

# Einführung in die Computerlinguistik

## Definitions and Examples

Hinrich Schütze & Robert Zangeneid

Centrum für Informations- und Sprachverarbeitung, LMU München

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Sprache

ein System von Zeichen (Wortschatz) und Regeln (Grammatik) zur Mitteilung von Bedeutungen

## Token

konkretes Vorkommen eines Zeichens (Laut, Buchstabe, Wort, Satz, Text, ...) an einem Ort zu einer Zeit

## Type

Klasse von Token, die von ihren Sprechern und Hörern nicht unterschieden werden und daher als gleich wahrgenommen werden

## Computational linguistics

Computational linguistics is the scientific study of models and methods for automatic processing of natural language.

Computational linguistics is an interdisciplinary field that shares a large part of its subject matter with computer science and linguistics. However, computational linguists also work on theories, models and methods that are not part of core linguistics or core computer science.

## Text corpus

A corpus (plural corpora) or text corpus is a large and structured set of texts, nowadays usually electronically stored and processed.

## Heaps' law

$$M = kT^b$$

$M$  is the size of the vocabulary,  $T$  is the number of tokens in the collection. Typical values for the parameters  $k$  and  $b$  are:  
 $30 \leq k \leq 100$  and  $b \approx 0.5$ .



## Morphologie

Lehre von der Bildung und Struktur von Wörtern; und davon wie Oberflächenmodifikationen Verwendung & Bedeutung des Wortes verändern

## Gebrauch / use

Ein Wort, mit dessen Hilfe über etwas gesprochen wird (also über dessen Denotat), heißt “gebraucht”.

“Es ist fraglich, ob Hans ein guter Vater für das Kind ist.”

## Erwähnung / mention

Ein Wort, über das gesprochen wird, heißt “erwähnt”.

“Es ist fraglich, ob ‘Hans’ ein guter Name für das Kind ist.”

## Verteilung / distribution

Verteilung eines Zeichens  $Z$ ; Menge der Kontexte, in denen  $Z$  vorkommt

## Paradigmatische Sprachachse

Beziehung von einem Zeichen (Wörtern, Wortformen) zu anderen Zeichen des gleichen Paradigmas, Ebene der Ersetzung

## Syntagmatische Sprachachse

Beziehung von einem Zeichen (Wortformen) zu Zeichen in seinem Kontext (in einem konkreten Satz), Ebene der Kombination

Die Studentin sitzt in der Vorlesung  
Ein Student lernt im Seminar  
Hans liest im Hörsaal

↑  
paradigmatisch  
↓

← syntagmatisch →

## Wohlgeformtheit / Grammatikalität

Ein sprachlicher Ausdruck  $A$  aus einer Sprache  $L$  heißt wohlgeformt, wenn er (laut Intuition der Sprecher von  $L$ ) ein gültiger Ausdruck von  $L$  ist.



## Semiotisches Dreieck: Symbol

Ausdrucksseite des sprachlichen Zeichens: signifiant (Saussure),  
Signifikant, Bezeichnendes

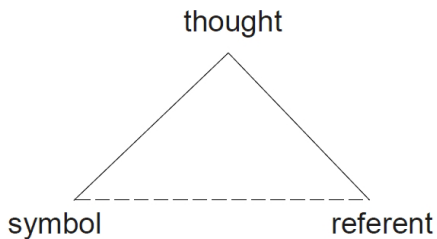
## Semiotisches Dreieck: Thought

Inhaltsseite des sprachlichen Zeichens: signifié (Saussure), Signifikat (Morris), Bezeichnetes, Sinn (bei Frege), Bedeutung (außer bei Frege), Intension (Carnap)

## Semiotisches Dreieck: Referent

Etwas, das seinen Platz in der außersprachlichen Wirklichkeit hat:  
Gegenstand, Ereignis etc.: Denotat (Morris), Extension (Carnap),  
Bedeutung (bei Frege)

## Semiotisches Dreieck



## Word shape

Replace all lower case letters with 'x', replace all upper case letters with 'X', replace all digits letters with '9', "deduplicate" any sequence of  $n > 1$  identical characters. "submarine" ->

"xxxxxxxxxx" -> "x"; "London" -> "Xxxxxx" -> "Xx"; "half-filled" -> "xxxx-xxxxxx" -> "x-x"; "LMU" -> "XXX" -> "X"; "2015" -> "9999" -> "9"; "3.1415" -> "9.9999" -> "9.9"; "Yahoo!" -> "Xxxxx!" -> "Xx!"

## Onomatopoeia

(Lautmalerei) the formation of a word from a sound associated with what is named (e.g., cuckoo, sizzle) – exception to the rule that symbols are arbitrary

## What is a word?

The concept of word is a prototype / family resemblance concept based on orthographic / graphemic, phonological, morphological, morphosyntactic and semantic criteria. These criteria are neither necessary nor sufficient.

## Lexikalische Funktion

eine syntaktisch-semantische Korrelation zwischen linguistischen Ausdrücken, die systematisch und häufig in einer Sprache auftritt

$$f(X) = Y$$



## Zwei Arten von lexikalischen Funktionen

Syntagmatische LFs: kombinieren X und  $f(X)$ , z.B., MAGN

Paradigmatische LFs: ersetzen X durch  $f(X)$ , z.B. SYN

## LF $V_0$ : Syntaktisches Derivat

Gleiche Bedeutung, andere Wortart

$V_0$ (Interesse) = sich interessieren

$V_0$ (Aggression) = angreifen

## LF Conv: Konversivum

Gleiche Bedeutung, andere Reihenfolge der Argumente

Person1 kauft Auto2 von Person3 für Betrag4.

Person3 verkauft Auto2 an Person1 für Betrag4.

1234 → 3214

## Stützverb

Ein Verb, das im Kontext des Argumentes **semantisch fast bedeutungsleer ist** und **fast nur eine syntaktische Funktion** hat.  
**idiomatisch und einzelsprachspezifisch**

Beispiele: "Widerstand **leisten**", "put up resistance"  
auch: Funktionsverb, support verb, light verb

## To paraphrase

to express the meaning of (the writer or speaker or something written or spoken) using different words

## Kompositionalität

Eine Phrase wird **kompositional** (oder auch kompositionell) genannt, wenn sich ihre Bedeutung vollständig aus der Bedeutung der Töchter und der Kompositionalitätsregel ergibt.

