Natural Language Processing

Timo Baumann

OSTBAYERISCHE
TECHNISCHE HOCHSCHULE
REGENSBURG

Principle of Compositionality

- Speech sounds (or letters) form words
- Words follow each other, forming sequences
- sequences of words have an inner structure (syntax)
- meaning (semantics) can be derived from the structure

today

- consolidation:
  - Parsing (=syntacic analysis) with CFGs
  - Chomsky-Normal-Form
- compositional semantics
  - meaning representation with first-order logic
  - lambda calculus for meaning composition

Syntax and Parsing: Sentence structure and its analysis

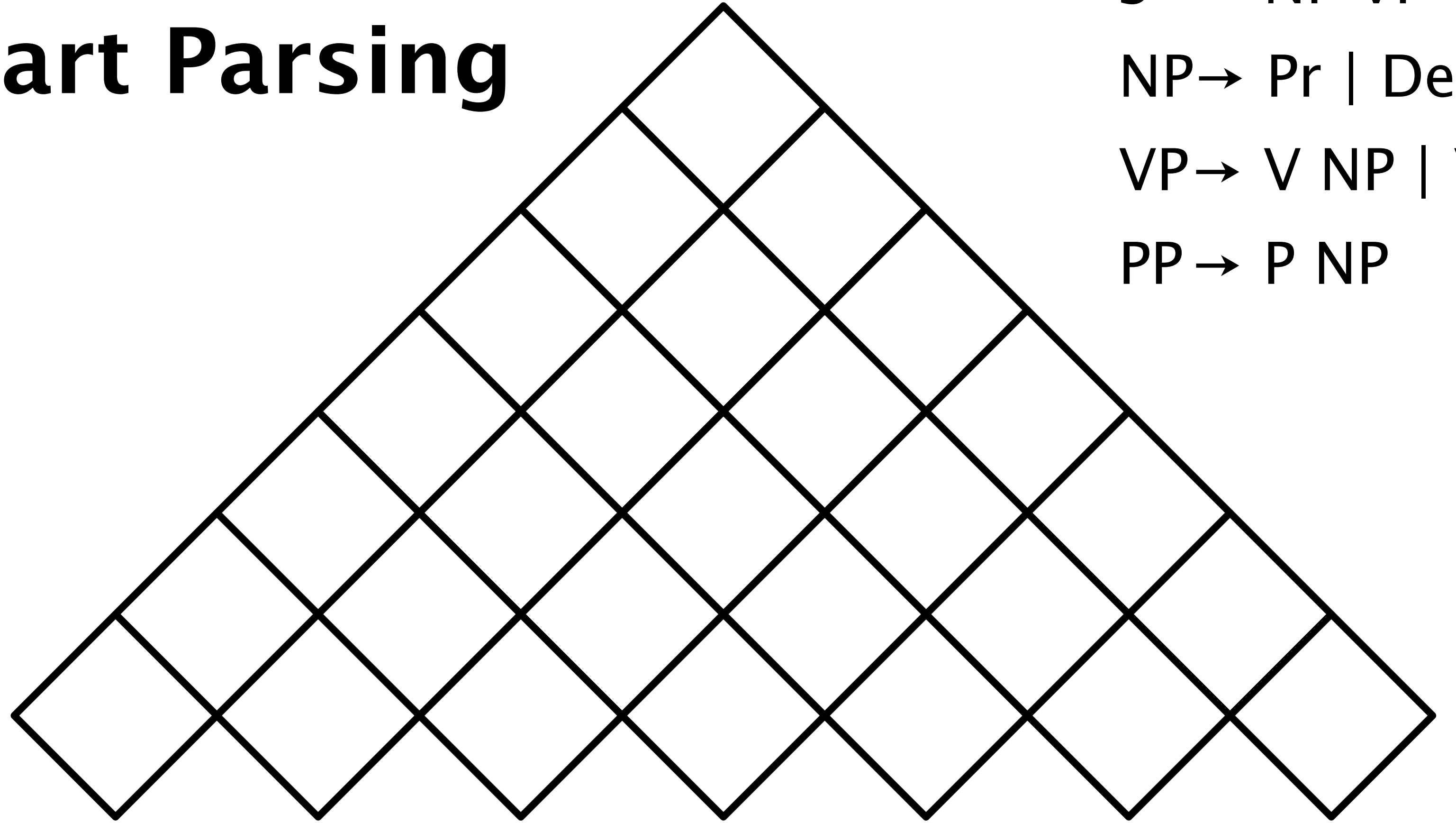
Chart Parsing

S → NP VP

NP→ Pr | Det N | NP PP

VP→ V NP | VP PP

PP→ P NP



Chomsky-Normal-Form

- Rules follow one of two forms:

  A → a (produces exactly one terminal)

  A → B C (produces exactly 2 non-terminals)

- Linguistics/NLP also knows other rules:
  - VP → V NP NP (“kauft dem Kind ein Eis”)

    “buys the child an ice-cream”

conjunctions and ellipses

“kauft dem Kind ein Eis und sich einen Eiskaffee”

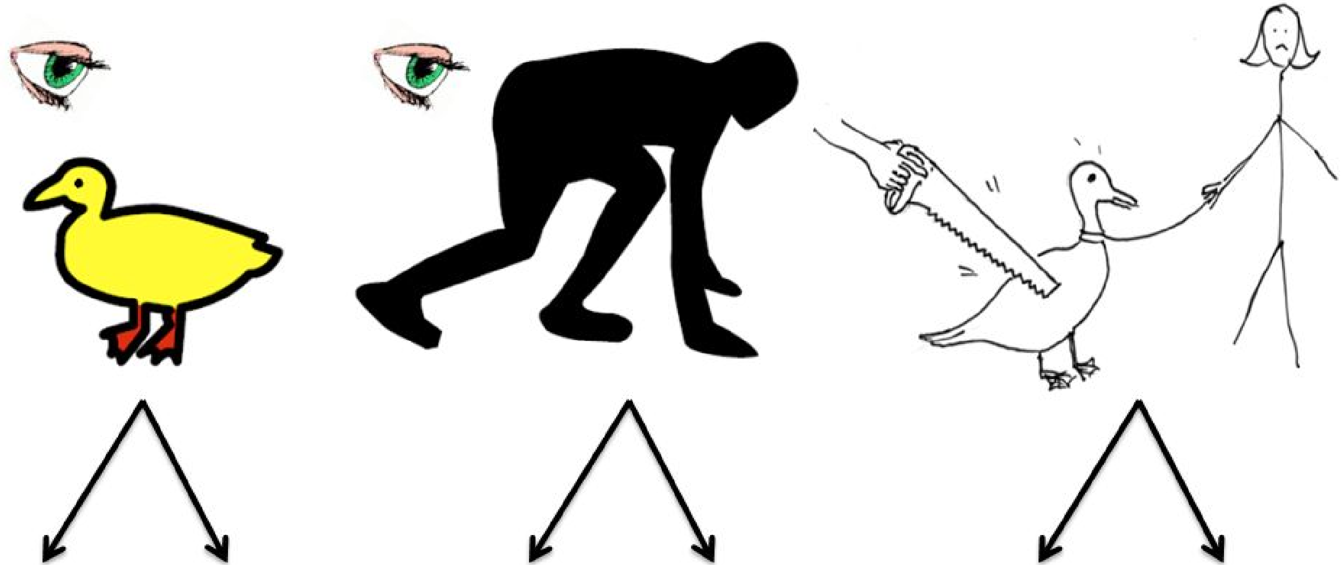
VP	→	V NP NP	(ditransitive verbs)
VP	→	VP und VP	(coordination)
V	→	ϵ	(omission of material)

normalize a grammar to be in CNF

- remove terminals from “mixed” (T/NT) rules by replacing each T with a new NT
  - VP → VP Conj VP   Conj → und
- rules with >2 NTs (on the right-hand-side) can be split into multiple rules:
  - VP → VP X   X → Conj VP
- ϵ productions: remove the NT that produces ϵ from all RHSs (it if always produces ϵ) or create alternate RHSs without the symbol
  - VP → V NP NP   VP → NP NP
- chain rules: for unary rules A → B , B → w add direct rule A → w
  - this can be ignored because it's easy to extend CKY to handle such unary rules (although, strictly speaking, they are non-CNF rules)

