Einführung in die Computerlinguistik
Semantik

Hinrich Schütze

Center for Information and Language Processing

2018-11-19

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

Outline

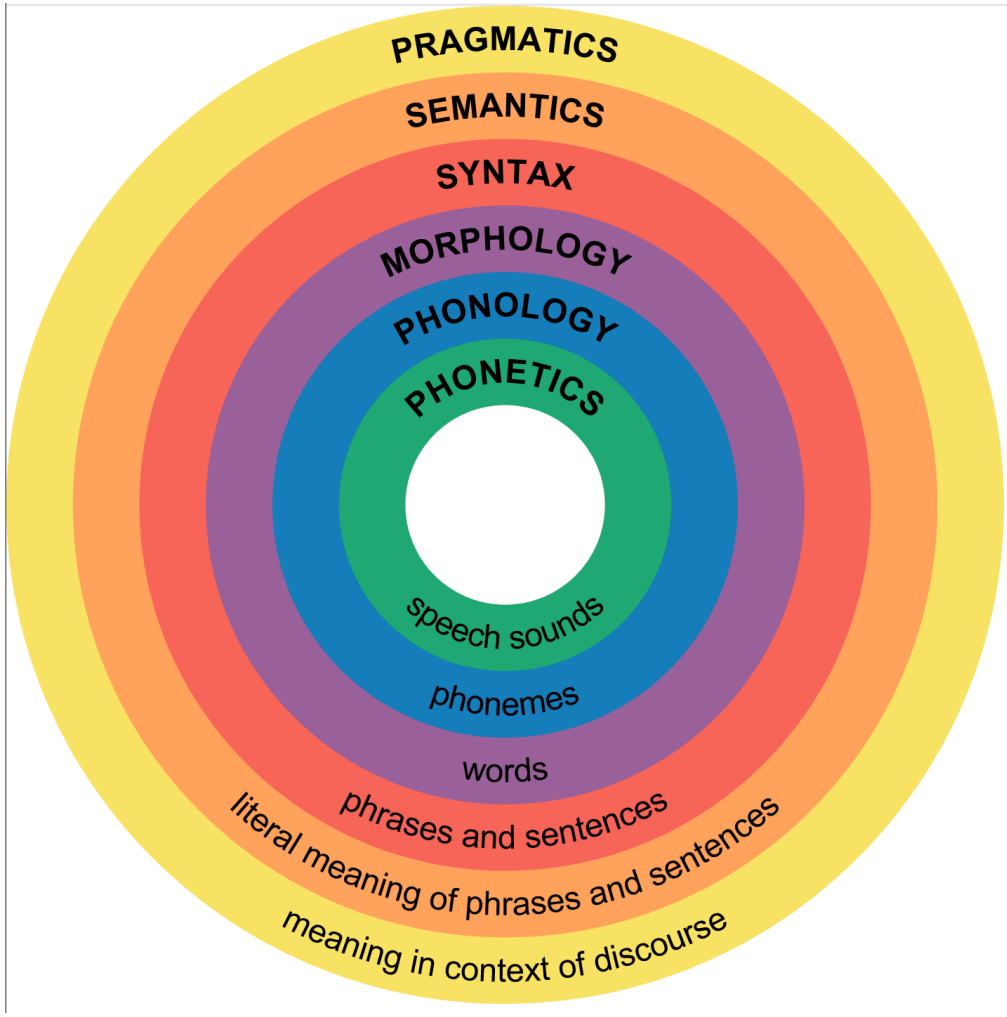
1. Intro
2. Lexikalische Semantik
3. Textähnlichkeit
4. Logik
5. Zeit&Aspekt
6. Kompositionalität
7. Referenz

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 sẽma = Zeichen, Merkmal
- wörtlich:
  Lehre von der Bedeutung
- (der Sprache)


PRAGMATICS
SEMANTICS
SYNTAX
MORPHOLOGY
PHONOLOGY
PHONETICS
speech sounds
phonemes
words
phrases and sentences
literal meaning of phrases and sentences
meaning in context of discourse

Kaffee machen

Motivation

Machen Sie mir
bitte
eine Tasse Kaffee!

Sprache/Text

$\Rightarrow$

Verarbeitung

Motivation

Machen Sie mir
bitte
eine Tasse Kaffee!

Sprache/Text

$\Rightarrow$

Verarbeitung

Motivation

Machen Sie mir
bitte
eine Tasse Kaffee!

Sprache/Text

$\Rightarrow$

Verarbeitung

Motivation

Machen Sie mir
bitte
eine Tasse Kaffee!

Sprache/Text

$\Rightarrow$

Verarbeitung

Motivation

Machen Sie mir
bitte
eine Tasse Kaffee!