## Free/compositional combinations: Examples

Mary gave Peter the book.

The car is red.

## Non-Free/Non-compositional combinations: Examples

He kicked the bucket.

Er steht auf dem Schlauch.



## Thematic hierarchy and canonical realization

Agent	>	Theme	>	Goal	>	Oblique
subject		object		indirect object		prepositional phrase
Mary		the car		to Peter		for ten dollars

Mary sold the car to Peter for ten dollars.

- Agent > Experiencer > Goal/Source/Location > Theme (Grimshaw 1990)
- Agent > Theme > Goal > Oblique (Larson 1988)
- Agent > Benefactive > Recipient/Experiencer > Instrument > Theme/Patient > Location

## Konventionalität

Typischerweise gibt es in einer Sprache viele verständliche Möglichkeiten, einen Sachverhalt auszudrücken, aber nur einige wenige, **die konventionellen Ausdrucksmöglichkeiten**, werden standardmäßig verwendet.

## Beispiel für Konventionalität

“ich bin mit dem Auto gefahren / gekommen” (konventionell)  
vs. “Ich habe das Auto genommen” (nicht-konventionell)

“I took the car” / “I drove” (konventionell)  
vs. “I drove/came with the car” / “I drove the car”  
(nicht-konventionell)

## Phraseme

Kollokationen, Idiome, Quasi-Idiome, Pragmateme

## Kollokation

Kombination von A und B; A: Standardbedeutung; B: geänderte Bedeutung in Abhängigkeit von A

“starker Raucher”

## Idiom

Kombination von A und B; weder die Bedeutung von A noch die Bedeutung von B kommen in "dominanter Position" vor  
"Er steht auf dem Schlauch."

## Quasi-Idiom

Kombination von A und B; zur Kombination der Bedeutungen kommt noch **zusätzliche Bedeutung** dazu; und/oder "syntaktisch abhängiges Lexem nimmt dominante Position ein"

"bacon and eggs" (hinzu kommt: *Streifen* von Speck)

"start a family" (conceive a first child with one's spouse, starting a family)

## Pragmatem

Kombination von A und B; Signifikat ist regulär; aber die vom Sprecher beschriebene Situation erfordert einen ganz bestimmten Ausdruck

“mindestens haltbar bis”



## Freie Kombination / Verbindung

Kombination von A und B; Signifikat ist regulär; die vom Sprecher beschriebene Situation erfordert **nicht** einen ganz bestimmten Ausdruck

“Der Tisch ist rund.”

Why are traditional lexicons difficult to use in computational linguistics?

One reason is that their definitions are circular.

## Wierzbicka's axiomatic approach

Three levels: (i) semes, semantic primitives, words with a simple, elementary meaning, (ii) more complex intermediate concepts (e.g., "sky"), (iii) all other words of the language

## Examples for semes

Quantifiers:

ONE, TWO, SOME, ALL, MUCH~MANY, LITTLE~FEW

Space:

WHERE~PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE,  
INSIDE

# Summary

- Konventionalität
- Beispiel für Konventionalität
- Phraseme
- Kollokation
- Idiom
- Quasi-Idiom
- Pragmatem
- Freie Kombination / Verbindung
- Why are traditional lexicons difficult to use in computational linguistics?
- Wierzbicka's axiomatic approach
- Examples for semes

## Die zwei Bereiche der Morphologie

Flexionslehre

Wortbildungslehre

## Morphem

Kleinste bedeutungstragende Einheit

## Lexikalisches Morphem

trägt lexikalische Bedeutung, z.B. {sag}



Grammatisches Morphem (Grammem, funktionales Morphem)

hat rein grammatische Funktion, z.B. {te} ( $\rightarrow$  *sagte*)

## Freies (ungebundenes) Morphem

kann ohne Vorhandensein anderer Morpheme ein Wort bilden; z.B.  
{Garten}, {Zwerg}

## Gebundenes Morphem

kann nicht selbständig ein Wort bilden, z.B. {s} als Genitiv Singular

## 4 Arten von Affixen

- Suffixe; z.B. {sam} (z.B. *aufmerksam*)
- Präfixe; z.B. {auf}
- Infixe; (z.B. engl. *fan-bloody-tastic*)
- Zirkumfixe (diskontinuierlich); z.B. {ge- -t} (→ *gesagt*)

## Grammatikalisierung

Wort geht in Richtung grammatisches Morphem

Beispiel: {te} ( $\rightarrow$  *sagte*), Herkunft wahrscheinlich etwas wie "er sagen tat".

## Lemma – lexikographische Definition

Grundform eines Wortes (Zitierform): “Überschrift” im Eintrag eines Wörterbuchs

Lemma:

Vereinfachte Verwendungweise in der Computerlinguistik

Grundform+Wortart

Beispiele: "boom+v" (das Verb "boom"),  
"boom+n" (das Nomen "boom")

## WordNet: "boom"

### Noun

- **S: (n) boom**, [roar](#), [roaring](#), [thunder](#) (a deep prolonged loud noise)
- **S: (n) boom** (a state of economic prosperity)
- **S: (n) boom**, [bonanza](#), [gold rush](#), [gravy](#), [godsend](#), [manna from heaven](#), [windfall](#), [bunce](#) (a sudden happening that brings good fortune (as a sudden opportunity to make money)) *"the demand for testing has created a boom for those unregulated laboratories where boxes of specimen jars are processed like an assembly line"*
- **S: (n) boom**, [microphone boom](#) (a pole carrying an overhead microphone projected over a film or tv set)
- **S: (n) boom** (any of various more-or-less horizontal spars or poles used to extend the foot of a sail or for handling cargo or in mooring)

### Verb

- **S: (v) boom**, [din](#) (make a resonant sound, like artillery) *"His deep voice boomed through the hall"*
- **S: (v) smash**, [nail](#), **boom**, [blast](#) (hit hard) *"He smashed a 3-run homer"*
- **S: (v) thunder**, **boom** (be the case that thunder is being heard) *"Whenever it thunders, my dog crawls under the bed"*
- **S: (v) boom**, [boom out](#) (make a deep hollow sound) *"Her voice booms out the words of the song"*
- **S: (v) boom**, [thrive](#), [flourish](#), [expand](#) (grow vigorously) *"The deer population in this town is thriving"; "business is booming"*



## Traditional lexicon: “boom”

**boom**<sup>1</sup> | bum |

noun

a loud, deep, resonant sound: *the deep boom of the bass drum.*

- the characteristic resonant call of the bittern.

verb [no obj.]

make a loud, deep, resonant sound: *thunder boomed in the sky* | *her voice boomed out*.