Ambiguity in language

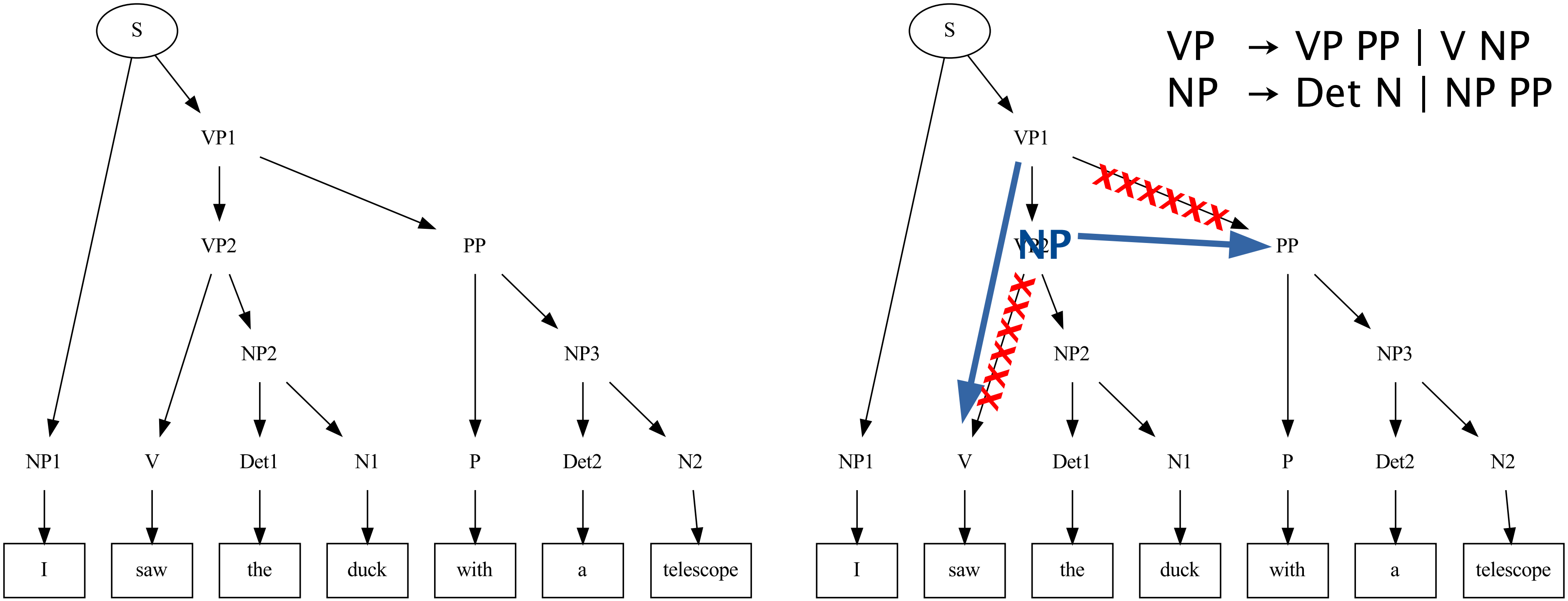
"I saw her duck with a telescope…"



