans ist ein Kind. $\text{kind}(\text{Hans})$

Einstelliges Prädikat \approx Eigenschaft

Hans mag Maria. $\text{m\ddot{o}gen}(\text{Hans}, \text{Maria})$

Zweistelliges Prädikat \approx Relation

Individuenvariablen

$\text{m\ddot{o}gen}(x, y)$

Ein Wert (Hans, Maria, ...) kann x und y zugewiesen werden, dann kann der Wahrheitswert berechnet werden.

Existenzquantor \exists

$\exists x \ (\text{kind}(x) \wedge \text{mögen}(x, \text{Maria}))$

Für ein x , für das gilt: x ist ein Kind und x mag Maria.
Ein (irgendein nicht näher bestimmtes) Kind mag Maria.

Allquantor \forall

$\forall x \ \text{mögen}(x, \text{Maria})$

Für jedes x gilt: x mag Maria.
Alle mögen Maria.

$\forall x \ (\text{kind}(x) \rightarrow \text{mögen}(x, \text{Maria}))$

Für jedes x , das ein Kind ist, gilt: x mag Maria.
Alle Kinder mögen Maria.

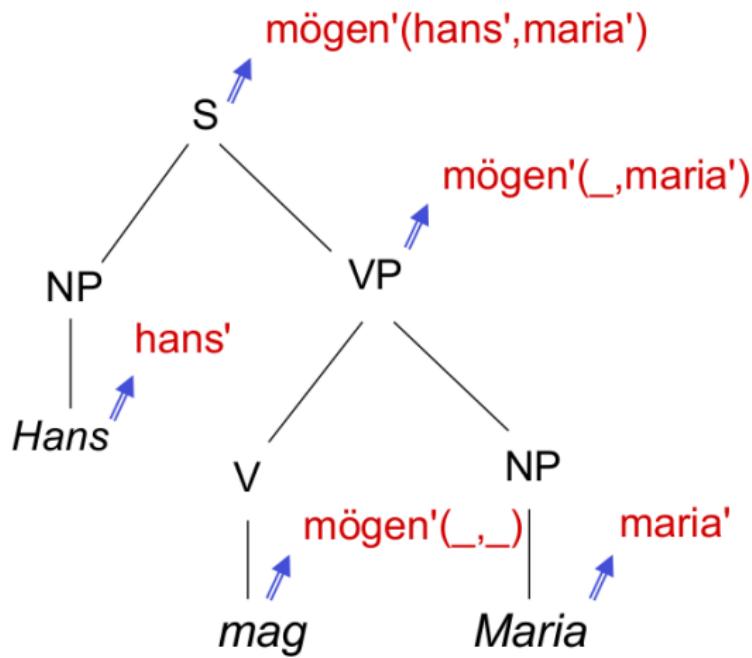
Syntax der Prädikatenlogik (vereinfacht)

- Term: Variable oder Individuenkonstante
- atomare Formel: $P(t_1, \dots, t_n)$
(t_i sind Terme, P ist ein Prädikatsymbol)
- oft wird als atomare Formel auch zugelassen:
 $(t_i = t_j)$
- A ist eine Formel von PL, wenn eine der folgenden Bedingungen erfüllt ist:
 - (i) A ist eine atomare Formel
 - (ii) B und C sind Formeln von PL und A ist:
 $\neg B$, $(B \wedge C)$, $(B \vee C)$, $(B \rightarrow C)$, $(B \leftrightarrow C)$,
 $\forall x A$, $\exists x A$

Semantik-Konstruktion

Entlang der Konstituentenstruktur werden komplexe semantische Ausdrücke "kompositionell" aus einfacheren Ausdrücken zusammengesetzt.

Zu jedem Lexikoneintrag und jeder syntaktischen Regel wird eine semantische Komponente hinzugefügt.



Semantik-Konstruktion: Herausforderung

Jede Teilnehmerin präsentiert ein Papier.

