

---

# Efficient Estimation of a Log-Linear Parsing Model

Stephen Clark  
Computing Laboratory  
University of Oxford

July 2006

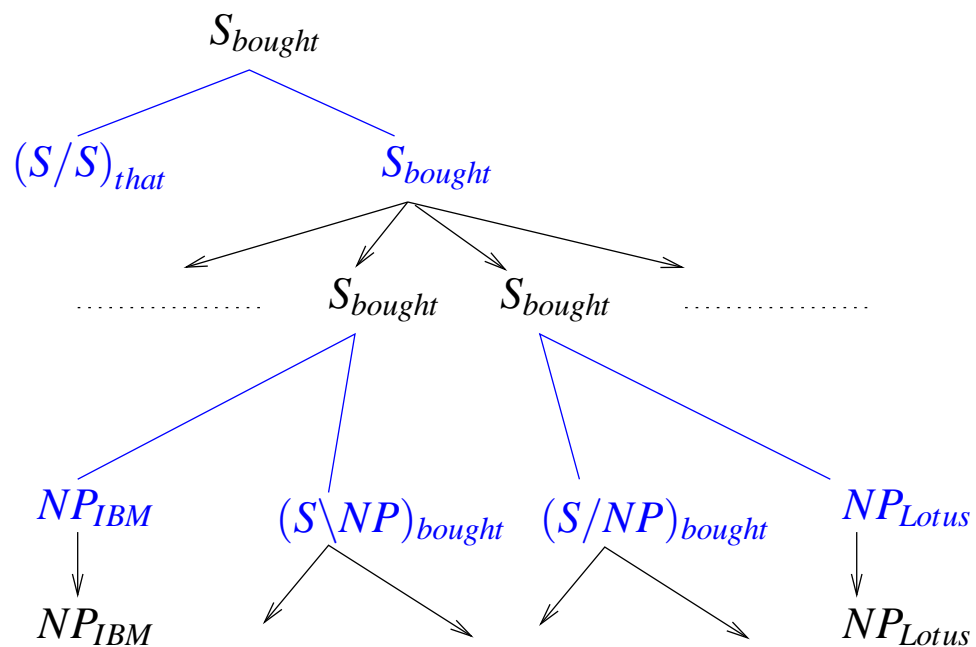
---

# Packed Charts

- A derivation typically shares many parts with other derivations
- Features in our models are *local* to rule applications
- A *packed chart* can efficiently represent all derivations for a sentence
  - for the purposes of model estimation and decoding