Slide credit: Dhruv Batra, figure credit: Liang Huang

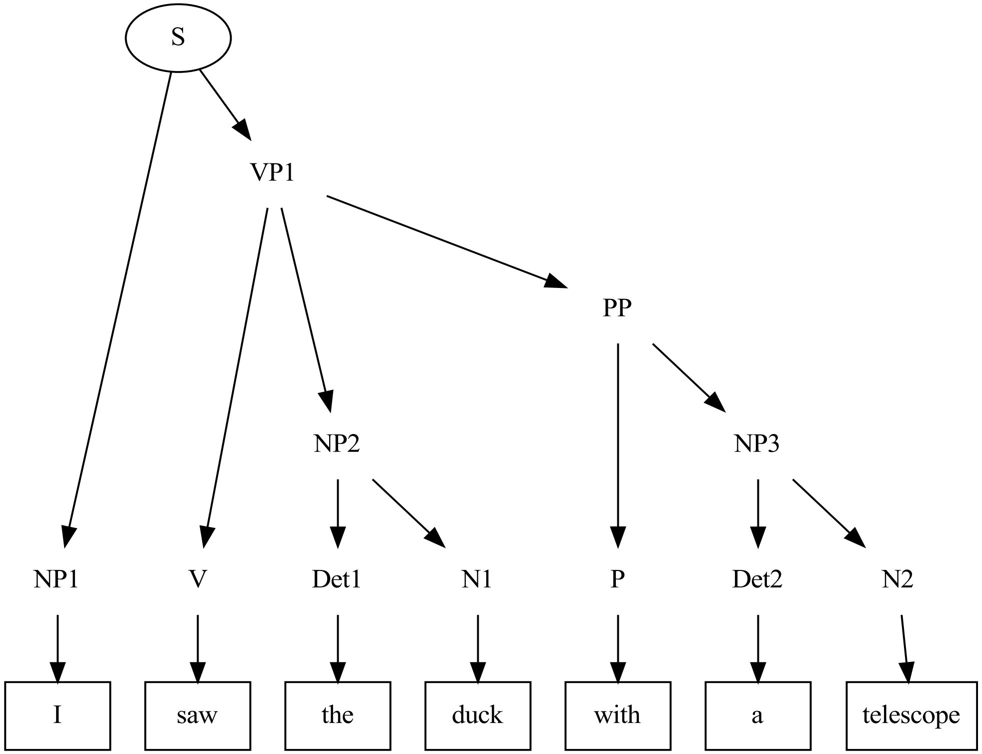
ambiguity in derivation trees

VP → VP PP | V NP
NP → Det N | NP PP

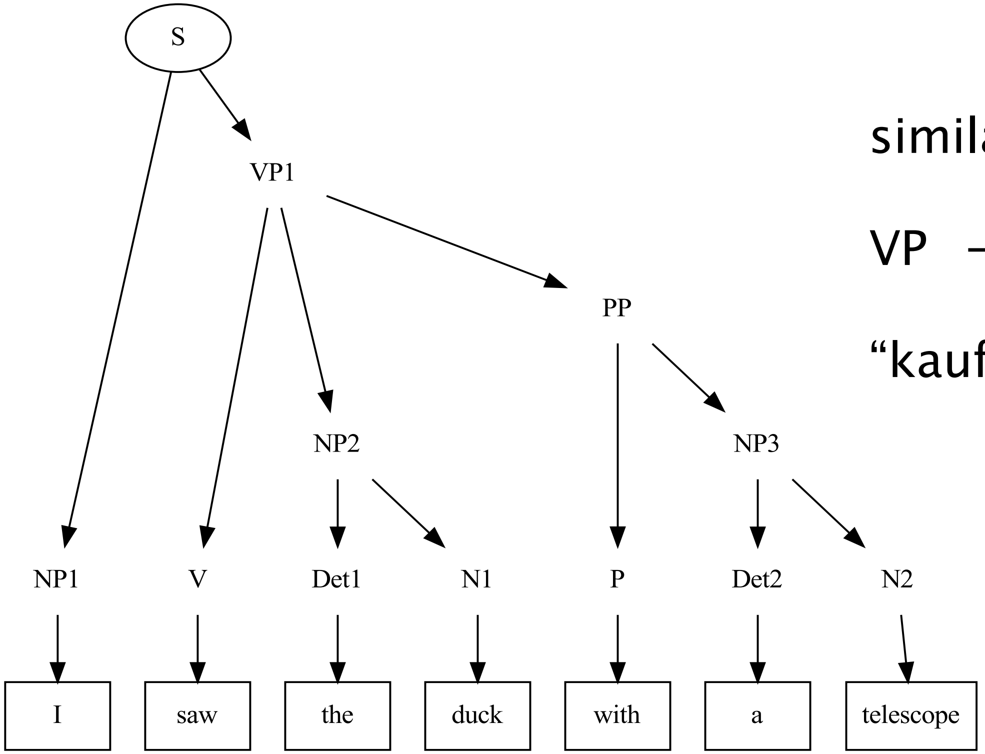


→ encode ambiguity by flatening trees

VP → V NP PP



→ encode ambiguity by flatening trees



VP → V NP PP

similarly:

VP → V NP NP

“kauft dem Kind ein Eis”

Parsing strategies

- (purely) top-down:
  - expand S until we find the full sentence (hopefully)
- bottom-up (Cocke-Kasami-Younger 1967)
  - check what can be combined (e.g. NP → ... → “the telescope”), hoping that we'll eventually get an S at the top
- mixed: left-corner parsing (Earley 1970)
  - expand S (top-down), such that the first (then second, third, ...) words fit the derived structure

S → NP VP

NP→ Pr | Det N | NP PP

VP→ V NP | V | VP PP

PP→ P NP

I saw her duck with the telescope

limits of CFGs for syntax in NLP

- language contains structures which go beyond CFGs
  (if you're looking for examples of highly complex language,
  Swiss German is often the language of choice)
  - see separate slide set
- agreement is difficult to model in CFGs
  - e.g.: subject-verb have to agree in person and number
    → solution: unification based grammars
- subcategorization, e.g. arity (number of open semantic slots) of verbs
  - *John found. *John disappeared the ring.

summary

- parse trees capture syntactic structure and help in deducing semantics / meaning
- CFGs are useful to describe the grammar of natural language (by and large)
- efficient computation of parse trees via dynamic programming

applications of CFGs

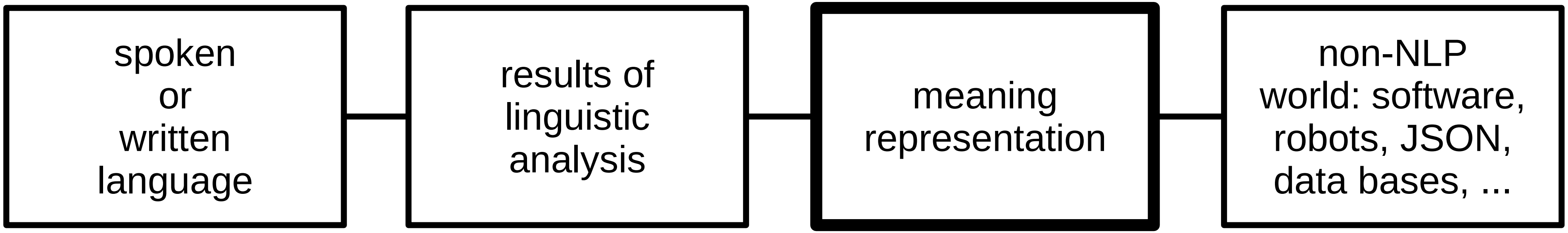
- most NLP applications profit from syntactic structure (it's, however, not always clear whether automatically derived structure yields a benefit)
- speech recognition grammars that can readily be semantically interpreted (e.g. in command&control applications)
- beyond NLP (e.g. parsing programming languages)

Compositional Semantics
The meaning of sentences.

The meaning of sentences.

- expressed in a meaning representation language
- should allow to express everything(?)
  that can be written/spoken in a sentence
- represents knowledge about a part of the world
  - typically: objects, their properties, interrelations

meaning representation: linking natural language and the world



requirements for meaning representations

- represent knowledge about the world
- obtain knowledge from language
  - e.g. for a robot that interacts via language
- Question Answering
  - e.g. queries towards a knowledge base
- ambiguous and vague expressions
  - variability in linguistic expressivity
- fact checking
  - are statements in a text true according to our knowledge?
- argument checking
  - is a text coherent or does it contain contradictions? (regardless of our knowledge of the world)

model-based semantics

- model: simplified representation of the world (or a part thereof = domain): objects, properties, relations
- today: (simplified) predicate / first-order logic
- deduction that is “sound” (i.e., what can be derived is true, what is true can be derived)

