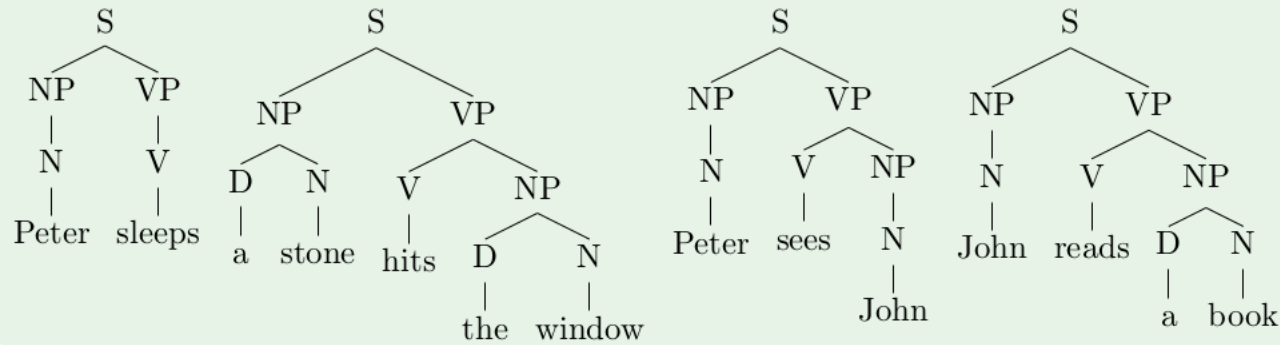


- Baumbanktraining

- ▶ benötigt eine manuell erstellte Baumbank.
- ▶ Die Regelhäufigkeiten werden gezählt.
- ▶ Die Regelwahrscheinlichkeiten werden mit relativen Häufigkeiten geschätzt:

$$p(A \rightarrow \alpha) = \frac{f_{A \rightarrow \alpha}}{\sum_{\beta} f_{A \rightarrow \beta}}$$



Extract grammar rules, their frequency, and $p(\text{rule})$

$S \rightarrow \text{NP VP}$, $f = 4$, $p = 4 / 4 = 1$

$\text{NP} \rightarrow n$, $f = 4$, $p = f(\text{NP} \rightarrow n) / f(\text{NP} \rightarrow n) + f(\text{NP} \rightarrow \dots) + \dots$

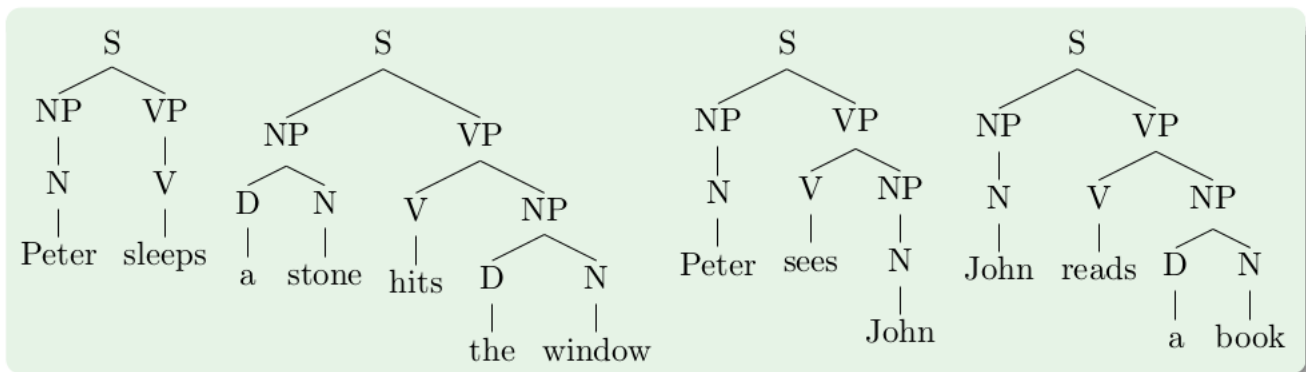
$N \rightarrow \text{peter}$, $p = f(N \rightarrow \text{peter}) / f(N \rightarrow \text{peter}) + f(N \rightarrow \dots) + \dots$

- Baumbanktraining

- ▶ benötigt eine manuell erstellte Baumbank.
- ▶ Die Regelhäufigkeiten werden gezählt.
- ▶ Die Regelwahrscheinlichkeiten werden mit relativen Häufigkeiten geschätzt:

$$p(A \rightarrow \alpha) = \frac{f_{A \rightarrow \alpha}}{\sum_{\beta} f_{A \rightarrow \beta}}$$

Baumbank-Training



Extraktion der Grammatikregeln und Regelhäufigkeiten:

S → NP VP	4	1
VP → V NP	3	0.75
VP → V	1	0.25
NP → D N	3	0.43
NP → N	4	0.57

D → a	2	0.67
D → the	1	0.33
V → sleeps	1	0.25
V → hits	1	0.25
V → sees	1	0.25
V → reads	1	0.25

N → Peter	2	0.29
N → John	2	0.29
N → stone	1	0.14
N → window	1	0.14
N → book	1	0.14

Nachteil von Baumbank-Training?

Teuer

Welche Alternativ kann man verwenden?

-unüberwachtes Training → EM

- Baumbanken müssen manuell erstellt werden.
- Der Aufbau einer Baumbank ist mühsam und teuer.
- Was tun, wenn keine Baumbank zur Verfügung steht?