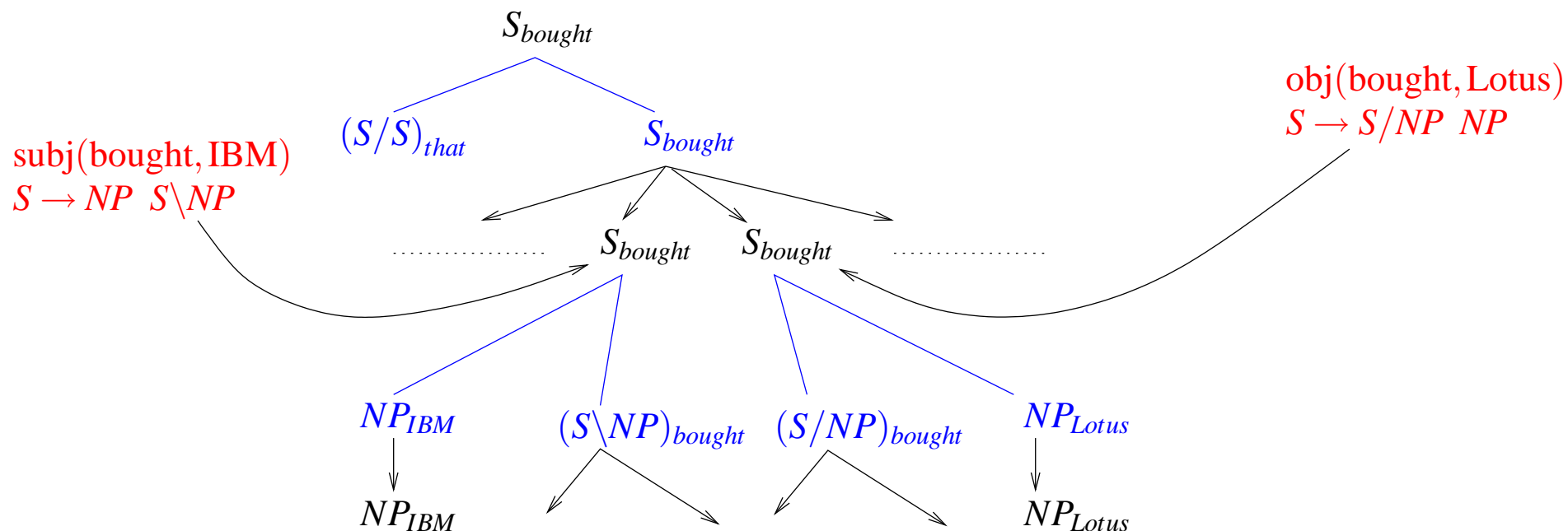
# Packed Charts

*I said that IBM bought Lotus*

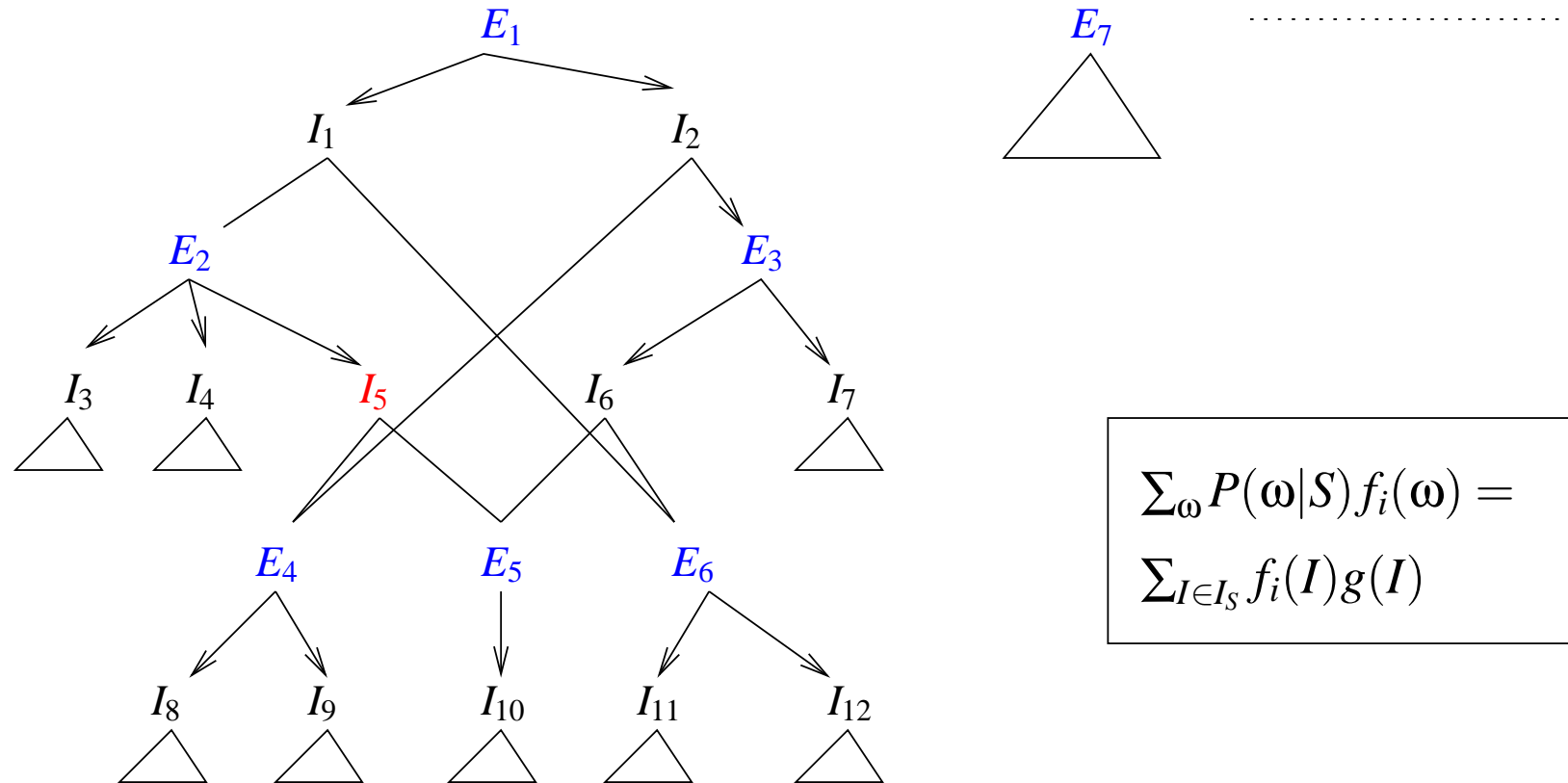


# Features in Packed Charts

*I said that IBM bought Lotus*

