

Syntax natürlicher Sprachen

Tutorium

Phrasenstrukturgrammatik

Sarah Anna Uffelmann

27.10.2023

Termine

Am 03.11. findet kein Tutorium statt!

Syntaktische Kategorien

„Eine syntaktische Kategorie umfasst eine Menge von **sprachlichen Einheiten**, die **bestimmte Eigenschaften gemeinsam** haben und deren Eigenschaften **relevant** sind **für die Beschreibung syntaktischer Strukturen**.“ (Dürscheid)

Verschiedene Varianten der syntaktischen Kategorisierung, etwa

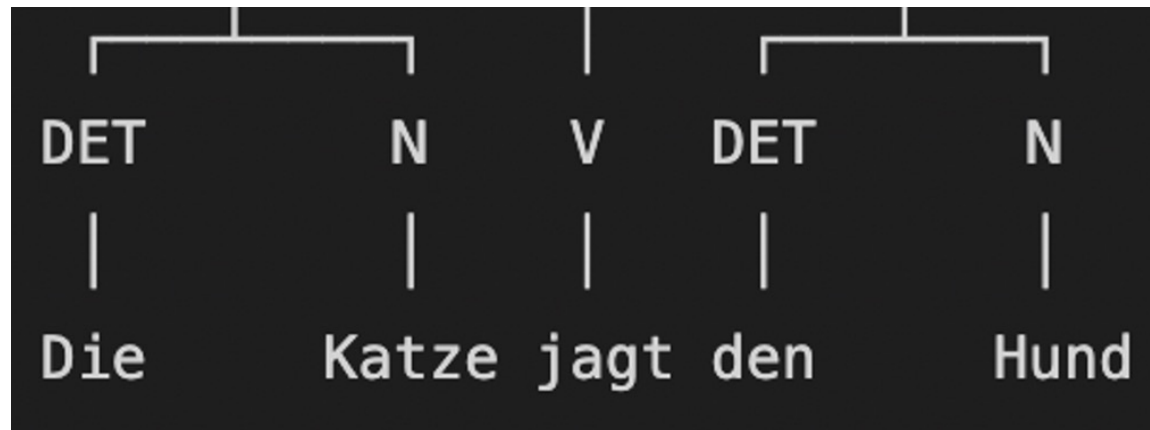
- nach Wortarten
- nach Phrasen bzw. nach Konstituenten

Lexikalische Kategorien: Wortarten, Parts of Speech (POS)

Wort -> atomare syntaktische Einheit

Wortart -> Klasse von Wörtern mit gemeinsamen Eigenschaften

Im Syntaxbaum sind Wörter **Terminale** und Wortarten **Präterminale**
(= Nichtterminale eine Ebene über den Terminalen):



Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

phrasenbildend, bilden je den Kopf einer Phrase

- **Nomen**: Substantive, Eigennamen, nominalisierte Adjektive **deklinierbar**

Hund

Shakespeare

(das) Schöne

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

phrasenbildend, bilden je den Kopf einer Phrase

- Nomen: Substantive, Eigennamen, nominalisierte Adjektive
- **Verben**: intransitiv, transitiv, ditransitiv

konjugierbar

schlafen

sehen

schenken

Er schläft.

Er sieht ein Haus.

Er schenkt seiner Schwester ein Buch.

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

phrasenbildend, bilden je den Kopf einer Phrase

- Nomen: Substantive, Eigennamen, nominalisierte Adjektive
- Verben: intransitiv, transitiv, ditransitiv

- **Adjektive** grün

deklinierbar & komparierbar

Arten des Gebrauchs:

attributiv -> das gute Buch

prädikativ -> das Buch ist gut

adverbial -> das Essen schmeckt gut

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

phrasenbildend, bilden je den Kopf einer Phrase

- Nomen: Substantive, Eigennamen, nominalisierte Adjektive
- Verben: intransitiv, transitiv, ditransitiv
- Adjektive
- **Adverbien** gern vorhin überall

nicht flektierbar

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- phrasenbildend, bilden je den Kopf einer Phrase
- Nomen, Verben, Adjektive, Adverbien

(2) Nominale Begleiter und Proformen

- **Pronomen**: vertritt eine Nominalphrase (Proform)

deklinierbar

Personalpronomina: **ich, du, er, sie, es** ...

