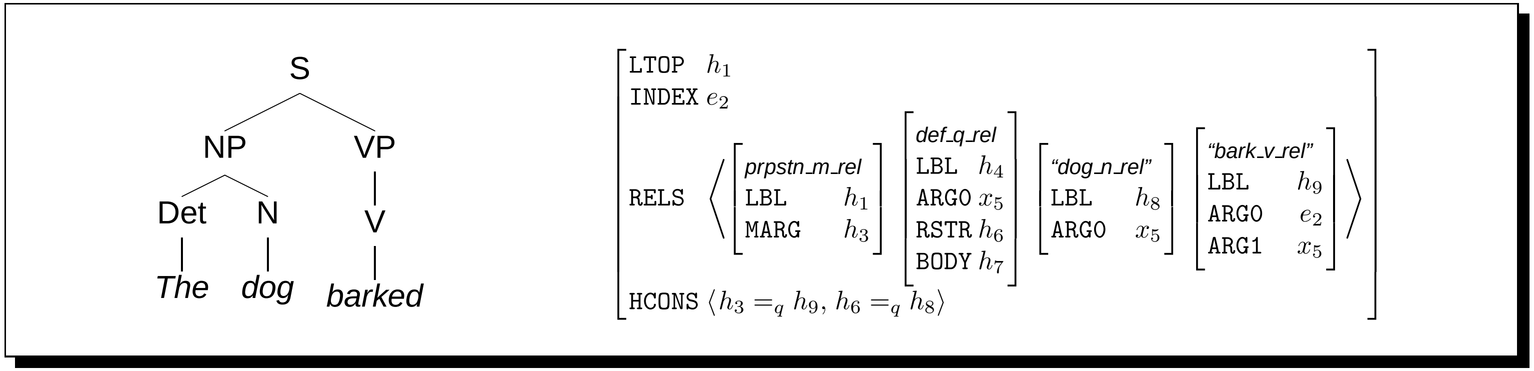


Computational Linguistics (INF2820 — Stochastics)

$$P(S \longrightarrow NP\ VP) = 1.0;\ \ P(NP \longrightarrow Det\ N) = 0.6$$

Stephan Oepen

Universitetet i Oslo

oe@ifi.uio.no

Ambiguity Resolution Remains a (Major) Challenge

The Problem

- With broad-coverage grammars, even moderately complex sentences typically have multiple analyses (tens or hundreds, rarely thousands);
- unlike in grammar writing, exhaustive parsing is useless for applications;
- identifying the 'right' (intended) analysis is an 'AI-complete' problem;
- inclusion of (non-grammatical) sortal constraints is generally undesirable.

Typical Approaches

- Design and use statistical models to select among competing analyses;
- for string S, some analyses T_i are more or less likely: maximize $P(T_i|S)$;

$\rightarrow$ Probabilistic Context Free Grammar (PCFG) is a CFG plus probabilities.

Probability Theory and Linguistics?

The most important questions of life are, for the most part, really only questions of probability. (Pierre-Simon Laplace, 1812)

Probability Theory and Linguistics?

The most important questions of life are, for the most part, really only questions of probability. (Pierre-Simon Laplace, 1812)

Special wards in lunatic asylums could well be populated with mathematicians who have attempted to predict random events from finite data samples. (Richard A. Epstein, 1977)

Probability Theory and Linguistics?

The most important questions of life are, for the most part, really only questions of probability. (Pierre-Simon Laplace, 1812)

Special wards in lunatic asylums could well be populated with mathematicians who have attempted to predict random events from finite data samples. (Richard A. Epstein, 1977)

But it must be recognized that the notion 'probability' of a sentence is an entirely useless one, under any known interpretation of this term. (Noam Chomsky, 1969)

Probability Theory and Linguistics?

The most important questions of life are, for the most part, really only questions of probability. (Pierre-Simon Laplace, 1812)

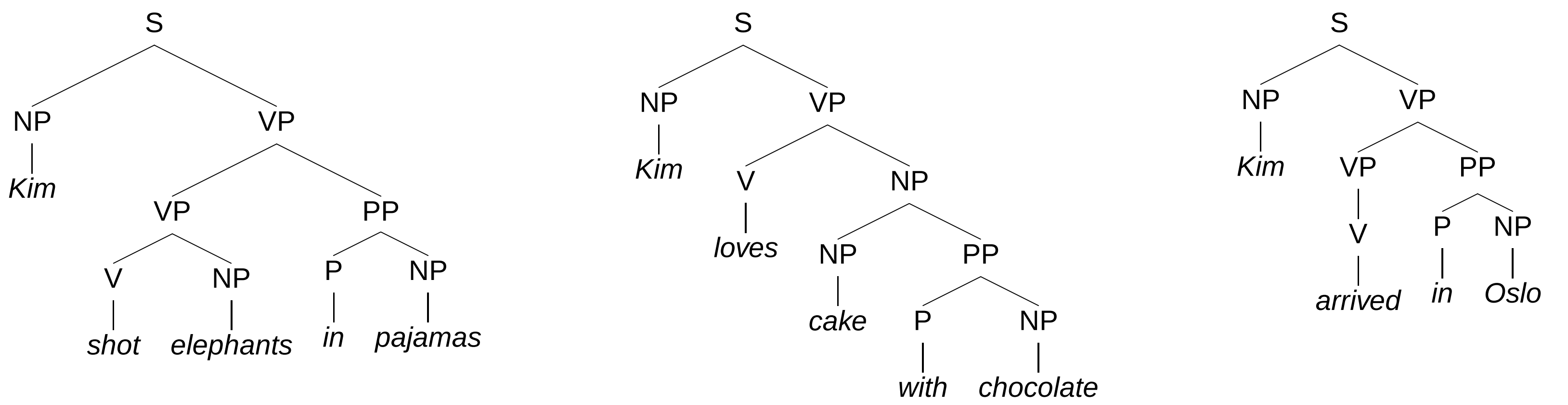
Special wards in lunatic asylums could well be populated with mathematicians who have attempted to predict random events from finite data samples. (Richard A. Epstein, 1977)

But it must be recognized that the notion 'probability' of a sentence is an entirely useless one, under any known interpretation of this term. (Noam Chomsky, 1969)

Every time I fire a linguist, system performance improves. (Fredrick Jelinek, 1980s)

Basics of Probability Theory

Probabilistic Context-Free Grammars (Simplified)



P(RHS|LHS)	CFG Rule
	S → NP VP
	VP → VP PP
	VP → V NP
	PP → P NP
	NP → NP PP
	VP → V

- Estimate rule probability from observed distribution;

→ conditional probabilities:

$$P(\text{RHS}|\text{LHS}) = \frac{C(\text{LHS}, \text{RHS})}{C(\text{LHS})}$$

Formally: Probabilistic Context-Free Grammars

- Formally, a *context-free grammar* (CFG) is a quadruple: $\langle C, \Sigma, P, S \rangle$

...

- P is a set of category rewrite rules (aka *productions*), each with a conditional probability P(RHS|LHS), e.g.

...

NP $\rightarrow$ Kim [0.6]
NP $\rightarrow$ snow [0.4]

...

- for each rule '$\alpha \rightarrow \beta_1, \beta_2, ..., \beta_n$' $\in P$: $\alpha \in C$ and $\beta_i \in C \cup \Sigma$; $1 \leq i \leq n$;

...

- for each $\alpha \in C$, the probabilities of all rules R '$\alpha \rightarrow$ ...' must sum to 1.