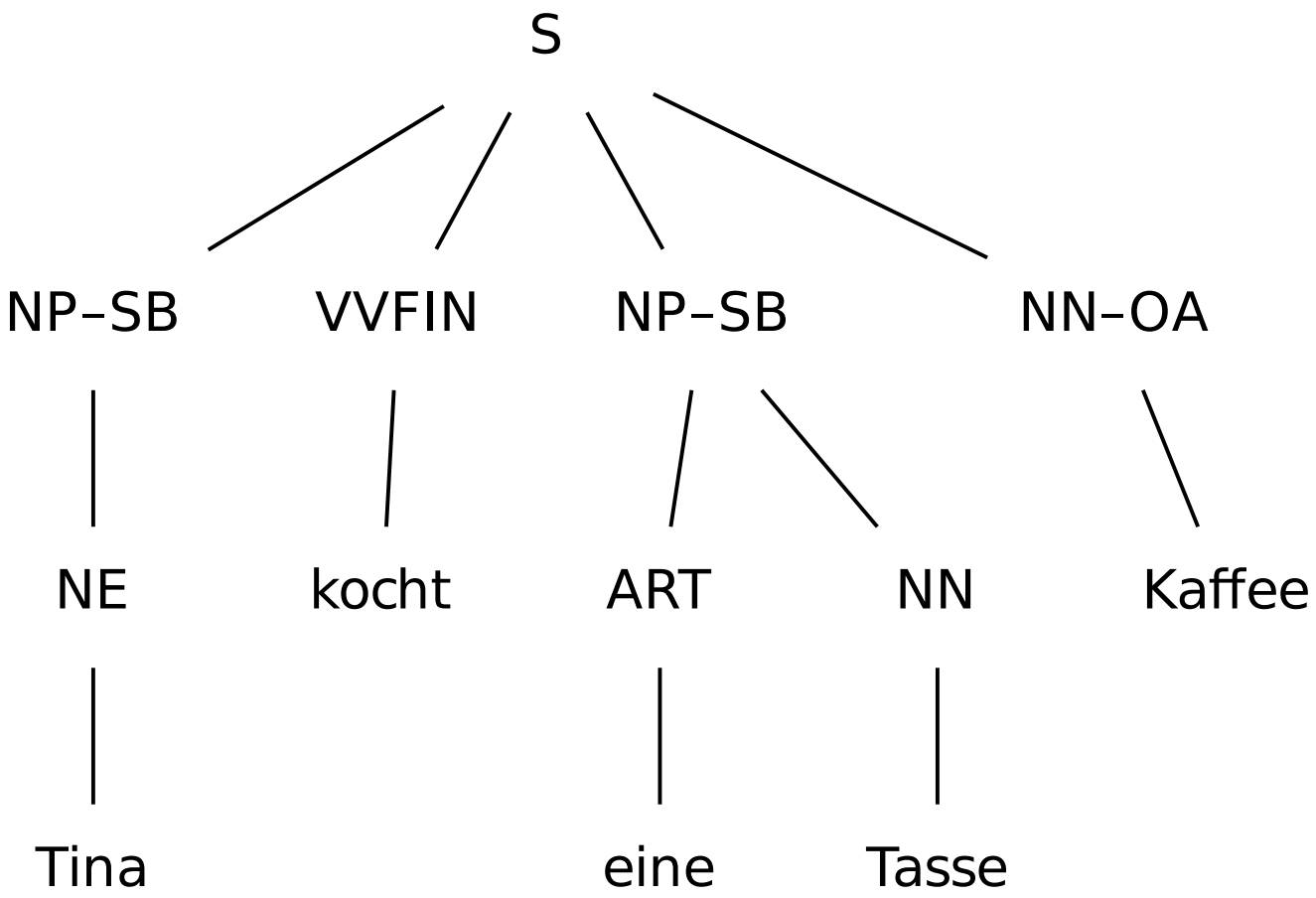
Sprache/Text

$\Rightarrow$

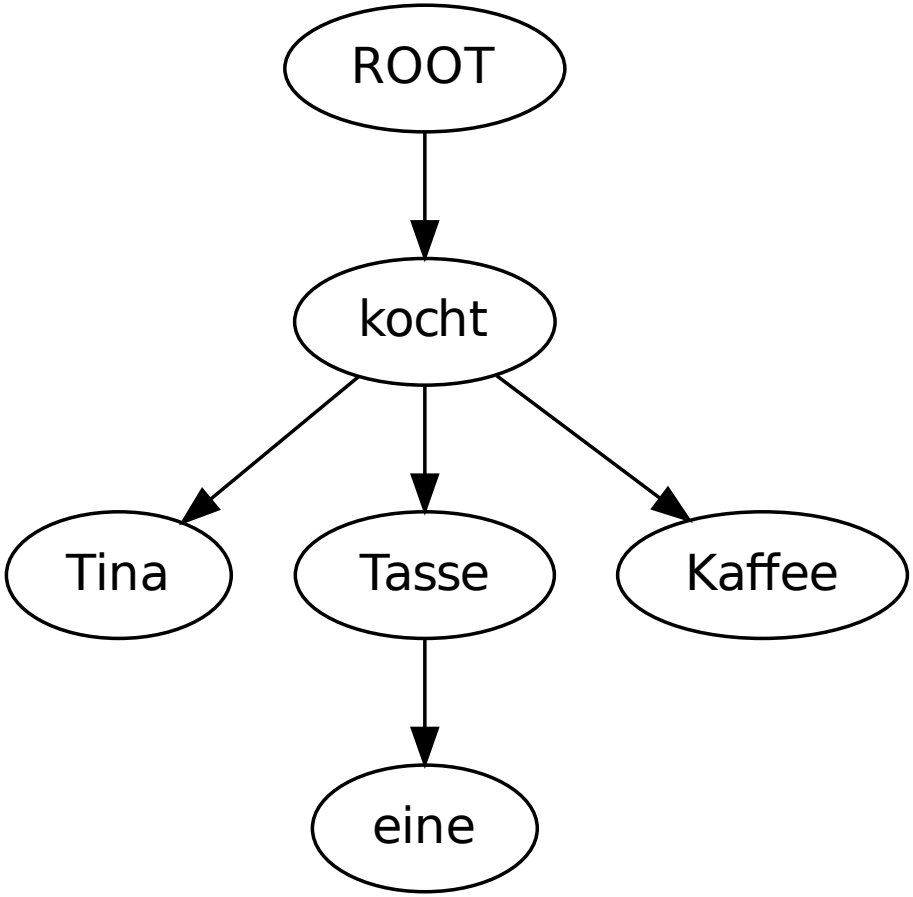
Verarbeitung

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

Konstituenten-Baum?



Dependenz-Baum?



Konstituenz/Dependenz:
nur Syntax.
Wir brauchen auch Semantik.

Semantikrepräsentation: WordNet

WordNet

Noun (46){07945759} ¡noun.food¿[13] S: (n) coffee#1 (coffee%1:13:00::), java#2 (java%1:13:00::) (a beverage consisting of an infusion of ground coffee beans) “he ordered a cup of coffee”

`https://wordnet.princeton.edu/`

Semantikrepräsentation: Semantische Rollen

PropBank

agree.01
Arg0: Agreer
Arg1: Proposition
Ex: [$_{Arg0}$ The goup] *agreed* [$_{Arg1}$ it wouldn't make an offer . . .]

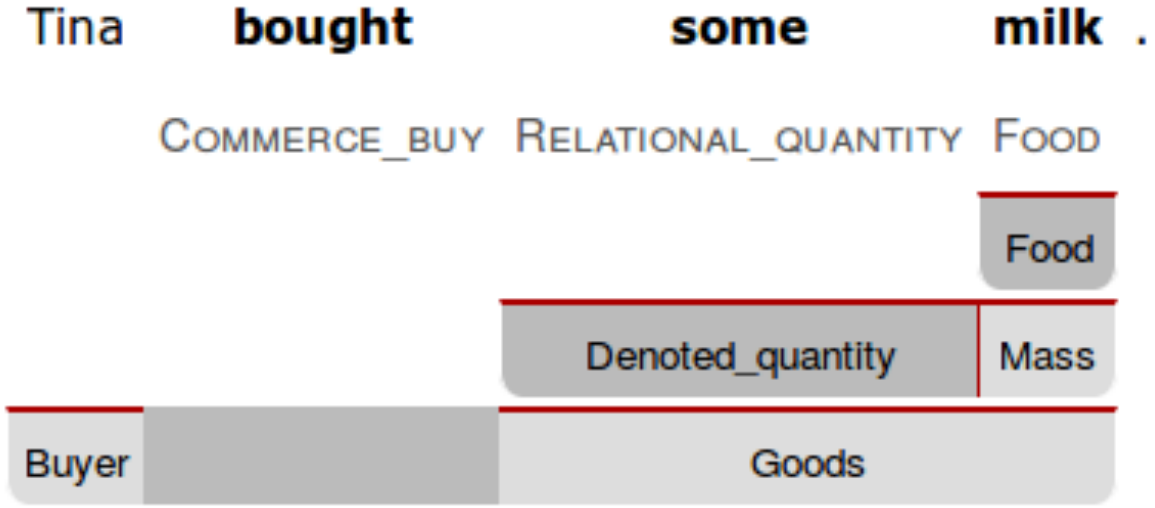
Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. Computational linguistics, 31(1), 71-106.
`http://propbank.github.io`

