

Syntax

General Linguistics

Jennifer Spenader, February 2006

(Most slides: Petra Hendriks)

Levels of language

- Text/Dialogue ⇒ Pragmatics (lecture 11)
- Sentences ⇒ Syntax (lectures 5 en 6)
 Sentence semantics (lecture 10)
- Words ⇒ Morphology (lecture 4)
 Lexical semantics (lecture 9)
- Syllables ⇒ Phonology (lecture 3)
- Sounds ⇒ Phonetics (lecture 2)

Structure of the lecture

1. Parts-of-speech: Syntactic building blocks
2. Constituent level
 1. Phrases, recognizing phrases
3. What structure do constituents have?
4. What structure do sentences have?

Goal: Understanding Language

- To understand language we have to interpret it semantically
- Semantics are often reached through syntax
 - many believe there is a close (if not 1-to-1) relationship between syntax and semantics
- Syntactic structures are built up according to the parts of speech of the words in the sentence
- What parts of speech are there and how do we identify them?

Grammars

Grammars in syntax are different from traditional grammars

- Traditional grammars are prescriptive (*voorschrijvend*), syntax is descriptive and explanatory
- School grammars look at exceptions, syntax looks at rules

Grammars

Grammars in syntax are different from traditional grammars

- Traditional grammars are prescriptive (*voorschrijvend*), syntax is descriptive and explanatory
- School grammars look at exceptions, syntax looks at rules

What is a part of speech?

- Different criteria can be used to divide words into parts-of-speech
 - meaning (semantic criteria)
 - e.g. a noun is a person, place or thing
 - form (morphological criteria)
 - A noun can be preceded by “de” or “het”
 - function (functional criteria)
 - A noun can function as the subject of a sentence

Schoolhouse Rock

- Conjunction Junction, what's your function?
- Hooking up word and phrases and clauses



What word classes are generally recognized?

- N (noun): veer, tandenborstel
- V (verb): zijn, eten, denken
- A (adjective): gekke, dronken
- Adv (adverb): gek, dronken
- P (preposition): in, op, van
- Det (determiner): de, een, sommige, drie
- Deg (degree element): erg, heel
- Pro (pronoun): ik, hij, hem
- COMP (complementizer): dat, omdat, die
- Conj (conjunction): en, of
- PRT (partikel): te, op (in *opbellen*)

Parts-of-speech in the eye of the beholder

- Even within the same language there are different view about which parts-of-speech should be recognized
 - Recognizing very fine distinctions can be useful for some tasks, but a hindrance in others
- Almost all languages a class that can be considered “noun” and “verb”
- However, David Gil claims Riau Indonesion only has one word class

Linguistic Field Work

David Gil

Riau Province, Indonesia



Tagging

- For applications, text analysis often requires identifying the parts-of-speech of every word
- TAGGING (or POS-tagging)
 - see CL Demo (www.ling.su.se)
- Different corpora have used different POS-tags

– Corpus	#of tags
– Brown	179
– Penn treebank	45
– CLAWS1	132
– CLAWS2	166
– CLAWS5	65
– London-Lund	197

Difficulties of tagging

- The same surface form often maps to several words in different parts-of-speech
- This is why tagging isn't a trivial task (even though we are pretty good at it now)
- How often are forms ambiguous?
 - In the Brown corpus 11.5 % of all **word types** are ambiguous
 - = 40 % of all **tokens**

# of tags	1	2	3	4	5	6	7
# of word types	35 340	3760	264	61	12	2	1

- If David Gil is right and Riau Indonesian only has one part-of-speech, how difficult will tagging be?

Constituents

- Jan is op het hoedje met de veer gaan zitten.
 - op het hoedje met de veer
 - het hoedje met de veer
 - het hoedje
 - met de veer
 - de veer
- Constituent: woordgroep

Constituent structure

- Constituent $\neq$ sentential part (*zinsdeel*) in traditional grammar
- Constituents can contain other constituents
- Constituent structure is a hierarchical structure

Constituency tests

The following tests can be used to determine if a sequence of words is a constituent or not:

- Substitution test
- Movement test
- Coordination test (Nevenschikkingstest)

