

Einführung in die Computerlinguistik

Phrasenstrukturgrammatik

Alex Fraser / Robert Zangenfeind

Center for Information and Language Processing

2019-11-11

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth unter Zuhilfenahme von Materialien aus Vorlesungen von Prof. Dr. Tania Avgustinova erstellt. Fehler und Mängel sind ausschließlich meine Verantwortung.

- 1 Intro
- 2 Konstituenten
- 3 Konstituententypen
- 4 Phrasenstrukturgrammatik

- was ein grammatikalisch korrekter Satz ist?
 - was ein Satzglied ist?
 - wie Wörter zu Satzgliedern zusammengefügt werden können (so dass das Ergebnis grammatikalisch korrekt ist)?
 - wie Satzteile kombiniert und angeordnet werden können?
-
- Dies ist die Sichtweise der Phrasenstrukturgrammatik.
 - Vergleiche: Sichtweise der Abhängigkeitsgrammatik.

- Verschiedene Definitionsversuche:
 - ① Eine grammatische Kette von Wörtern
 - ② Die Grundeinheit des Diskurses
 - ③ verschiedene Definitionen anhand prosodischer Kriterien
 - ④ Drückt eine Proposition aus
- Proposition: Aussage, die einen Wahrheitswert hat.
 - Inhalt, den ein Satz (in einem Kontext) hat
 - mit Satzsubjekt (darüber wird etwas ausgesagt)
 - und Prädikation (sagt etwas über das Satzsubjekt aus)
- entscheidendes Merkmal: Prädikativität; wird ausgedrückt durch: Tempus, Modus (Verb in finiter Form)
- stellt die zeitliche Korrelierung der Information zur Wirklichkeit und den Realitätsbezug her; z.B. *Hans hustet.*
- Satzsubjekt und Prädikation können weiter modifiziert werden:
Der nette Hans hustet.
Hans hustet oft.

Satz: Schwierigkeiten der Definition

- Beispiele:
 - *Der Junge liest.*
 - *Es wird dunkel.*
 - *In Brasilien wird gut Fussball gespielt.*
 - *Vorsicht!*
- Grenzen der Prädikativität: Fragen, Befehle
- Trennen von Semantik: *Colorless green ideas sleep furiously.*
- Syntaktische Theorien sollten aus der Menge aller Wortketten diejenigen auswählen, die wohlgeformte Sätze sind.
 - *Hans sieht uns.*
 - **Hans uns liest sieht.*
 - **Hans gesehen uns.*
- Traditionelle Computerlinguistik: Algorithmen/Parser, die auf solchen syntaktischen Theorien basieren

- Der Satz ist zusammengesetzt aus Phrasen, die in syntaktischen Beziehungen zueinander stehen.
- Genauso sind größere Phrasen/Syntagmen zusammengesetzt aus kleineren Phrasen/Syntagmen, die in syntaktischen Beziehungen zueinander stehen.
- Wortfügungen / subordinierende Syntagmen:
kleines Kind, schnell laufen
- Wortreihungen / koordinierende Syntagmen:
Mutter oder Sohn, kommen und sehen
- analog: Satzfügungen (subordinierend) & -reihungen (koordinierend)
- Dies ist die (nicht von allen geteilte) Grundannahme der Phrasenstrukturgrammatik.

- Konstituentengrammatik (auch Phrasenstrukturgrammatik): Satz ist nur mittelbar aus Wörtern gebaut.
- Elemente der Grammatik sind **Konstituenten**: Gruppen von Wörtern (oder nur ein Wort)
- Die Beziehungen zwischen diesen Gruppen sind wichtig, es besteht eine abgestufte Zusammengehörigkeit.
- Phrasenstrukturgrammatik
 - rekursives System komplexer Strukturen
 - beruht ausschließlich auf der Form immer wiederkehrender Muster

- **Schachtelstruktur:**
[[[Die] [Katze]] [[schläft] [nicht]]]
- **Terminale Konstituenten:** Konstituenten, die nicht weiter zerlegt werden können,
- **Unmittelbare Konstituenten** einer Konstituente **A**:
Konstituenten, aus denen **A** unmittelbar zusammengesetzt ist.
[The voters]₁ [sought the safety of the centre]₂

Konstituententests: Motivation

- Intuitiv klare Gruppen?
- Beispiel: *Der kleine Junge fährt mit dem neuen Fahrrad, das ihm der Opa geschenkt hat, zum Kindergarten.*
- Konstituenten (eine Auswahl): [Der kleine Junge] fährt [mit [dem neuen Fahrrad], [das ihm [der Opa] geschenkt hat,]] [zum Kindergarten].
- Nicht-Konstituenten: “Junge fährt”, “das ihm der”
- Wörter sind nicht nur aneinandergereiht, sondern gruppiert.
- Hierarchische Strukturen
- → Tests sind nötig, um die Gruppen / syntaktischen Einheiten zu ermitteln.
- Konstituententests: unterstützende Hilfsmittel zur Bestimmung der Konstituenten eines Satzes
- Können sich widersprechen: oft mehrere Analysemöglichkeiten.

