Computational Linguistics

1. Natural Language and the Computer
2. Words and Wordforms
3. Phrases and Sentences
4. Discourse: Texts and Dialogs

Phrases and Sentences

- phrases and sentences are more than admissible sequences of words
- they have an internal structure (syntax) and a meaning (semantics)
- the meaning of a phrase/sentence can be
  - compositional: combining the meaning contributions of their components (words)
  - holistic: cannot be obtained from simpler components
    $\rightarrow$ holophrases

Phrases and Sentences

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Phrases and Sentences

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Language Models

- grammar-based
  - descibing well-formed utterances
  - prediction of the possible wordforms next in an utterance
- probabilistic/connectionist
  - estimating the probability of a (partial) utterance
  - prediction of the probability distribution for the next wordform

Grammar-based Language Models

- often used in spoken-language dialog systems (e.g. VoiceXML)
- simplest case: word-pair grammar: bigrams without probability
- more often: context-free rules without recursion

Probabilistic Language Models

- based on n-gram probability distributions, e.g. trigrams
- probability of a (partial) wordform sequence

$$P(w_{1 \ldots n}) = \prod_{i=1}^{n} P(w_i | w_{i-2} w_{i-1})$$

- probability distribution for the next wordform

$$P(w_n | w_{1 \ldots n-1}) = \frac{P(w_{1 \ldots n})}{P(w_{1 \ldots n-1})}$$

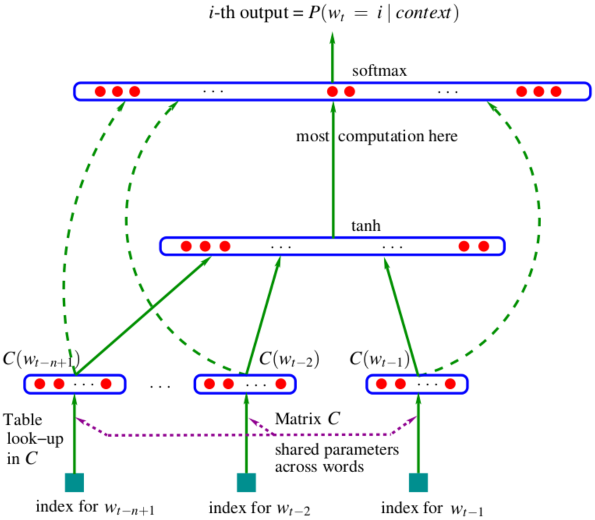
Probabilistic Language Models

- training by maximum likelihood estimation on unannotated corpus data

$$P(w_i|w_{i-2}w_{i-1}) = \frac{c(w_{i-2}w_{i-1}w_i)}{c(w_{i-2}w_{i-1})}$$

- dealing with data sparseness: backoff, smoothing, interpolation
- measuring the predictive power: perplexity
  - approximated by the testset perplexity

Continous-Space Language Models



Bengio et al. 2003

Continous-Space Language Models

- number of free parameters grows linearly with the size of the vocabulary and the window
- interpolation with a trigram model
- results on the AP news corpus (14M/1M/1M tokens)

	n	direct	mixture	validation	test
MLP10	6	no	yes	104	**109**
Del. Int.	3			126	132
Back-off KN	3			121	127
Back-off KN	4			113	119
Back-off KN	5			112	**117**

Phrases and Sentences

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Phrases and Sentences

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Chunking

- TODO

Structural Descriptions

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Structural Descriptions

1. Language models
2. Chunking
3. Structural descriptions
4. Parsing with phrase structure grammars
5. Probabilistic parsers
6. Parsing with dependency models
7. Principles and Parameters
8. Unification-based grammars
9. Semantics construction

Structural Descriptions

- broad consensus:

  trees are necessary and sufficient to capture relevant syntactic relationships

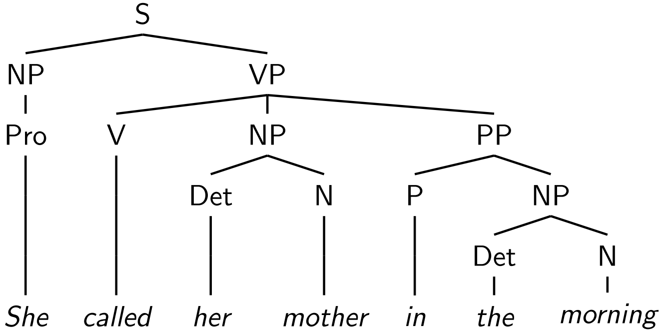
- two types of syntactic trees:

  - phrase structure trees
  - dependency trees

Structural Descriptions

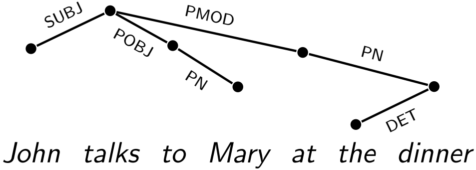
- phrase structure trees:

  typed constituents of a sentence are broken down/combined into sucessively smaller/larger constituents



Structural Descriptions

- dependency trees:
  wordforms (subtrees) are subordinated with a typed relationship under other wordforms



Phrase Structure

- Phrase structure
- Dependency structure
- Trees as structural descriptions?
- Levels of adequacy

Phrase Structure

- basic units: constituents
- constituent structure can be described by means of a context free grammar
  - non-terminal symbols: S, NP, VP, PP, ...
  - terminal symbols: *waits, for, in, the, John, Mary, park*

  context free rules:

  NT-Symbol $\rightarrow$ $\{$T-Symbol $|$ NT-Symbol$\}$*

Phrase Structure

- rules can be applied
  - generatively: produce sentences that are licensed by the grammar
  - analytically: check whether a sentence is licensed by the grammar
- recursion:
  - constituents can be embedded into other constituents
  - constituents can be embedded into a constituent of the same type
  - recursion can be indirect

Phrase Structure

- the phrase structure tree is a byproduct of the derivation process (recursive rule application)

  $\rightarrow$ close relationship between
  - rule structure
  - structural description
  - rule application (analysis/generation)

- rules can be extracted from a given phrase structure tree

Phrase Structure

- lexical insertion rules, preterminal rules, lexicon

N $\rightarrow$ *mother*
N $\rightarrow$ *morning*
Pro $\rightarrow$ *she*
P $\rightarrow$ *in*
Det $\rightarrow$ *the*
Det $\rightarrow$ *her*
V $\rightarrow$ *called*

Phrase Structure

- structure-building rules, grammar

  S → NP VP
  VP → V NP VP
  VP → V NP
  VP → V PP
  PP → P NP
  NP → Det N

- first constraint on the possible form of rules
  - lexicon

    PT-Symbol → T-Symbol
  - grammar

    NT-Symbol → {NT-Symbol | PT-Symbol}*

Phrase Structure

- recursive rules:
  potentially infinitely many sentences can be generated
  $\rightarrow$ creativity of language competence
- goal of linguistic modelling:
  specification of additional constraints on the possible rule forms

Phrases and Phrasal Categories

- phrasal categories: distributional type (purely structural perspective)
- phrasal categories are derived from lexical ones by adding additional constituents

  N $\Rightarrow$ NP
  V $\Rightarrow$ VP
  A $\Rightarrow$ AP
  ADV $\Rightarrow$ ADVP
  P $\Rightarrow$ PP

- lexical core: head of the phrase
  - determines crucial syntactic properties of the phrase

Phrases and Phrasal Categories

Morphological evidence

- phrasal inflection in English (only noun phrases)

  possessive genitive

  This crown is $[_{NP}$ the king]'s.
  ** This crown is$[_{N}$P the $[_{N}$ king]'s].*
  This crown is $[_{NP}$ the $[_{N}$ king] of England]'s.
  ** This crown is $[_{NP}$ the $[_{N}$ king]'s of England].*
  ** This crown is $[_{AP}$ very handsome]'s.*

Phrases and Phrasal Categories

Semantic evidence

- *explanation* of structural ambiguities
- e.g. scope ambiguity

The President could not ratify the treaty.

The President [$_M$ *could not*] *ratify the treaty.*
The President could [$_{VP}$ *not ratify the treaty*].

The President [$_M$ *simply could not*] *ratify the treaty.*
The President could [$_{VP}$ *simply not ratify the treaty*].

- explanation depends on phrasal categories, e.g. VP

Phrases and Phrasal Categories

Phonological evidence

- phonological contraction disambiguates

The President couldn't ratify the treaty.