Constituent tests

- Are these constituents?
 - Use substitution, movement and coordination tests!
- 1. [The tragedy] upset the entire family.
- 2. I love [peanut butter and jelly]
- 3. Jane [left town]
- 4. They hid [in the cave]
- 5. The frogs [stared at] the duck.
- 6. [Don't eat] your sandwich during class!
- 7. Answer the question [without looking in the book]

Syntactic categories

- NP (noun phrase): Jan, het hoedje
- VP (verb phrase): is gek, heeft het hoedje opgegeten
- AP (adjective phrase): gekke, dronken, erg dronken
- AdvP (adverb phrase): gek, dronken, erg dronken
- PP (prepositional phrase): met de veer, van kaas
- S (sentence)

Form vs. function

- Subjects and objects almost always NPs
- PPs are often attributive descriptions or prepositional phrases
- Prepositional phrases are always PPs.
- AP's modify nouns

How are constituents built?

- Head:
 - The element around which the constituent is built
- Specifier:
 - The element that specifies the meaning of the head
- Complement:
 - Information about elements that are implied by the head



NP



N



kaas

VP



V



slaapt

AP



A



gek

PP

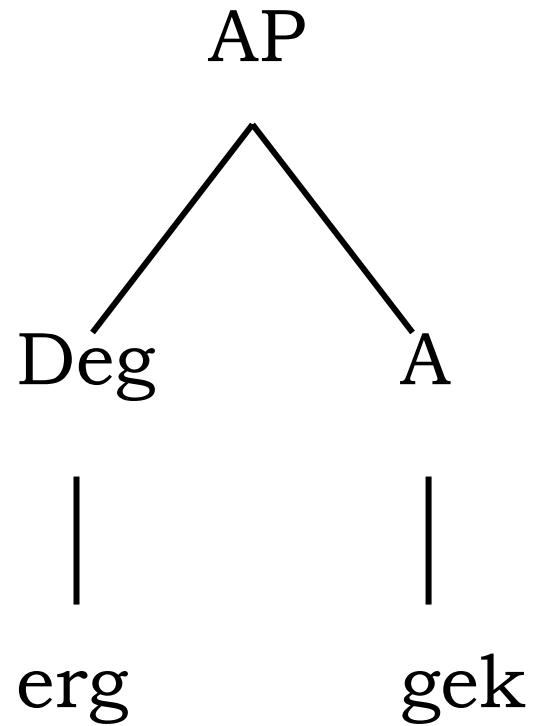
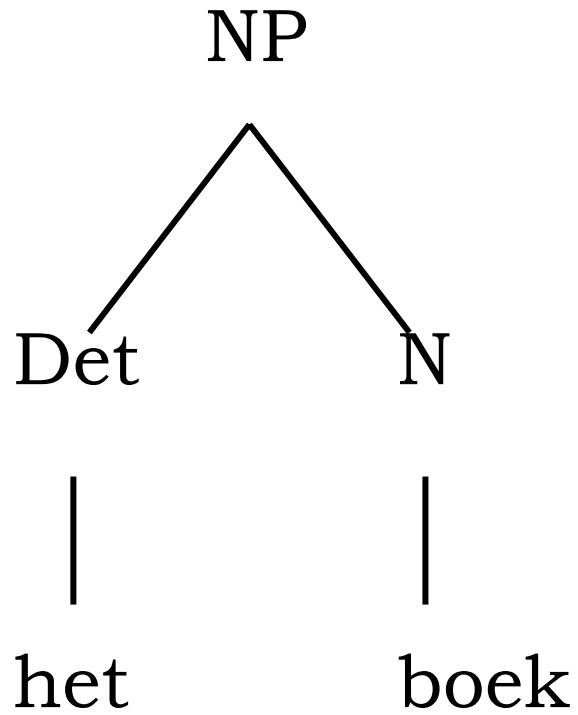


