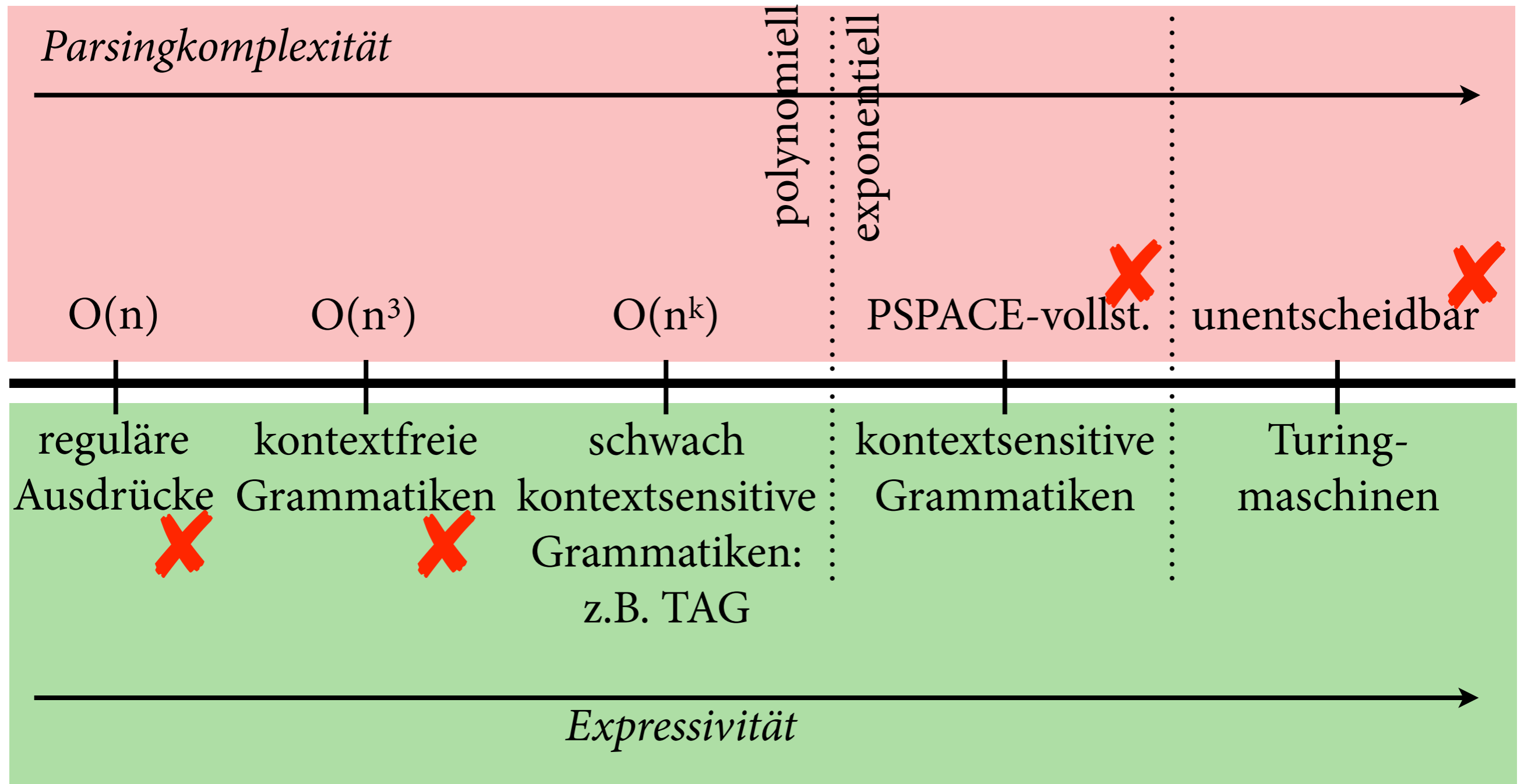


Kombinatorische Kategorialgrammatik

Vorlesung “Grammatikformalismen”
Alexander Koller

12. Mai 2015

Natürliche Sprachen in der Chomsky-Hierarchie



Kategorialgrammatiken

- Phrasenstrukturgrammatik als Gr.theorie:
 - ▶ große Konstituenten aus kleinen Konstituenten zusammensetzen
 - ▶ Beziehung zwischen V und NP: müssen zusammen auftreten, damit man VP daraus bauen kann
- Alternative Gr.theorie: *Kategorialgrammatik*
 - ▶ Ausdrücke haben *Kategorien*
 - ▶ Kategorie gibt an, mit welchen Kategorien Ausdruck kombiniert werden muss, um größeren Ausdruck zu bauen
⇒ Funktor-Argument-Struktur

Beispiel

Hans isst ein Käsebrötchen

Beispiel

Hans isst ein Käsebrötchen
NP

Beispiel

Hans isst ein Käsebrötchen
NP N

Beispiel

$\frac{\text{Hans}}{\text{NP}}$ isst $\frac{\text{ein}}{\text{NP/N}}$ $\frac{\text{Käsebrötchen}}{\text{N}}$