- [with direct speech] say in a loud, deep, resonant voice: *the imperative “Silence!” boomed out by Ray himself.*
- (of a bittern) utter its characteristic resonant call.

DERIVATIVES

**boom-y** adjective

ORIGIN late Middle English (as a verb): ultimately imitative; perhaps from Dutch *bommen* ‘to hum, buzz.’

**boom**<sup>2</sup> | bum |

noun

a period of great prosperity or rapid economic growth: *a boom in precious metal mining* | [as modifier] : *a boom economy.*

verb [no obj.]

enjoy a period of great prosperity or rapid economic growth: *business is booming* | *the popularity of soy-based foods has boomed in the last two decades.*

DERIVATIVES

**boom-let** | 'bʊmlɪt | noun ,

**boom-y** adjective

ORIGIN late 19th cent. (originally US): probably from **BOOM**<sup>1</sup>.

**boom**<sup>3</sup> | bum |

noun

a long pole or rod, in particular:

- a spar pivoting on the after side of the mast and to which the foot of a vessel's sail is attached, allowing the angle of the sail to be changed.
- [often as modifier] a movable arm over a television or movie set, carrying a microphone or camera: *a boom mike.*
- a long beam extending upward at an angle from the mast of a derrick, for guiding or supporting objects being moved or suspended.
- a floating beam used to contain oil spills or to form a barrier across the mouth of a harbor or river.
- a retractable tube for inflight transfer of fuel from a tanker airplane to another airplane.

ORIGIN mid 16th cent. (in the general sense ‘*beam, pole*’): from Dutch, ‘*beam, tree, pole*’; related to **BEAM**.

## Traditional lexicon: Two lemmata, same parts-of-speech (1)

boom<sup>1</sup> | bum |

noun

a loud, deep, resonant sound: *the deep boom of the bass drum.*

- the characteristic resonant call of the bittern.

verb [no obj.]

make a loud, deep, resonant sound: *thunder boomed in the sky* | *her voice boomed out* .

- [with direct speech] say in a loud, deep, resonant voice: *the imperative "Silence!" boomed out by Ray himself.*
- (of a bittern) utter its characteristic resonant call.

DERIVATIVES

**boom•y** adjective

ORIGIN late Middle English (as a verb): ultimately imitative; perhaps from Dutch *bommen* 'to hum, buzz.'

## Traditional lexicon: Two lemmata, same parts-of-speech (2)

boom<sup>3</sup> | bum |

noun

a long pole or rod, in particular:

- a spar pivoting on the after side of the mast and to which the foot of a vessel's sail is attached, allowing the angle of the sail to be changed.
- [ often as modifier ] a movable arm over a television or movie set, carrying a microphone or camera: *a boom mike*.
- a long beam extending upward at an angle from the mast of a derrick, for guiding or supporting objects being moved or suspended.
- a floating beam used to contain oil spills or to form a barrier across the mouth of a harbor or river.
- a retractable tube for inflight transfer of fuel from a tanker airplane to another airplane.

ORIGIN mid 16th cent. (in the general sense *'beam, pole'*): from Dutch, *'beam, tree, pole'*; related to [BEAM](#).

# WordNet vs. Traditional lexicon

	# lemmata	# basic-form+POS	# senses
trad. lexicon	3	2	10
WordNet	2	2	10

WordNet: “deep prolonged loud noise” and “a state of economic prosperity” are grouped together for the entry of the **same basic-form+POS** combination.

Traditional lexicon: “a loud deep resonant sound” and “a period of great prosperity or rapid economic growth” are assigned to **separate lemmata**.

# Summary

- Die zwei Bereiche der Morphologie
- Morphem
- Lexikalisches Morphem
- Grammatisches Morphem (Grammem, funktionales Morphem)
- Freies (ungebundenes) Morphem
- Gebundenes Morphem
- 4 Arten von Affixen
- Grammatikalisierung
- Lemma – lexikographische Definition
- Lemma – vereinfachte Verwendungweise in der Computerlinguistik

# Outline

- 1 October
- 2 11/2**
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Feature semantics

Feature semantics represents a word as a **vector of features** or as a bundle of features. Features are basic meaning elements like “carnivore” and “animate”.

	moon	silver fox	fruitbat	owl	homo sapiens
animate	no	yes	yes	yes	yes
furry	no	yes	yes	no	no
silver	yes	yes	no	no	no
can-fly	no	no	yes	yes	no
carnivore	no	yes	no	yes	yes

## Connotation

an idea or feeling that a word invokes in addition to its literal or primary meaning

Example: the word “discipline” has connotations of punishment and repression



## Polysemy

One linguistic form (one lemma), but two different, related meanings.

Example: school (building) vs. school (institution)

## Cohyponymy

Two nouns are cohyponyms if they are hyponyms of the same hypernym.

“apple” and “pear” are cohyponyms.

## Meronymy / Holonymy

Part-whole relationship / Whole-part relationship

finger is a meronym of hand, hand is a holonym of finger

## Troponymy

presence of a “manner” relation between two lexemes  
“nibble” is a troponym of “eat”

## Scope of a quantifier or operator

the extent of the formula that the quantifier or operator applies to

Example:  $\forall x [[\exists y A(x, y)] \vee B(x)]$

Scope of " $\forall x$ " is " $[[\exists y A(x, y)] \vee B(x)]$ "

Scope of " $\exists y$ " is " $A(x, y)$ "

## Scope ambiguity: Example

In Rome, all men were not free.

Translation 1:  $\forall x[M(x) \rightarrow \neg F(x)]$

Translation 2:  $\neg \forall x[M(x) \rightarrow F(x)]$

## Limitations of logic for the semantics of natural language

- The meaning of “pragmatic” function words like “bei weitem”, “noch” is not captured.
- The relationship between propositions is difficult to capture in first-order logic (“obwohl”, “aber”).
- Implicatures are difficult to capture (“Nicht nur Hans ist im Haus”).
- Many natural-language quantifiers (“most”) and modals (“must”, “can”) cannot be translated into the basic form of predicate logic.
- The natural-language meaning of operators (“if - then”) and quantifiers (“all”) is not the same as the logical meaning.