*Jede **Teilnehmerin** präsentiert ein **Papier***

$$\forall x (\text{teilnehmerin}'(x) \rightarrow \exists y (\text{papier}'(y) \wedge \text{präsentieren}'(x,y)))$$

Zum Aspekt

Verbklassen bei Vendler: innere temporale Struktur von Verben
(Zeno Vendler: Verbs and times. In: The Philosophical Review, Vol. 66, No. 2 (Apr., 1957), S. 143-160.)

- **State:** Zustand

Ich wohnte in München.

- **Activity:** Aktivität ohne bestimmten Endpunkt

Sie fuhr Fahrrad.

- **Accomplishment:** Aktivität mit bestimmtem Endpunkt

Er flog nach New York.

- **Achievement:** Ereignis geschieht “in einem Moment”.

Er erreichte New York.

Interpretation von Tempus: Präsens

(nach H. Reichenbach: Elements of symbolic logic. New York 1947)

- **Sprechzeit S** (speech time)
- **Referenzzeit R** (reference time)
- **Ereigniszeit E** (event time)

Präsens

Anna kocht (gerade) einen Tee.

E
R
S



$$E = R = S$$

Interpretation von Tempus: Präteritum

- Sprechzeit S (speech time)
- Referenzzeit R (reference time)
- Ereigniszeit E (event time)

Präteritum

Anna kochte einen Tee.

E

R

S



$$E = R < S$$

Interpretation von Tempus: Plusquamperfekt

- Sprechzeit S (speech time)
- Referenzzeit R (reference time)
- Ereigniszeit E (event time)

Plusquamperfekt

Anna hatte einen Tee gekocht.



Kompositionnalität

Kompositionalitätsprinzip (Gottlob Frege, 1848-1925)

Die Bedeutung eines komplexen Ausdrucks ist eine Funktion der Bedeutungen seiner Teile und der Art ihrer Kombination.

Beispiel

Der Hund biss den Mann.

Bedeutung setzt sich zusammen aus:



Semantikkonstruktion . . .

wird benötigt, um die genaue Art der Zusammensetzung zu bestimmen.

Vollverb vs. Stützverbkonstruktion

Vollverb

Semantik klar, Bedeutung ist eine Hauptbedeutung aus dem Wörterbuch, **kompositionell** mit weiteren Teilen des Satzes

Beispiel:

Susi nimmt einen Keks. → Susi + nehmen + Keks

Stützverbkonstruktion (mit Stützverb, "Funktionsverb", engl. light verb, support verb und prädikativem Wort)

Bedeutung des Verbs: keine (bzw. "schwach"), prädikatives Nomen trägt die Bedeutung

Beispiel:

Susi nimmt ein Bad. ≠ Susi + nehmen + Bad

Idiom

Wortverbindung, deren Gesamtbedeutung sich nicht aus den Bedeutungen der einzelnen Wörter ableiten lässt.
(Spezialfall: Komposita als **morphologische Idiome**)

Beispiele:

rotes Tuch, die Rechnung ohne den Wirt machen,

Hochzeit [morphologisches Idiom]

Idiome sind nicht kompositionell!

Diverse Beispiele

Idiom? Stützverbkonstruktion? Vollverb?

