Structured Language Modeling using Stochastic Tree Adjoining Grammars

AT&T Student Research Day

October 29, 1999

Anoop Sarkar

Dept. of Computer and Information Science

University of Pennsylvania

Prefix Probabilities

- Language model: given a string $a_1, \ldots, a_{i-1}$, a_i can be any word in the vocabulary Σ, what is $P(a_i \mid a_1, \ldots, a_{i-1})$?

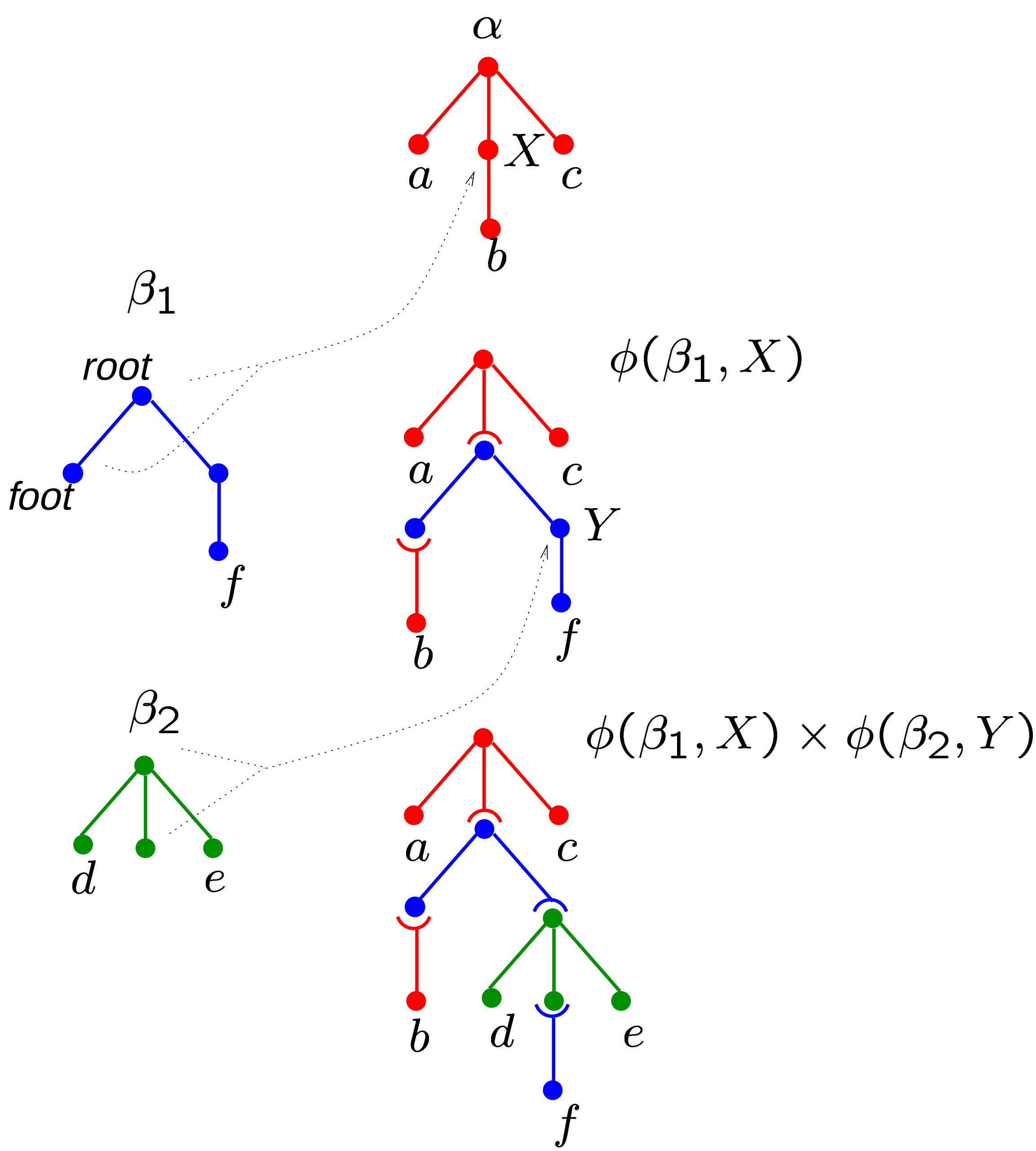
- Standard techniques use trigram models:

$$P(a_i \mid a_{i-2}, a_{i-1})$$

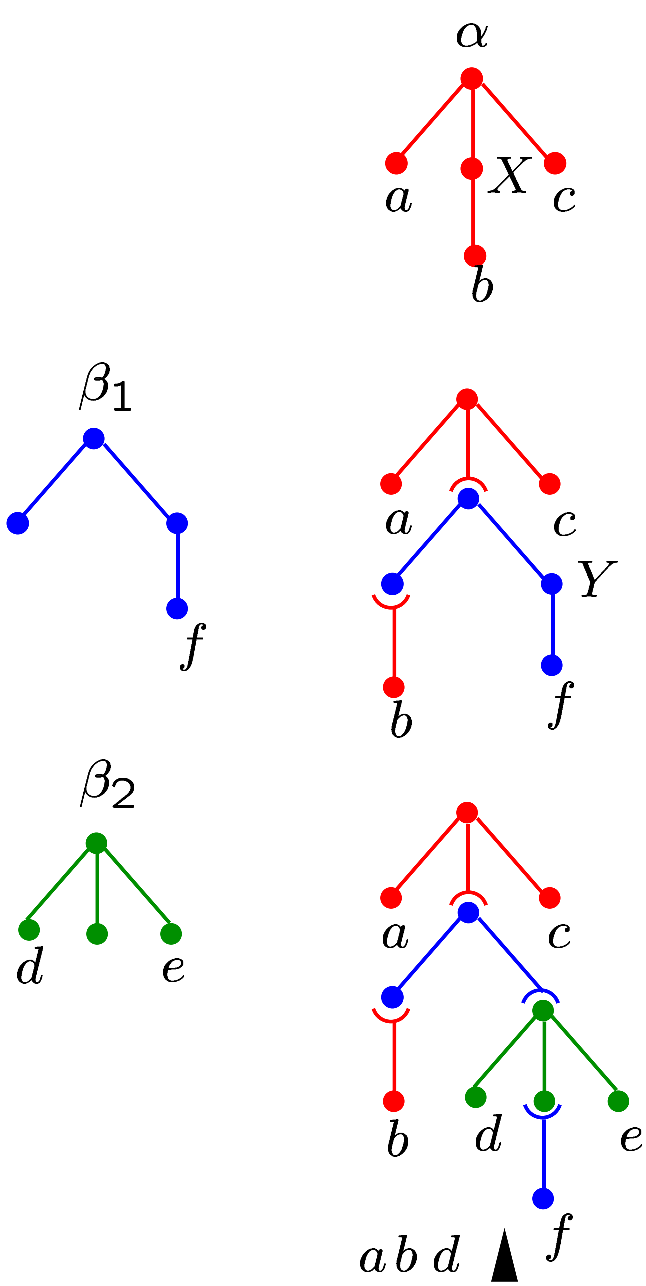
- A stochastic grammar can be used by computing the prefix probability:

$$\sum_{w \in \Sigma^*} P(a_1, \ldots, a_i w)$$

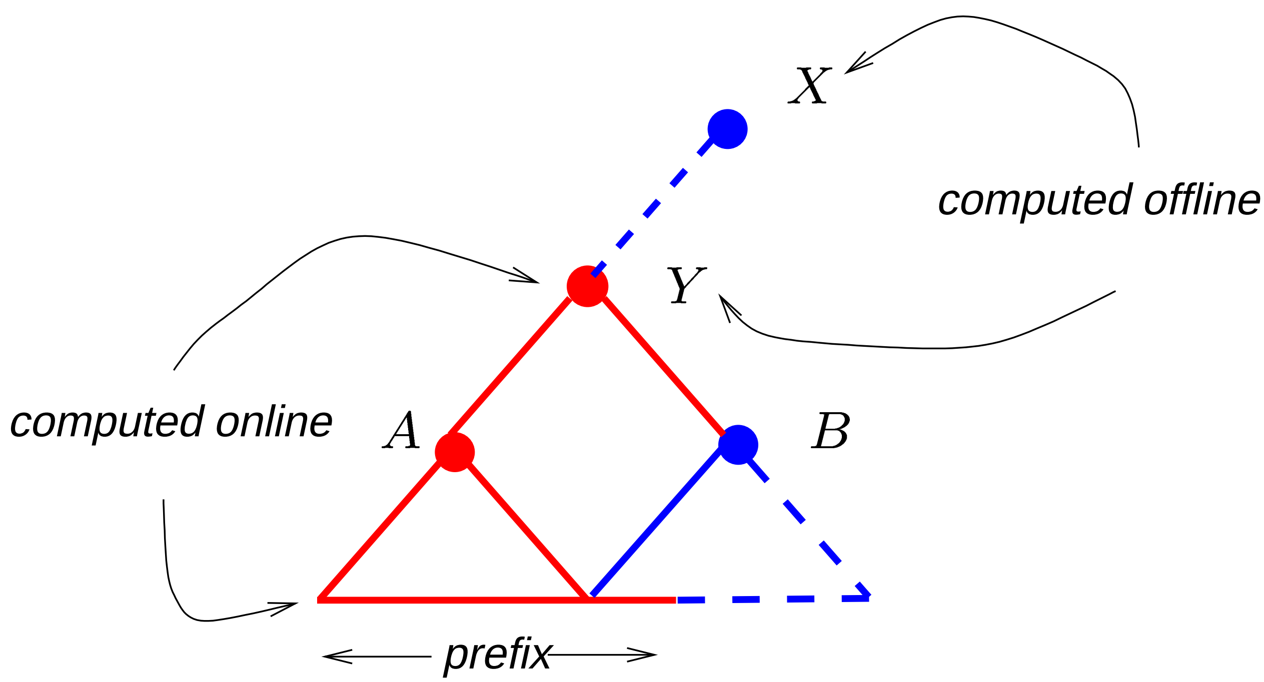
Stochastic Tree Adjoining Grammars

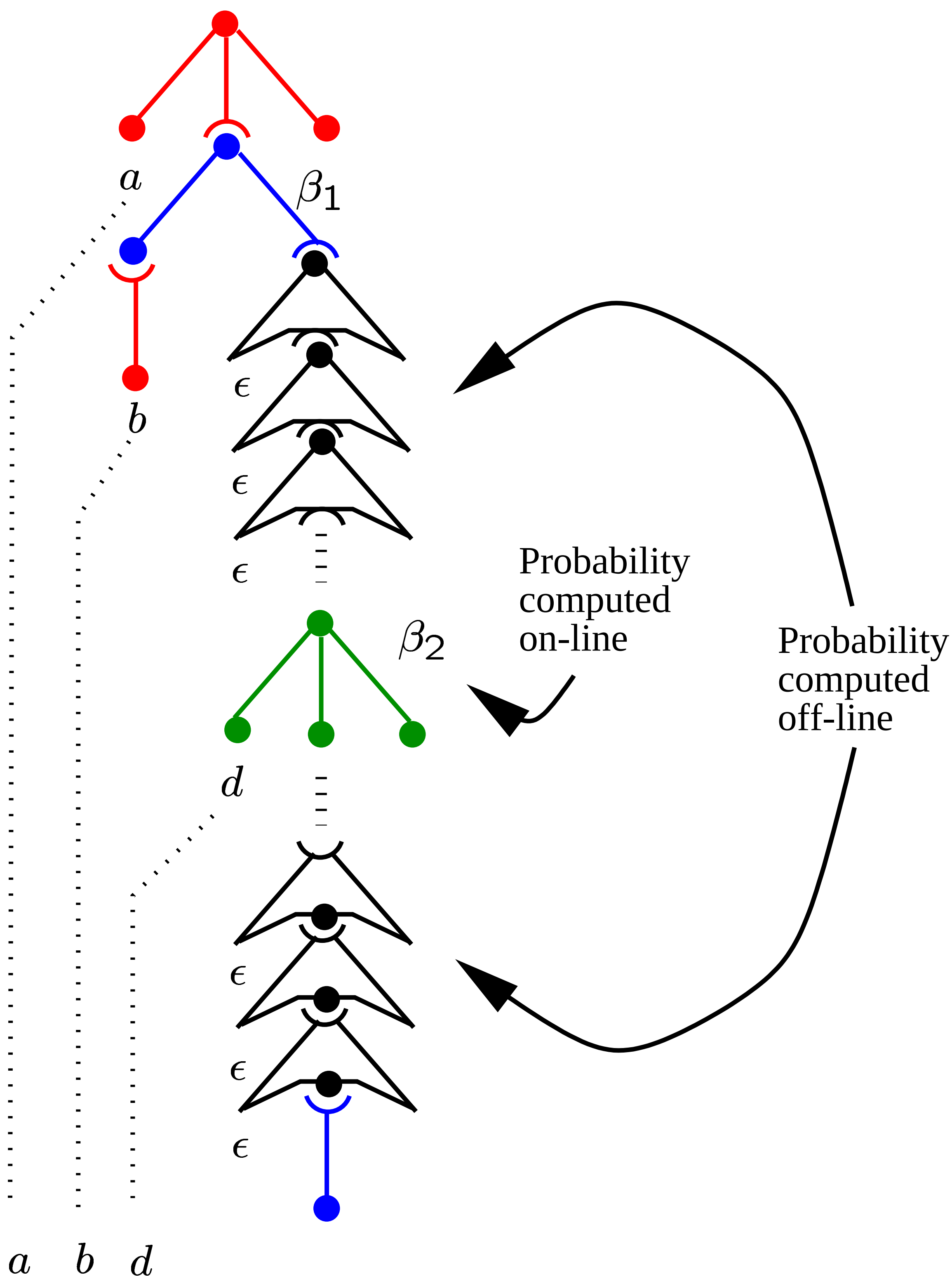


Let prefix $= abd$



Prefix Pr for CFGs





a
β_1
b
ϵ
ϵ
ϵ
Probability computed on-line
β_2
Probability computed off-line
d
ϵ
ϵ
ϵ
a b d