# Summary

- Feature semantics
- Connotation
- Homonymy
- Polysemy
- Cohyponymy
- Meronymy / Holonymy
- Troponymy
- Scope of a quantifier or operator
- Example of scope ambiguity
- Limitations of logic for the semantics of natural language



# Outline

- 1 October
- 2 11/2
- 3 11/9**
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Homography

Two lemmata that are written the same,  
but have two different, unrelated meanings.

Example:

Lachen (plural of “Lache” ‘puddle’)

vs. Lachen (‘laughing’)

Note: these two words are homographs,  
but they are not homophones

## Homophony

Two lemmata that are pronounced the same, but have two different, unrelated meanings.

Example:

Lärche ('larch', a conifer))

vs. Lerche ('lark')

Note: these two words are homophones, but they are not homographs

## Homonymy

Definition 1: homonym = homograph

Definition 2: homonym = homophone

Definition 3: homonym = homograph and homophone

(Definition 4: homonym = homograph or homophone)

Definition most commonly used in computational linguistics:

Definition 1: homonym = homograph

## Pragmatics

Study of: Actions performed by utterances, intentions motivating utterances, context-dependent meaning, meaning beyond true/false dichotomy

## Semantics vs. Pragmatics

### semantics

---

literal, primary meaning  
what was said literally?  
context-independent meaning  
truth conditions of meaning

### pragmatics

---

utterances as actions  
why was it said?  
context-dependent meaning  
meaning beyond true/false

## Deictic

word or expression whose meaning is dependent on the context in which it is used

personal pronouns (“he”), possessive pronouns (“my”), temporal adverbs (“now”), locative adverbs (“here”)

## Präsupposition

Sinn-Komponente, die der Hörer für wahr halten muss, damit der Satz für ihn einen Sinn ergibt

Beispiel: "Karl weiß nicht, dass Rom die Hauptstadt von Italien ist"  
hat die Präsupposition "Rom ist die Hauptstadt von Italien"



## Generalized conversational implicature C of S

Ordinarily, when somebody says S,  
they also commit to the truth of C.

It would be

misleading to utter S (without explicitly denying C)  
if C were false.

“John criticized Harry for writing the letter” implicates:

“Harry wrote the letter”

defeasible: “but since the letter was actually written by Mary,  
it was quite unfair of John.”

## Speech acts

### Locutionary act:

the act of uttering the words

Example: "İz6nt it rılı hat ın hır"

### Illocutionary act:

the act the utterance is intended to perform,  
e.g., question, assertion, exclamation

Example: "Isn't it really hot in here?" is a question.

### Perlocutionary act:

the rhetorical act intended by the speaker  
in performing the illocutionary act

Example: Speaker wants hearer to open a window.

## Gricean maxims

Maxim of quality

Maxim of quantity

Maxim of relevance

Maxim of manner

# Summary

- Homography
- Homophony
- Homonymy
- Pragmatics
- Semantics vs. Pragmatics
- Deictic
- Präsupposition
- Generalized conversational implicature C of S
- Speech acts
- Gricean maxims

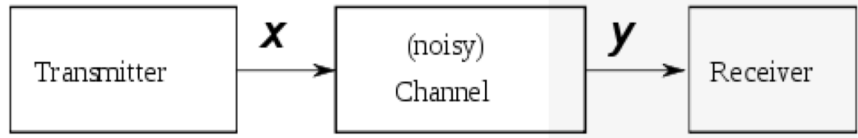
# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13**
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Statistical Natural Language Processing

Statistical Natural Language Processing (StatNLP) uses methods of supervised, semisupervised and unsupervised learning to address tasks that involve written or spoken (human) language.

### Noisy channel (picture)



## Noisy channel (formula)

Message “word sequence” generated by  $P(\text{word-sequence})$

Message is transmitted by channel according to

$P(\text{evidence}|\text{word-sequence})$

Receiver observes “evidence”

Receiver “decodes” by finding word-sequence that maximizes

$P(\text{evidence}|\text{word-sequence})P(\text{word-sequence})$



## Noisy channel (example)

Message: English “word sequence” generated by  $P(\text{word-sequence})$

Message is transmitted by channel according to translation model

English  $\rightarrow$  French

$P(\text{evidence}|\text{word-sequence})$

Receiver observes “evidence”, the French sentence

Receiver “decodes” by finding English word-sequence that maximizes  $P(\text{evidence}|\text{word-sequence})P(\text{word-sequence})$

## Independence

Two events  $A$  and  $B$  are independent iff  $P(A \cap B) = P(A)P(B)$

## Simplest way of estimating a probability $P(A)$

Estimate it as a relative frequency

Count the number of events  $A$

Count the total number of events

Estimate  $P(A) = \frac{\text{count}(A)}{\text{count}(\text{everything})}$

## Testing for independence

Estimate  $P(A)$ ,  $P(B)$ ,  $P(AB)$

Compare  $P(A)P(B)$  with  $P(AB)$

$P(AB) \gg P(A)P(B)$ : This indicates  $A$  and  $B$  are strongly dependent.

$P(AB) \approx P(A)P(B)$ : This indicates  $A$  and  $B$  are independent.

$P(AB) \ll P(A)P(B)$ : This indicates  $A$  and  $B$  are strongly dependent (negatively correlated).

## Conditional probability

The conditional probability is the **updated probability of an event given some knowledge**.

Definition:  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  ( $P(B) > 0$ )

## Bayes' theorem

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

## How to automatically estimate a (word) translation model

i.e.,  $P(\text{evidence}|\text{word-sequence})$

Find parallel corpus (e.g., Canadian Hansards), sentence-align it, word-align it, then estimate word-word translation probabilities

# Summary

- Statistical Natural Language Processing
- Noisy channel (picture)
- Noisy channel (formula)
- Noisy channel (example)
- Independence
- Simplest way of estimating a probability  $P(A)$
- Testing for independence
- Conditional probability
- Bayes' theorem
- How to automatically estimate a (word) translation model



# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16**
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## How to estimate $P(e_i|f_j)$

Word-align parallel corpus

Count  $C(f_j)$  the number of alignment links of  $f_j$

Count  $C(f_j - e_i)$  the number of alignment links connecting  $f_j$  and  $e_i$

$$P(e_i|f_j) = C(f_j - e_i)/C(f_j)$$

## Empty cept $e_0$

An imaginary English word that is the “cause” (translation) of all unaligned French words.

