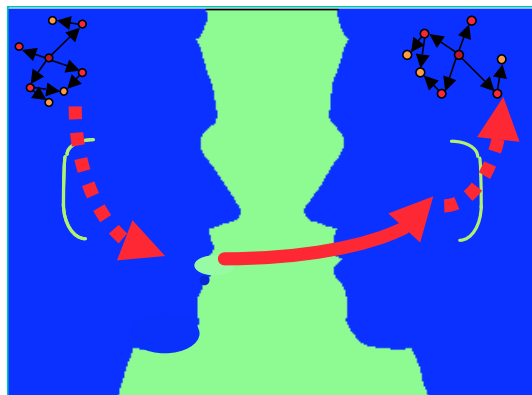

Vorlesung Grammatikformalismen:

Die Rolle der Grammatik

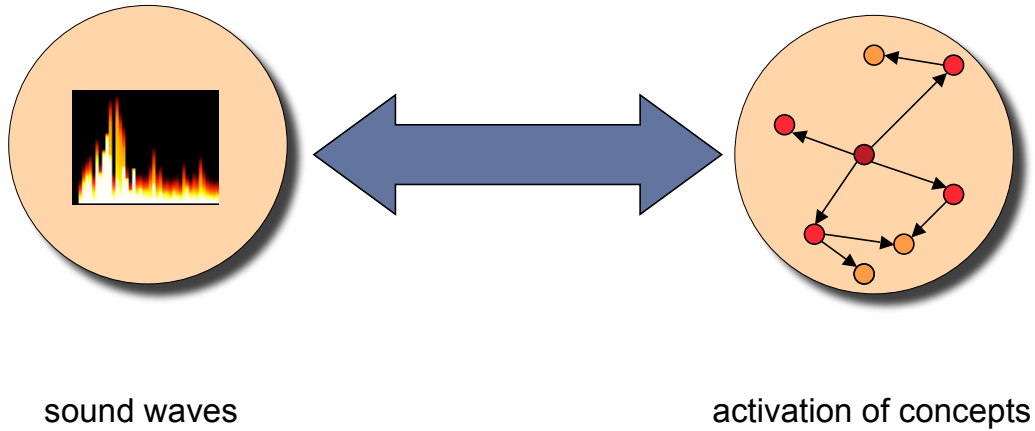
Ersetzungsregelgrammatiken

Hans Uszkoreit

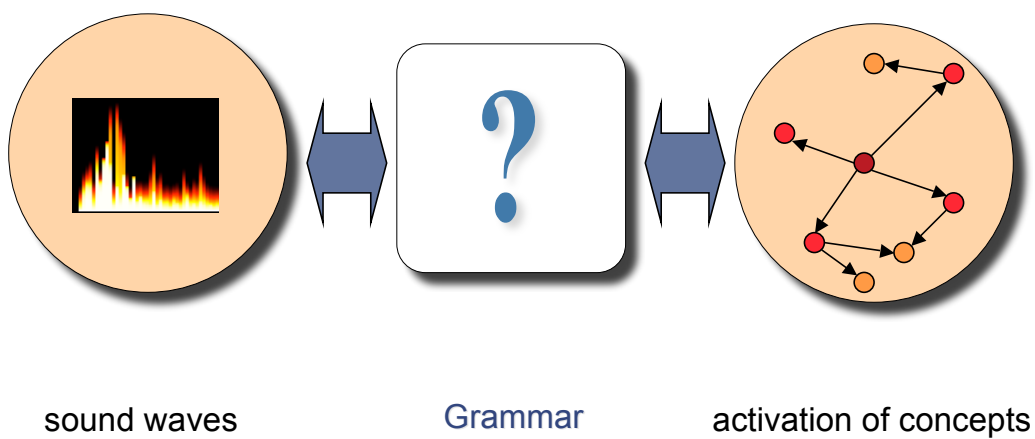
WHAT HAPPENS IN BETWEEN?