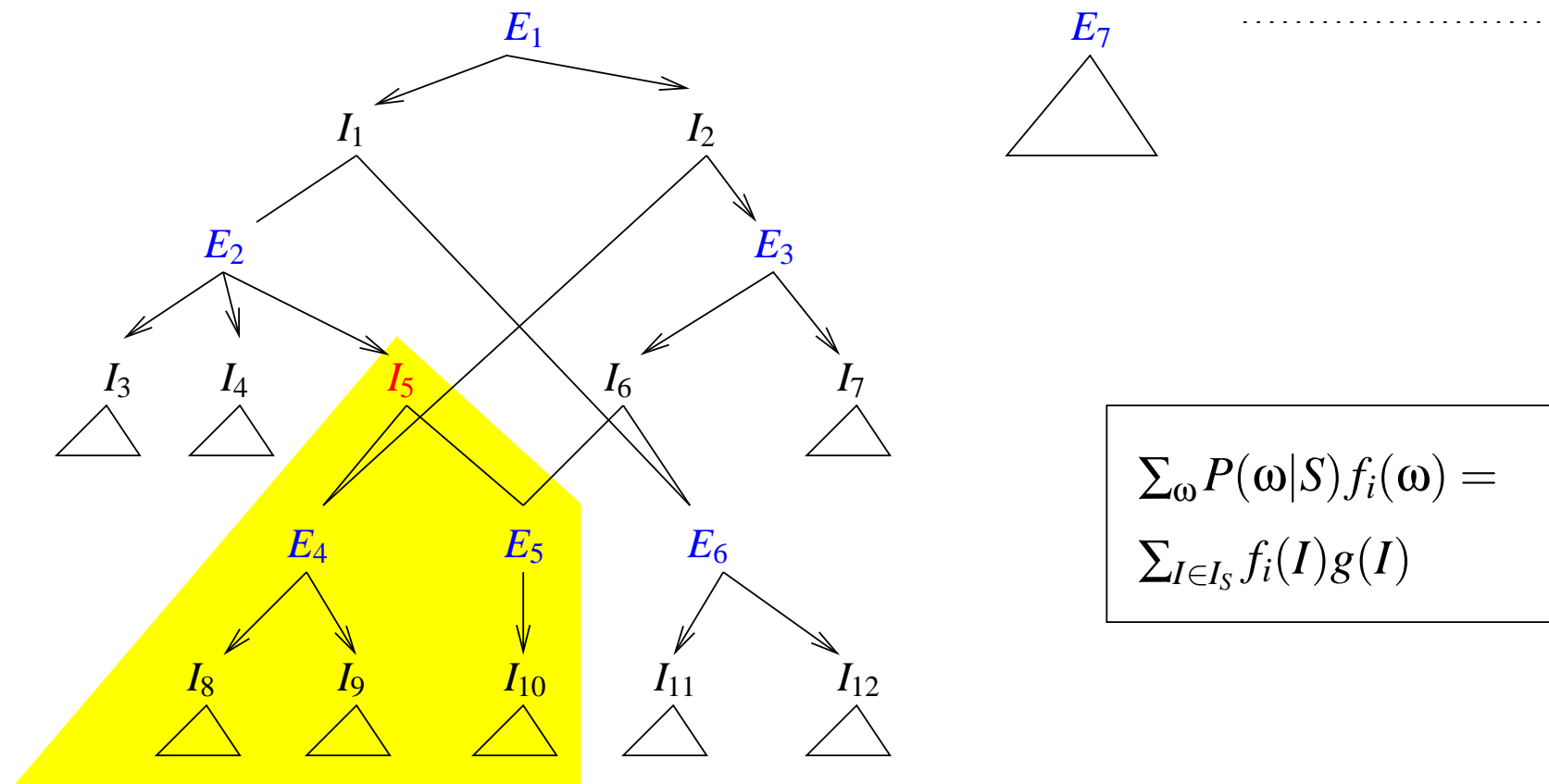


# Calculating Feature Expectations

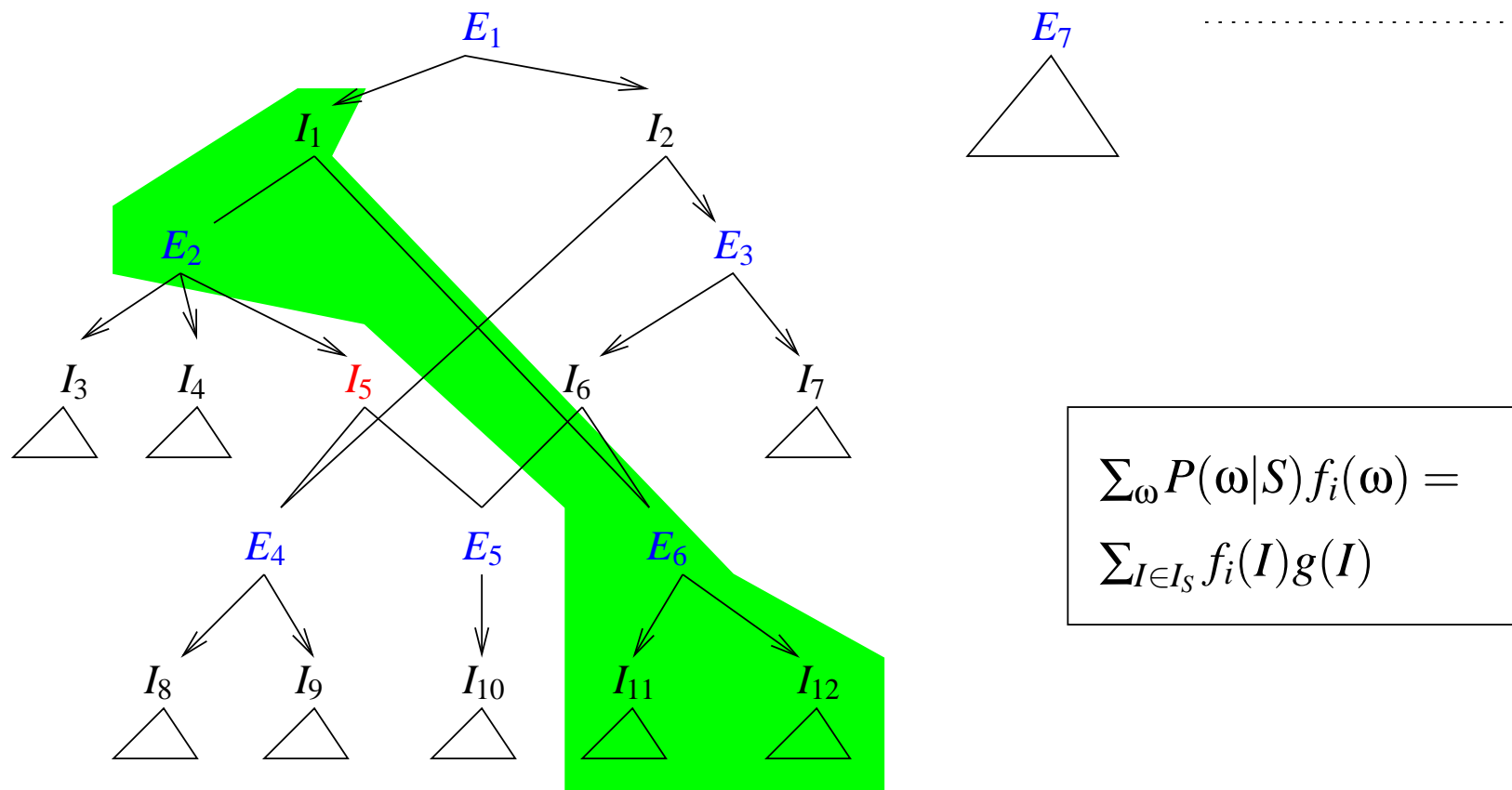


$$\sum_{\omega} P(\omega|S) f_i(\omega) = \sum_{I \in I_S} f_i(I) g(I)$$