## IBM model: Assumption of independence of word translations

For any two translation pairs  $f_j - e_h$  and  $f_k - e_i$ : The probability of  $f_j - e_h$  being correct is independent of the probability of  $f_k - e_i$  being correct.

## Independence of word translations: Example where it fails

Ambiguity of source. "suit" → "Anzug" vs. "suit" → "Prozess"  
depends on translations of other words

## IBM model: Assumption that translation is word-by-word

Each French word is generated by exactly one English word (possibly by the imaginary word  $e_0$ ).

## Assumption that translation is word-by-word: Example where it fails

*m*-to-1 alignments: “lower house” → Unterhaus.  
“Unterhaus” is a rare translation of “lower”  
and a rare translation of “house”.

## IBM model: Assumption of word-alignment independence

A particular alignment has a fixed probability, independent of the words being aligned.



## Assumption of word-alignment independence: Example where it fails

“Siehst Du!” ↔ “You see”

crossed alignments – low probability in general, but the only good translation in this case

## Why is morphology a challenge for machine translation?

If a rare morphological form was not observed in the parallel training corpus, then it cannot be generated as a translation by a purely statistical approach.

Example: “flanierst” is currently not correctly translated by Google Translate

IBM model: What does  $p(\langle 4, 2, 1, 3 \rangle)$  mean?

This is the probability of an alignment in which  $f_1$  is the translation of  $e_4$ ,  $f_2$  is the translation of  $e_2$ ,  $f_3$  is the translation of  $e_1$  and  $f_4$  is the translation of  $e_3$ .

“wütend machte er mich” – “he made me angry”

# Summary

- How to estimate  $P(e_i|f_j)$
- Empty cept  $e_0$
- IBM model: Assumption of independence of word translations
- Independence of word translations:  
Example where it fails
- IBM model: Assumption that translation is word-by-word
- Assumption that translation is word-by-word:  
Example where it fails
- IBM model: Assumption of word-alignment independence
- Assumption of word-alignment independence:  
Example where it fails
- Why is morphology a challenge for machine translation?
- IBM model: What does  $p(< 4, 2, 1, 3 >)$  mean?

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23**
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Statistical language model

A statistical language model is a probability distribution  $P_{\text{LM}}$  over word sequences. For each word sequence  $w_1, \dots, w_n$ , it computes the probability  $P_{\text{LM}}(w_1, \dots, w_n)$  of this word sequence.

## Bigram model $P_{\text{bigram}}$

$$P_{\text{bigram}}(w_1, \dots, w_n) = P(w_1) \prod_{i=2}^n P(w_i | w_{i-1})$$

## Maximum likelihood estimate of $P(w_2|w_1)$

Relative frequency:

$$P_{\text{ML}}(w_2|w_1) = \frac{C(w_1 w_2)}{C(w_1)}$$



## Main problem with the maximum likelihood estimate?

$P_{\text{ML}}(w_2|w_1) = 0$  for  $C(w_1 w_2) = 0$

→ A sequence not observed in the training set is deemed to be impossible.

## Laplace estimate of $P(w_2|w_1)$

$$P_{\text{laplace}}(w_2|w_1) = \frac{C(w_1 w_2) + 1}{C(w_1) + |V|}$$

## What are pros and cons of the Laplace estimate?

Pro: No zeros!

Con: Too much probability mass is reallocated from seen events to unseen events.

Result: Probability estimates for many seen events are way too low.

## Part-of-speech tagging

Part-of-speech tagging is the process of disambiguating the syntactic category of a word in context.

## Brown corpus tag set

AT	article	RB	adverb
BEZ	the word "is"	RBR	comparative adverb
IN	preposition	TO	the word "to"
JJ	adjective	VB	verb, base form
JJR	comparative adj.	VBD	verb, past tense
MD	modal	VBG	verb, present participle, gerund
NN	sng. / mass noun	VBN	verb, past participle
NNP	sng. proper noun	VBP	verb, non-3rd person sng. present
NNS	plural noun	VBZ	verb, 3rd sng. present
PERIOD	. : ? !	WDT	wh-determiner: "what", "which"
PN	personal pronoun		

## Two sources of information for POS tagging

- (i) “bias”, probabilities of different POS tags for a word
- (ii) context of POS tags to the left and right

## Examples for two sources of information for POS tagging

- (i) "bias": "put" as a verb is much more likely than "put" as a noun
- (ii) context: for a JJ/NN ambiguity in the context "AT \_ VBZ", NN is much more likely than JJ

# Summary

- Statistical language model
- Bigram model  $P_{\text{bigram}}$
- Maximum likelihood estimate of  $P(w_2|w_1)$
- Main problem with the maximum likelihood estimate
- Laplace estimate of  $P(w_2|w_1)$
- Pros and cons of the Laplace estimate
- Part-of-speech tagging
- Brown corpus tag set
- Two sources of information for POS tagging
- Examples for two sources of information for POS tagging



# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27**
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## POS tagging: Context parameters

$$P(t_{i+1}|t_i)$$

Example: P(NN | JJ)

## Context parameters: Estimation

$$\hat{P}_{\text{ml}}(t^k | t^j) = \frac{C(t^j t^k)}{C(t^j)}$$

$$\hat{P}_{\text{laplace}}(t^k | t^j) = \frac{C(t^j t^k) + 1}{C(t^j) + |T|}$$

## POS tagging: Word bias (emission) parameters

$$P(w_i | t_i)$$

Example:  $P(\text{take} | VB)$

NOT:  $P(t_i | w_i)$

## Word bias (emission) parameters: Estimation

$$\hat{P}_{\text{ml}}(w|t) = \frac{C(w:t)}{C(t)}$$

$$\hat{P}_{\text{laplace}}(w|t) = \frac{C(w:t) + 1}{C(t) + |V|}$$

## Decision tree

A decision tree is a tree. Each non-leaf node represents a test on an attribute and each branch descending from this node represents the outcome of the test (the value of the attribute). Each leaf node represents a decision or classification.

## Decision tree: Application to input $x$

Start at the root. Recursively do the following: test  $x$  on the attribute  $A$  of the current node and go down the branch that corresponds to the value that  $x$  has on  $A$ . When you arrive at a leaf, make the (classification) decision that the leaf is annotated with.