Problem

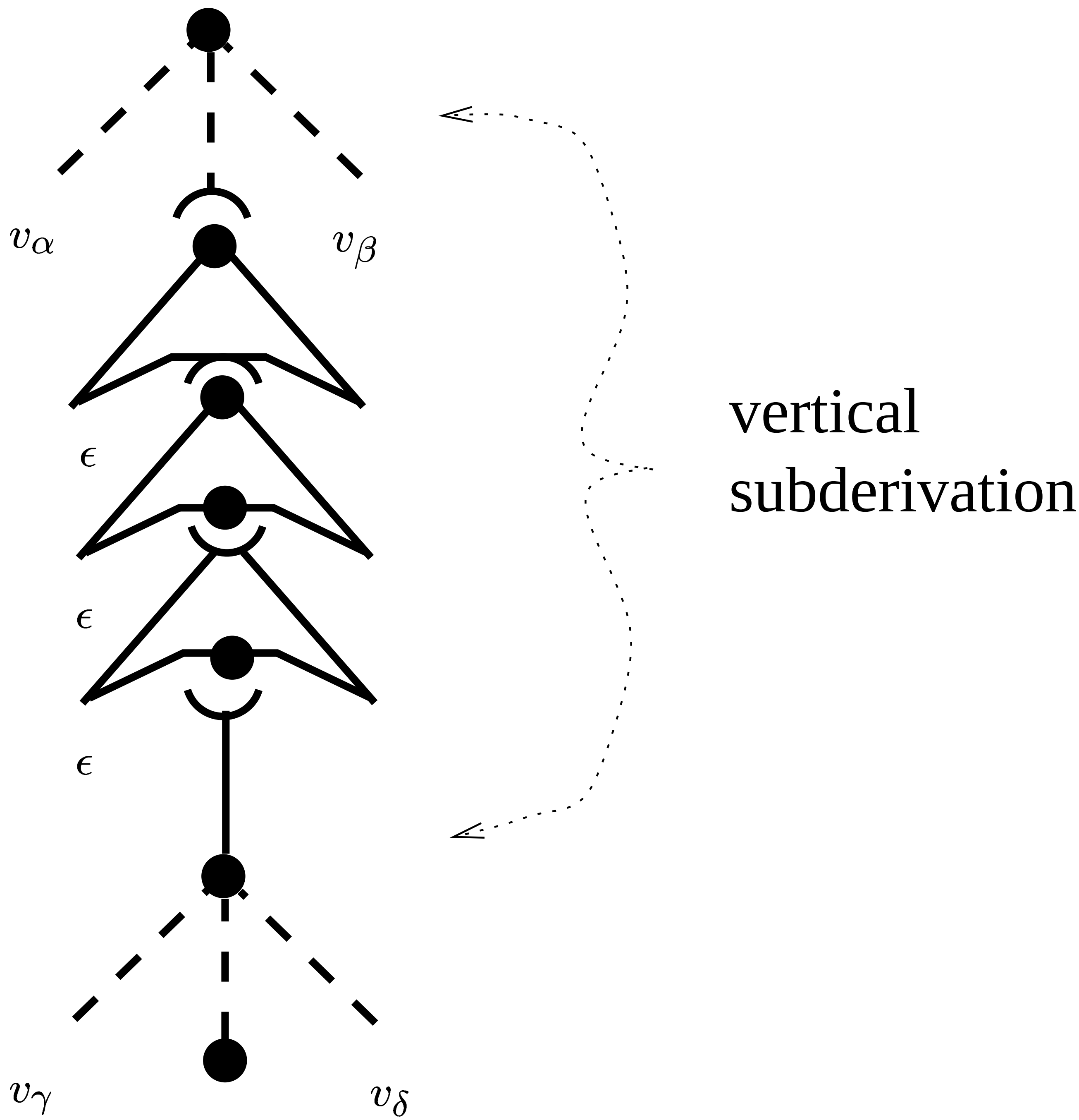
- Derivations are a combination of two kinds of subderivations:

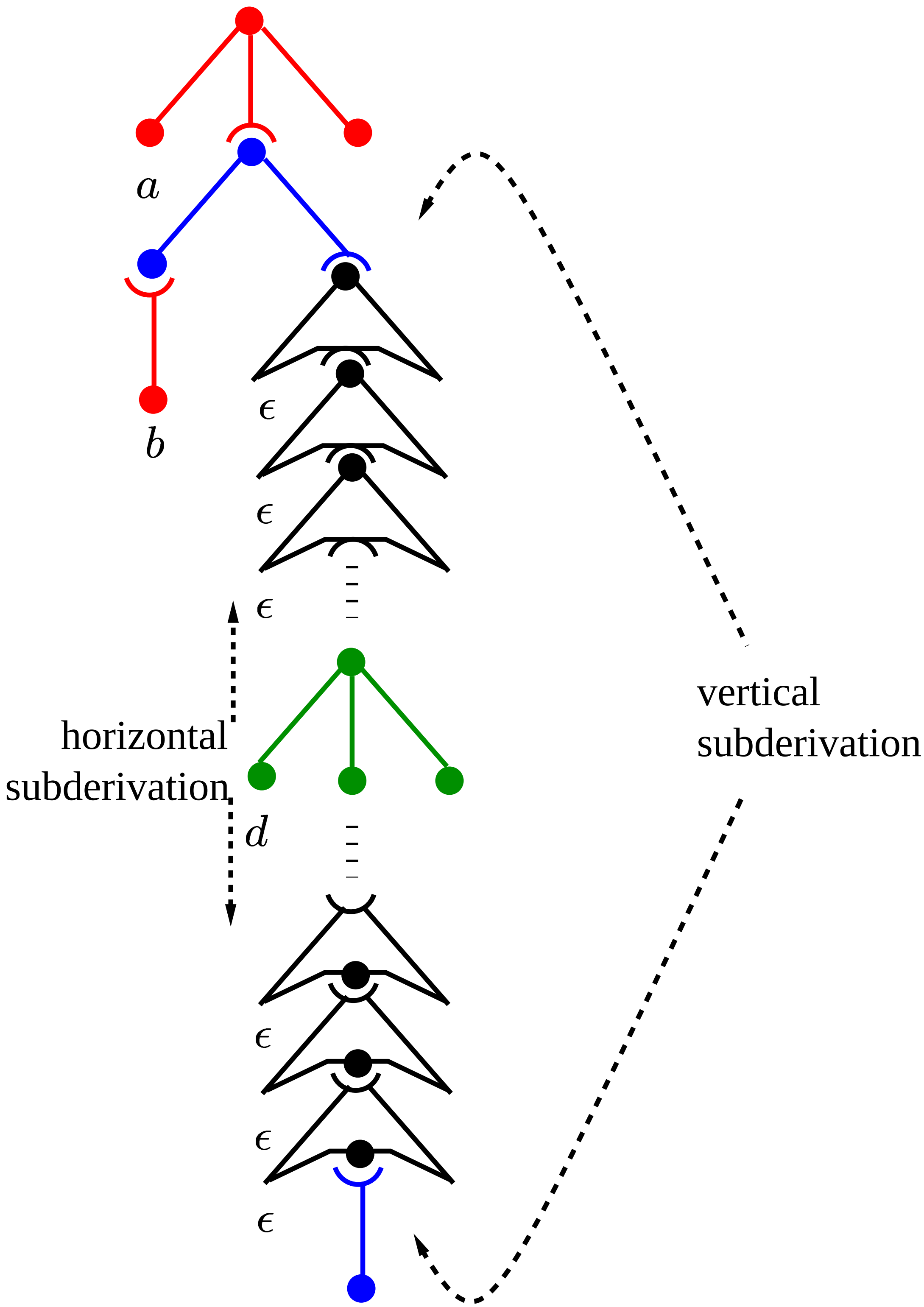
  1. potentially unbounded subderivations, independent of input

  2. bounded subderivations, depend on input symbols

- Problem: how to partition derivations uniquely into subderivations.

- Without unique partitions, algorithm will return incorrect probabilities.


v_α
v_β
ϵ
ϵ
ϵ
v_γ
v_δ
vertical
subderivation



a
b
ϵ
ϵ
ϵ
horizontal
subderivation
d
ϵ
ϵ
ϵ
vertical
subderivation

Offline Probability Computation

- Probability for jumping from one tree to another which contributes to the prefix are computed offline.

- These paths between trees are potentially unbounded.

- Are there closed form solutions instead of approximations.