Ersetzungstest

Wortfolgen, die sich gegenseitig ersetzen lassen, sodass ein Satz grammatisch bleibt, sind möglicherweise Konstituenten.

Beispiel

Die langen Abende ...

Goethe und Kohl versetzen mich in Melancholie

Alle ...

- Außer der Ersetzung darf nichts geändert werden.
- Beispiel: *Du kochst.*
- *du* lässt sich nicht ersetzen: *ich koche, er kocht, wir kochen*

- Was sich pronominalisieren lässt, ist eine Konstituente, vgl.:
- *Tine und Jakob wohnen hier. Sie wohnen hier.*

Variante: *Das Kind will einen Pudding essen. Das will ich auch.*

- Da sich bestimmte Ketten nicht pronominalisieren lassen, ist dieser Test nicht immer anwendbar:

Ich möchte nach Paris fahren.

Der Teil “möchte nach Paris fahren” kann nicht pronominalisiert werden – ist aber syntaktische Einheit!

- Wonach sich fragen lässt, ist eine Konstituente, vgl.:
- *Hans rasiert sich mit dem alten Piratenmesser.*
- *Womit rasiert sich Hans? Mit dem alten Piratenmesser.*

Koordinationstest

Was sich koordinieren lässt, ist eine Konstituente.

Beispiele

*John ran **up the hill** and **up the stairs**.*

*John rang up **his mother** and **his sister**.*

(“up his mother” ist keine Konstituente)

*Willst du lieber **Ball spielen** oder **Bier trinken**?*

- Die so bestimmten Konstituenten sind vom selben Typ.
- Ausnahmen ...

Verschiebetest / Permutationstest

Was verschoben werden kann, ist eine Konstituente.

Beispiel

Gestern hat Peter seinem Lehrer *den Aufsatz* gegeben.

Peter hat *gestern* seinem Lehrer *den Aufsatz* gegeben.

Den Aufsatz hat Peter seinem Lehrer *gestern* gegeben.

- Eine bloße **Einfügung** gilt nicht als Verschiebung, vgl.:
- *Peter baut zwei Häuser.*
- *Peter **entwirft und** baut zwei **oder drei** Häuser.*
- "baut zwei" ist keine Konstituente!
- Es ist aber kein Problem im Sinne des Tests, wenn andere Teile des Satzes auch verschoben werden, nicht nur die zu testende Konstituente (vgl.o.).

Weglasstest / Tilgung

Konstituenten können weggelassen werden.

Beispiele

- Weglassen von syntaktisch optionalen Ergänzungen:
Maria singt in der Badewanne. → *Maria singt.*
- Weglassen in elliptischen Konstruktionen:
Peter liebt den Winter, aber Karl hasst den Winter. → *Peter liebt, aber Karl hasst den Winter.*

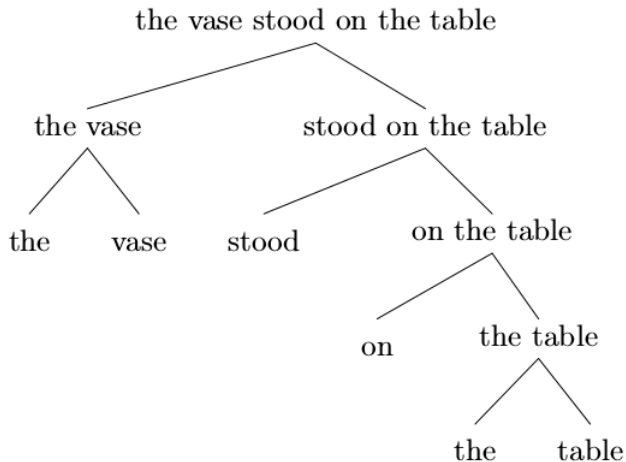
- Strikte Interpretation: nur direkt anwendbar auf den Satz, wie er vorliegt.
- Also z.B. nicht anwendbar auf:
Peter liebt den Winter.
- Großzügigere Interpretation: man darf den Satz ändern/erweitern.
- Beispiel:
Peter liebt den Winter.
→ *Peter liebt den Winter, aber Karl hasst den Winter.*
→ *Peter liebt, aber Karl hasst den Winter.*