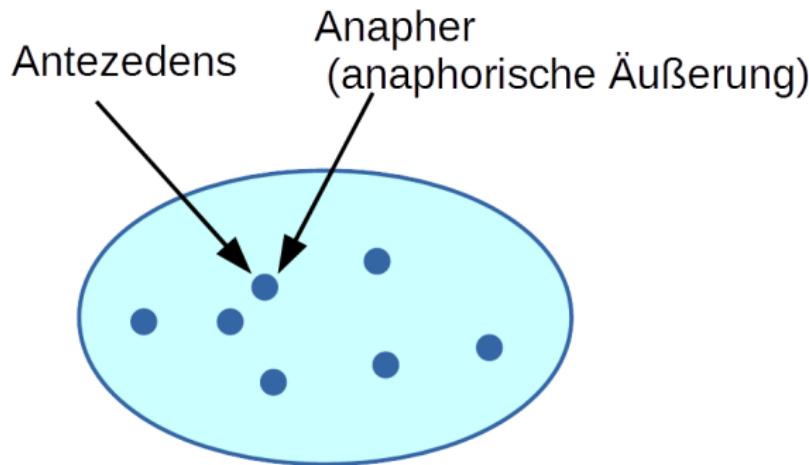
- ① *mit dem Bus fahren*
- ② *aus der Haut fahren*
- ③ *in Kontakt treten*
- ④ *etw. zur Anwendung bringen*
- ⑤ *eine Sprache lernen*
- ⑥ *jdn. auf die Palme bringen*

Anaphorische Referenz

- aus dem Altgriechischen *ana* = zurück *phero* = tragen
- Anaphorische Ausdrücke zeigen auf Referenten, die im sprachlichen Kontext schon eingeführt wurden:
Maria ... sie ... die[se] Frau ...
- Zwei Arten von anaphorischer Referenz:
 - Koreferenz (bei identischem Referenten)
 - Bridging (assoziative Beziehung zwischen den Referenten)

Anaphorische Referenz: Koreferenz (1)

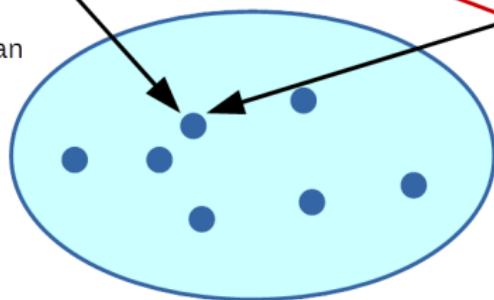
- Koreferenz = identischer Referent
- typischerweise durch Pronomen oder definite Nominalphrasen



Anaphorische Referenz: Koreferenz (2)

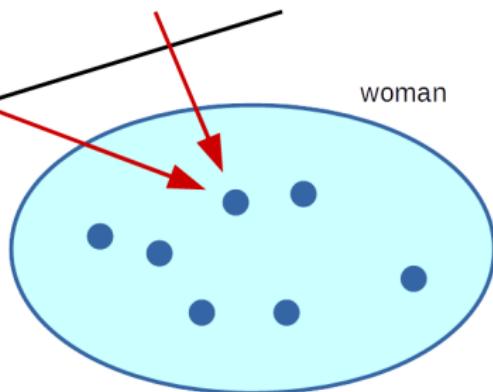
Hans liebt Maria.

man



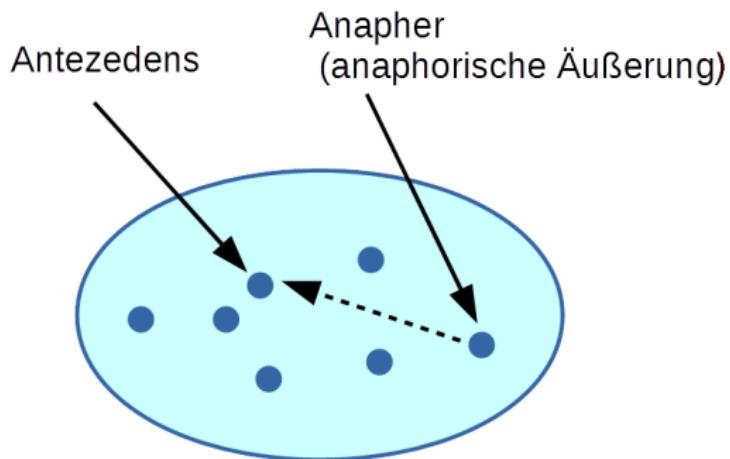
Und sie liebt ihn.