## Learning a decision tree

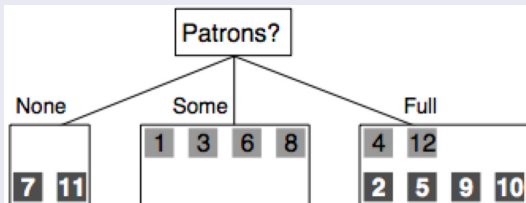
Recursively construct subtrees for a set  $S$  of training examples. Each subtree is constructed by (i) creating a node  $n$  for the subtree, (ii) selecting the most important / useful attribute  $A$  for  $S$ , (iii) attaching one branch per value of  $A$  and attaching it to  $n$ , (iv) recursively creating subtrees for each of the nodes that the branches lead to. (Stopping criteria omitted.)



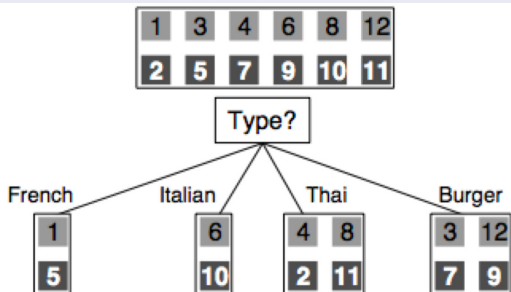
## Importance / usefulness of an attribute

Degree to which the attribute reduces uncertainty about the decision

## Example of important / useful attribute



## Example of useless attribute



# Summary

- POS tagging: Context parameters
- Context parameters: Estimation
- POS tagging: Word bias (emission) parameters
- Word bias (emission) parameters: Estimation
- Decision tree
- Decision tree: Application to input  $x$
- Learning a decision tree
- Importance / usefulness of an attribute
- Example of important / useful attribute
- Example of useless attribute

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30**
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Feature engineering in NLP

feature = attribute

feature engineering = designing/inventing attributes

A feature/attribute is any easily computable property of a word, sentence or text that is likely to help in making a decision on a task.

Examples for tasks: What type of named entity is this (person vs. location)? Is an email spam or not? Is the review of a smart phone positive or negative?

## Knowledge engineering

the craft and science of building and maintaining knowledge bases

The term is mainly used for manual approaches.

## Cyc(orp)

Knowledge engineering company that aims to build a comprehensive knowledge base of common sense knowledge.



## Deduction (logical inference)

deriving logical conclusions from premises

## Induction (inductive reasoning)

inferring a generalization from a set of observations

## Abduction (abductive reasoning)

finding the simplest and most likely explanation for a set of observations; inference to the best explanation

Example: His car got a scratch when he left the parking lot.

## Example for abduction

“Four Venezuelan firefighters who were traveling to a training course in Texas were killed when their sport utility vehicle drifted onto the shoulder of a highway and struck a parked truck”

Explanation inferred by abduction: “the firefighters were in the car” (which is not said explicitly)

## Uses of knowledge in NLP

Any type of natural language understanding requires knowledge, often employing techniques like abduction for processing knowledge.

# Summary

- Feature engineering in NLP
- Knowledge engineering
- Cyc(orp)
- Deduction (logical inference)
- Induction (inductive reasoning)
- Abduction (abductive reasoning)
- Example for abduction
- Uses of knowledge in NLP

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7**
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Lemmatisierung

Zuordnung der Wortformen zu Lemmata



## Tokenisierung

Aufteilung in Token

## Flexion

Anpassen der Wortformen an Umgebung im Satz

## Deklination

Flexion von Substantiven, Adjektiven, Pronomina, Numeralen und Artikeln

## Zur Deklination von Adjektiven

starke Deklination: nach Nullartikel und “dessen, wessen, manch, solch, welch, ...”

schwache Deklination: nach bestimmtem Artikel und “derjenige, derselbe, dieser, jener, jeder, ...”

gemischte Deklination: nach unbestimmtem Artikel und “kein, mein, ...”

# Konjugation

## Flexion von Verben

## Kongruenz (engl. agreement bzw. congruence)

signalisiert syntagmatische Zusammenhänge zwischen Wörtern  
Übereinstimmung von Flexionsmorphemen

## Rektion (engl. government)

signalisiert syntagmatische Zusammenhänge zwischen Wörtern  
prädikatives Wort weist einem von ihm regierten Wort ein  
Flexionsmorphem (Kasus) zu

## Morphologische Dependenz

Die Wortform  $w_2$  ist von der Wortform  $w_1$  in einem gegebenen Satz morphologisch abhängig, wenn mindestens ein Grammem (Flexionsmorphem) von  $w_2$  durch  $w_1$  bestimmt wird

kann symmetrisch sein

antireflexiv

eine Wortform kann von mehreren Wortformen abhängen



# Summary

- Lemmatisierung
- Tokenisierung
- Flexion
- Deklination
- starke, schwache, gemischte Deklination
- Konjugation
- Kongruenz
- Rektion
- Morphologische Dependenz

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14**
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25

## Wortarten

Lexikalische Kategorien

Part of speech

POS

## Linguistische Kriterien zur Klassifizierung der Wortart

morphologische Kriterien (nutzen Flektierbarkeit)

syntaktisch distributionelle Kriterien (Kontext des Wortes)

## Entscheidungsfragen bei Heringer zu Wortartenklassifizierung

ist das Wort flektierbar?

konjugierbar?

deklinierbar?

komparierbar?

artikelfähig?

kann es allein ein Satzglied sein?

satzbildend?

verlangt es einen bestimmten Kasus?