- Die in Spaltsätzen und Pseudospaltsätzen abgespaltenen Elemente sind Konstituenten, vgl.:
- *John Miller used my toothbrush.* ⇒
*It was **John Miller** who used my toothbrush.*
*It was **my toothbrush** what John Miller used.*
- *His dishonesty annoys me.* ⇒
*What annoys me is **his dishonesty**.*

- Was einer Aktiv-Passiv-Konversion unterliegt, ist eine Konstituente, vgl.:
- *John used my toothbrush.*
- *My toothbrush was used by John.*

Konstituentenanalyse für: *The vase stood on the table*

Konstituente	Test
<i>The vase</i>	Pronominalisierung: <i>It stood on the table.</i>
<i>stood on the table</i>	Koordinationstest: <i>The vase stood on the table and looked splendid.</i>
<i>on the table</i>	Frage­test: <i>Where did the vase stand?</i>
<i>the table</i>	Pseudospaltsatz: <i>What the vase stood on was the table.</i>



Fazit?

- In der Phrasenstrukturgrammatik gibt es das Konzept des **Kopfes**.
- Vergleiche: Dependenzgrammatik.
- (Fast) jede Konstituente enthält einen obligatorischen Kopf.
- Der Kopf bestimmt den Phrasentyp der Konstituente **nach außen**.
- Der Kopf strukturiert die Konstituente **intern** – er verlangt z.B. obligatorische Argumente oder Determinatoren.
- **A syntactic head determines the internal structure and external distribution of the constituent it projects.** (Emily Bender 2013:63)

(i) Nominalphrasen (NP)

- *das Kind*
- *die sehr frechen Kinder*
- *Kinder, die in München leben*

(ii) Adjektivphrasen (AP)

- *Der Wagen ist neu.*
- *Die Rostlaube ist ziemlich alt.*
- *Leo ist seiner Frau treu.*
- *Heinz ist auf seine Tochter stolz.*

(iii) Verbalphrasen (VP)

- *Eva hustet.*
- *Eva besucht die Nachbarn.*
- *Eva schenkt ihrem Mann ein Buch.*
- *Eva glaubt, dass Peter krank ist.*

(iv) Adverbialphrasen (AdvP)

- *heute früh*
- *gestern Morgen*
- *sehr oft*

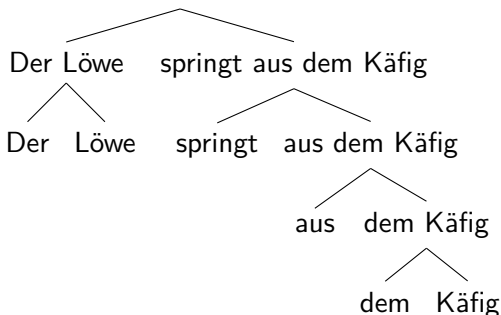
(v) Präpositionalphrasen (PP)

- *Wir fahren mit dem Fahrrad.*
- *Sie stehen vor der Tür.*
- *Wir spazieren die Straße entlang.* (hier mit Akk wg. Postposition)
- *Sie kommen von dort.*
- *Wir brachten sie bis vor die Tür.*

- Im Konstituentenbaum können die Wortketten nicht nur gruppiert werden, sondern hierarchisch durch die Symbole der syntaktischen Kategorien ersetzt werden.
- Außerdem wird eine zusätzliche Ebene eingefügt, in der die Wörter durch ihre jeweilige Wortkategorie ersetzt werden.
- Ein Satz (S) kann dann in seine unmittelbaren Konstituenten (NP + VP) und weiter bis zur Ebene der Wortkategorien zerlegt werden.