Indefinitpronomina: **einer, jede** ...

Demonstrativpronomina: **dieser, jener**, ...

Fragepronomina: **wer, was**,

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- phrasenbildend, bilden je den Kopf einer Phrase
- Nomen, Verben, Adjektive, Adverbien

(2) Nominale Begleiter und Proformen

- Pronomen: vertritt eine Nominalphrase (Proform)
- **Determinativ**: nominaler Modifikator
Artikel, Quantifizierer, attributiv gebrauchte Pronomina

der
eine

alle
keiner

dein
diese

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- Nomen, Verben, Adjektive, Adverbien

(2) Nominale Begleiter und Proformen

- Pronomen, Determinativ

(3) Funktionswörter

- **Adposition**: Präposition, Postposition

phrasenbildend !

auf
in

halber

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- Nomen, Verben, Adjektive, Adverbien

(2) Nominale Begleiter und Proformen

- Pronomen, Determinativ

(3) Funktionswörter

- Adposition: Präposition, Postposition
- **Konjunktion**: koordinierend, subordinierend

und

weil

Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- Nomen, Verben, Adjektive, Adverbien

(2) Nominale Begleiter und Proformen

- Pronomen, Determinativ

(3) Funktionswörter

- Adposition: Präposition, Postposition
- Konjunktion: koordinierend, subordinierend

- **Partikel:**
 - Negationspartikel: **keinesfalls**
 - Intensitätspartikel: **sehr**
 - Modalpartikel: **ja, doch**
 - Diskurspartikel: **weil**
 - Ausdrucksartikel: **ähm**
juhu, oje

Tagsets

Zum Labeln syntaktischer Kategorien in NLP haben wir verschiedene Tagsets zur Auswahl.

POS-Tagsets:

vgl. Vorlesungsnotebook

- **Universal POS-Tagset:** <https://universaldependencies.org/u/pos/>
 - **NLTK-Version (reduziert):** <http://www.nltk.org/book/ch05.html#tab-universal-tagset>
 - **Penn Treebank POS-Tagset:** https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html
 - **TIGER/STTS-POS-Tagset:** https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/hagen/STTS_Tagset_Tiger
-

Konstituenten-Tagsets:

- **Penn Treebank Constituent Tags:** <http://surdeanu.cs.arizona.edu//mihai/teaching/ista555-fall13/readings/PennTreebankConstituents.html>
 - **TIGER Konstituenten Labels:** https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/hagen/Tiger_Knotenlabels
-

Dependency-Tagsets:

- **UD Dependency Labels:** <https://universaldependencies.org/u/dep/all.html>
 - **TIGER Dependency Labels:** https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/hagen/DDB_edge
-

Tagsets – Beispiele (Auszüge)

Universal POS tags

- [ADJ](#): adjective
- [ADP](#): adposition
- [ADV](#): adverb
- [AUX](#): auxiliary
- [CCONJ](#): coordinating conjunction
- [DET](#): determiner
- [INTJ](#): interjection
- [NOUN](#): noun
- [NUM](#): numeral
- [PART](#): particle
- [PRON](#): pronoun
- [PROPN](#): proper noun
- [PUNCT](#): punctuation
- [SCONJ](#): subordinating conjunction
- [SYM](#): symbol
- [VERB](#): verb
- [X](#): other

Penn Treebank POS tags

12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural

27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present

Tiger / STTS POS tags

VVFIN	finites Vollverb
VAFIN	finites Voll- oder Kopulaverb
VMFIN	finites Modalverb
VVINFIN	infinites Vollverb
VAINFIN	infinites Hilfsverb oder Kopulaverb
VMINFIN	infinites Modalverb
VVIMP	Vollverb im Imperativ
VAIMP	Kopulaverb im Imperativ
VVPP	partizipiales Vollverb (Partizip II)
VAPP	partizipiales Hilfs-/Kopulaverb (Partizip II)
VMPP	partizipiales Modalverb (Partizip II)
VVIZU	Vollverb/Partikelverb im "zu"-Infinitiv

Konstituenz

Wie kommt man darauf, Sätze in Phrasen zu unterteilen?