P

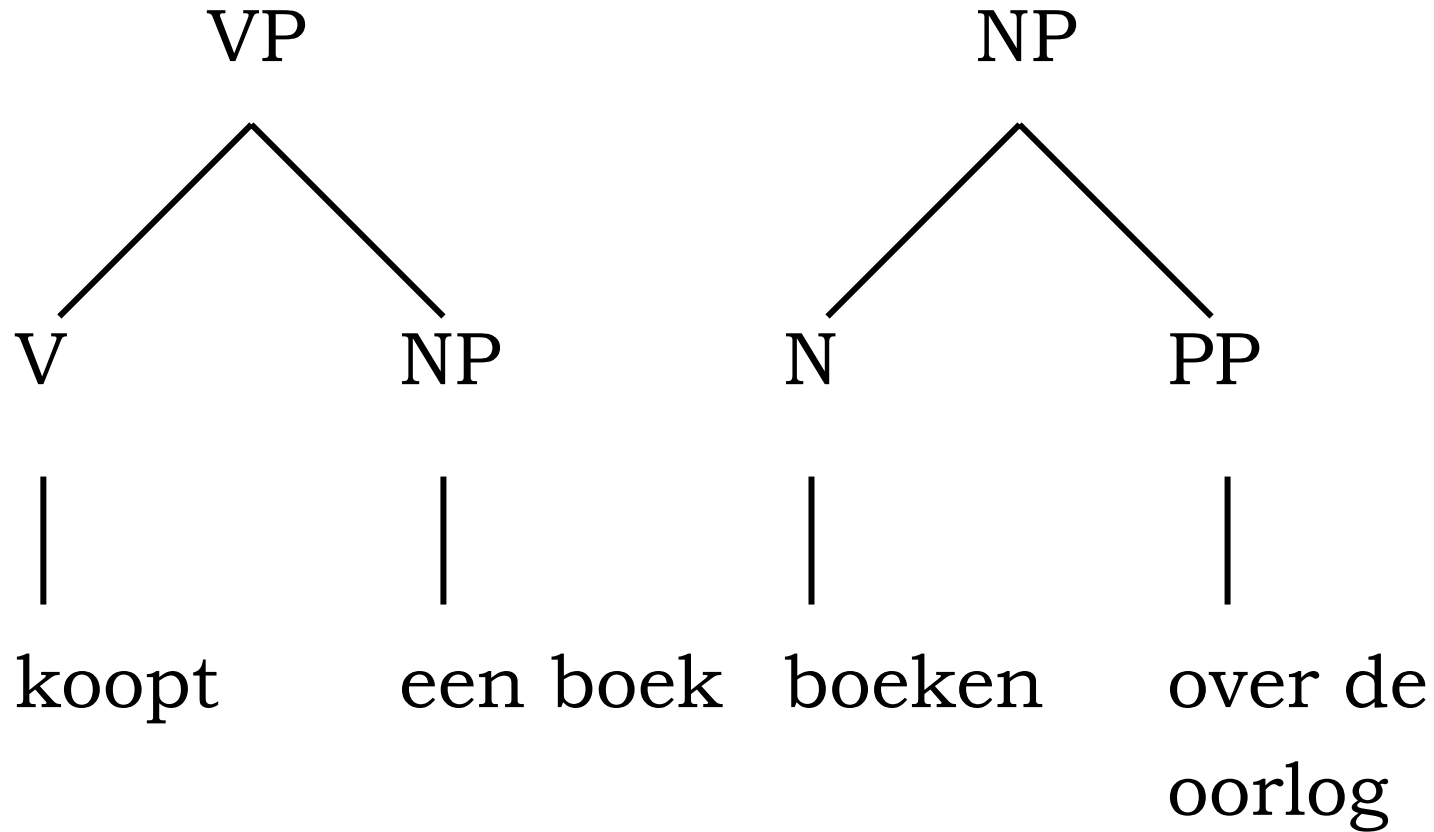


beneden

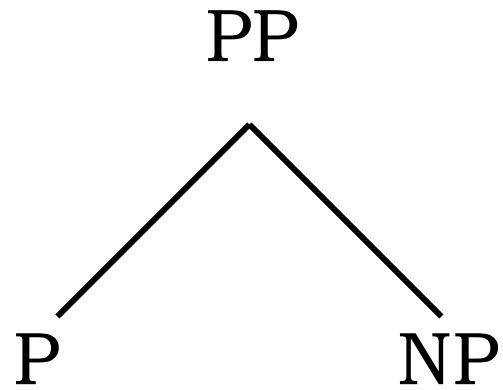
Specifiers



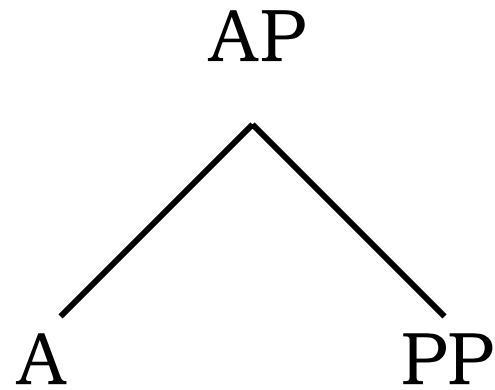
Complements



Complements



in
het huis



jaloers
op zijn zus

Rewrite rules

- NP → (Det), N, (PP)
- VP → (Aux), V, (NP)
- AP → (Deg), A, (PP)
- PP → (Deg), P, (NP)

Generalisation:

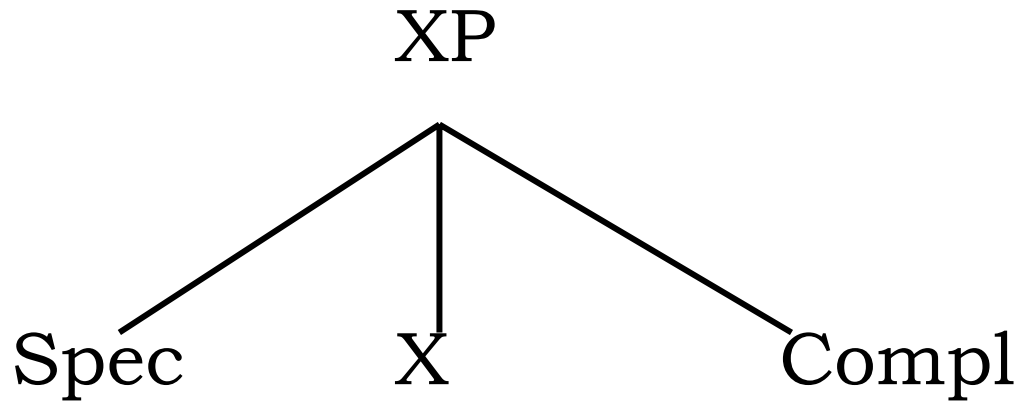
- XP → (Specifier), X, (Complement)

Linguistic analysis

- De kikker is gek.
 - De kikker: noun phrase
 - is: copula (*koppelwerkwoord*)
 - gek: adjective
- ➔ One level of linguistic analysis is the identification of parts of speech (word types)
- Knowing the part-of-speech of a word tells you a bit about its function