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/fndrupal`



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 :quant 1
:unit (m / mug)))
:ARG3 (i / i))
```

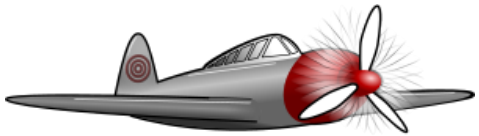
http://amr.isi.edu

Example: Thanks to Tim O'Gorman.

Anforderung an Semantikrepräsentationen

Verifizierbarkeit

Es muss möglich sein, eine Aussage mit einem Zustand der Welt zu verbinden, sodass wir testen können, ob ein Satz wahr ist.



"Das Flugzeug hat einen Propeller."

$\Downarrow$

Repräsentation: `part-of(Propeller, Flugzeug))`

$\Downarrow$

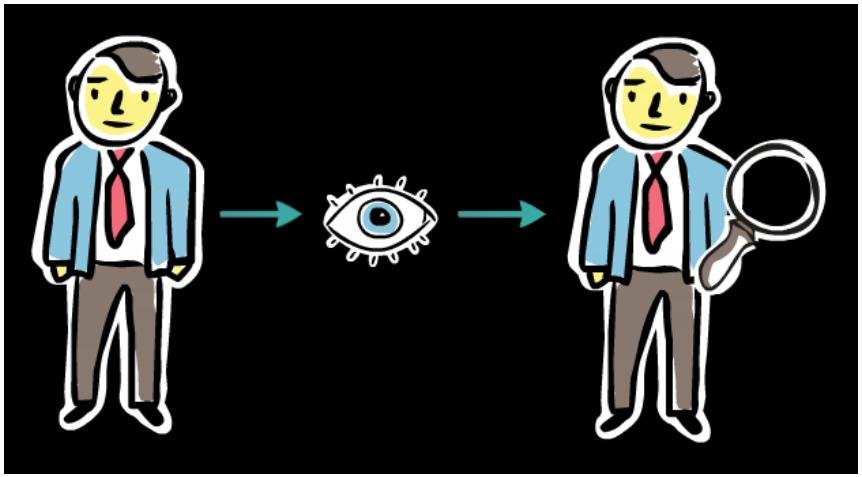
Kommt dies in der Wissensbasis (*knowledge base*) vor?

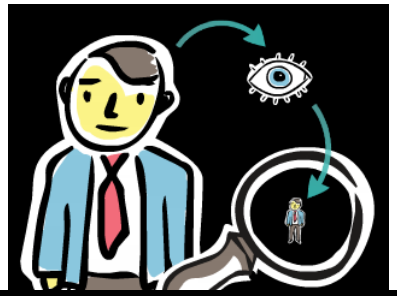
Anforderung an Semantikrepräsentationen

Eindeutigkeit = Nicht ambig

Sprache ist ambig, aber wir wollen eine eindeutige Semantikrepräsentation.

"Hans sah den Mann mit der Lupe."





Anforderung an Semantikrepräsentationen

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?"
$\Rightarrow$ Egal, wie die Frage gestellt wird, wir wollen eine Repräsentation für unsere Datenbankabfrage.
Zum Beispiel: `BietetAn(Almaz,VeganesEssen)` $\rightarrow$ true/false?
Paraphrasen erkennen

Anforderung an Semantikrepräsentationen

Inferenzen

Alle Dackel sind Hunde.
Alle Hunde mögen Wurst.
Fiffi ist ein Dackel.
$\Rightarrow$ Fiffi mag Wurst. wahr oder falsch?

Wir wollen nicht nur das direkt gesagte repräsentieren, sondern auch neue Aussagen / neues Wissen davon ableiten.

Anforderung an Semantikrepräsentationen

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`

Outline

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

Wort-Bedeutungs-Relation

Beispiel

Wort: Bank

Konzepte:

Homonym = ein Wort, steht für verschiedene

Word-Bedeutungs-Relation

Weitere Relationen zwischen Konzepten ($\neq$ zwischen Wörtern)

1. Synonymie: zwei Begriffe bedeuten dasselbe
2. Hypernymie: ein Begriff ist ein Überbegriff eines anderen
3. Hyponymie: ein Begriff ist ein Unterbegriff eines anderen
4. Homophonie: zwei Begriffe klingen gleich, werden aber unterschiedlich geschrieben
5. Meronymie: ein Begriff verweist auf ein Teil eines anderen Begriffs
6. Holonymie: ein Begriff verweist auf ein Ganzes, von dem der andere Begriff ein Teil ist

a. Dach, Haus
b. Hand, Finger
c. Mensch, Lebewesen
d. Leere, Lehre
e. Geige, Violine
f. Kuchen, Marmorkuchen

WordNet

- Eine 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 + Anwendungsbeispiele

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

- WordNet: 170.000 lexikalische Einträge (Wörter) / 120.000 Synsets
- WordNet-Versionen gibt es für etwa 45 Sprachen