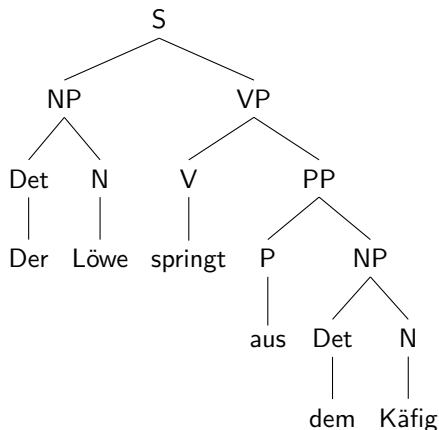
Konstituentenbaum (einfach)

Der Löwe springt aus dem Käfig



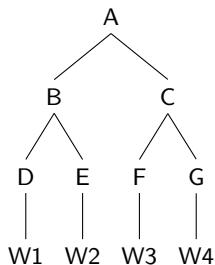
(Bsp. aus K.H. Ramers: Einführung in die Syntax. Paderborn 2007:29)

Konstituentenbaum (mit Kategorien)



alternative Darstellung:

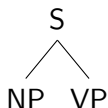
[S [NP [Det Der] [N Löwe]] [VP [V springt] [PP [P aus] [NP [Det dem] [N Käfig]]]]]



- Variablen A–G: Kategoriale Symbole
- speziell A–C: phrasale Kategorien
- speziell D–G: präterminale Kette; Wortarten
Einzelne Wortformen sind nicht unmittelbare Konstituenten einer Phrase → präterminale Kette nötig
- W1–W4: terminale Kette; konkrete Wortformen
- (Unmittelbare) **Dominanz**: Ein Knoten dominiert (unmittelbar) die (nächst-)tieferen Knoten.

- Satz → Nominalphrase + Prädikat
Nominalphrase → Artikel + Nomen etc.
- Viele syntaktische Theorien basieren auf solchen Regeln.
- setzt voraus: Klassifizierung der beteiligten syntaktischen Einheiten (da Regeln Variablen enthalten, die auf Klassen Bezug nehmen)
- Syntaxtheorien nehmen an, dass Sprachen eine große, aber endliche Zahl von grammatischen Regeln haben.
- Diese begrenzte Anzahl ermöglicht die Bildung von potentiell unbegrenzt vielen Sätzen.
- Beispiel: theoretisch unbegrenzte Möglichkeit der Hinzufügung von Elementen wie Relativsätzen
- Hierarchische Organisation der Konstruktionen in vielen syntaktischen Theorien

- Auch: Context-Free Grammar (CFG).
- Beschreibt die Regeln der Konstituentenstruktursyntax.
- Zentrales Element der generativen Grammatik (N. Chomsky).
- In generativer Grammatik werden Beziehungen der unmittelbaren Dominanz als Ersetzungsregeln zur Erzeugung von Sätzen (Konstituentenstrukturen) aufgestellt, z.B.:



- Regel (rewriting rule, Phrasenstrukturregel) hierzu:
- $S \rightarrow NP VP$
- zu lesen als: “expandiere S zu .../ ersetze S durch .../ S dominiert unmittelbar .../ schreibe S um zu ...”
- S: Startsymbol (“Satz”)

Allgemeines Schema einer PS-Regel:

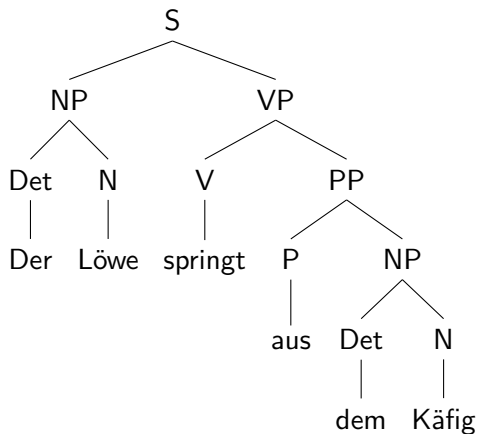
$X \rightarrow (W) Y (Z)$

(W–Z: Kategorialsymbole)

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

PS-Regeln sind kontextfrei, d.h. genau ein Nichtterminalsymbol (links) wird expandiert zu beliebiger Folge von Nichtterminal- und Terminalsymbolen (rechts).

Beispiel für Phrasenstrukturregeln (1)



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

$VP \rightarrow V PP;$ $PP \rightarrow P NP$

Beispiel für Phrasenstrukturregeln (2)

$S \rightarrow NP VP;$ $NP \rightarrow Det N$
 $VP \rightarrow V PP;$ $PP \rightarrow P NP$

Mit diesen Regeln kann obige Baumstruktur bis zur Ebene der Wortkategorien (präterminale Kette) generiert werden.

Außerdem möglich:

- *Die Katze sitzt auf der Bank*
- *Der Bauer arbeitet auf dem Feld*
etc.