woman



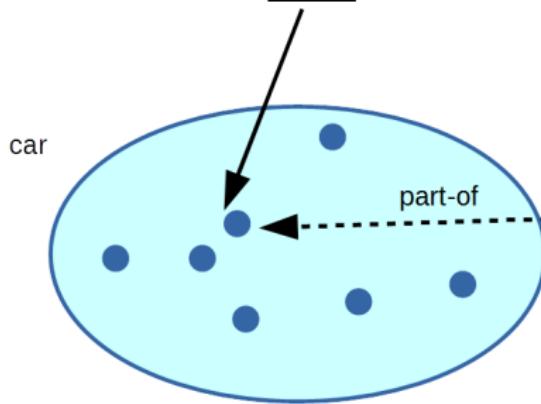
Anaphorische Referenz: Bridging (1)

- Bridging = Assoziative Referenz:
- der Referent der Anapher kann auf Grund einer assoziativen Beziehung (z.B. Teil-Ganzes, Eigenschaften etc.) aus dem Antezedens geschlossen werden.
- typischerweise durch definite Nominalphrasen

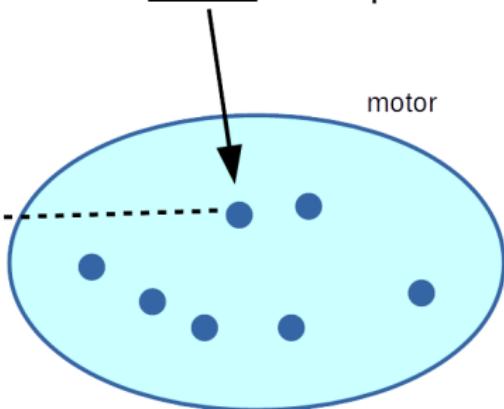


Anaphorische Referenz: Bridging (2)

Hans hat ein Auto.



Der Motor ist kaputt.



Singleton

A mention referring to a referent is called a singleton if this referent is not mentioned again in the text. (= single reference to this referent)

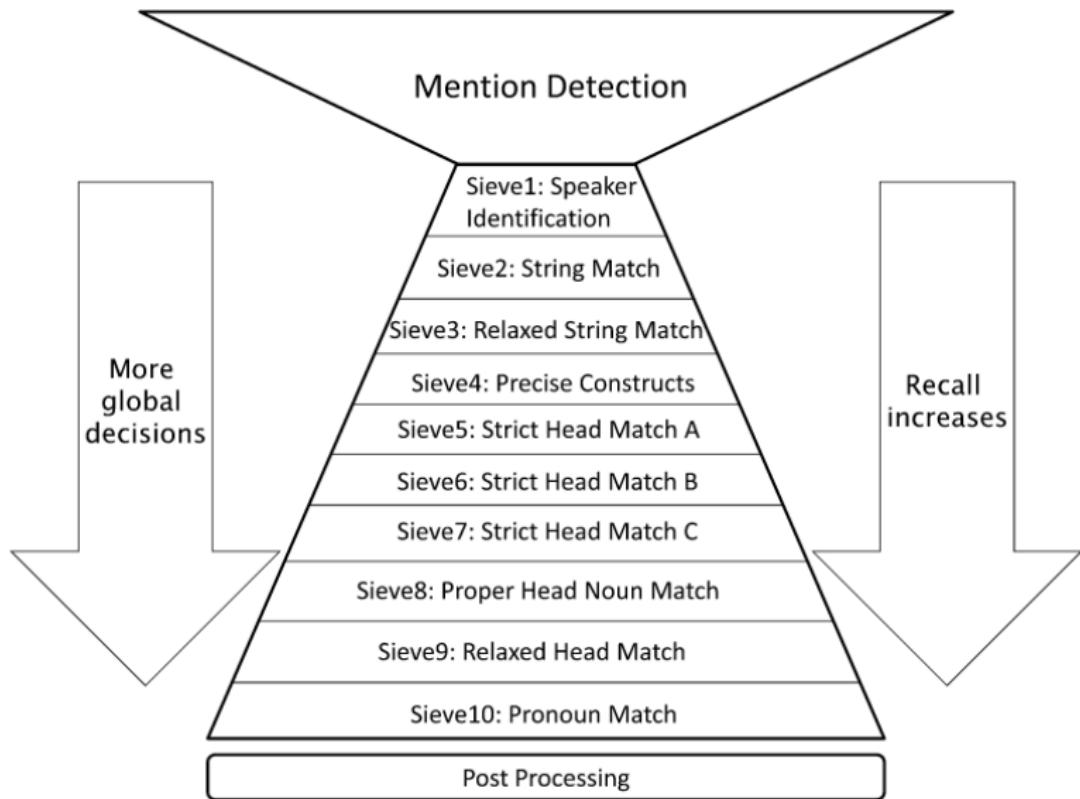
Anaphernauflösung = coreference resolution