- WordNet wird in vielen sprach- und informationstechnologischen Anwendungen erfolgreich genutzt, insbesondere als Grundlage von Inferenz

https://wordnet.princeton.edu

Word Sense Disambiguation (WSD)

Wortbedeutung im Kontext

Schwester Maria sprach ein Gebet.
Meine Schwester ärgert oft meinen kleinen Bruder.
Woher weiß ich, 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)

Word Sense Disambiguation: Trainingskorpus

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. ...

Word Sense Disambiguation: Trainingskorpus

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. ...

Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

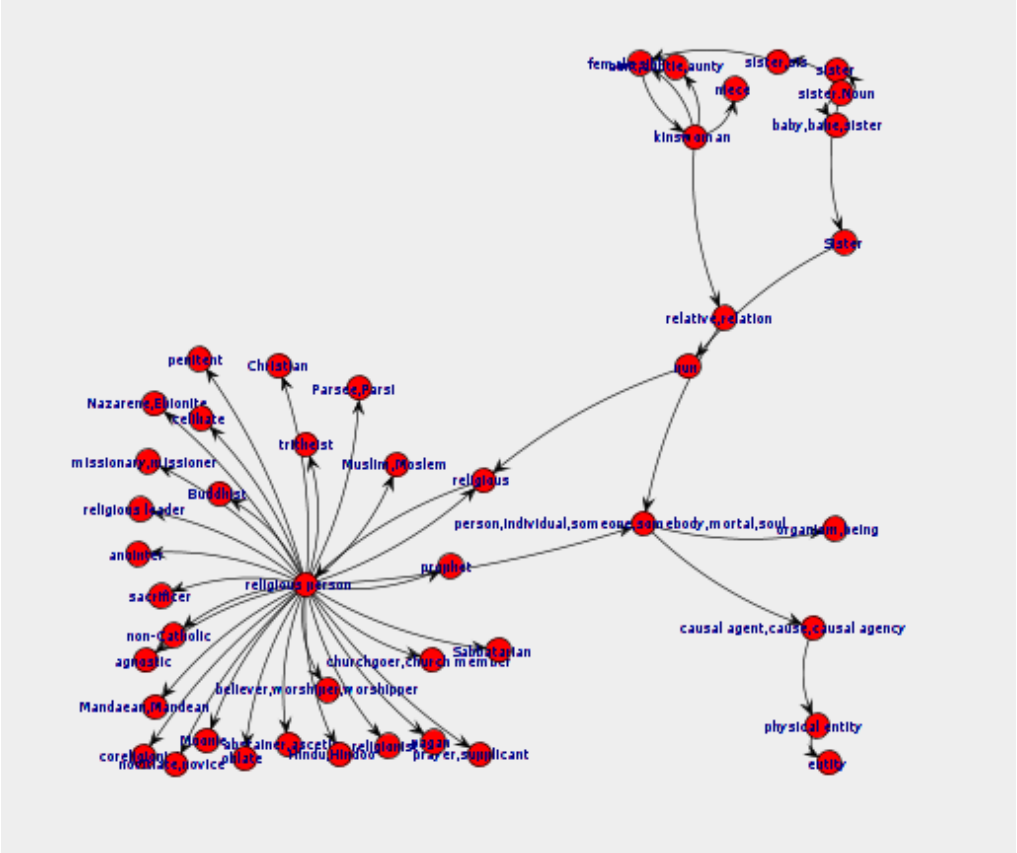
Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

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: overlapp of zero words
- Select the sense with the greatest overlap

WSD: Pfad-basierte Algorithmen (WordNet-basiert)

Ähnlichkeit zwischen zwei Konzepten $\propto$ Distanz im Graph

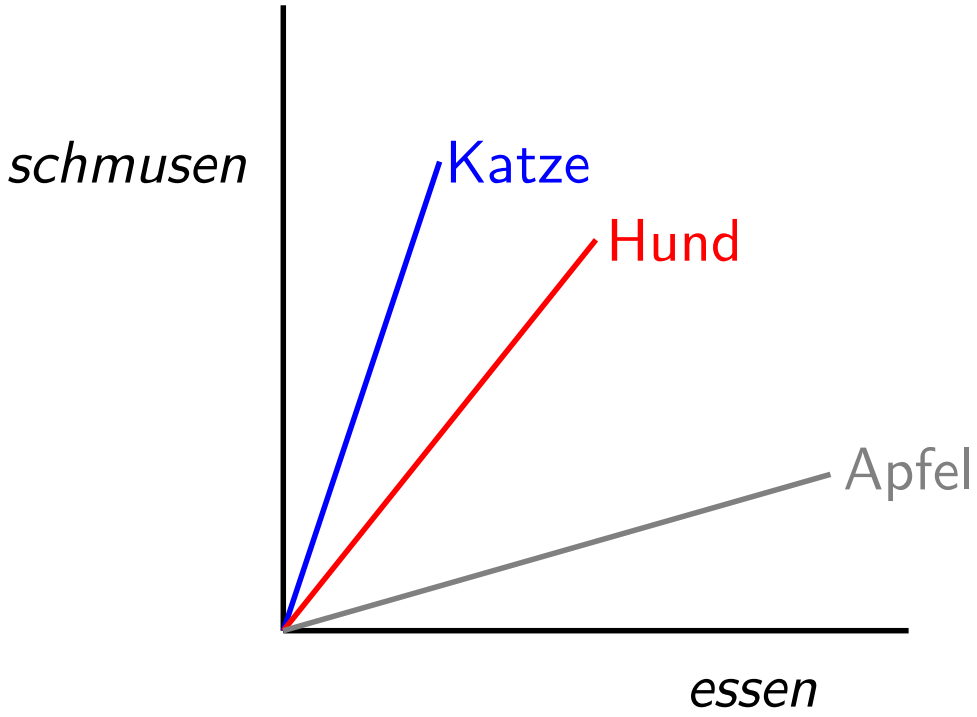


WordNet browser

WSD: Distributionelle Methoden

- Wörter ($\neq$ Konzepte) werden als Vektoren dargestellt
- Dimensionen = Wörter des Vokabular
- Einträge = wie häufig das Kontextwort vorkam (evtl. lemmatisiert, normalisiert)
- Ähnlichkeit von zwei Wörtern: Winkel zwischen den Kontextvektoren (je kleiner der Winkel, desto ähnlicher)

	schmusen	essen
Katze	100	30
Hund	40	100
Apfel	1	100
...	...	...
...	...	...
...	...	...
...	...	...



Übung: Lesk-Algorithmus

Disambiguieren Sie "state" im Satz "Dry ice is the solid state of carbon dioxide."

- state, province (the territory occupied by one of the constituent administrative districts of a nation) "his state is in the deep south"
- state (the way something is with respect to its main attributes) "the current state of knowledge"; "his state of health"; "in a weak financial state"
- state (the group of people comprising the government of a sovereign state) "the state has lowered its income tax"
- state, nation, country, land, commonwealth, res publica, body politic (a politically organized body of people under a single government) "the state has elected a new president"; "African nations"; "students who had come to the nation's capitol"; "the country's largest manufacturer"; "an industrialized land"
- state of matter, state ((chemistry) the three traditional states of matter are solids (fixed shape and volume) and liquids (fixed volume and shaped by the container) and gases (filling the container)) "the solid state of water is called ice"