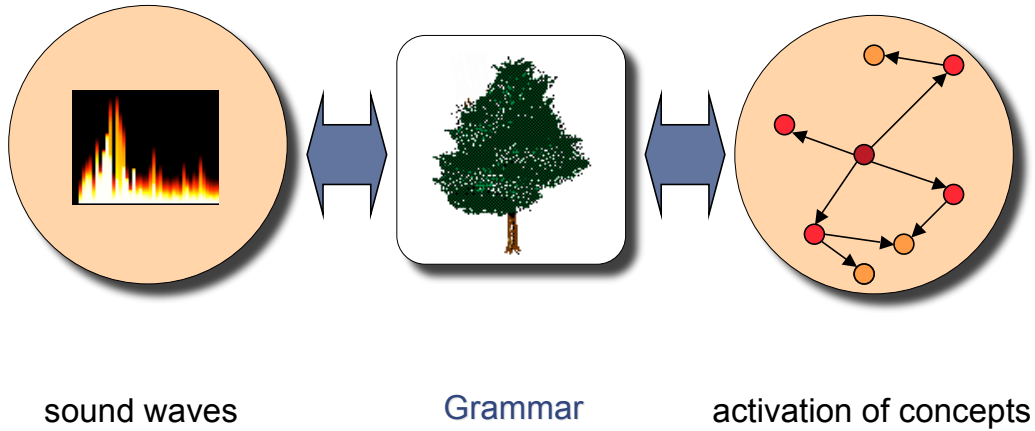
WHAT HAPPENS IN BETWEEN?



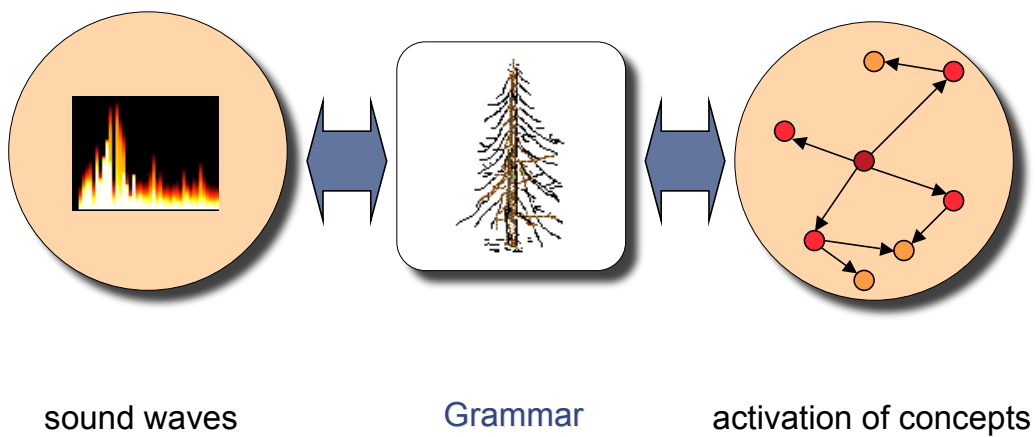
WHAT HAPPENS IN BETWEEN?



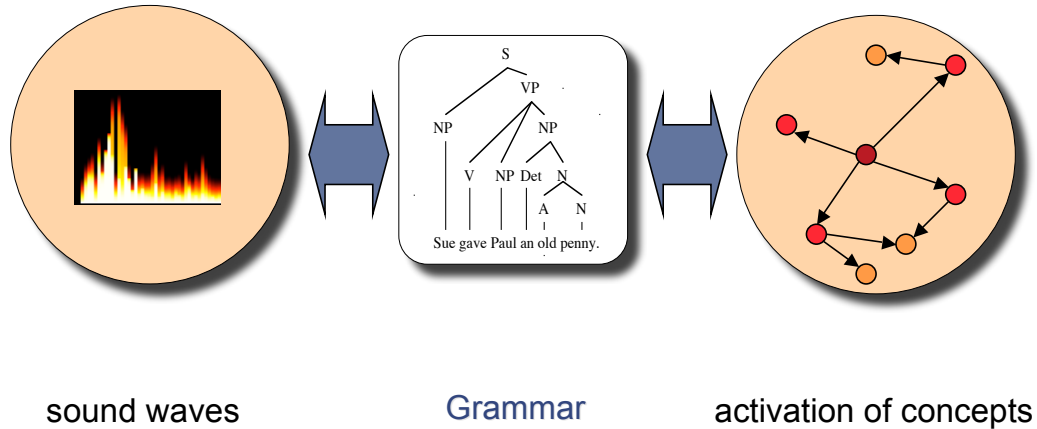
WHAT HAPPENS IN BETWEEN?