Wir können Teile eines Satzes vertauschen:

(1a) Die Katze jagt den Hund.

(1b) Den Hund jagt die Katze.

Aber wir können nicht beliebig Wörter vertauschen:

(1c*) Den Katze die Hund jagt. \rightarrow nicht wohlgeformt / ungrammatisch

Wörter stehen also nicht isoliert im Satz, sondern wir können Einheiten bilden:
Phrasen

Phrasen sind Konstituenten

Phrasenstrukturgrammatik

(Kontextfreie Grammatik, **CFG** (= Context-free Grammar))

- folgt dem Ansatz der **Konstituenz**
- **Konstituente**: Bestandteil einer syntaktischen Struktur;
- sprachliche Einheit, die Teil einer größeren sprachlichen Einheit ist (Dürscheid)
- **Phrase**: Wortgruppe, die syntaktisch eine Einheit bildet (Dürscheid)
- Alle Phrasen sind Konstituenten, aber nicht alle Konstituenten sind Phrasen.

Was ist ein Beispiel für eine Konstituente, die keine Phrase ist?

Phrasenstrukturgrammatik

Was ist ein Beispiel für eine Konstituente, die keine Phrase ist?

→ Einzelne Wörter eines Satzes.

In **Der Hund jagt die Katze** ist „Hund“ eine Konstituente, aber keine Phrase.

Vorsicht: Einzelne Wörter können auch Phrasen sein!

In **Er jagt die Katze** ist „Er“ sowohl Konstituente als auch Phrase.

(Phrasenbildende) Wortarten des Deutschen

(1) Lexikalische Hauptkategorien:

- Nomen, Verben, Adjektive, Adverbien

NP

VP

ADJP

ADVP

(2) Nominale Begleiter und Proformen

- Pronomen, Determinativ

(3) Funktionswörter

- Adposition, Konjunktion, Partikel

PP

Phrasenkategorien des Deutschen

- NP – Nominalphrase: "der **Elefant**"
- VP – Verbalphrase: "ins Kino **gehen**"
- PP – Präpositionalphrase: „**in** meinem Schlafanzug“
- ADJP – Adjektivphrase: „sehr **großer**“
- ADVP – Adverbialphrase: „sehr **oft**“

IC-Analyse

IC = immediate Constituents = unmittelbare Konstituenten

Die unmittelbaren Konstituenten eines Satzes sind die längsten Wortfolgen, in die man einen Satz (bzw. eine übergeordnete Konstituente) zerlegen kann.

Die Katze jagt den Hund.