- state (a state of depression or agitation) "he was in such a state you just couldn't reason with him"
- country, state, land (the territory occupied by a nation) "he returned to the land of his birth"; "he visited several European countries"
- Department of State, United States Department of State, State Department

Outline

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

Textähnlichkeitsmaß: Jaccard

- Grundidee des Lesk-Algorithmus: die Definition aus wählen, die zum Kontext die größte Ähnlichkeit hat.
- Maß für Textähnlichkeit: 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:

$$\textsc{jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

  $(A \neq \emptyset$ or $B \neq \emptyset)$
- $\textsc{jaccard}(A, A) = 1$
- $\textsc{jaccard}(A, B) = 0$ 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?

Outline

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

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

Aussagesatz (Proposition) $\rightarrow$ Wahrheitswert

Beispiel

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)

Beispiele

Semantik der Aussagenlogik

- Die Semantik einer Aussage ist ihr Wahrheitswert.
- Die Semantik einer komplexen Aussage wird rekursiv aus der Semantik der Junktoren und der Teilaussagen berechnet.
- Semantik der Junktoren: Wahrheitstabelle

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 = wahr / true F = falsch / false

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.	$\neg p$		

Negation: “nicht”, “kein”

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

Negation: “nicht”, “kein”

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

Negation: “nicht”, “kein”

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.	$p \vee q$				

ausschließendes vs nicht ausschließendes oder

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

ausschließendes vs nicht ausschließendes oder

Implikation: (kein perfekt passendes deutsches Wort)

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

Wenn p, dann muss q auch eintreten.

Aber q kann auch ohne p eintreten $\rightarrow$ kein Rückschluss möglich.

Äquivalenz: “genau dann wenn”

Tim trinkt Kaffee.	p	T	T	F	F
Anna kocht Tee.	q	T	F	T	F
Tim trinkt Kaffee genau dann, wenn Anna Tee kocht.	$p \leftrightarrow q$				

Äquivalenz: “genau dann wenn”

Tim trinkt Kaffee.	p	T	T	F	F
Anna kocht Tee.	q	T	F	T	F
Tim trinkt Kaffee genau dann, wenn Anna Tee kocht.	$p \leftrightarrow q$	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.	$p \leftrightarrow q$				

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.	$p \leftrightarrow q$	F	T	T	F

Prädikatenlogik (erster Stufe) – First-Order Logic