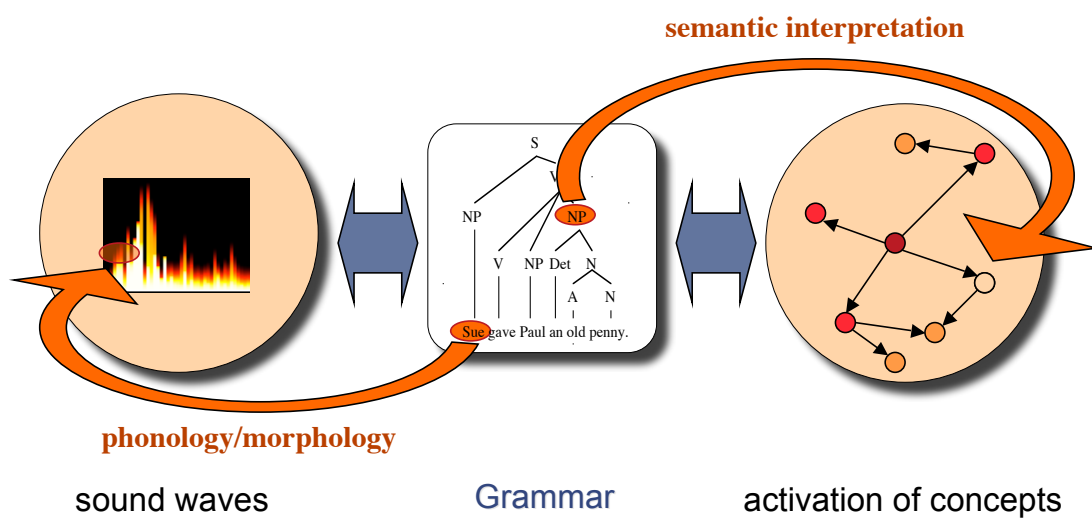
WHAT HAPPENS IN BETWEEN?



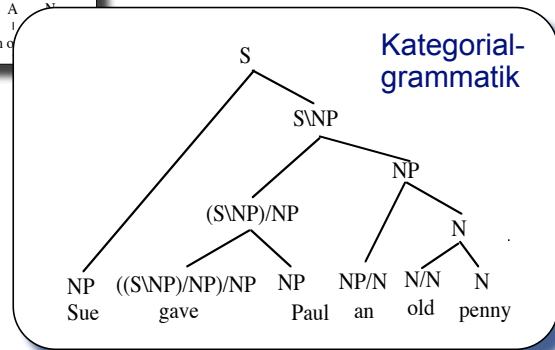
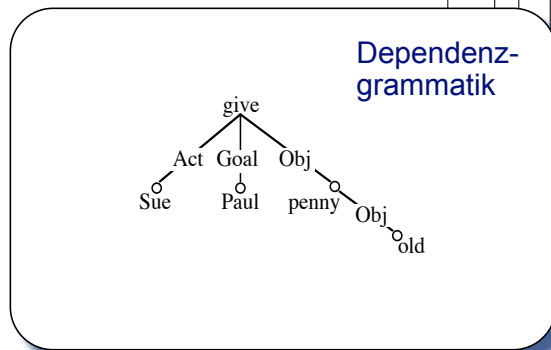
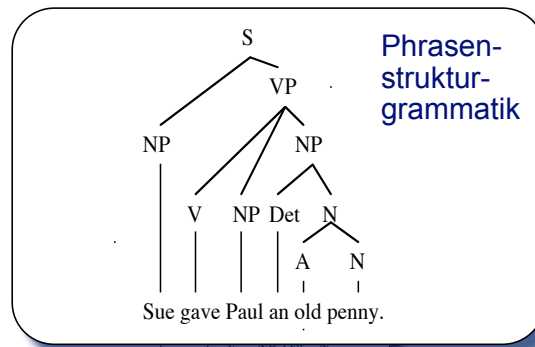
WHAT HAPPENS IN BETWEEN?



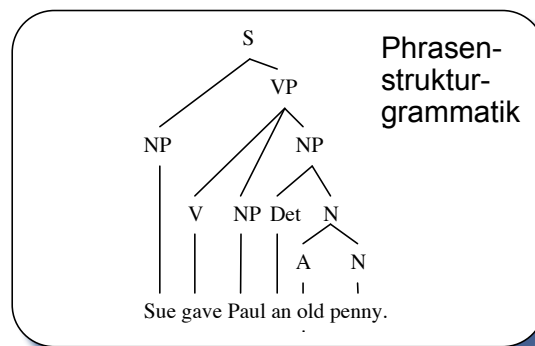
WHAT HAPPENS IN BETWEEN?