-> unmittelbare Konstituenten: „Die Katze“ und „jagt den Hund“

Die Katze

-> unmittelbare Konstituenten: „Die“, „Katze“

jagt den Hund

-> unmittelbare Konstituenten: „jagt“, „den Hund“

... usw. bis zur Wortebene

Phrasenstrukturgrammatik

$S \rightarrow NP VP$

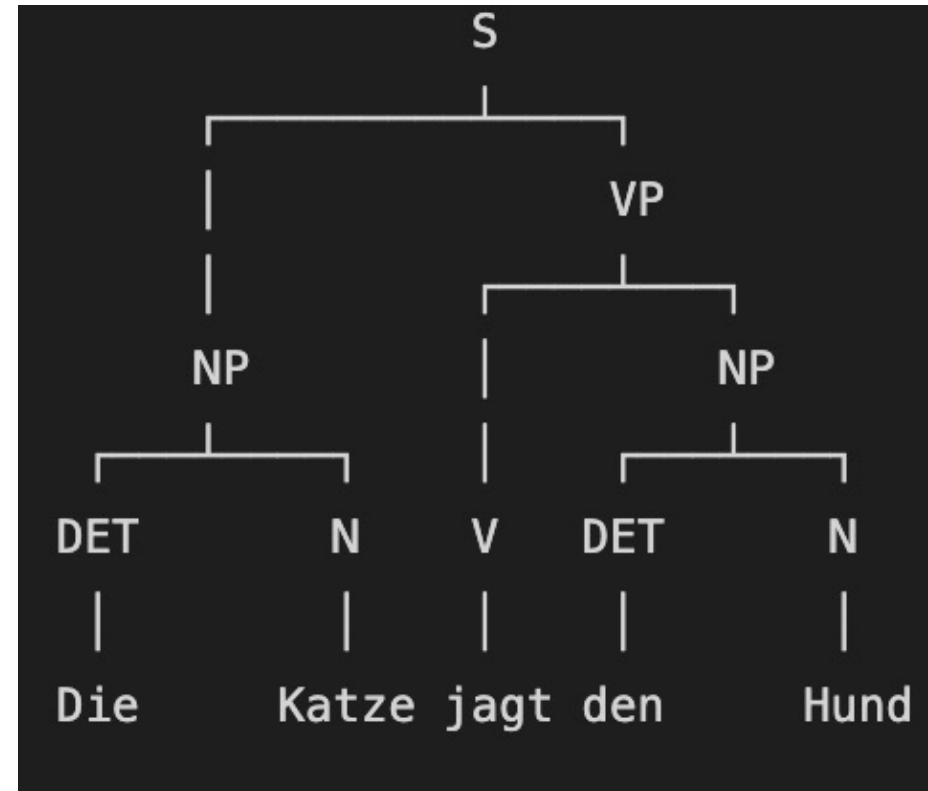
$NP \rightarrow Det N$

$VP \rightarrow V NP$

$Det \rightarrow „die“ \mid „den“$

$N \rightarrow „Katze“ \mid „Hund“$

$V \rightarrow „jagt“$



Phrasenstrukturgrammatik

syntaktische Regeln