TAG Derivations and Branching Processes

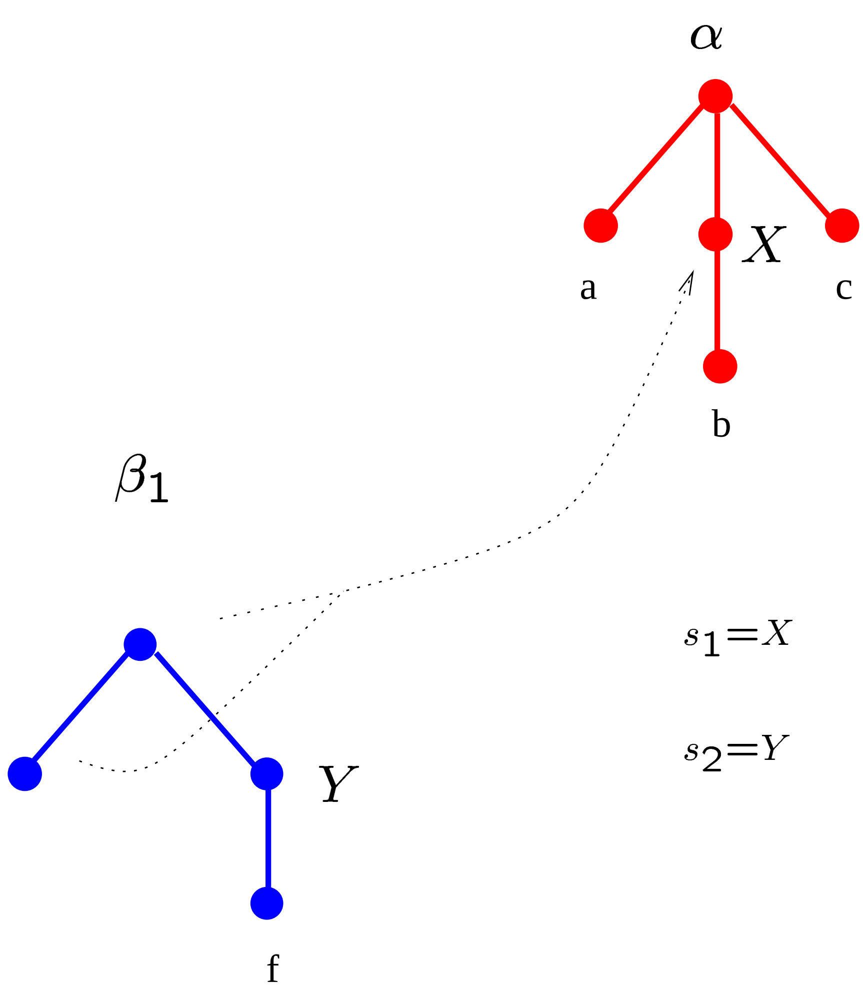
- There is an initial set of objects in the 0-th generation which produces with some probability a first generation.
- The first generation in turn with some probability generates a second, and so on.
- We will denote by vectors $Z_0, Z_1, Z_2, \ldots$ the 0-th, first, second, $\ldots$ generations.

TAG Derivations and Branching Processes

- The size of the n-th generation does not influence the probability with which any of the objects in the $(n+1)$-th generation is produced.
- $Z_0, Z_1, Z_2, \ldots$ form a Markov chain.
- The number of objects born to a parent object does not depend on how many other objects are present at the same level.
- We associate a generating function for each level Z_i.

Adjunction Generating Function

$$g_1(s_1, s_2) = \phi(\beta_1, X) \cdot s_2 + \phi(\mathit{NA}, X)$$



$s_1 = X$

$s_2 = Y$

Level generating functions

$$\begin{aligned} G_0(s_1, \ldots, s_k) &= s_1 \\ G_1(s_1, \ldots, s_k) &= g_1(s_1, \ldots, s_k) \\ G_n(s_1, \ldots, s_k) &= G_{n-1}[g_1(s_1, \ldots, s_k), \ldots, \\ &\quad\; g_k(s_1, \ldots, s_k)] \end{aligned}$$

- we can express $G_i(s_1, \ldots, s_k)$ as a sum

$$D_i(s_1, \ldots, s_k) + C_i$$

  where C_i is a constant and $D_i(\cdot)$ a polynomial with no constant terms.

- Closed-form solutions to these equations compute the necessary offline probabilities

- A stochastic TAG will be consistent iff

$$lim_{i \to \infty} D_i(s_1, \ldots, s_k) \to 0$$

Summary

- An algorithm for computing prefix probabilities from a Stochastic Tree Adjoining Grammar.

- Probabilities for future histories computed offline by exploiting the theory of branching processes.

- *Problem:* Estimation of parameters in a structured language model requires labeled data. n-gram models train from easily available unlabeled data.

- *Current Research:* Estimation algorithms for STAGs which can combine labeled and unlabeled sources of data.