# The Inside Score $\phi(I_5)$

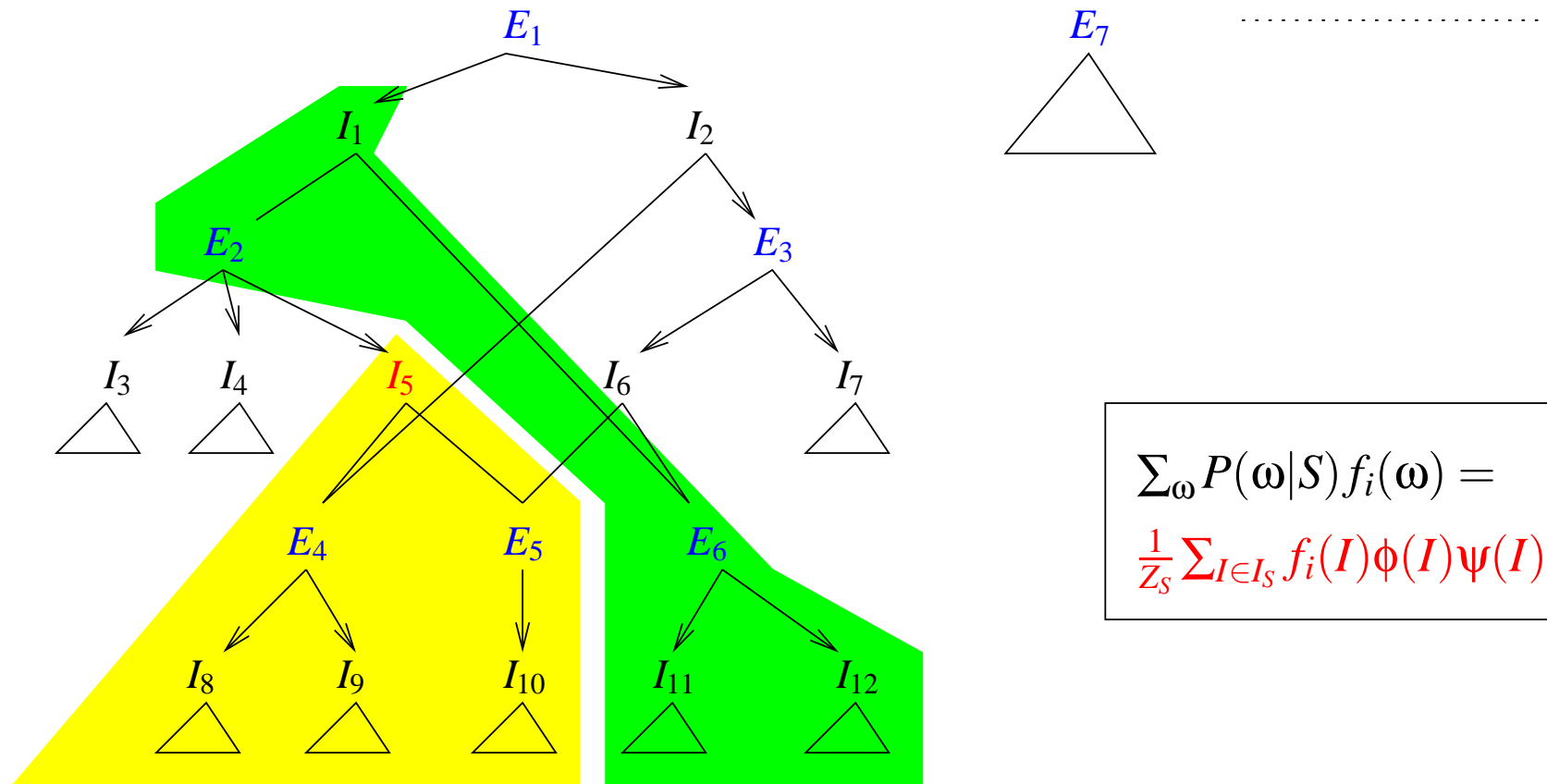


# The Outside Score $\psi(I_5)$



$$\sum_{\omega} P(\omega|S) f_i(\omega) = \sum_{I \in I_S} f_i(I) g(I)$$

# Feature Expectations