S \rightarrow NP VP

NP \rightarrow Det N

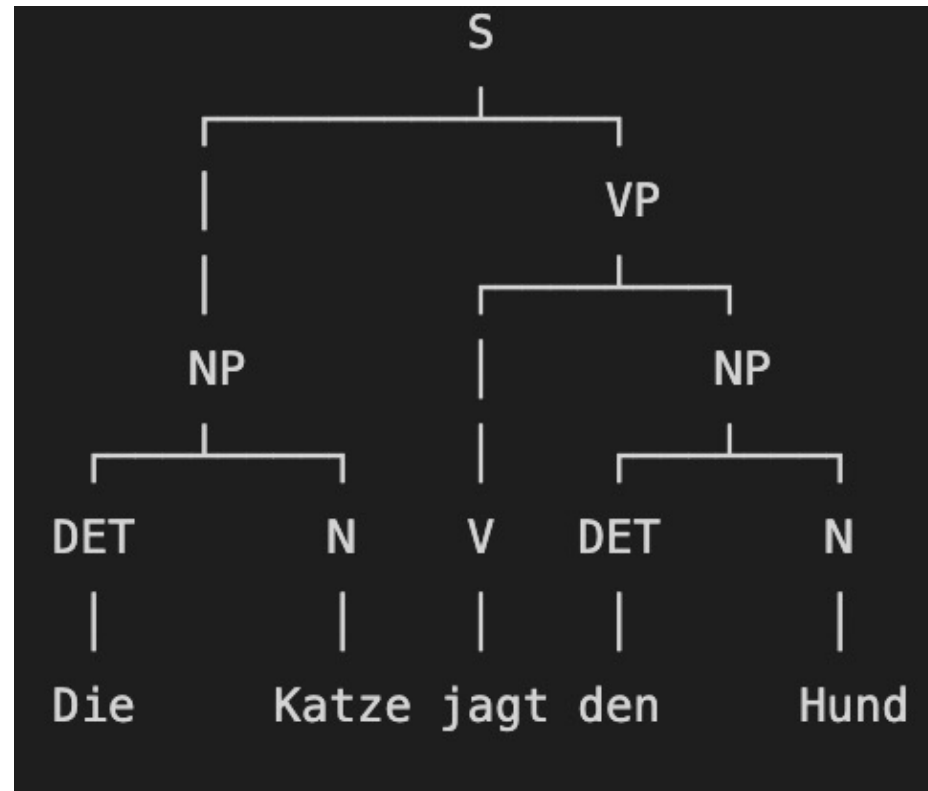
VP \rightarrow V NP

lexikalische Regeln

Det \rightarrow „die“ | „den“

N \rightarrow „Katze“ | „Hund“

V \rightarrow „jagt“



Phrasenstrukturgrammatik

syntaktische Regeln

S \rightarrow NP VP Startsymbol

NP \rightarrow Det N

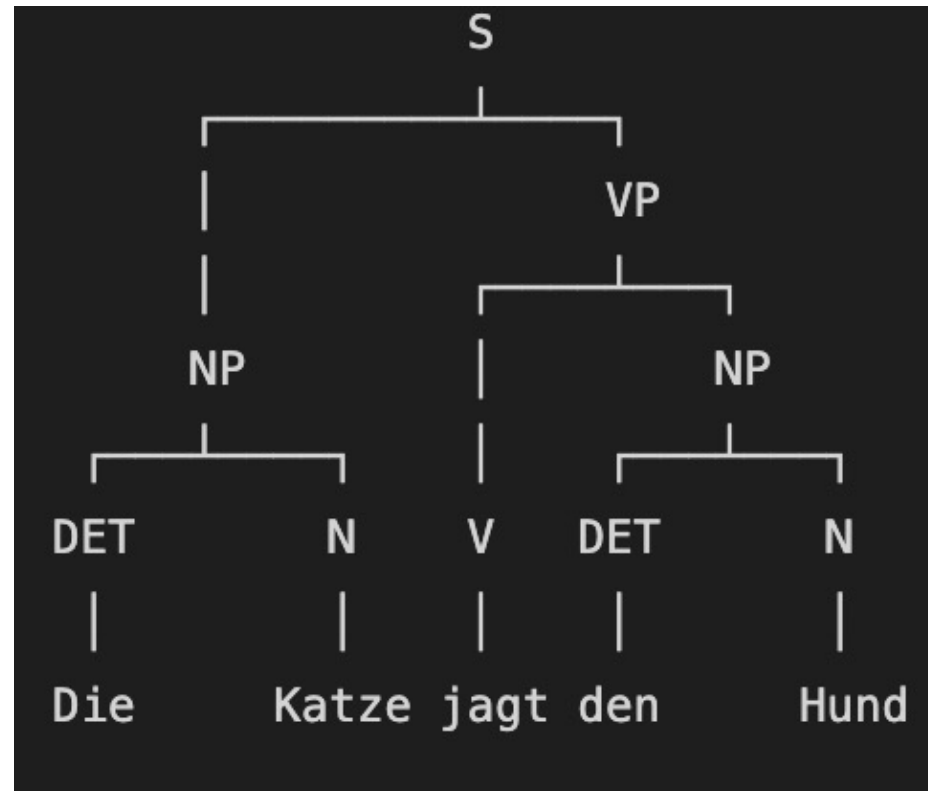
VP \rightarrow V NP

lexikalische Regeln

Det \rightarrow „die“ | „den“

N \rightarrow „Katze“ | „Hund“

V \rightarrow „jagt“



Phrasenstrukturgrammatik