first-order logic

- terms represent elements of the domain
  - constants (Peter, Klaus, ...) and variables (x, y, ...)
- predicates express properties (unary predicates) and relations (binary, ...)
  - human(x), likes(Peter, Klaus), likes(Peter, icecream), ...
  - predicates work on terms (but not on other predicates; hence first-order not higher-order logic)
- logical connectives: $\wedge$ $\vee$ $\neg$ $\Rightarrow$ (and, or, not, implies)
- quantifiers

quantifiers in first-order logic

- two ways of introducing a variable:
  - refer to any (but an existing) object in the domain
    → existential quantifier ∃
  - refer to all objects in the domain
    → universal quantifier ∀

example domain: restaurants

- “Gibt es ein vegetarisches Restaurant in der Altstadt?”
  $\exists x\ \mathrm{Restaurant}(x) \land \mathrm{Küche}(x, \mathrm{Vegetarisch}) \land \mathrm{In}(x, \mathrm{Altstadt})$
- “Restaurants nahe der Donau sind teuer oder schlecht.”
  $\forall x\ \mathrm{Restaurant}(x) \land \mathrm{Nah}(x, \mathrm{Donau}) \Rightarrow (\mathrm{teuer}(x) \lor \mathrm{schlecht}(x))$

example domain: restaurants II

- “die Mensa ist nicht laut.”

  $\neg$laut(Mensa)

- “Manche Leute mögen Sushi.”

  $\exists$x menschlich(x) $\wedge$ mag(x, Sushi)

logic-based (Montague)-semantics

(Richard Montague, say: Mon tuh gyoo)

- the *denotation* (=meaning) of a sentence is the set of constraints that are needed such that the sentence becomes true
- “Manche Leute mögen Sushi.”
  is true, iff
  $\exists x$ menschlich(x) $\wedge$ mag(x, Sushi)
  is true.

logic-based (Montague)-semantics

(Richard Montague, say: Mon tuh gyoo)

- the *denotation* (=meaning) of a sentence is the set of constraints that are needed such that the sentence becomes true

- "Manche Leute mögen Sushi."
  Augensuppe??
  is true, iff
  $\exists x$ menschlich(x) $\wedge$ mag(x, Sushi)
  Augensuppe??
  is true.

compositionality of meaning

- constituents determine the meaning of a sentence
  - constituents are determined by syntax (context-free – or at least close!)
- we'll need to be able to express the (partial) meaning of constituents and be able to compose meaning from “sub-meanings”
- words are the smallest semantic units
  → humans store their meaning in a mental lexicon
- grammar rules come with additional rules for composing the meaning of their constituting constituents

partial meaning: Lambda calculus

- extends first-order logic with “partial formulae”
- Lambda creates an anonymous (unary) function
  - λx. expression
  - examples:
    - λx. mag(x, Sushi)
    - λy.λx. mag(x, y)
- Lambda reduction: apply the anonymous function
    - λy.λx. mag(x, y)(Sushi) = λx. mag(x, Sushi)
    - λy.λx. mag(x, y)(Sushi)(Timo) = mag(Timo, Sushi)

semantic analysis

- general approach:
  - generate syntactic parse tree
  - look up meaning for each word (expression from lambda calculus)
    - many words will have multiple meanings
  - each grammar rule also has its meaning (in lambda calculus)
  - reconstruct the full formula (bottom-up)
  - potentially: resolve formula against world knowledge

Example

(Alan Black)

semantics: summary

- meaning representation is binding link between language and knowledge processing
  - language → meaning representation: analysis

    meaning representation → language: generation
- represent formulae in first-order language
  - limitations such as a closed world assumption

Vielen Dank, Ihre Fragen?

`timo.baumann@oth-regensburg.de`

weiterführende Literatur

- Jurafsky and Martin (2009): Kapitel 15
- Eisenstein (2018): Natural Language Processing: Kapitel 12
- Carstensen et al. (2004): Abschnitt 3.6

Lehrziele

- die Studierenden erkennen die immense strukturelle Ambiguität natürlicher Sprache
- die Studierenden verstehen die Grundideen des Lambda-Kalküls und wie diese zur Modellierung der kompositionalen Semantik eingesetzt werden können
- den Studierenden sind die Beschränkungen der vorgestellten Modelle, mit Blick auf Kompositionalität und die Behandlung von Mehrdeutigkeit bewusst