Beispiel

Hans	isst	ein	Käsebrötchen
<hr/>		<hr/>	<hr/>
NP		NP/N	N
		<hr/>	<hr/>
		NP	

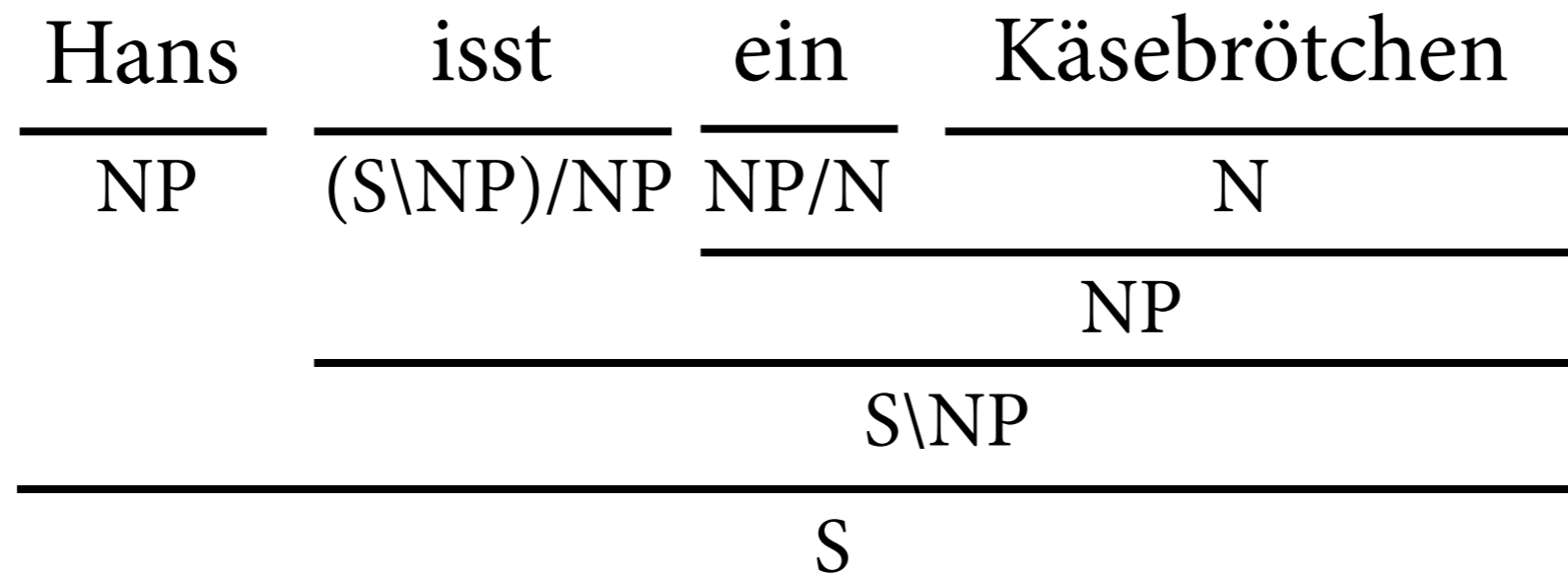
Beispiel

Hans	isst	ein	Käsebrötchen
<hr/>	<hr/>	<hr/>	<hr/>
NP	(S\NP)/NP	NP/N	N
		<hr/>	
		NP	

Beispiel

Hans	isst	ein	Käsebrötchen
<hr/>	<hr/>	<hr/>	<hr/>
NP	(S\NP)/NP	NP/N	N
		<hr/>	
		NP	
	<hr/>		
	S\NP		

Beispiel



Kategorien

- *Atomare* Kategorien: S, NP, N
(mit diesen dreien kommt man schon ziemlich weit)
- *Funktionale* Kategorien: Wenn A und B Kategorien sind, dann auch A/B und $A \setminus B$.
 - ▶ A/B : braucht ein B *zur Rechten*, um ein A zu werden
 - ▶ $A \setminus B$: braucht ein B *zur Linken*, um ein A zu werden
 - ▶ A ist *Funktor*; $/B$ bzw. $\setminus B$ sind *Argumente*.

Geschichte der KG

- Ideen von Kazimierz Ajdukiewicz (1935).
- Formalisiert von Joachim Lambek (1958).
- Sehr beliebt unter holländischen Logikern.
- Wesentliche Version von KG in der CL:
Kombinatorische Kategorialgrammatik (CCG)
(Mark Steedman, späte 1980er).

CCG

- CCG-Grammatik besteht aus:
 - ▶ *Lexikon*: weist jedem Wort eine endliche Menge von Kategorien (= *lexikalische* Kategorien) zu.
 - ▶ *Kombinationsregeln*: Applikations- und Kompositionsregeln, evtl. mit grammatikspezifischen Einschränkungen.
- Grundregel in CCG: Applikation.

$$\frac{X/Y \quad Y}{X} >$$

Vorwärts-Applikation

$$\frac{X \quad Y \setminus X}{Y} <$$

Rückwärts-Applikation

CCG-Ableitungen

- Eine *Ableitung* in CCG ist ein Beweisbaum, so dass gilt:
 - ▶ an den Blättern stehen von links nach rechts lexikalische Kategorien der Wörter w_1, \dots, w_n
 - ▶ die inneren Knoten enthalten die Kategorien, die bei Anwendung von Kombinationsregeln auf Kinder herauskommen.
- Ein String ist in der *Sprache* einer CCG-Grammatik, wenn es eine Ableitung für ihn gibt.

Soweit ich erkennen kann, kann man in OpenCCG nicht angeben, was für eine Kategorie für den ganzen Ausdruck herauskommen soll (z.B. S).

Beispiel

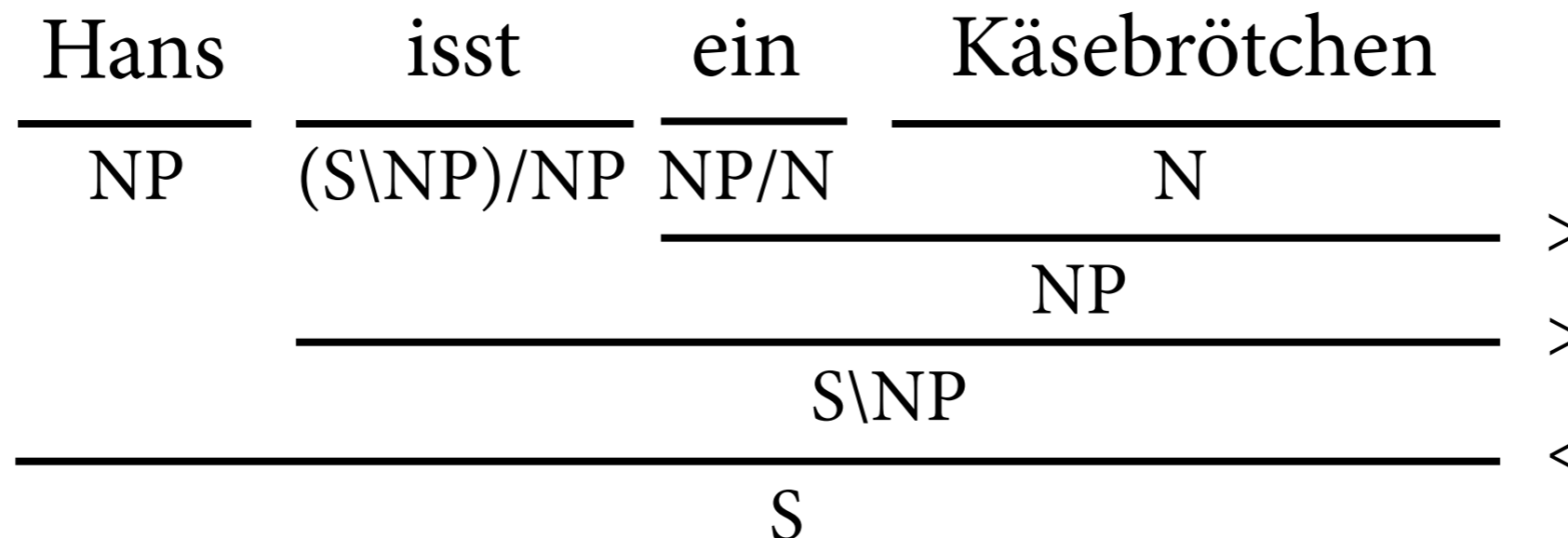
Lexikon:

Hans: NP

ein: NP/N

isst: (S\NP)/NP

Käsebrötchen: N



Beispiel: Modifikation

Lexikon:

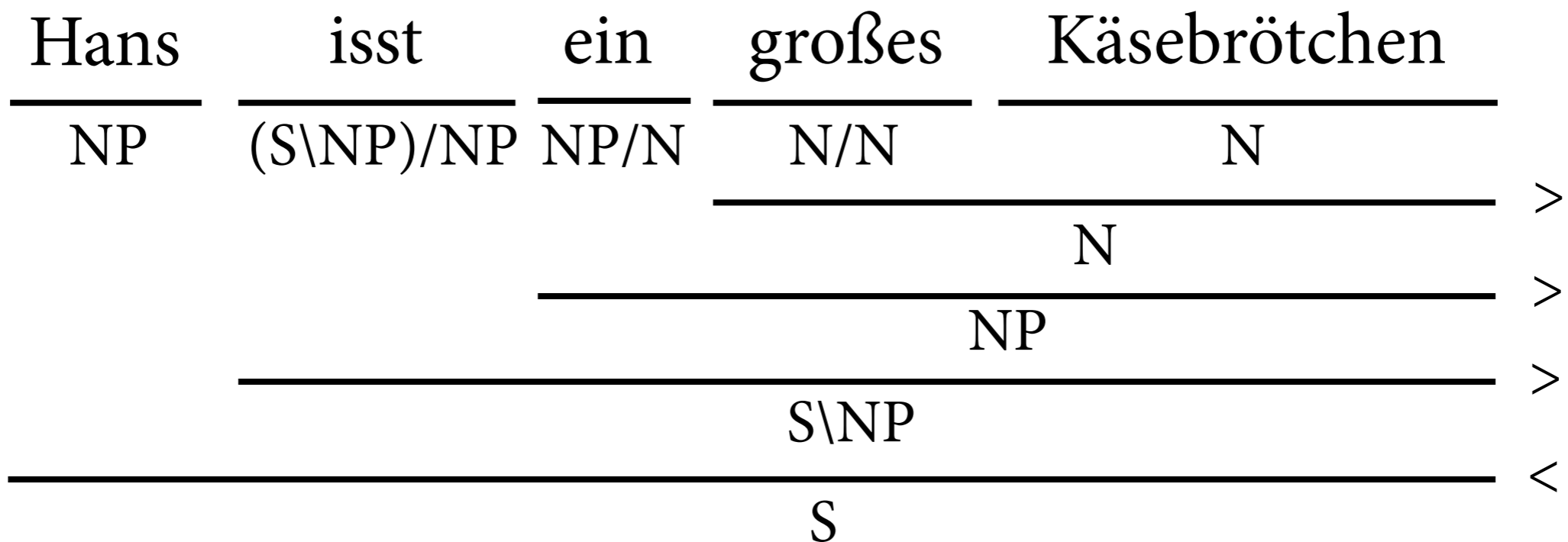
Hans: NP

ein: NP/N

isst: (S\NP)/NP

Käsebrötchen: N

großes: N/N



Beispiel: Raising

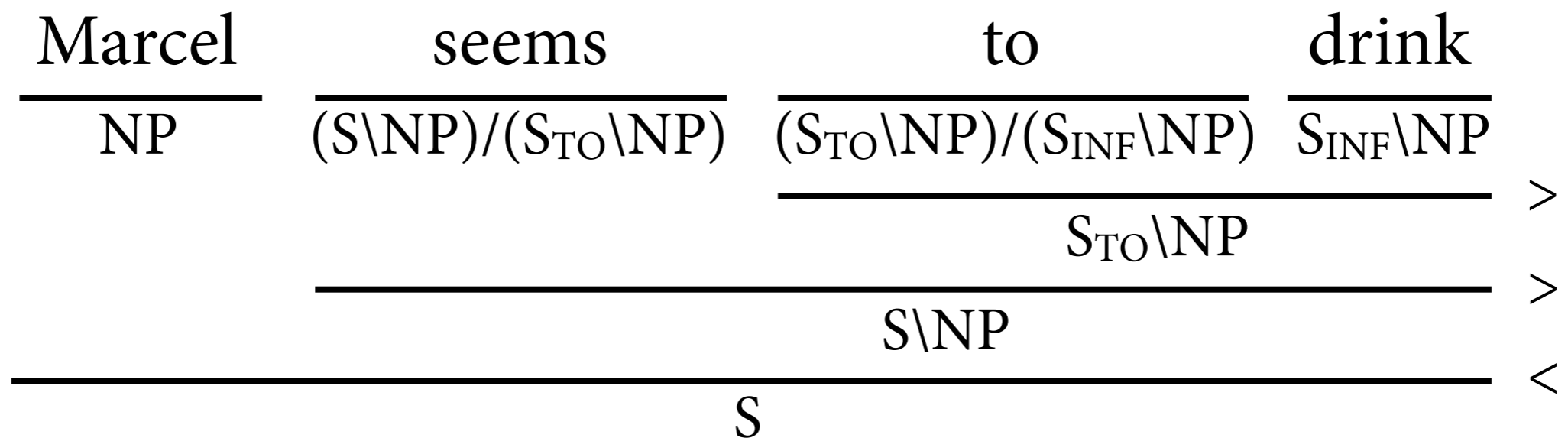
Lexikon:

Marcel: NP

seems: $(S \backslash NP) / (S_{TO} \backslash NP)$

drink: $S_{INF} \backslash NP$

to: $(S_{TO} \backslash NP) / (S_{INF} \backslash NP)$



Satzeinbettung

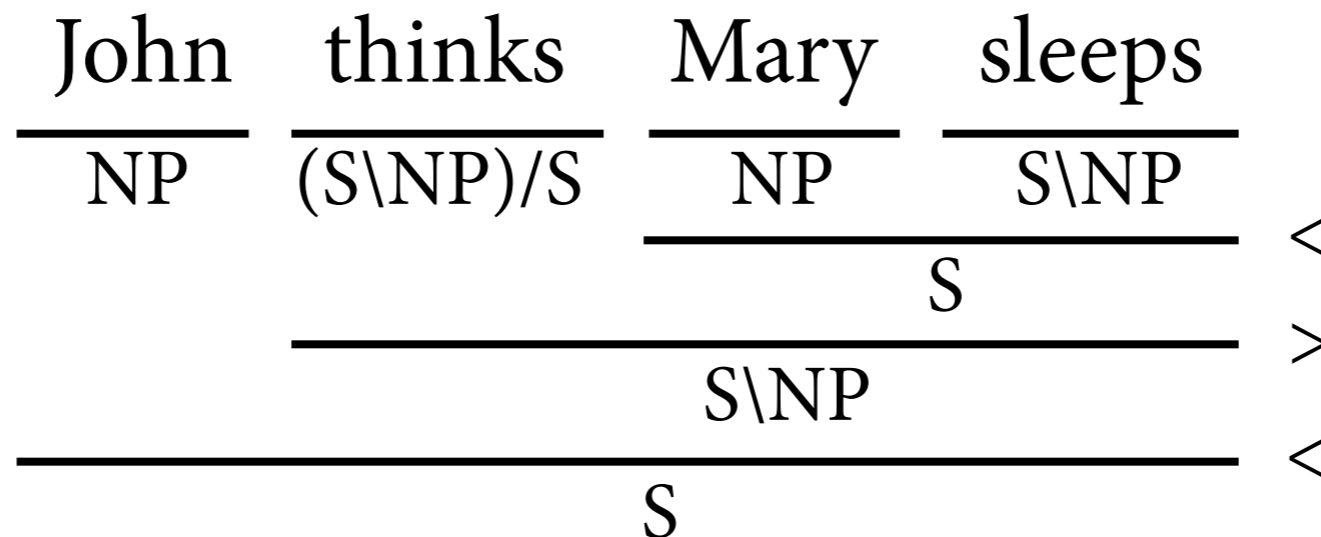
Lexikon:

John: NP

thinks: (S\NP)/S

Mary: NP

sleeps: S\NP



Fernabhängigkeiten

Lexikon:

John: NP

Mary: NP

who: NP_{wh}

think: (S\NP)/S

likes: (S\NP)/NP

does: (S\NP_{wh})/(S/NP)

<u>who</u>	<u>does</u>	<u>John</u>	<u>think</u>	<u>Mary</u>	<u>likes</u>
NP _{wh}	(S\NP _{wh})/(S/NP)	NP	(S\NP)/S	NP	(S\NP)/NP

Fernabhängigkeiten

Lexikon:

John: NP

Mary: NP

who: NP_{wh}

think: (S\NP)/S

likes: (S\NP)/NP

does: (S\NP_{wh})/(S/NP)

<u>who</u>	<u>does</u>	<u>John</u>	<u>think</u>	<u>Mary</u>	<u>likes</u>
NP _{wh}	(S\NP _{wh})/(S/NP)	NP	(S\NP)/S	NP	(S\NP)/NP

?

Type-Raising

- In CCG kann man die Kategorie eines Ausdrucks *anheben* und so die Funktor-Argument-Richtung umkehren.

$$\frac{X}{Y/(Y \setminus X)} >^T$$

$$\frac{X}{Y \setminus (Y/X)} <^T$$

- Macht für Applikation im wesentlichen keinen Unterschied:

$$\frac{A \quad B \setminus A}{B} >$$

$$\frac{\frac{A}{B/(B \setminus A)} >^T \quad B \setminus A}{B} <$$

Komposition

- Mit den *Kompositions*-Regeln kann man zwei Kategorien verbinden und dabei noch Argumente “übriglassen” und weitergeben.

Harmonische Komposition:

$$\frac{X/Y \quad Y/Z}{X/Z} >B$$

$$\frac{Y\backslash Z \quad X\backslash Y}{X\backslash Z} <B$$

Crossed Composition:

$$\frac{X/Y \quad Y\backslash Z}{X\backslash Z} >Bx$$

$$\frac{Y/Z \quad X\backslash Y}{X/Z} <Bx$$

Type-Raising + Komposition

- Durch die Kombination von Type-Raising und Komposition kann man Argumente von Kategorien “zu früh” abbinden.

$$\frac{\frac{\text{Mary}}{\text{NP}}}{\text{S}/(\text{S}\backslash\text{NP})} > \text{T} \quad \frac{\text{likes}}{(\text{S}\backslash\text{NP})/\text{NP}} > \text{B}$$

$$\text{S}/\text{NP}$$

Fernabhängigkeiten

Lexikon:

John: NP

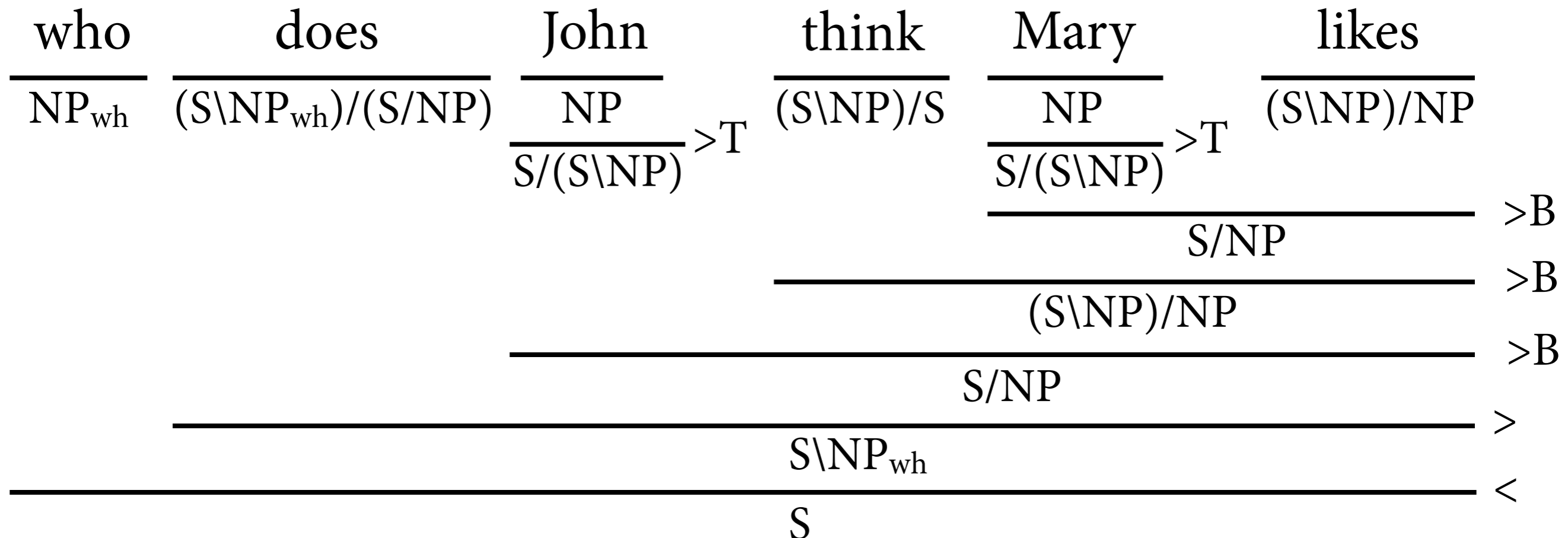
Mary: NP

who: NP_{wh}

think: (S\NP)/S

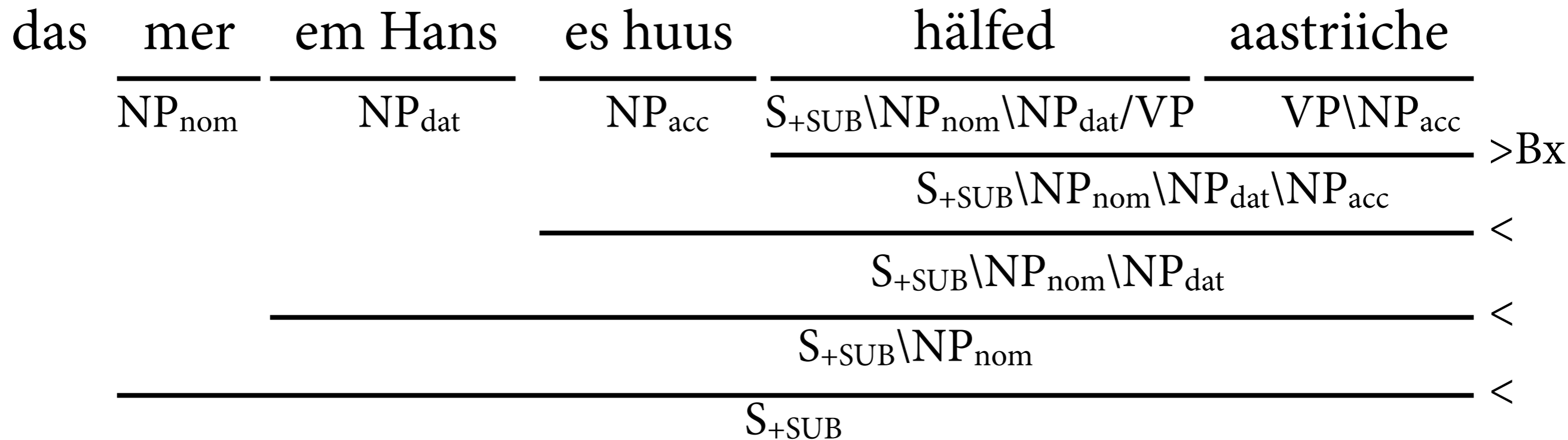
likes: (S\NP)/NP

does: (S\NP_{wh})/(S/NP)



Schweizerdeutsch

- Crossed composition erlaubt Modellierung von cross-serial dependencies:



(“VP” = Abkürzung für $S \setminus NP_{nom}$)

Kompositionen höheren Grads

- Mit normaler Komposition kann man ein einziges Argument weitergeben:

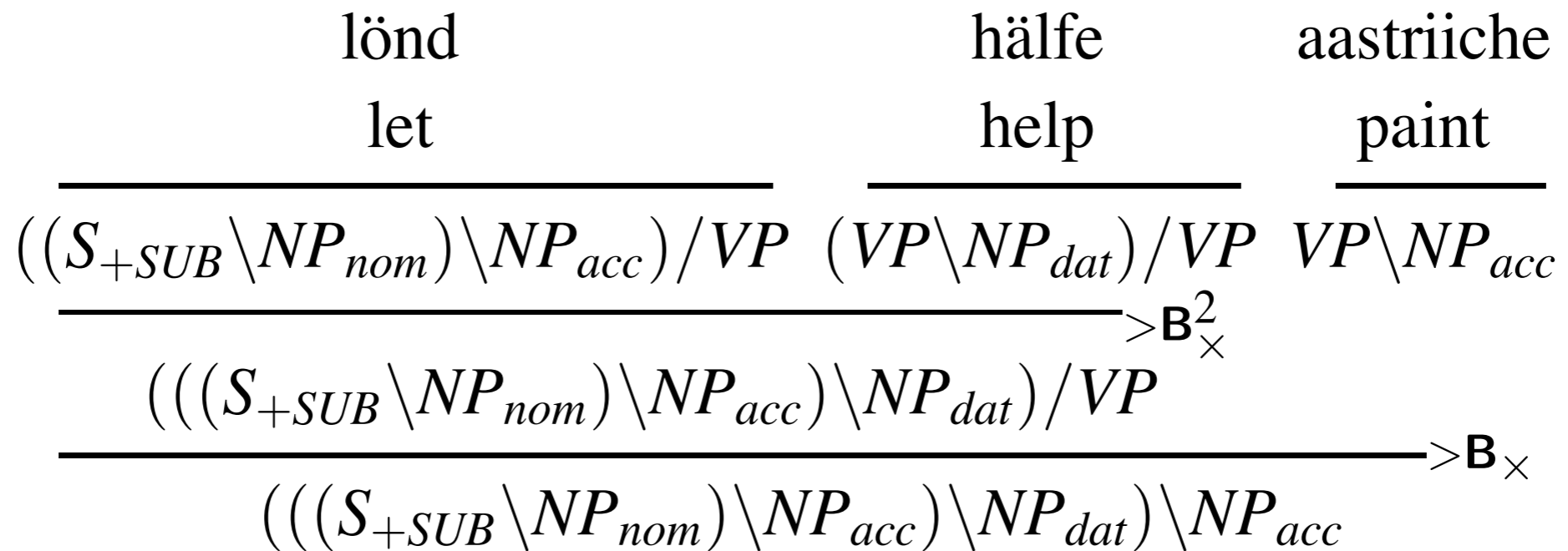
$$\frac{X/Y \quad Y/Z}{X/Z} >B \qquad \frac{Y\backslash Z \quad X\backslash Y}{X\backslash Z} <B$$