The President [$_M$ couldn't] ratify the treaty.
** The President could[$_{VP}$ n't ratify the treaty].*

Phrases and Phrasal Categories

Syntactic evidence: syntax tests and distributional criteria

- cleft transformation

  It was [the girl] that called her father in the morning.
  It was [her father] that the girl called in the morning.
  It was [in the morning] that the girl called her father.
  **It was [her father in the morning] that the girl called.*

- constituent questions and stand-alone test

  Who called her father in the morning? The girl.
  Whom the girl called in the morning? Her father.
  When the girl called her father? In the morning.
  **Whom the girl called in the morning? Her father in the morning.*
  What did the girl do? Call her father in the morning.

Phrases and Phrasal Categories

Syntactic evidence: syntax tests and distributional criteria

- coordination

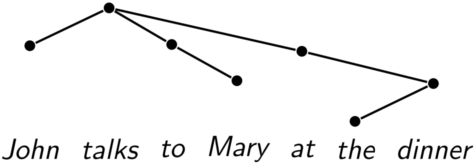
  The girl called [$_{XP}$ her father] and [$_{XP}$ her mother].
  **The girl called [$_{XP}$ her father] and [$_{YP}$ in the morning]. (XP≠YP)*
  The girl [$_{XP}$ called her father] and [$_{XP}$ met her mother].
  **The girl called [$_{XP}$ her father] and [$_{YP}$ met her mother]. (XP≠YP)*

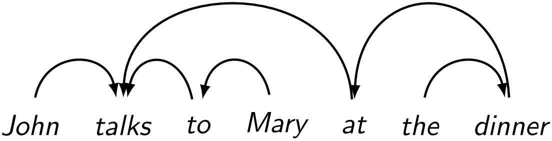
- substitution by a pronoun

  [She] called her father in the morning.
  The girl called [him] in the morning.
  The girl called her father [then].
  The girl did [so].

Dependency Structure

- subordination of wordforms (modifier) under other wordforms (modifiee)

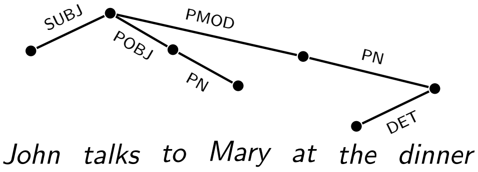




- the modifiee roughly corresponds to the head
- alternative view: subordination of partial trees under wordforms

Dependency Structure

- edges can be annotated with syntactic functions (subordination/dependency relations)



Dependency Structure

- (weak) distributional tests
  - deletion: if a wordform can only appear together with another one, it has to be attached to/depends on the other one
  - substitution: two subtrees that cannot be substituted for each other have to attached with a different label
  - coordination: subtrees that can be coordinated should be attached with the same label

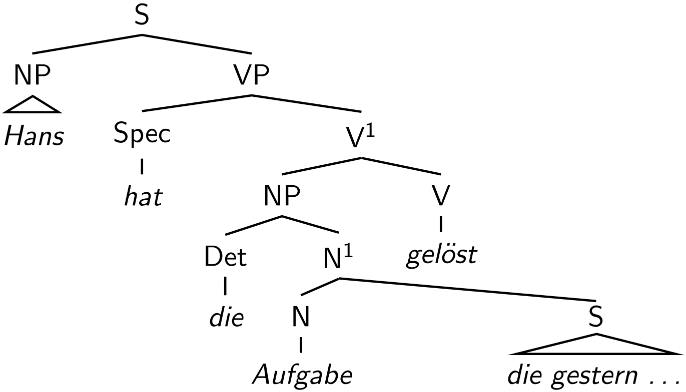
Dependency Structure

- examples of dependency relations

SUBJ	subject of a verb
OBJA	accusative object of a verb
OBJD	dative object of a verb
OBJC	a finite verb in a subordinate clause modifying the verb in a main clause
OBJP	a preposition (of a prepositional phrase) modifying a verb
PP	prepositional modifier of a verb or a noun
REL	a relative pronoun modifying a noun
DET	a determiner modifying a noun
AUX	a full verb modifying an auxiliary
ADV	an adverbial modifying a verb
	...

Trees as Structural Descriptions?