Zur Generierung der terminalen Kette *Der Löwe springt aus dem Käfig* sind noch weitere Regeln nötig:

Det → der

N → Löwe

V → springt

P → aus

Det → dem

N → Käfig

- zu lesen als: *Käfig* ist ein Exemplar der Menge der Nomen
- diese Regeln entsprechen einem Lexikon
- die Wörter können in die durch PS-Regeln erzeugten Strukturbäume eingesetzt werden (lexikalische Einsetzung)

Rekursion ist möglich, d.h. phrasale Gruppen können Konstituenten derselben phrasalen Kategorie enthalten:

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

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

→ komplexe NP:

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

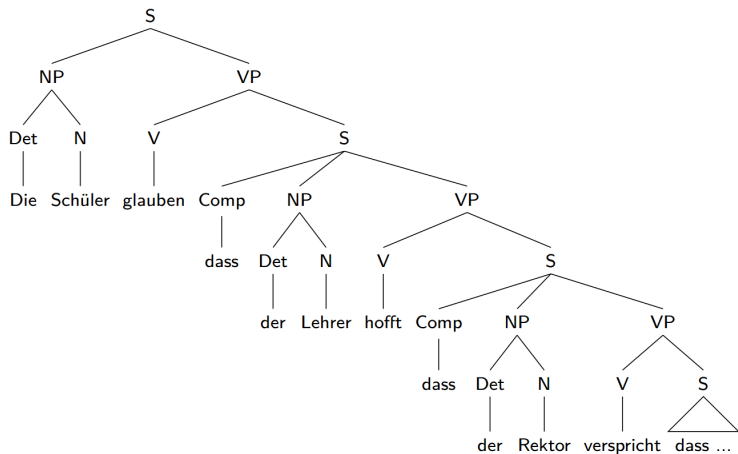
(ii) Ausgabe einer Regel dient als Eingabe einer “früheren” Regel:

$S \rightarrow (Comp)\ NP\ VP$

$VP \rightarrow V\ S$

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

Rekursion (2)



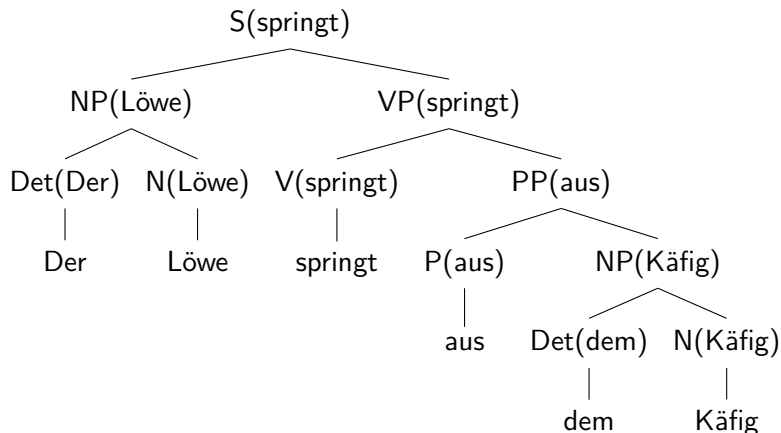
$S \rightarrow (\text{Comp}) \text{NP VP}$

$\text{VP} \rightarrow \text{V S}$

...

- In der Phrasenstrukturgrammatik gibt es das Konzept des **Kopfes**.
- Vergleiche: Dependenzgrammatik.
- (Fast) jede Konstituente enthält einen obligatorischen Kopf.
- Der Kopf bestimmt den Phrasentyp der Konstituente **nach außen**.
- Der Kopf strukturiert die Konstituente **intern** – er verlangt z.B. obligatorische Argumente oder Determinatoren.
- **A syntactic head determines the internal structure and external distribution of the constituent it projects.** (Emily Bender 2013:63)

Beispiel für Phrasenstrukturregeln mit Kopf-Annotation



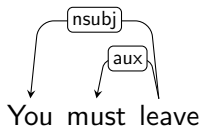
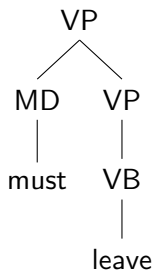
- **Perkolation** (“Durchfließen”): Ein Kopf wird je nach Phrasentyp von einer Ebene zur nächsten weitergereicht.
- Welcher Kopf perkoliert (weitergereicht) wird, wird in einer Extra-Tabelle angegeben, z.B.:
S \rightarrow NP VP; head(S) = head(VP)
NP \rightarrow Det N; head(NP) = head(N)
VP \rightarrow V PP; head(VP) = head(V)
PP \rightarrow P NP; head(PP) = head(P)
- Für terminale Regeln ist der Kopf immer das terminale Symbol.
- Wichtiger Begriff: **perkolierter Kopf**