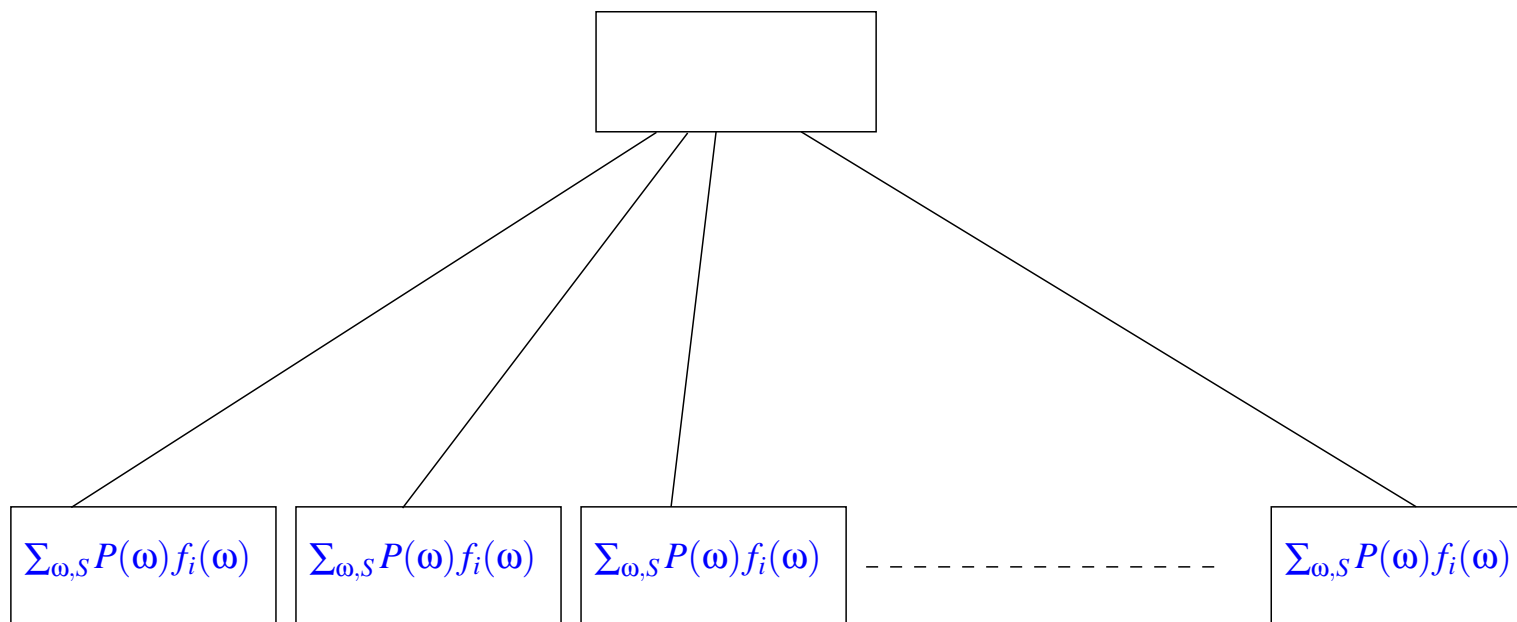


---

## Estimation in Practice

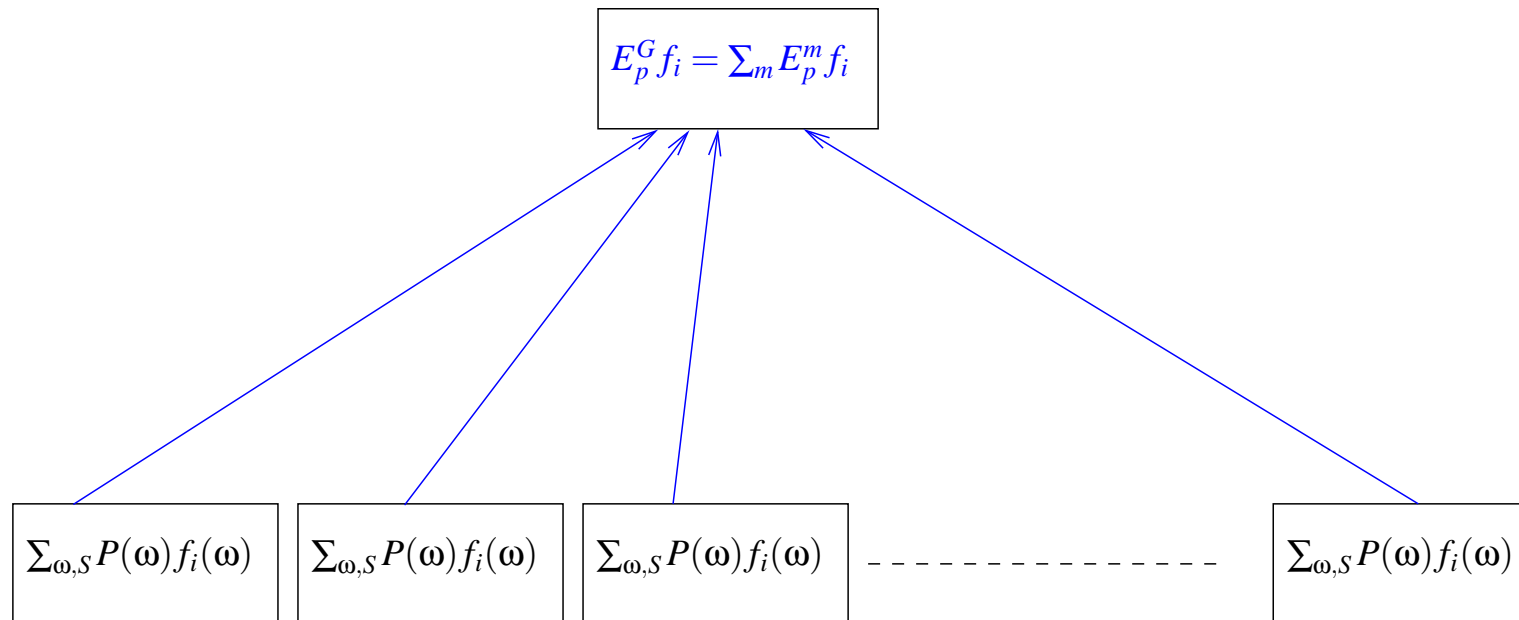
- Parse sentences from sections 2-21 of CCGbank
  - Maximum Entropy supertagger assigns categories to words