- constituents can be split
  $\rightarrow$ non-projective structures/discontinuous constituents

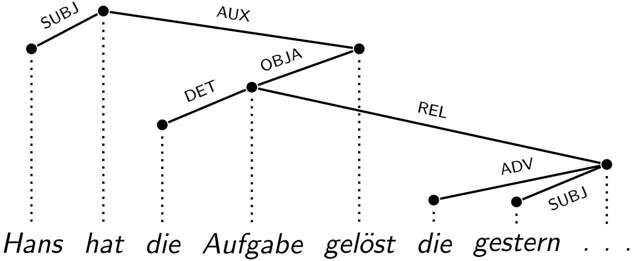


Trees as Structural Descriptions?

- non-projective structures cannot be generated by a context-free grammar
  - approximation by means of projective trees or
  - using additional formal mechanisms, e.g movement or transformation

Trees as Structural Descriptions?

- dependency structures suffer from the same (representational) problem



- but non-projective trees can be produced by more local attachment operations
- generating non-projective trees usually results in exponential parsing effort

Trees as Structural Descriptions?

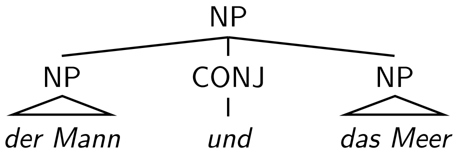
- amount of non-projectivity varies from language to language

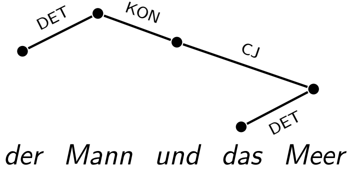
language	amount of nonprojective dependencies	amount of nonprojective sentences
Dutch	5.4	36.4
German	2.3	27.8
Czech	1.9	23.2
Slovene	1.9	22.2
Portuguese	1.3	18.9
Danish	1.0	15.6

measured on the CoNLL-X Shared Task data (Kübler 2010)

Trees as Structural Representations?

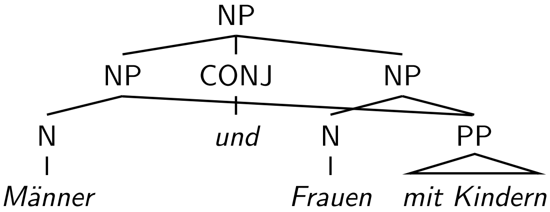
- problem with dependency trees: representing coordination

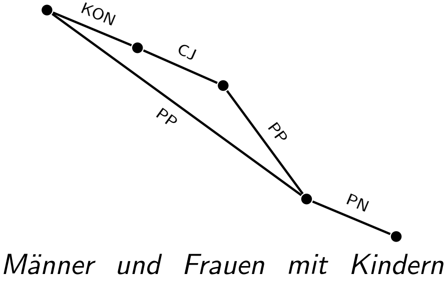




Trees as Structural Descriptions?

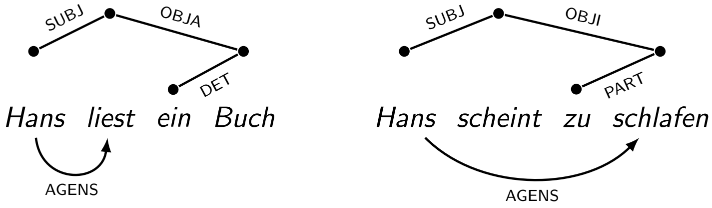
- elliptical constructions: shared constituent coordination





Trees as Structural Descriptions?

- syntax and semantics exhibit different structural relationships
- e.g. raising verbs



syntactic subject $\neq$ logical subject

Levels of Adequacy

- observational adequacy

  specification of all well-formed sentences of a language

  - formally explicit
  - sound and complete
  - no consideration of semantic aspects
  - Chomsky (1957)

- descriptive adequacy

  additionally: specification of structural descriptions, that correspond in a pricipled manner with the intuitions of a speaker of the language

  - connection of linguistic structures with meanings
  - Chomsky (1965)

Levels of Adequacy

- explanatory adequacy

  additionally: specification of at few as possible, universal principles that mirror psychologically plausible assumptions about language processing in humans

  - allows to derive predictions
  - explains language acquisition phenomena
  - CHOMSKY (1981)