syntaktische Regeln

$S \rightarrow NP VP$ Nicht-Terminale

$NP \rightarrow Det N$

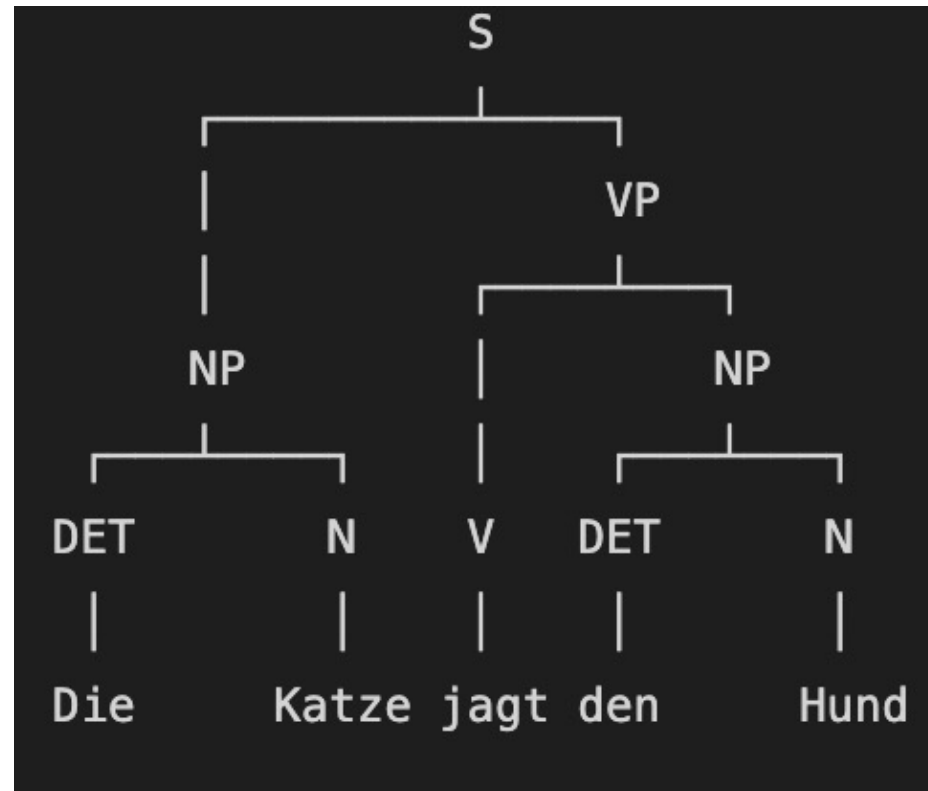
$VP \rightarrow V NP$

lexikalische Regeln

$Det \rightarrow \text{„die“} \mid \text{„den“}$

$N \rightarrow \text{„Katze“} \mid \text{„Hund“}$

$V \rightarrow \text{„jagt“}$



Phrasenstrukturgrammatik

syntaktische Regeln

$S \rightarrow NP VP$

$NP \rightarrow Det N$

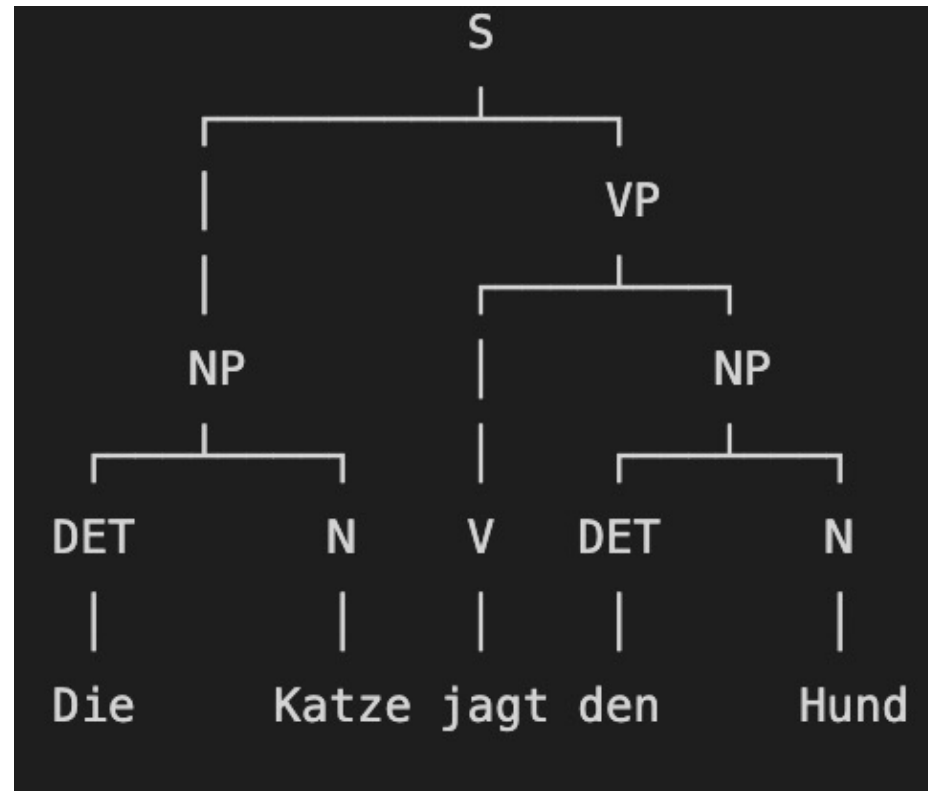
$VP \rightarrow V NP$

lexikalische Regeln Terminale

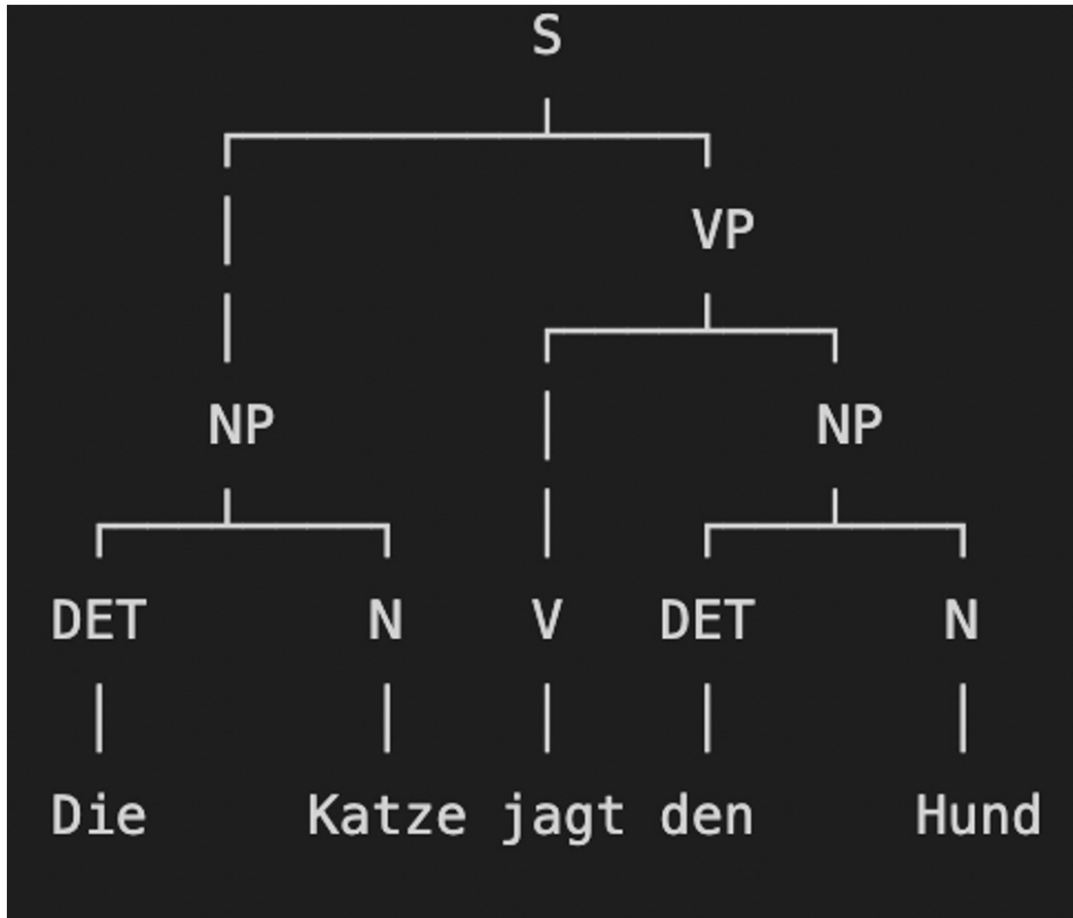
$Det \rightarrow \text{„die“} \mid \text{„den“}$

$N \rightarrow \text{„Katze“} \mid \text{„Hund“}$

$V \rightarrow \text{„jagt“}$



Begrifflichkeiten



S ist **Mutterknoten** von NP und VP

S ist **dominant** zu NP und VP

NP ist **Mutterknoten** von DET und N

NP ist **Tochterknoten** von S

VP ist **Tochterknoten** von S

NP und VP sind **Geschwisterknoten**

NP und VP sind **Ko-Konstituenten**
(gemeinsam konstituieren sie S)

(Mit NP ist hier immer die NP „Die Katze“ gemeint!)