## Stemming

Wortstamm wird der Wortform zugeordnet

# Summary

- Wortarten
- Linguistische Kriterien zur Klassifizierung der Wortart
- Entscheidungsfragen bei Heringer zu Wortartenklassifizierung
- Stemming

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18**
- 12 12/21
- 13 1/18
- 14 1/25



## Wortbildung

transparente Wortbildung: kompositionell, segmentierbar (plus "semantischer Mehrwert"), z.B. *Arbeiterstadt*

lexikalisierte Wortbildung: Wort ist Teil des Wörterbuchs, z.B. *Hochzeit*

## Komposition

mindestens 2 Lexeme → Kompositum, z.B. *Handelsschule*  
Basis (meist) rechts: Kopf

## XN-Komposita

2 Elemente werden verkettet, evtl. mit (bedeutungslosem) Fugenelement

endozentrisch: B ist Oberbegriff für AB, z.B. *Mausefalle*

exozentrisch: weder A noch B ist Oberbegriff für AB, z.B. *Sündenbock, Katzenauge*

Possessivkompositum: *Großmaul*

Kopulativkompositum: *Fürstbischof*

## Derivation

Basis + Affix(e): *sprach-lich, be-sprechen, Ge-red-e*

Derivationsmorphem: Morphem, das zur Bildung neuer Lexeme dient

phonologische Varianten der Basis: Ablaut: *trink-en – Trank*

Affixe können grammatikalische Merkmale der Basis verändern

Suffix weist (z.B.) Genus zu (bei Nomen), z.B. *Spiel > Spieler*

## Kurzwortbildung

Anfangssegment erhalten: *Abitur* → *Abi*

Endsegment erhalten: *Autobus* → *Bus*

Segmente auswählen: *Auszubildender* → *Azubi*

Anfangsbuchstaben auswählen: *Lastkraftwagen* → *Lkw*

## Syntax

vgl.a. "Syntaktik": in der Semiotik von Morris: Verknüpfungen und Beziehungen von Zeichen

3-Teilung: Syntax – Semantik – Pragmatik

4-Teilung: Phonologie – Morphologie – Syntax – Semantik

## Gegenstandsbereich der Syntax

begrenzte Zahl von Regeln → potentiell unbegrenzt viele Sätze  
Klassifizierung der beteiligten syntaktischen Einheiten  
Hierarchische Organisation der Konstruktionen (Syntagmen, Sätze); vgl. z.B. Wortstellung bei *die Freunde ihrer Söhne – die Söhne ihrer Freunde*

## Satz

Grundeinheit der Syntax

Prädikativität (entscheidendes Merkmal des Satzes): Tempus,  
Modus (Relation zur Wirklichkeit)

vgl. Syntagma: auch nichtprädikative Konstruktionen

bildet relativ geschlossene Äußerung, Sinneinheit:

Mitteilungsabsicht des Sprechers; Hörer versteht etwas



## Satzbau

dt. Satzbau: Zweigliedrigkeit (Satzsubjekt, Prädikat) wesentlich, auch ohne sem. Subjekt wichtig, z.B. *Es donnert* (vgl. russ. *Gremit*)  
zwei konkurrierende Grundprinzipien: Konstituenz und Dependenz  
Syntax muss alle (syntaktisch) wohlgeformten (akzeptablen?)  
Sätze einer Sprache erzeugen/beschreiben

# Summary

- Wortbildung (Komposition, Derivation, Kurzwortbildung)
- Syntax
- Regeln
- Satz
- Prädikativität (Tempus, Modus)

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21**
- 13 1/18
- 14 1/25

## Konstituententests

Tests zum Feststellen syntaktischer Einheiten

Wörter sind nicht einfach aneinandergereiht, sondern gruppiert in hierarchischen Strukturen

## Permutationstest

Konstituente: Wörter, die zusammen umgestellt werden können, so dass der Satz grammatisch bleibt

z.B. *An der Ecke steht ein Haus. – Ein Haus steht an der Ecke.*

## Substitutionstest

Konstituente: Wortgruppe, die durch eine andere ausgetauscht werden kann

z.B. *Ein Haus steht an der Ecke. – Karl steht an der Ecke.*

Variante Pronominalisierungstest:

Pronomen (allgemeiner: Proform) dient als Ersatz

z.B. *Es/jemand steht an der Ecke.*

Variante Fragetest:

Fragepronomen als Ersatz (z.B. *wer, wen, was?*)

## Koordinationstest

Konstituente: Satzelement, das sich mit einem anderen koordinieren lässt

z.B. Konjunktion *und*: *Peter kauft [neue Kartoffeln] und [saftige Äpfel].*

## Eliminierungstest

Konstituente: Wortgruppe, die zusammen weggelassen werden kann

z.B. *Maria singt in der Badewanne.* – *Maria singt.*



## Europäische Grammatik-Tradition

zwei gleichberechtigte Zentren im Satz:

nominales Subjekt

verbales Prädikat

## Klassifikation von Konstituenten/Phrasen

jede Wortgruppe enthält als obligatorischen Kern ein Wort einer bestimmten Wortart, z.B.

Nominalphrasen (NP), z.B. *Kinder, die in Stuttgart leben*

Adjektivphrasen (AP), z.B. *... auf seine Tochter stolz*

Verbalphrasen (VP), z.B. *Eva besucht die Nachbarn.*

Präpositionalphrasen (PP), z.B. *Wir fahren mit dem Fahrrad.*

Adverbialphrasen (AdvP), z.B. *gestern Morgen*

## Phrasenstrukturgrammatik (PSG)

Beziehungen der unmittelbaren Dominanz als Regeln zur Erzeugung von Sätzen (Konstituentenstrukturen)

z.B. Regel (rewriting rule, Phrasenstrukturregel):  $S \rightarrow NP \quad VP$

## Allgemeines Schema einer PS-Regel:

$$X \rightarrow (W) \quad Y \quad (Z)$$

(W, X, Y, Z: Kategorialesymbole)

- X: Eingabesymbol (genau ein solches steht links des Pfeils: Nichtterminalsymbol!)
- W, Y, Z: Ausgabesymbole
- Y: obligatorisches Element
- W, Z: fakultative Elemente

PS-Regeln sind kontextfrei

## Rekursion

Phrasale Gruppen können Konstituenten derselben phrasalen Kategorie enthalten

(i) Ausgabe einer Regel dient als Eingabe der gleichen Regel:

z.B.  $NP \rightarrow Det\ N\ NP$

z.B. *Die Farbe der Augen der Tochter des Pianisten . . .*

(ii) Ausgabe einer Regel dient als Eingabe einer anderen Regel:

z.B.  $VP \rightarrow V\ S; S \rightarrow Comp\ NP\ VP$

z.B. *Die Schüler glauben, dass der Lehrer hofft, dass der Rektor verspricht, dass . . .*

## Lexikonregeln

Zur Generierung der terminalen Kette, z.B.:

Det  $\rightarrow$  der

N  $\rightarrow$  Löwe

etc.