- Aussagenlogik:
  Aussagen werden als nicht analysierbare Einheiten behandelt.
- Prädikatenlogik:
  Wir stellen auch die interne Struktur von Aussagen dar.

Prädikatenlogik (erster Stufe) – First-Order Logic

Individuenkonstanten

`Hans, Maria, das Buch, ‘‘Dracula’’ ...`

Prädikate und Argumente

Hans ist ein Kind. `kind(Hans)`
Einstelliges Prädikat $\approx$ Eigenschaft

Hans mag Maria. `mag(Hans, Maria)`
Zweistelliges Prädikat $\approx$ Relation

Individuenvariablen

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

Prädikatenlogik (erster Stufe) – First-Order Logic

Existenzquantor

$\exists$`x (kind(x)` $\wedge$ `mögen(x, Maria))`

Es gibt ein x, sodass x ein Kind ist und Maria von x gemocht wird.

Ein (irgendein bestimmtes) Kind mag Maria.

Allquantor

$\forall$`x mögen(x, Maria)`

Für jedes x gilt, dass Maria von x gemocht wird.

Alle mögen Maria.

$\forall$`x (kind(x)` $\rightarrow$ `mögen(x, Maria))`

Für jedes x, das ein Kind ist, gilt, dass Maria von x gemocht wird.

Alle Kinder mögen Maria.

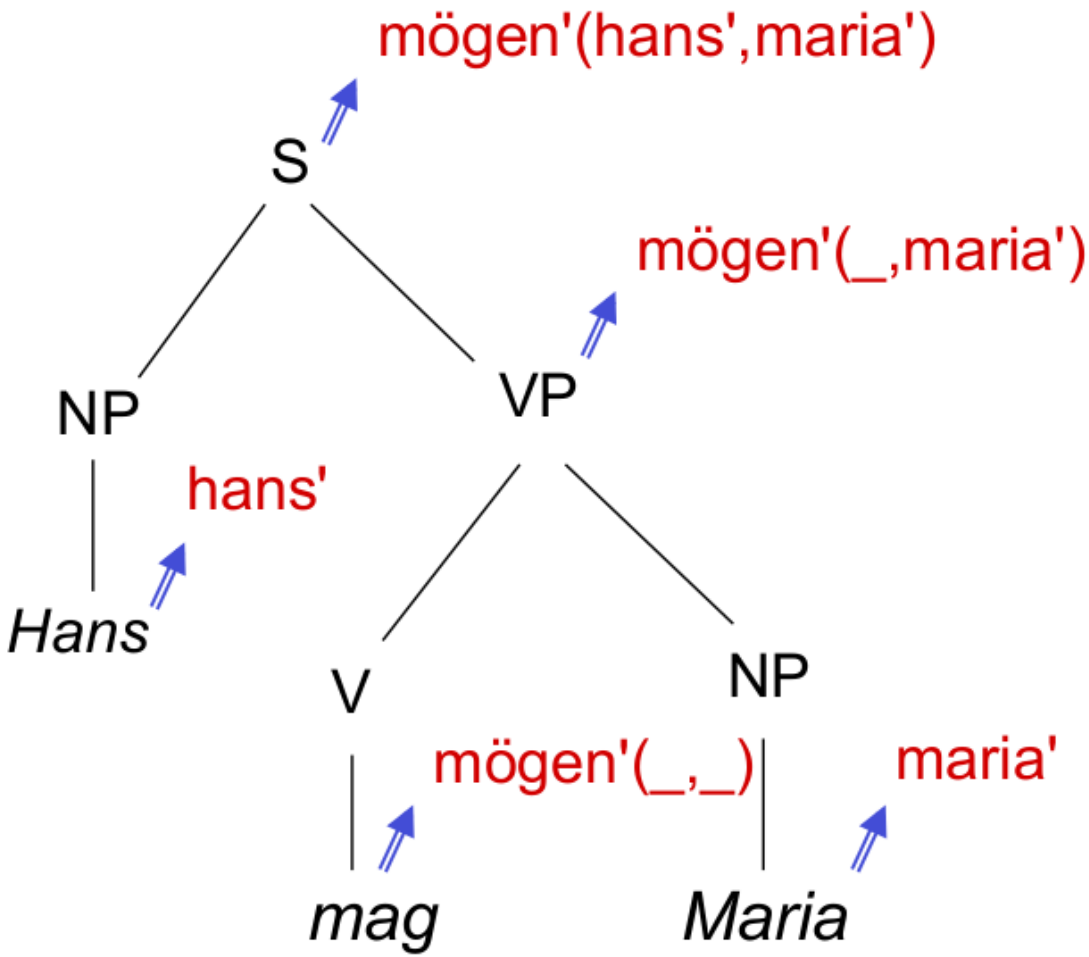
Syntax der Prädikatenlogik (vereinfacht)

- Term: Variable oder Individuenkonstante
- atomare Formel: $P(t_1, \ldots, t_n)$
  (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),
  $\forall xA$, $\exists xA$

Semantik-Konstruktion

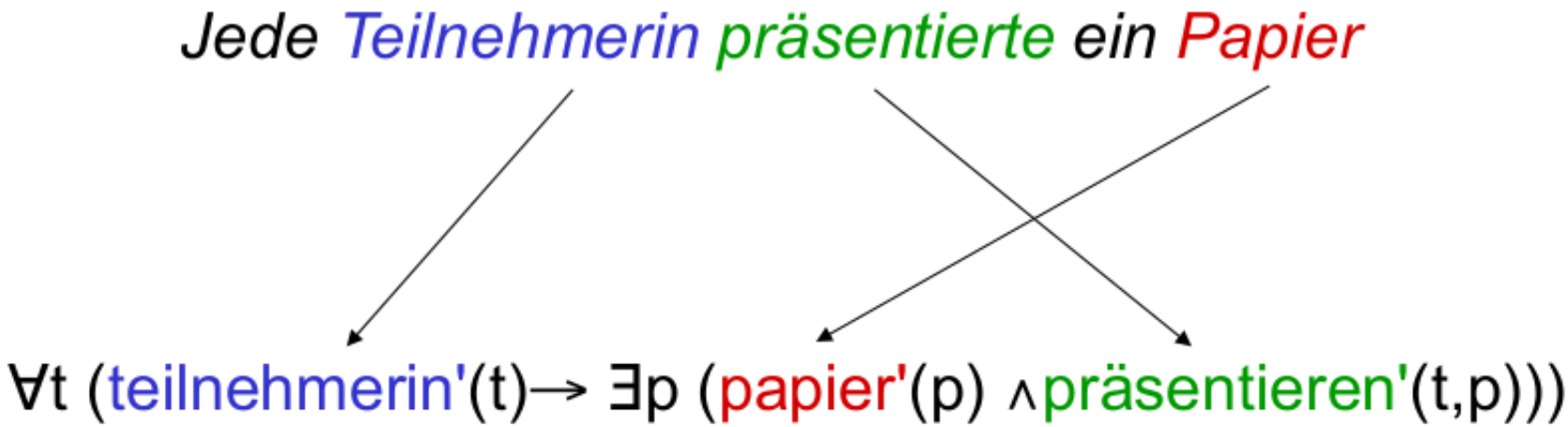
Wir bauen an der Konstituentenstruktur entlang komplexe semantische Ausdrücke "kompositionell" aus einfacheren Ausdrücken zusammen.

Für jeden Lexikoneintrag und jede syntaktische Regel fügen wir noch eine semantische Komponente hinzu.



Semantik-Konstruktion: Herausforderung

Jede Teilnehmerin präsentierte ein Papier.



Übung

Übersetzen Sie in die Prädikatenlogik:

1. Nemo ist ein Fisch.
2. Nemo schwimmt.
3. Nemo ist keine Schildkröte.
4. Ein Student träumt.
5. Jeder Student kennt ein Buch.

Beispiele

"Irgendein Kind mag Maria."

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

"Alle mögen Maria."

$\rightsquigarrow \forall x\ \texttt{mögen(x, Maria)}$

"Alle Kinder mögen Maria."

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

Outline

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

Aspekt

Vendler-Klassen: innere temporale Struktur von Verben / Sätzen
Zeno Vendler: Verbs and times. 1957.

- 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."

Zustände vs. Ereignisse

Zustände (STATE)

. . . beschreiben Eigenschaften, die sich über einen gewissen Zeitraum nicht ändern.

"Anne wohnte von 2011 bis 2016 in Saarbrücken."

Ereignisse (EVENT)

. . . beschreiben Veränderungen.

"Anne zog im Oktober nach München."