- Für manche komplizierten Konstruktionen möchte man mehr als ein Argument weitergeben:

$$\frac{X/Y \quad Y/Z/W}{X/Z/W} >B^2 \qquad \frac{X/Y \quad Y/Z/W/V}{X/Z/W/V} >B^3$$

(usw. für höhere $>B^n$; $<B^n$, $>B^n_x$, $<B^n_x$ analog)

Mehr Schweizerdeutsch



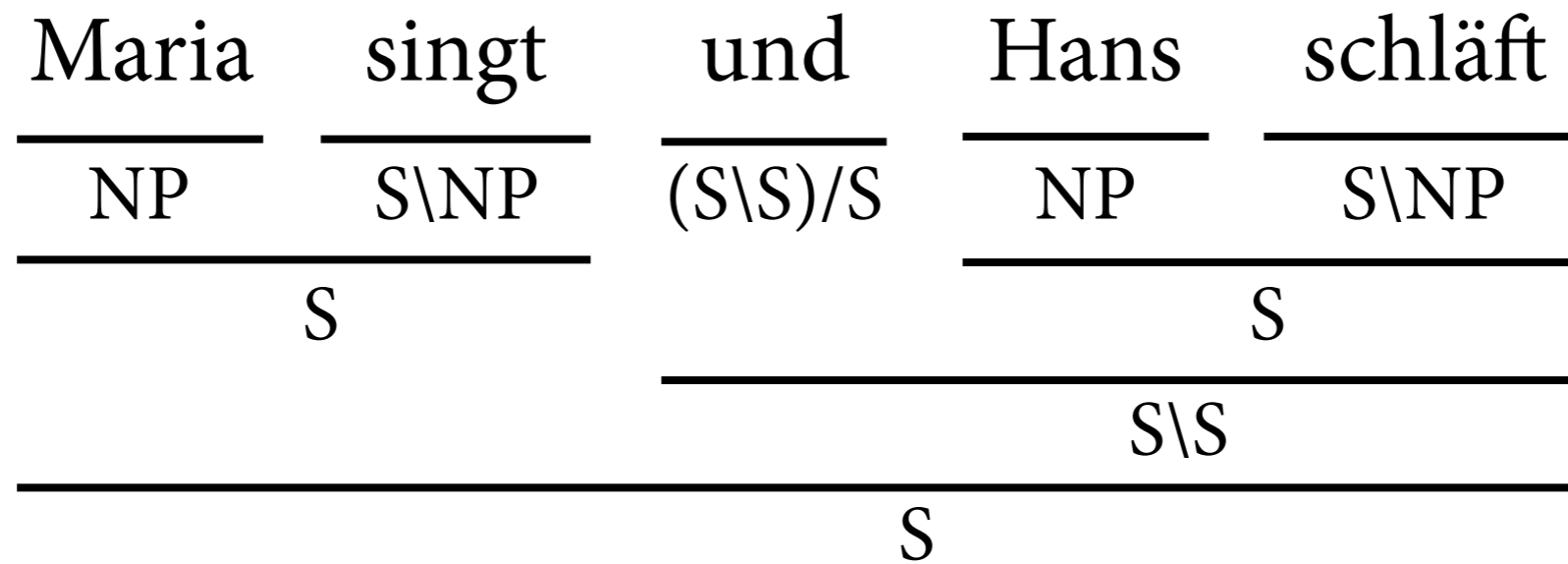
(“VP” = Abkürzung für $S \setminus NP_{nom}$)

Koordination

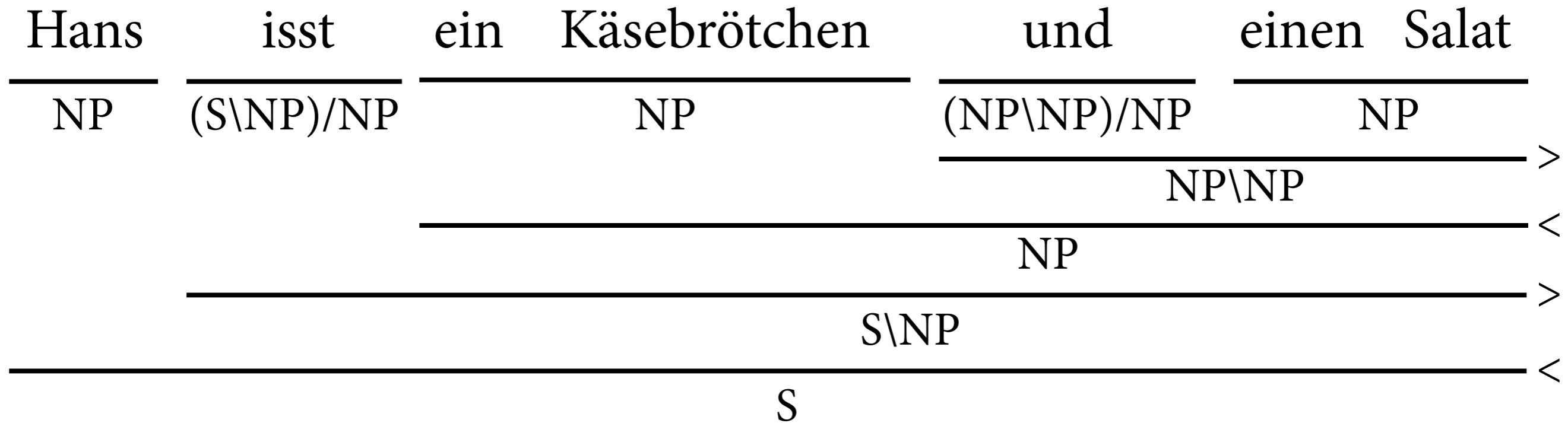
- Ein Highlight von CCG ist die Behandlung von Koordination.
- In anderen Grammatikformalismen schwierig, weil gleiche Konjunktion (“und”, “oder”) viele Kategorien koordinieren kann.
- CCG: Lexikoneintrag für Konjunktion sind alle Instanzen des folgenden Schemas:

und: $(X \backslash X) / X$

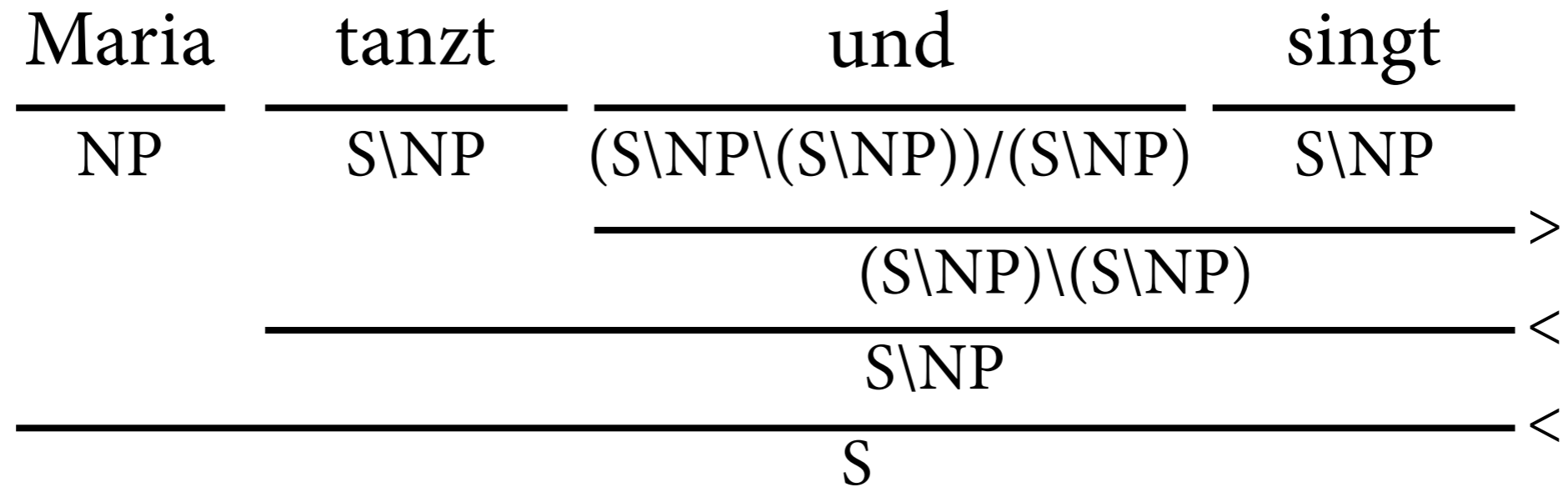
Koordination



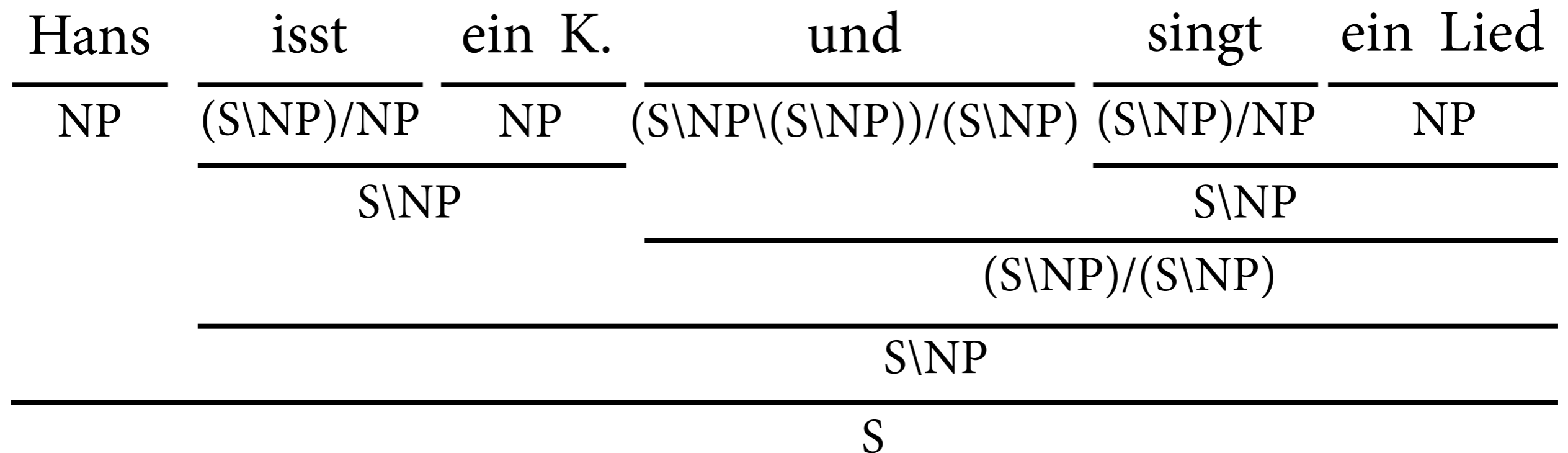
Koordination



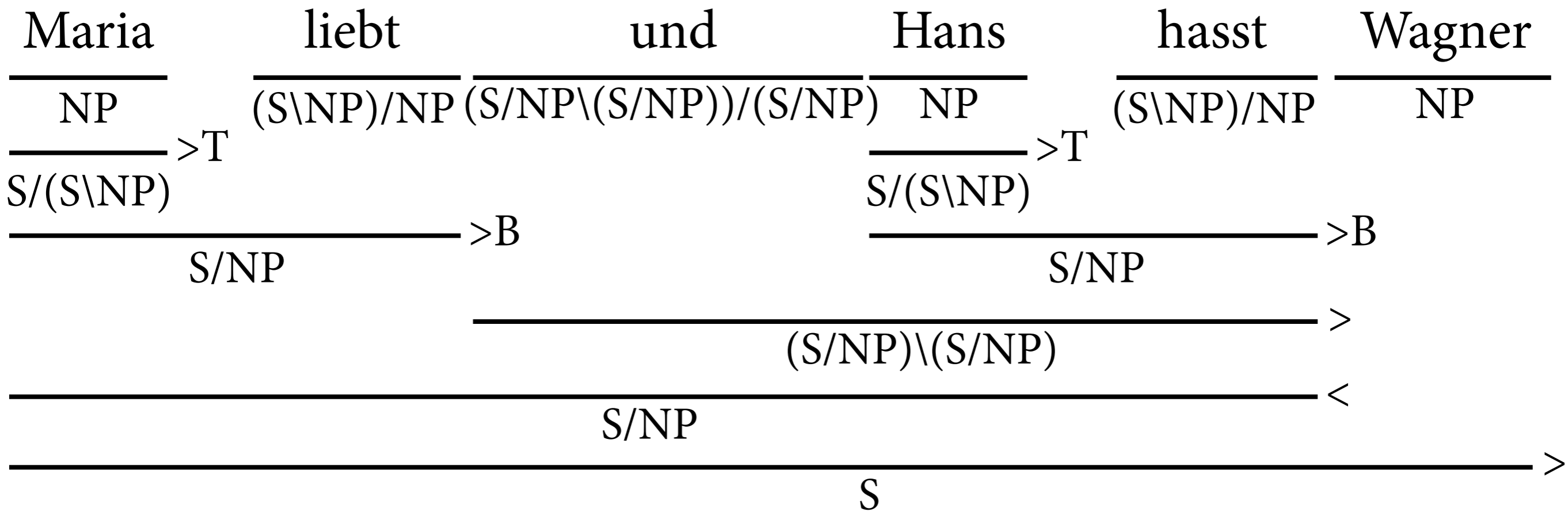
Koordination



Koordination



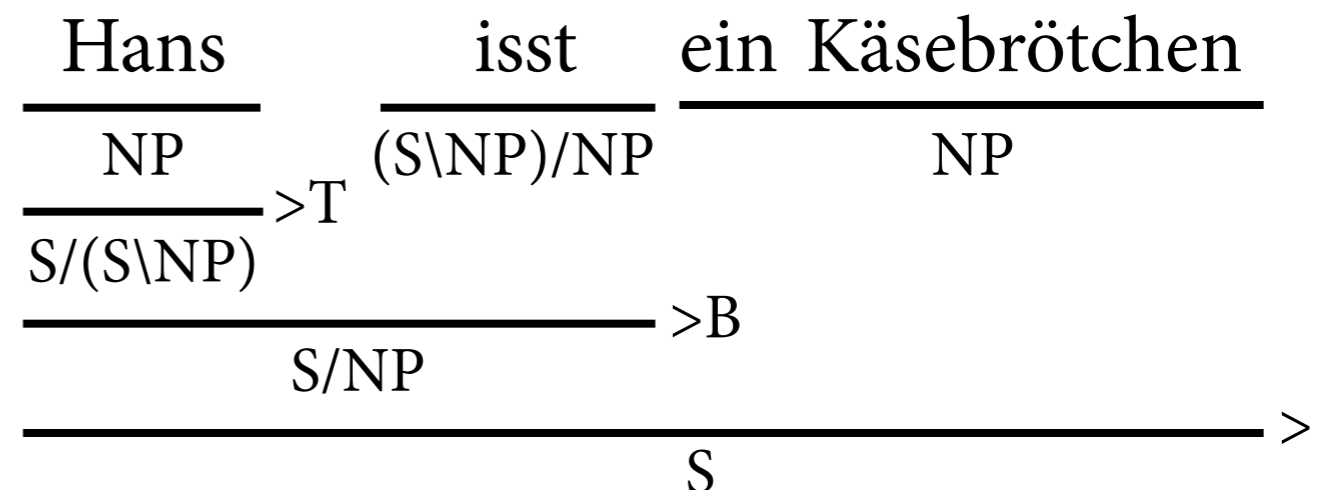
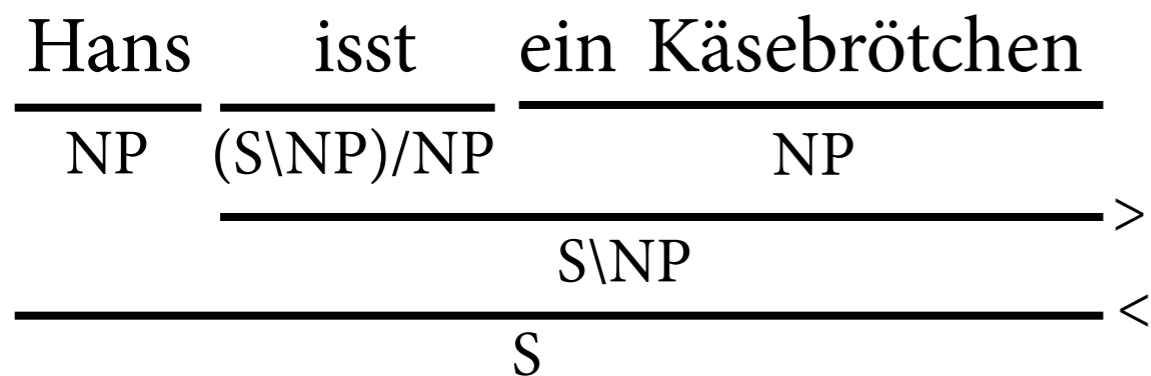
Koordination



(sog. "right node raising")

Spurious ambiguities

- CCG erlaubt “non-constituent coordination”, indem (wegen Komposition) viele Ausdrücke als Konstituenten zählen können.
- Preis, den man dafür bezahlt: Manche Sätze haben viel mehr Ableitungen, als man denkt. (Kann aber nützlich sein, z.B. für Informationsstruktur.)



Linguistische Grundprinzipien

- Vermeide lexikalische Ambiguität soweit möglich.
 - ▶ Verwendung der gleichen lexikalischen Kategorie in verschiedenen Kontexten durch Kombinatoren.
 - ▶ Allgemein viel weniger lexikalische, viel mehr derivationelle Ambiguität als in TAG.
- TAG-artige “extended domain of locality”: schon Lexikoneintrag legt Argumente fest.
- Ziel ist volle Lexikalisierung
(aber siehe Kuhlmann et al. 2015).

Zusammenfassung

- Kategorialgrammatik: Grammatiktheorie, die Funktor-Argument-Struktur von syntaktischen Valenzen annimmt.
- Haupt-Formalismus in der CL: CCG.
 - ▶ Applikation; Komposition: harmonisch, gekreuzt, höheren Grades; Type-Raising
- Behandlung von Fernabhängigkeiten, cross-serial dependencies, Koordination.