Regeln entsprechen Lexikon

## Subkategorisierung

Informationen über den zugelassenen syntaktischen Rahmen des Verbs, z.B.:

schenken V, [ \_\_\_ NP, NP]

# Summary

- Konstituententests
- Klassifikation von Konstituenten/Phrasen
- Phrasenstrukturgrammatik (PSG)
- PS-Regel
- Rekursion
- Lexikonregeln
- Subkategorisierung



# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18**
- 14 1/25

## Valenz

Eigenschaft eines Prädikats, andere Wörter an sich zu binden

Aktanten: wesentliche Mitspieler eines Verbs, eines prädikativen Nomens etc.

semantische Analyse des Prädikats (Bedeutungsexplikation)!  
(Aktanten: notwendig und hinreichend für die Beschreibung)

## Rektionsmodell (RM)

Darstellung der semantischen bzw. syntaktischen Aktanten  
 vgl. Valenzrahmen, Frame, Subkategorisierungsrahmen  
 z.B. *versprechen 1a*

I = A (Subjekt)	II = B (Inhalt)	III = C (Gegenspieler)
$S_{Nom}$	1. zu + $V_{Inf}$ 2. dass + Nebensatz 3. $S_{Akk}$ notwendig	$S_{Dat}$

## Dependenzsyntax

Verb steht im Zentrum des Satzes

alle Wortformen des Satzes hängen direkt oder indirekt vom Verb ab

hierarchische Baumstruktur

## Eigenschaften der syntaktischen Dependenz

antisymmetrisch (zwei Wortformen können nicht gegenseitig voneinander abhängen)

antireflexiv (Wortform kann nicht von sich selbst abhängen)

Namen für syntaktische Relationen (versch. Arten von synt. Relationen)

genau ein syntaktischer Herr (Regens) für jede Wortform (außer für den obersten Knoten)

jede Wortform kann eine, keine oder mehrere abhängige Wortformen (Dependenten) haben

Baumstruktur mit Knoten (nodes) und Pfeilen (Kanten, arcs, branches)

## Kriterien für die Existenz einer syntaktischer Relation zwischen zwei Wortformen

potentielle prosodische Einheit von w1 und w2: Wortformen können eine Phrase bilden

potentielle lineare Einheit von w1 und w2: lineare Position von w1 kann nicht bestimmt werden ohne Bezug auf w2

## Kriterien für die Richtung der syntaktischen Relation

wichtigstes Kriterium:  $w_1$  ist der syntaktische Herr von  $w_2$ , wenn die passive Valenz der ganzen Phrase mehr durch die passive Valenz von  $w_1$  bestimmt wird als durch diejenige von  $w_2$  (“Welches der beiden Wörter passt genau so gut in einen gegebenen Kontext wie die Wortverbindung?”)

zweitwichtigstes Kriterium:  $w_1$  ist der syntaktische Herr von  $w_2$ , wenn  $w_1$  die Flexion von weiteren Lexemen kontrolliert oder wenn die Flexion von  $w_1$  durch solche Lexeme kontrolliert wird

drittwichtigstes Kriterium:  $w_1$  ist der synt. Herr von  $w_2$ , wenn  $w_1 + w_2$  eine Art von  $w_1$  ist

## Kriterien für die Benennung syntaktischer Relationen mit einem bestimmten Namen

semantischer Kontrast

syntaktische Ersetzbarkeit (Quasi-Kunze-Eigenschaft)

Wiederholbarkeit der Relation beim gleichen syntaktischen Herrn



# Summary

- Valenz
- Rektionsmodell (RM)
- Dependenzsyntax
- Eigenschaften der syntaktischen Dependenz
- Kriterien für die Existenz einer syntaktischer Relation
- Kriterien für die Richtung der syntaktischen Relation
- Kriterien für die Benennung syntaktischer Relationen

# Outline

- 1 October
- 2 11/2
- 3 11/9
- 4 11/13
- 5 11/16
- 6 11/23
- 7 11/27
- 8 11/30
- 9 12/7
- 10 12/14
- 11 12/18
- 12 12/21
- 13 1/18
- 14 1/25**

## Klassifikation von komplexen Sätzen (1)

Satzreihe (Satzverbindung, Parataxe)

Hauptsatz + Hauptsatz

asyndetisch (ohne Bindeglied) oder syndetisch (mit koordinierender Konjunktion) oder als Parenthese

## Klassifikation von komplexen Sätzen (2)

Satzgefüge (Hypotaxe)

Hauptsatz (Matrixsatz) + Nebensatz (-sätze)

Subordination angezeigt durch: Konjunktion, Relativpronomen, Interrogativpronomen oder ohne einleitendes Wort

Zwischenstatus zwischen Wortgruppe und Nebensatz: satzwertiger Infinitiv oder satzwertiges Partizip

## Subklassifizierung der Nebensätze nach ihrer syntaktischen Funktion

Nebensätze mit Satzgliedstatus: Satzsubjekt, Satzergänzung, Adverbial oder Prädikativum

Gliedteilsätze (Erweiterung zu einem Bezugswort): syntaktische Funktion: Attribut

## Interne Struktur der NP

Linkserweiterungen des Nomens (z.B. Artikel + Adjektiv),  
Rechtserweiterungen des Nomens (z.B. Relativsatz)  
Abfolgerestriktionen, z.B. PP nicht vor Genitiv-NP; sächsischer  
Genitiv und Artikel schließen sich aus

## AP

Syntaktische Funktionen des Adjektivs: Attribut des Substantivs, Attribut d. Adjektivs, Prädikativum, Adverbial (subjekt-, prädikat- oder objektbezogen)

Interne Struktur der AP: verschiedene Erweiterungen der AP (z.B. Genitiv-NP)

## Satzstrukturen

Stellung des finiten Verbs (mit Person-, Tempus- und Numerusmerkmalen):

V/1, z.B. Kommt Peter heute? (Entscheidungsfrage)

V/2, z.B. Peter kauft sich ein Eis. (Aussagesatz)

V/E, z.B. (Er kommt nicht,) weil Peter keine Zeit hat. (Nebensatz mit Konjunktion)



# Summary

- Klassifikation von komplexen Sätzen (Satzreihe, Satzgefüge, satzwertiger Infinitiv oder satzwertiges Partizip)
- Subklassifizierung der Nebensätze (mit Satzgliedstatus oder Gliedteilsätze (als Attribut))
- Interne Struktur der NP
- Syntaktische Funktionen des Adjektivs
- Interne Struktur der AP
- Stellung des finiten Verbs