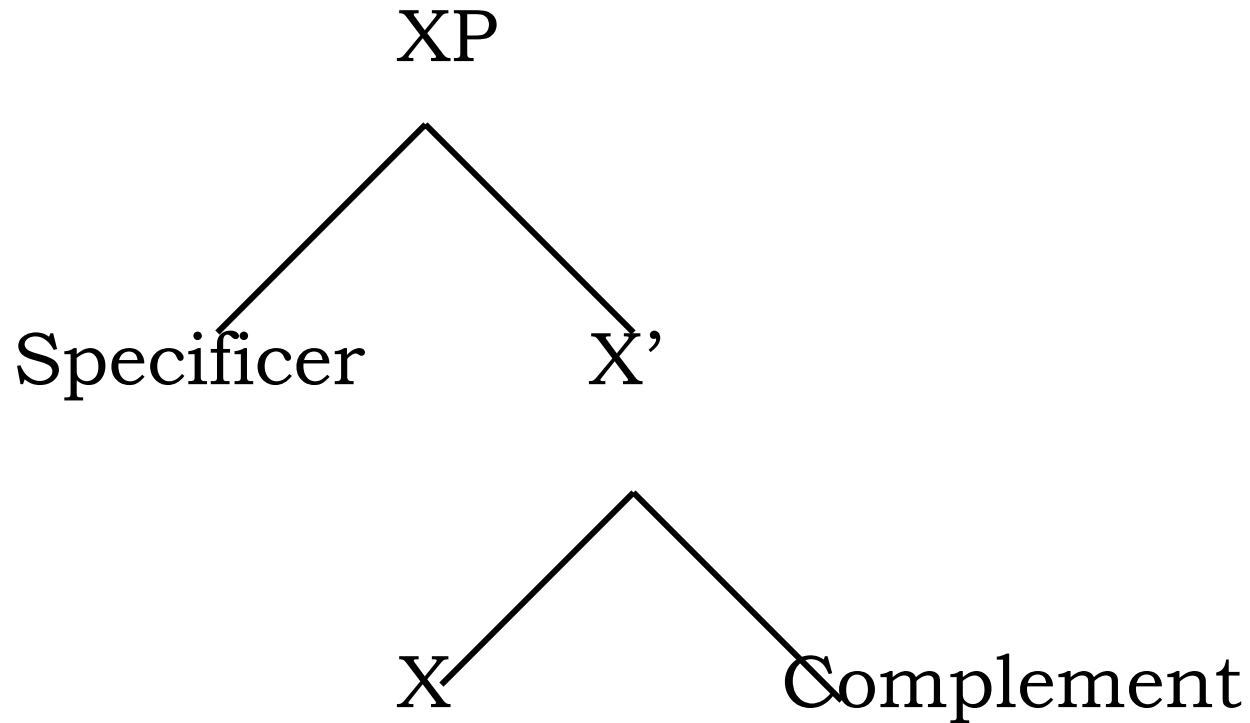


De structure van constituents



➔ However this is a simplification of what is actually binary branching

X-bar-structure



X-bar-rules

- $XP \rightarrow (\text{Specifier}), X'$
- $X' \rightarrow \text{Adjunct}, X'$
- $X' \rightarrow X, (\text{Complement})$

$X = N, V, A \text{ of } P$

X-bar-rules don't say anything about the order between the elements

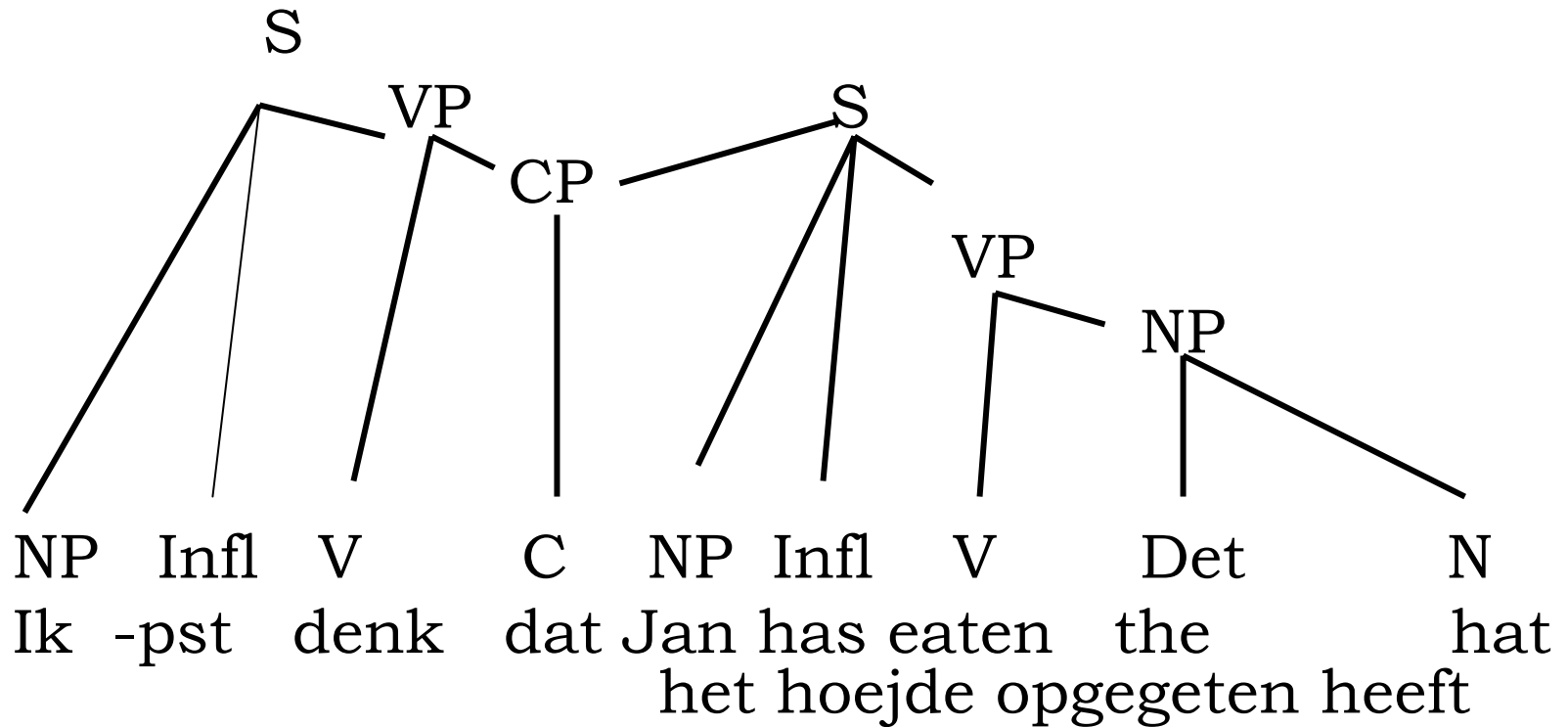
What kind of complements can verbs have?