Interpretation von Tempus nach Hans Reichenbach

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

Präsens

Anna kocht (gerade) einen Tee.

E
R
S

$E = R = S$

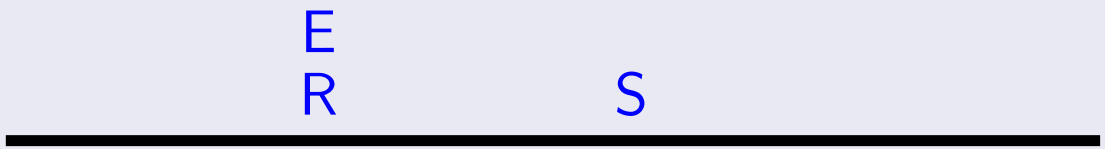
Repräsentation von Zeit

Interpretation von Tempus nach Hans Reichenbach

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

Präteritum

Anna kochte einen Tee.



$E = R < S$

Repräsentation von Zeit

Interpretation von Tempus nach Hans Reichenbach

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

Plusquamperfekt

Anna hatte einen Tee gekocht.

$E < R < S$

Übung

Interpretation von Tempus nach Hans Reichenbach: Zeichnen Sie auf der Zeitachse S, R und E ein – jeweils eine Zeitachse pro Satz.

1. Anna wird einen Tee kochen.
2. Anna wird einen Tee gekocht haben.

Outline

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

Kompositionalität

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 ...

brauchen wir, um die genau Art der Zusammensetzung zu bestimmen.

Vollverben vs. Stützverbkonstruktionen

Vollverb: Semantik klar, Bedeutung ist eine Hauptbedeutung aus dem Wörterbuch, kompositionell

Vollverben

Susi nimmt sich einen Keks. $\rightarrow$ Susi + nehmen + Keks

Stützverbkonstruktionen: Bedeutung des Verbs "schwach", Nomen (Argument) muss miteinbezogen werden.

Stützverbkonstruktion (light verbs, support verbs)

Susi nimmt ein Bad. $\neq$ Susi + nehmen + Bad

Idiom

Wortverbindung (Phrase oder Kompositum), deren Gesamtbedeutung sich nicht aus den Einzelbedeutungen der Wörter ableiten lässt

Beispiele:
“Friedhof”, “rotes Tuch”, “die Rechnung ohne den Wirt machen”
Idiome sind nicht kompositionell.

Aufgabe

Idiom? Stützverbkonstruktion? Vollverb?

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

Outline

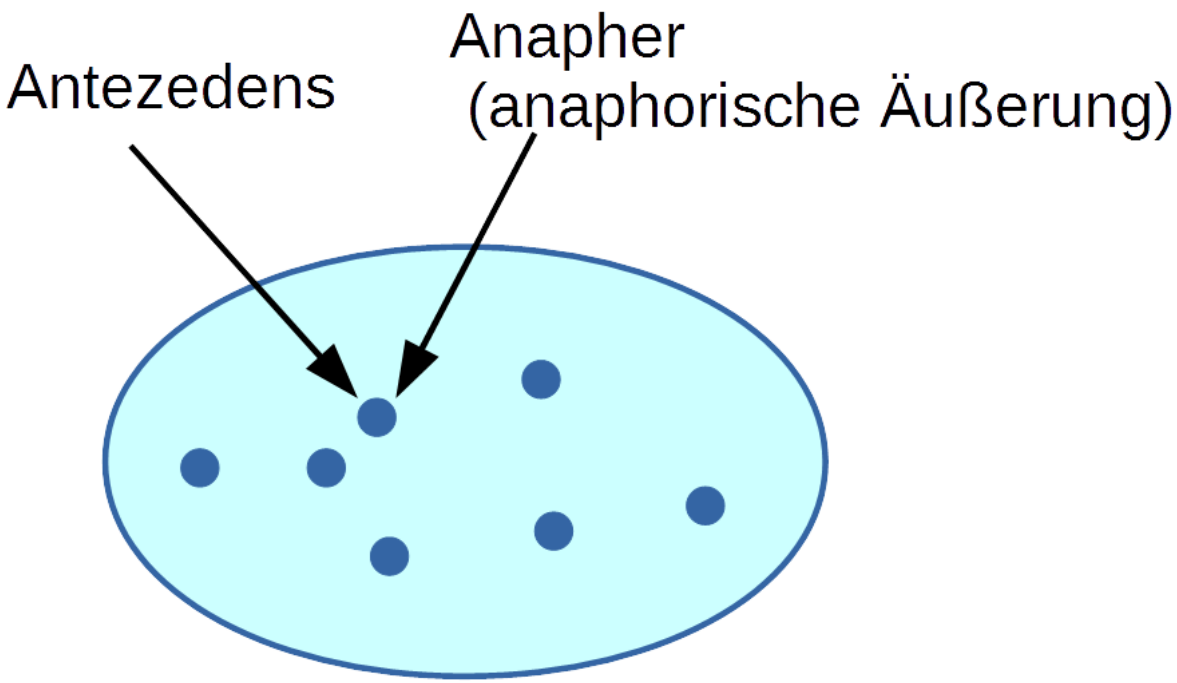
1. Intro
2. Lexikalische Semantik
3. Textähnlichkeit
4. Logik
5. Zeit&Aspekt
6. Kompositionalität
7. Referenz

Anaphorische Referenz

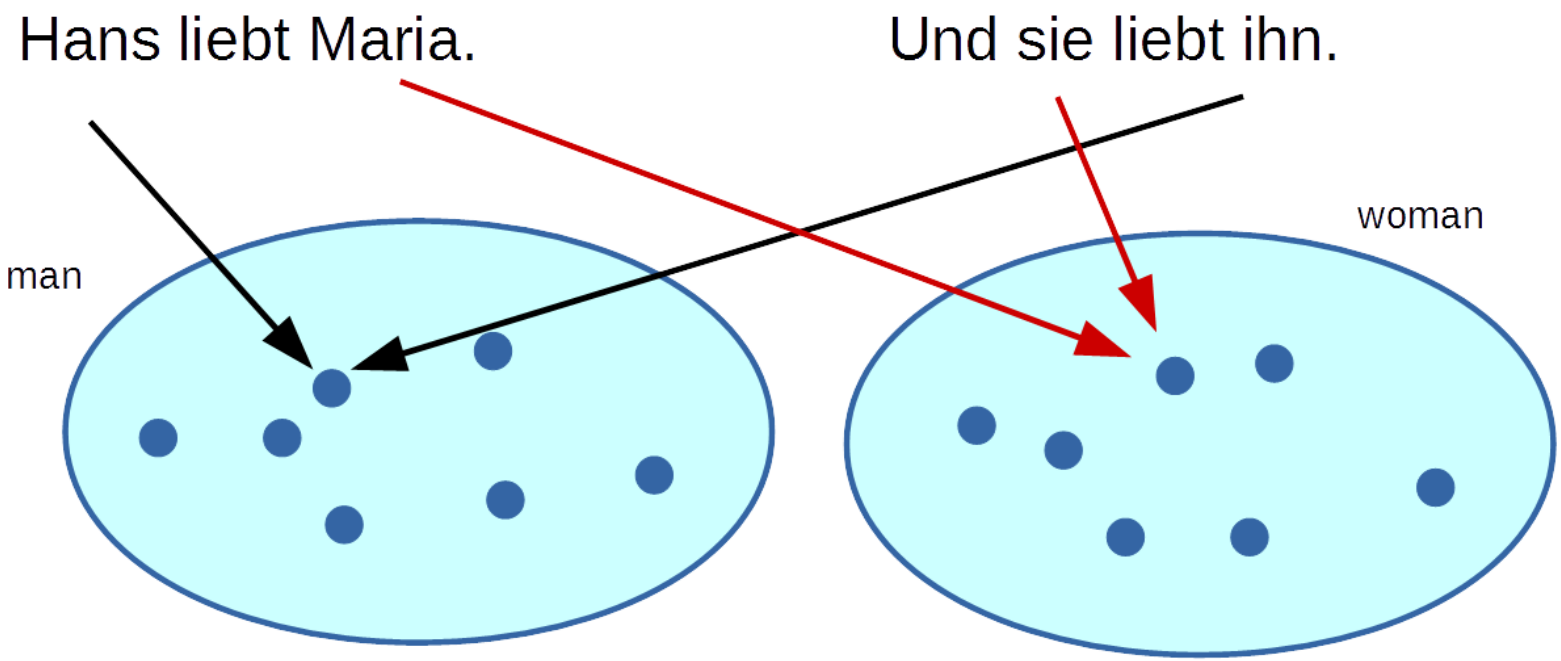
- aus dem Altgriechischen *ana* = zurück *phero* = tragen
- Anaphorische Ausdrücke zeigen auf Referenten, die im sprachlichen Kontext schon eingeführt wurden:
  Maria . . . sie . . . die(se) Frau . . .
- Zwei Arten von anaphorischer Referenz:
  - Koreferenz (bei identischem Referenten)
  - Bridging (assoziative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

Anaphorische Referenz: Koreferenz

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

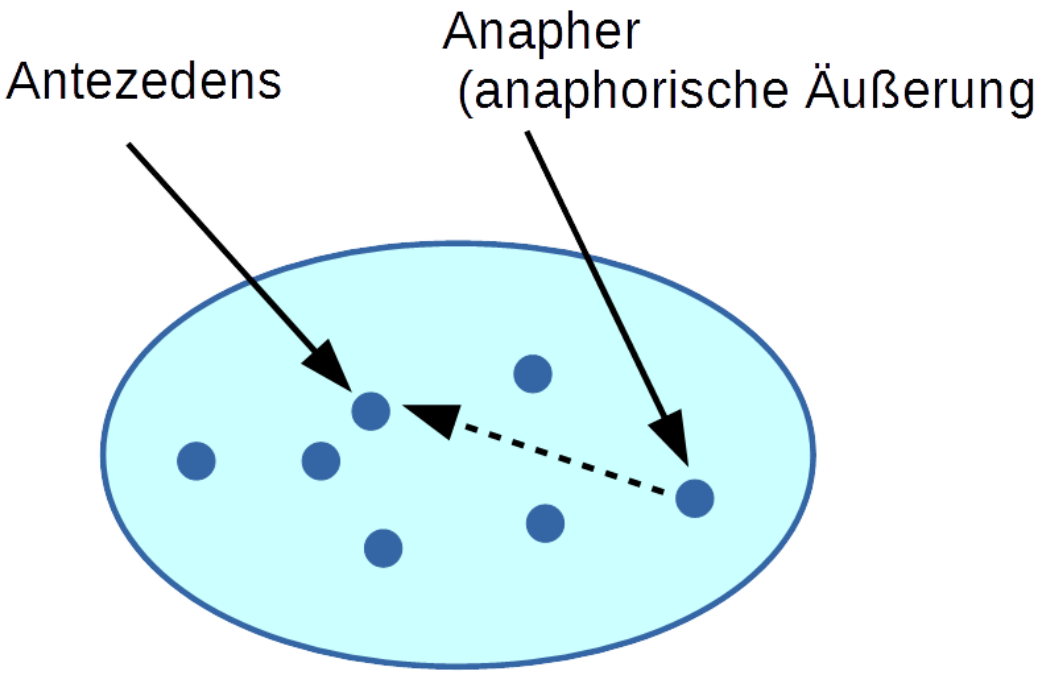


Anaphorische Referenz: Koreferenz

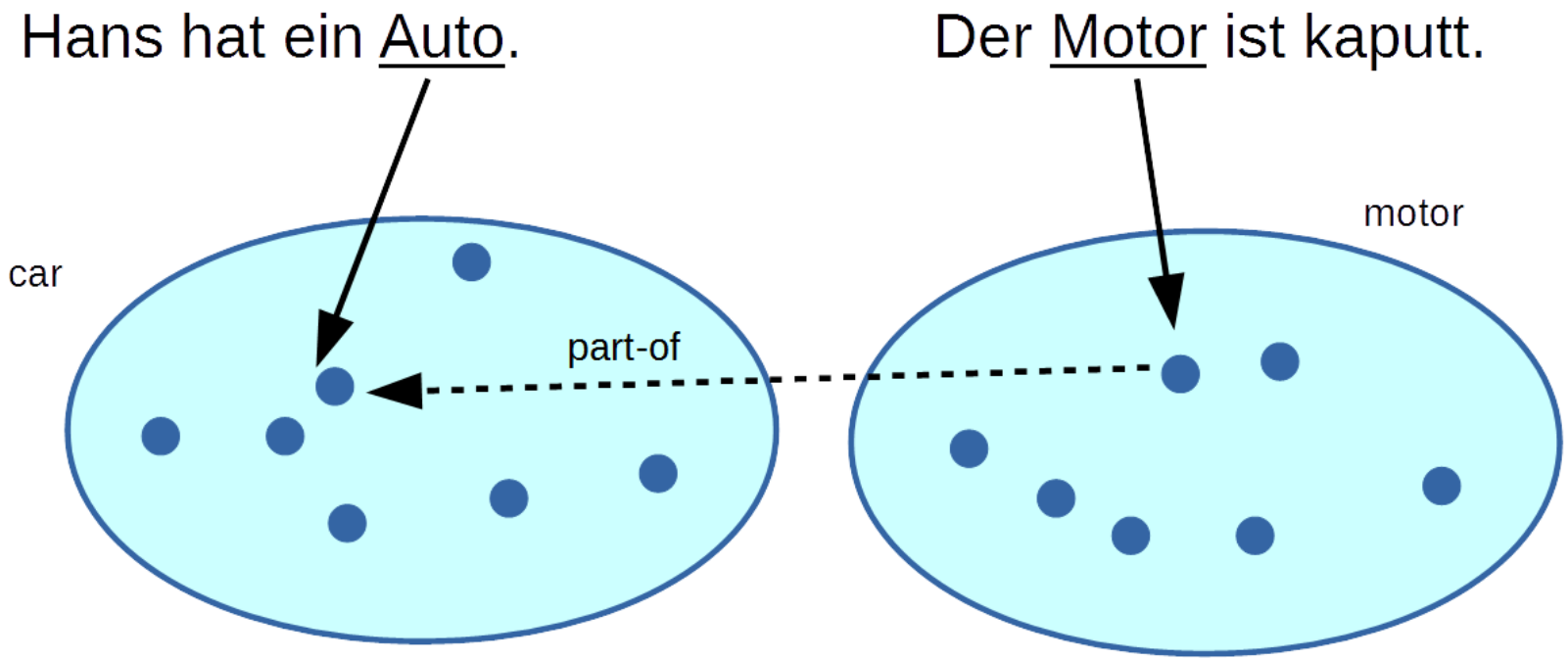


Anaphorische Referenz: Bridging