Konstituentenstruktur

Elemente der Konstituentenstruktur (Knoten):

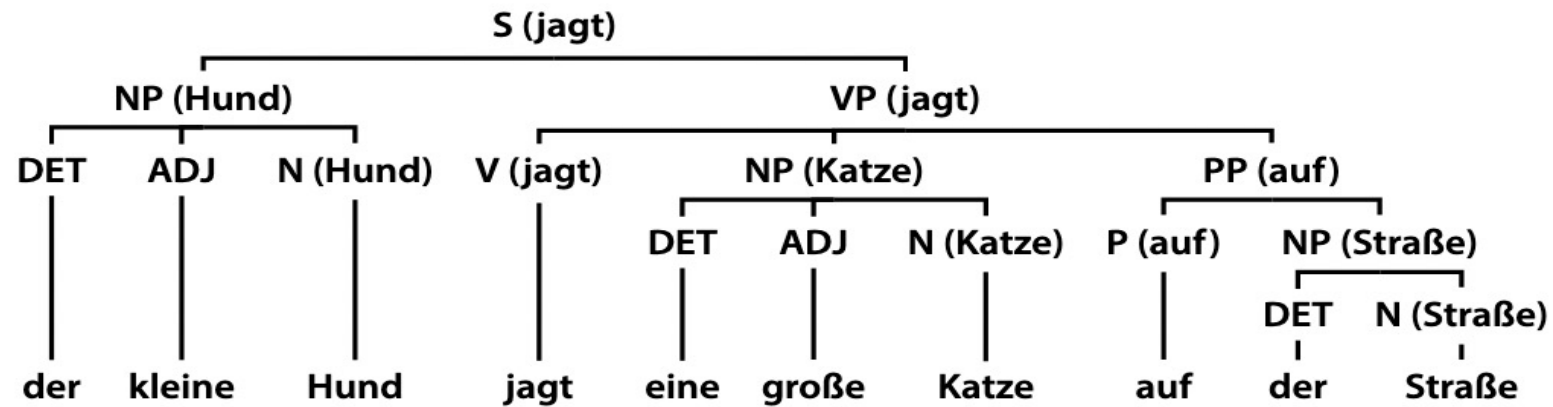
- Wörter = terminale Knoten
- Lexikalische Kategorien = Wortarten = Präterminale (nichtterminale) Knoten
- Syntaktische Kategorien = Phrasen = nichtterminale Knoten

Relationen der Konstituentenstruktur (Kanten):

- drückt eine Teil-Ganzen-Beziehung aus

Jede Phrase hat genau einen **Kopf**. Phrasenköpfe werden im Syntaxbaum nach oben weitergereicht. Dieses Prinzip nennt man **Kopf-Perkolation**.

Kopfperkolation



vgl. Vorlesungsfolien

Kopf einer NP: N (hier „Hund“, „Straße“ und „Katze“)

Kopf einer VP: V (hier „jagt“)

Kopf einer PP: P (hier „auf“)

Kopf von S = Kopf der VP = Verb (Kopf wird nach oben weitergereicht)

Regeltypen

syntaktische Regeln

links: syntaktische Kategorie, rechts syntaktische oder lexikalische Kategorie

$S \rightarrow NP VP$, $NP \rightarrow Det N$

lexikalische Regeln

links: lexikalische Kategorie, rechts Wörter

$N \rightarrow \text{„Katze“}$

rekursive Regeln

Nur bei syntaktischen Regeln; ein Symbol auf der rechten Seite entspricht dem Symbol auf der linken Seite

$NP \rightarrow Det N NP$

Regeltypen

Direkte Rekursion

NP -> Det N NP (rechtsrekursiv)

VP \rightarrow VP PP (linksrekursiv)

NP NP NP

„(der Hund) (der Frau) (des Nachbarn)“

Det N Det N Det N

Indirekte Rekursion

NP -> Det N PP

PP \rightarrow P NP

„(die Lampe) (auf dem Tisch) (neben der Tür)“

Det N P Det N P Det N

NP NP NP