- Ik zie jou.
- Ik zie de man met de verkijker.
- Ik ben op de grote markt.
- Hij kwam uit Tilburg.

Sentences within sentences

- Sentences can also function as verbal complements
- Ik denk **dat** Jan het hoedje opgegeten heeft.
 - dat Jan het hoedje opgegeten heeft: sentential object
- Words like “dat”, “of”, or “als” are complementizers, C
- They can form a Complementizer phrase

CP-structure



Automatic syntactic analysis

- Tagging: identify POS
- Syntactic analysis: Parsing
 - Very hard!
 - Syntacticians aren't in agreement! How can we find the right “parse”?
 - HPSG & LFG common grammatical formalisms
- NP-chunking & VP-chunking
 - Useful for Named-entity recognition, information retrieval, etc.

Problems for traditional grammars

- Wat denk je dat Jan opgegeten heeft?
 - je: subject
 - denk: persoonsvorm
 - dat Jan opgegeten heeft: objectzin
 - wat: ?
- ➔ What is the function of *wat* in the main clause (matrix clause) ?

What's happening here?

Wat can't be left out:

- *Jan heeft opgegeten.
- *Denk je dat Jan opgegeten heeft?

→ *opeten* needs an object

→ *wat* is the missing object from the embedded clause (*bijzin*).

But why is this sentences ungrammatical?:

- *Denk je dat Jan wat opgegeten heeft?

Problems with traditional grammars

- You can't specify that *wat* belongs to the embedded clause
- You can't explain why *wat* has to be at the beginning of the sentence. Why can't it just remain in the complement?

Movement

- Wat_i denk je dat Jan t_i opgegeten heeft?



- t = trace (spoor)
- i = index
- coinexation of *wat* en t
- Wh-movement (vraagwoordverplaatsing)

More movement

- Waar ging Jan op zitten?
- Wie vertelde dat Jan het hoedje opgegeten had?
- Wie vertelde Jan dat het hoedje opgegeten had?
- Wanneer zei je dat Jan het hoedje opgegeten had?

Movement in grammar

- Movement is a way to explain how one word fulfills more than one role
- Most modern grammatical formalisms make use of movement

Syntactic categories

- A constituent can be replaced with a constituent from the same category without it leading to ungrammaticality
- Ungrammaticality $\neq$ anomaly
 - Mijn tandenborstel is dronken.
 - Jan is dronken.
 - *Met mijn tandenborstel is dronken.

Colorless green ideas sleep
furiously



What evidence do we have that traces really exist?

- Teddy is the man who I want to succeed.
 1. Teddy is the man I want (Teddy) to succeed
 2. Teddy is the man I want to succeed (Teddy)
- BUT:
 - **Teddy is the man I wanna succeed.**
 - want to = wanna = assimilation (reduction)
 - This can't mean 1., only 2
 - Reduction applies after movement, but only to consecutive words, can't apply "over" a trace
 - "wanna" immediately disambiguates sentence as 2.

Next time....

- More syntax