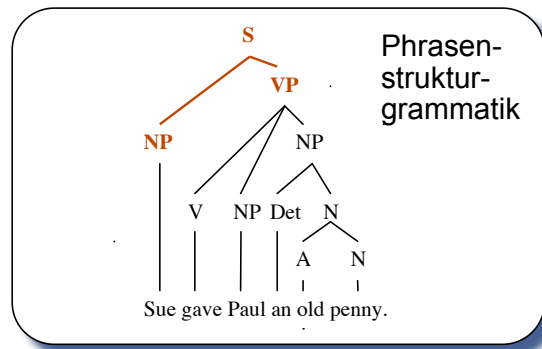
THREE TRADITIONS ☒



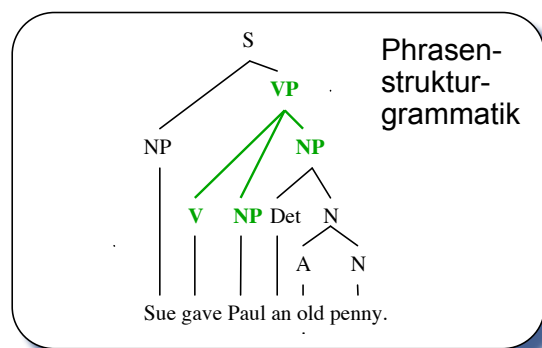
Grammatik



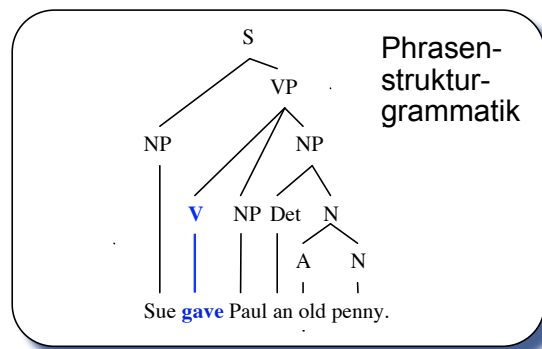
$S \rightarrow NP VP$



$S \rightarrow NP VP$



$S \rightarrow NP VP$
 $VP \rightarrow V NP NP$



$S \rightarrow NP VP$
 $VP \rightarrow V NP NP$

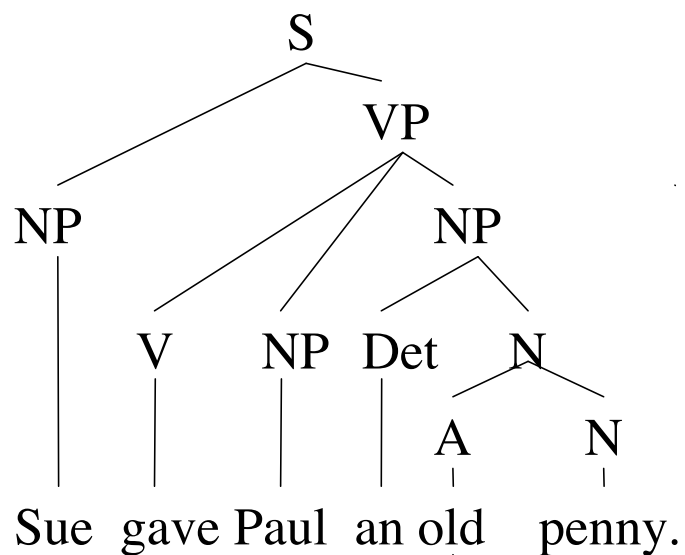
$V \rightarrow \text{gave}$

Symbolkonventionen

	Einzelexymbole	Ketten
nichtterminale	A, B, C, , X, Y, Z
terminale	a, b, c, , x, y, z
unspezifiziert	a, b, g, , f, c, y, w
Startsymbol	S	
die leere Kette	e	
Zahlen	..., i, j, k, l, m, n, ...	

Why Syntax

- Einen Hund hat dieser Mann gebissen.
- Ein Hund hat diesen Mann gebissen.
- This man has bitten a dog.
- A dog this man has bitten.
- A dog has bitten this man.
- Peter versprach Paul, die Akten zu bearbeiten.
- Peter überredete Paul, die Akten zu bearbeiten



Eine Sprache L über einem Alphabet (Vokabular) S ist eine Teilmenge von S^* .

Eine formale Grammatik G_L für eine Sprache L ist ein
Quadrupel $(V_N, V_T, \{S\}, P)$.

V_N - nichtterminales Vokabular, Hilfsalphabet

V_T - terminales Vokabular

$$(V_T \cap V_N = \emptyset, \quad L \subseteq V_T^*, \quad V = V_T \cup V_N)$$

$\{S\}$ - Einermenge mit dem Startsymbol, Axiomenmenge

P - Menge der Produktionen, Regelmenge

Menge von Regeln der Form $w_1 f w_2 \rightarrow w_1 y w_2$

meist geschrieben als $f \rightarrow y$

Ableitung

Relation "folgt":

Wenn $G = (V_N, V_T, \{S\}, P)$, dann folgt y aus f nach G genau dann, wenn es Ketten f_1, f_2, c, w gibt, so daß $f = f_1 c f_2$ und $y = f_1 w f_2$ und $c \rightarrow w \in P$.