- Automatische Anaphernauflösung ist eine Herausforderung für die automatische Sprachverarbeitung.
- Wichtig für viele Anwendungsgebiete:
 - Information Retrieval
 - Summarization
 - Machine translation
 - Dialogue Systems

Methoden zur Koreferenzauflösung

- Korpusannotation: Erstellen von Trainings- und Testdatensätzen (z.B. OntoNotes)
- Statistische Ansätze (z.B. Berkeley's System)
 - Nomen und Pronomen werden mit Hilfe von Merkmalen beschrieben
 - **Mention-Pair Models**: Algorithmus (maschinelles Lernen) wird trainiert, um zu entscheiden, ob sich zwei Nominalphasen auf dasselbe Objekt beziehen → Kettenbildung als 2. Schritt

Koreferenzauflösung: Regelbasierter Ansatz (Lee et al. 2013)



Koreferenzauflösung: Beispiel (1)

Nochmal die Terminologie: Anapher (*sie*), Antezedent (*Alice*)

Koreferenzketten

Alice fing an sich zu langweilen; *sie* saß schon lange bei *ihrer Schwester* am Ufer und hatte nichts zu tun. Das Buch, das *ihrer Schwester* las, gefiel *ihr* nicht; denn es waren weder Bilder noch Gespräche darin. "Und was nützen Bücher," dachte *Alice*, "ohne Bilder und Gespräche?"

Koreferenzauflösung: Beispiel (2)

Nochmal die Terminologie: Anapher (*sie*), Antezedent (*Alice*)

Koreferenzketten

Alice fing an sich zu langweilen; *sie* saß schon lange bei *ihrer Schwester* am Ufer und hatte nichts zu tun. Das **Buch**, das *ihrer Schwester* las, gefiel *ihr* nicht; denn es waren weder Bilder noch Gespräche darin. „Und was nützen **Bücher**,“ dachte *Alice*, „ohne Bilder und Gespräche?“

Schwierigkeit: generische Ausdrücke (*Buch* - *Bücher*?)

Zum Schluss: Besonders klausurrelevant

- Wortbedeutungs-Relationen, WordNet
- Word sense disambiguation, Lesk-Algorithmus
- Jaccard-Koeffizient
- Wahrheitsbedingungensemantik
- Syntax der Aussagenlogik
- Negation, Konjunktion, Disjunktion, Implikation, Äquivalenz
- Prädikatenlogik:
Prädikate, Variablen, Existenzquantor, Allquantor
- Übersetzung:
 - deutscher/englischer Satz in die Aussagenlogik
 - deutscher/englischer Satz in die Prädikatenlogik
- Kompositionalsprinzip
- Vollverb, Stützverbkonstruktion, Idiom
- Anaphorische Referenz: Koreferenz, Bridging
- Coreference resolution