Konstituentenbäume mit Kopf-Annotation können mit dem folgenden Schema in einfache Dependenzbäume überführt werden:

- 1 Identifizieren Sie für jede Konstituente die erzeugende Regel (bzw. die Kopf-Annotationen).
- 2 Dependenzrelationen bestehen zwischen dem Kopf der Konstituente, und den nicht-perkolierten Teilen der eingebetteten Konstituenten (auf der rechten Seite der Regel).
- 3 Der Kopf der Konstituente ist auch Kopf der Dependenzrelation, der nicht-perkolierte Teil ist Dependent.
- 4 Die Wortart des Dependents ist dann z.B. der Kantename der Dependenzrelation.

(vgl. dagegen UD-Bäume: folgende Folien)

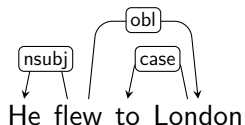
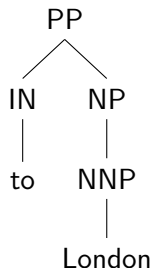
Auxiliar: (you) must leave



Konstituenzanalyse
(links): “must” ist der
Kopf der VP “must
leave”.

Stanford-UD-
Dependenzanalyse
(rechts): “leave” ist der
Kopf von “you must
leave”.

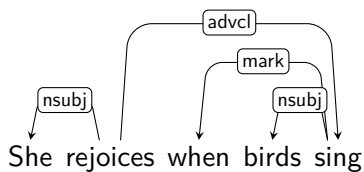
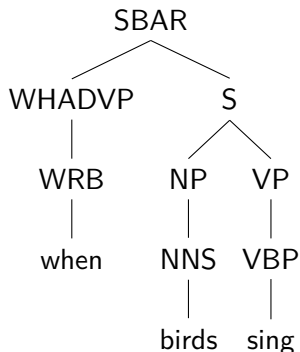
Präpositionalphrase: (He flew) to London



Konstituenzanalyse
(links): "to" ist der Kopf
der PP "to London".

Stanford-UD-
Dependenzanalyse
(rechts): "London" ist
der Kopf von "to
London".

Präpositionalphrase: (*she rejoices*) *when birds sing*



Konstituenzanalyse
(links): "when" ist
der Kopf des
Nebensatzes
"when birds sing"

Stanford-UD-
Dependenzanalyse
(rechts): "sing"
ist der Kopf von
"when birds sing".

Mehr Information nötig wg.:

(a) **Der Löwe schläft aus dem Käfig.*

(b) **Der Löwe springen aus den Käfig.*

- Lexikoneintrag muss mehr Informationen enthalten, um mögliche von nicht möglichen Kombinationen abzugrenzen:
 - für (a): Syntaktischer Rahmen des Verbs (Subkategorisierung)
 - (i) schlafen V, [___]
 - (ii) helfen V, [___ NP]
 - (iii) schenken V, [___ NP, NP]
 - für (b): Kongruenz bzgl. morphologischer Merkmale (Numerus, Person)

- Komplexere Kategorien einführen.
- Lexikoneinträge durch **Merkmale** ergänzen, und Phrasenstruktur-Mechanismus so erweitern, dass zusätzlich zu den Ersetzungsregeln Merkmalskongruenz gegeben sein muss.
- Übergenerierung zulassen, aber durch Wahrscheinlichkeitsmodell un plausible Möglichkeiten niedrig gewichten.

Zum Schluss: Besonders klausurrelevant

- Konstituente: Gruppierung von Wörtern, mit denen die Struktur eines Satzes beschrieben werden kann
- Konstituente: Wort oder Phrase
 - Der Satz ist auch eine Phrase
- Phrase = Kopf plus (optional) dominierte Wörter
- Syntaktische Kategorie des Kopfes entspricht syntaktischer Kategorie der Phrase.
- Konstituententests: Hilfsmittel zur linguistischen Analyse von Konstituentenstrukturen
- Konstituentenstruktur kann rekursiv sein.
- Kategoriensymbole ersetzen die Knoten im Baum → Konstituentenstruktur
- kontextfreie Grammatik: Beschreibungsformalismus zur Generierung und Erkennung wohlgeformter Sätze
- kontextfreie Regeln
- terminale vs. präterminale Kette