Notation: $\phi \Rightarrow_G \psi$

Ableitung:

Eine Folge von Ketten f_1, f_1, \dots, f_n ist eine Ableitung nach G g.d.w. $\phi_i \Rightarrow_G \phi_{i+1}$
für alle $i, 1 \leq i \leq n$.

Wenn es eine Ableitung nach G von f nach y gibt, können wir das

so schreiben: $\phi \xRightarrow[G]{*} \psi$

Die Relation ableitbar ist transitiv und wird als reflexiv definiert.

Die erzeugte Sprache

Die Sprache L: Eine Kette w ist nach G_L in der Sprache L g.d.w. die folgenden drei Bedingungen erfüllt sind:

1. $w \in V_T^*$
2. $S \xRightarrow[G]{*} w$
3. Es gibt kein c , so daß $w \xRightarrow[G]{*} c$ und $w \neq c$.

Man sagt auch G_L erzeugt die Sprache L . Die von G erzeugte Sprache L wird auch als $L(G)$ geschrieben.

Schwache Äquivalenz: Zwei Grammatiken G_1 und G_2 sind schwach äquivalent, wenn sie dieselbe Sprache erzeugen.

Typen von Grammatiken

Typ 0 (unbeschränkte Ersetzungsregelsysteme):

Jede Grammatik, die die Definition einer formalen Grammatik erfüllt, ist vom Typ 0.

Typ 1 (kontextsensitive Grammatiken):

Jede Produktion hat die Form $f Ay \rightarrow f wy$, wobei $A \in V_T$, $w \neq \epsilon$.

Typ 2 (kontextfreie Grammatiken):

Jede Produktion hat die Form $A \rightarrow w$, wobei $w \neq \epsilon$.

Typ 3 (reguläre Grammatiken):

Jede Produktion hat die Form $A \rightarrow x B$ oder $A \rightarrow x$, wobei $x \neq \epsilon$.

S
 NP VP
 DET ADJ N VP
 DET ADJ N V NP
 DET ADJ N V DET ADJ N

.
 .
 .

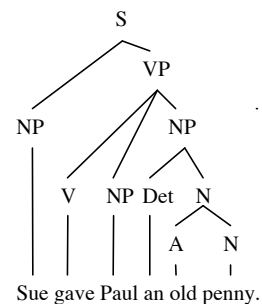
ein kleines Mädchen sucht ein kleines Mädchen

Bäume

der Begriff des syntaktischen Strukturbaums

kodierte Information

1. Die hierarchische Organisation der Teile eines Satzes in Konstituenten.
2. Der Zugehörigkeit jeder Konstituente in eine syntaktische (linguistische) Klasse (Kategorie).
3. Die lineare Abfolge der Konstituenten.

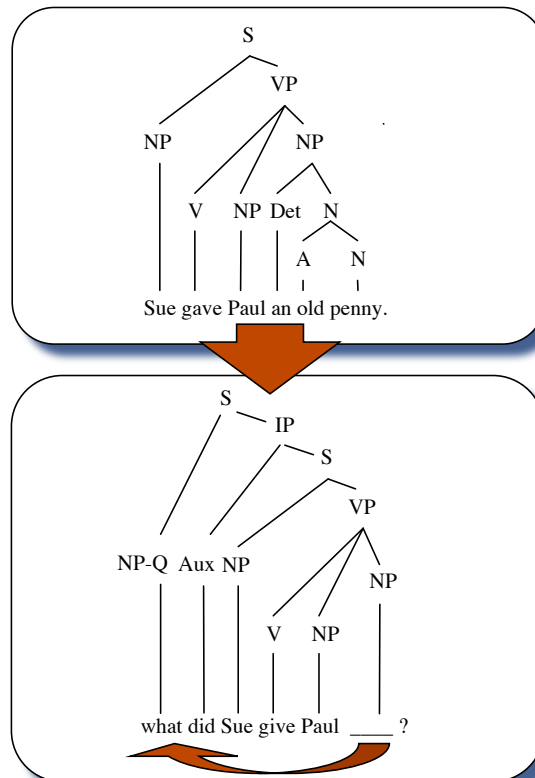


Konstituentenstrukturbaum: Quintupel (N, Q, D, P, L)

Bedingungen:

Vorlesung Einführung in die CL WS 04/05 © 2004 Hans Uszkoreit

Vorlesung Einführung in die CL WS 04/05 © 2004 Hans Uszkoreit



Transformations-
grammatik

Unifikationsgrammatik