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



Anaphorische Referenz: Bridging



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, zu entscheiden, ob sich zwei Nominalphasen auf dasselbe Objekt beziehen $\rightarrow$ Kettenbildung als 2. Schritt

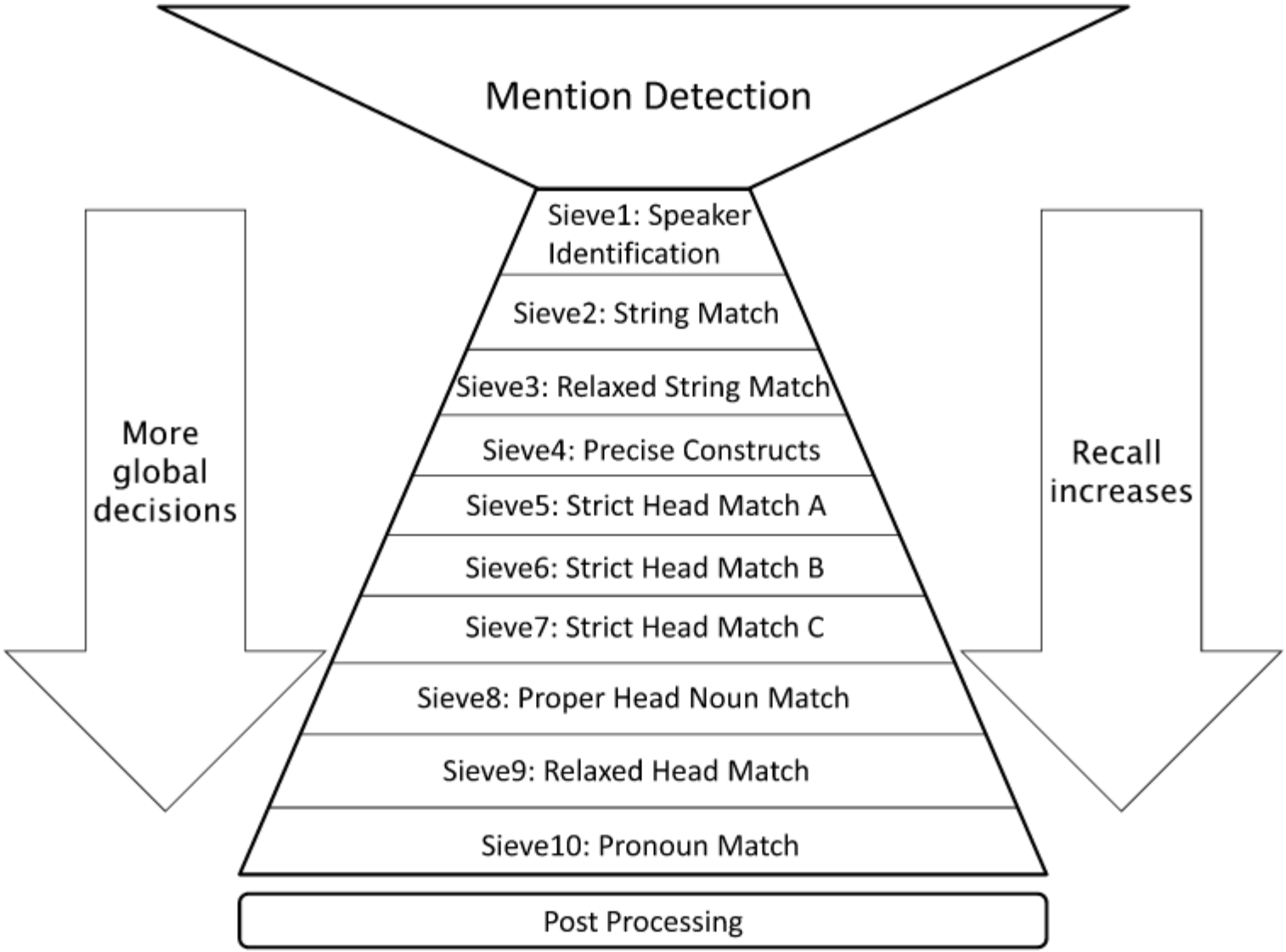
Word Sense Disambiguation: Merkmalsspezifikation

text: Satz	Beim Test Anlageberatung der Banken [bank2] löste kein Institut die einfache Frage nach einer sicheren Anlage wirklich gut.
kmale: Kontextwörter	*Anlageberatung, löste, sicheren Anlage, . . .*
kmalsvektor: Eintrag edes Wort im Vokabular	Angebot: 0, Anlage: 1, Anlageberatung: 1, Garten: 0, löste: 1, Rückenlehne: 0, schwingend: 0, sicheren: 1, . . .

Maschinelles Lernen

Algorithmus entscheidet, ob der Vektor mehr wie diejenigen für bank1 oder bank2 aussieht.

Koreferenzauflösung: Regelbasierter Ansatz



Koreferenzauflösung: Beispiel

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 ihre 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

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

Koreferenzketten

Alice fing an sich zu langweilen; sie saßschon lange bei ihrer Schwester am Ufer und hatte nichts zu tun. Das Buch, das ihre 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*?)

Übung

Welche Äußerungen referieren anaphorisch? Bestimmen Sie den jeweiligen Antezedenten!

"Cindy ist strohblond. Ihr Fell schimmert in der Sonne. Als Jeannette auf den Hof tritt, hebt die Haflingerstute den Kopf und blinzelt ihr von der Stalltür aus zu. Dann fängt es an zu regnen."

Übung

Entwickeln Sie einen Algrithmus und wenden Sie ihn an, der für jedes "referierende" Wort im folgenden Text den Antezedenten findet. Wie kann man die Bestimmung der Antezedenten in diesen Fällen automatisieren?

"A small maltese dog named Soshe has been reunited with her owners after spending a week on a floating sofa after floods, brought on by Hurricane Florence, devastated parts of North Carolina last month. Soshe's owner, who was out of the state at the time, was unable to return to her flooded home in the town of Burgaw. She made frequent calls to the Humane Society of Missouri (HSMO), who sent a team to North Carolina to help rescue animals trapped by the powerful storms. The HSMO team, who were using Pender County Animal Control as a base of operations, made two attempts to find the stranded dog but the flooding was so severe that it made locating the house extremely difficult. On their third attempt the motor on their boat stopped,

Zum Schluss: Besonders klausurrelevant

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