- Charts require  $\approx$  18 GB RAM
- Calculation of feature expectations is easily parallelised
- Parallel version of LBFGS using MPI
- Use an 18-node Beowulf cluster to perform the estimation
- Training takes  $\approx$  1,000 iterations, a few hours

# Parallel Estimation



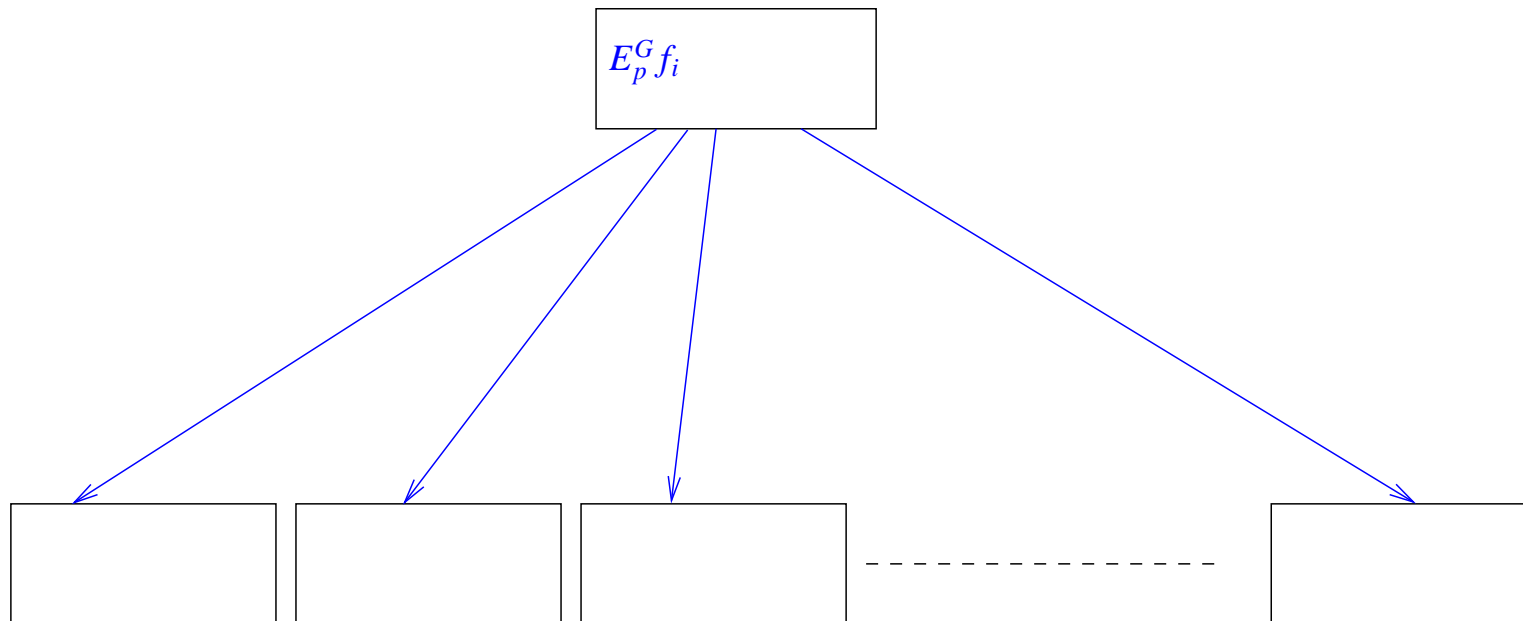
- Each machine calculates local expectations (using inside-outside)

# Parallel Estimation



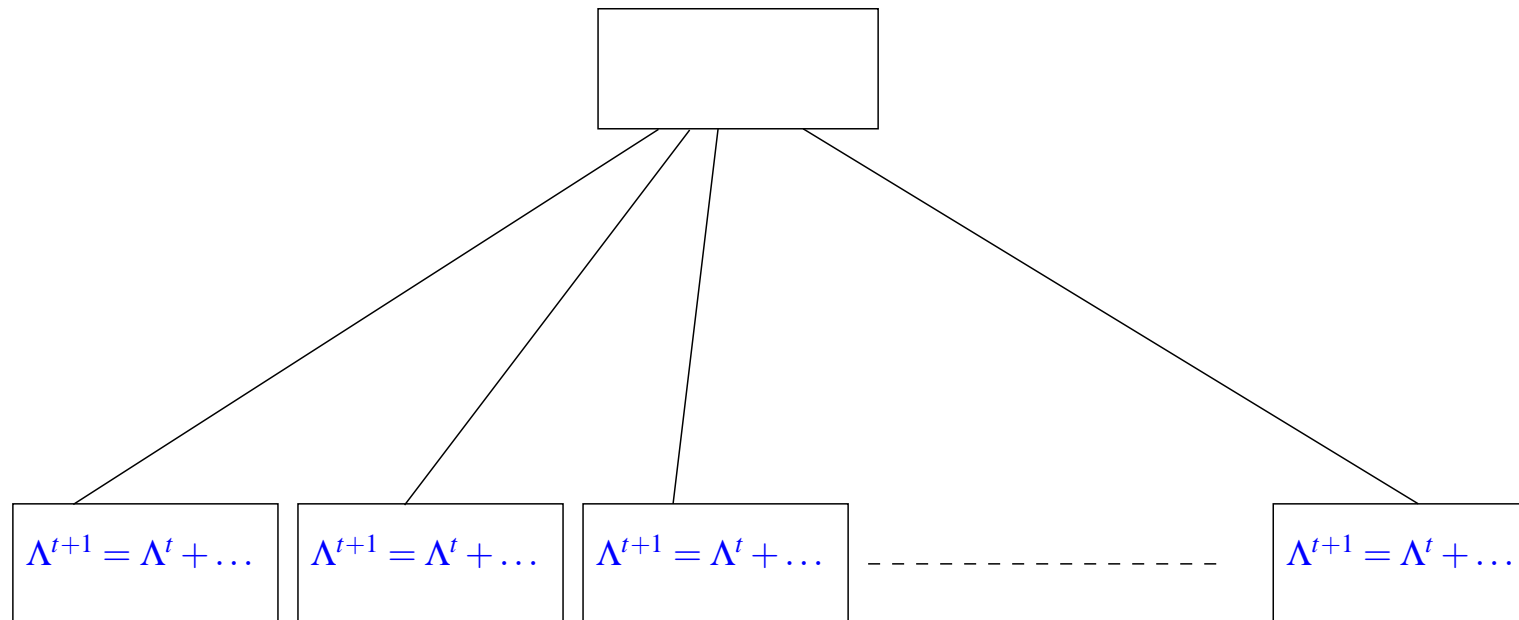
- Local expectations passed to root and summed

# Parallel Estimation



- Global expectation passed to each machine

# Parallel Estimation



- Each machine updates weights (using GIS, LBFGS)