Prefix Probabilities for Linear Indexed Grammars

TAG+ 4, IRCS/University of Pennsylvania

Mark-Jan Nederhof
(DFKI)

Anoop Sarkar
(UPenn)

Giorgio Satta
(Univ. di Padova)

Prefix Probabilities

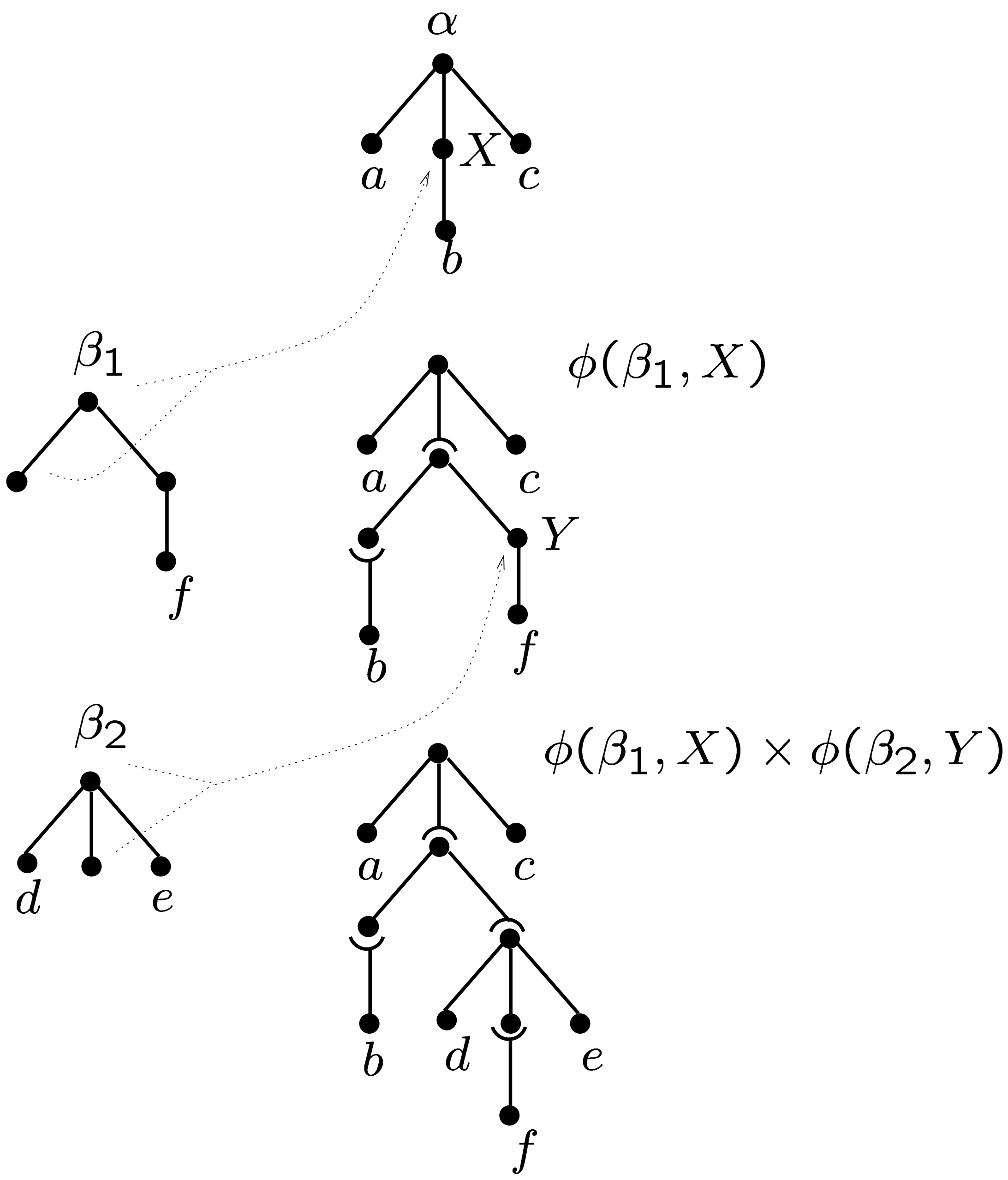
- Language model: given a string $a_1, \ldots, a_{i-1}$, a_i can be any word in the vocabulary Σ, what is $P(a_i \mid a_1, \ldots, a_{i-1})$?

- Standard techniques use trigram models: $P(a_i \mid a_{i-2}, a_{i-1})$

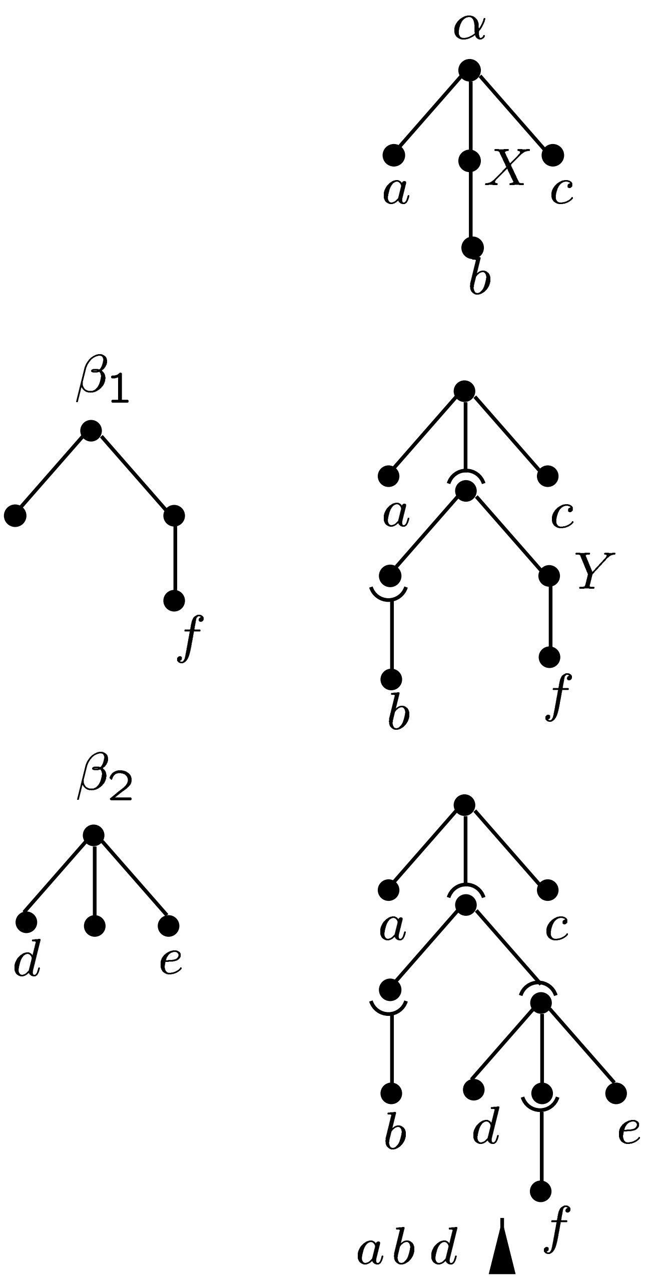
- A stochastic grammar can be used by computing the prefix probability:

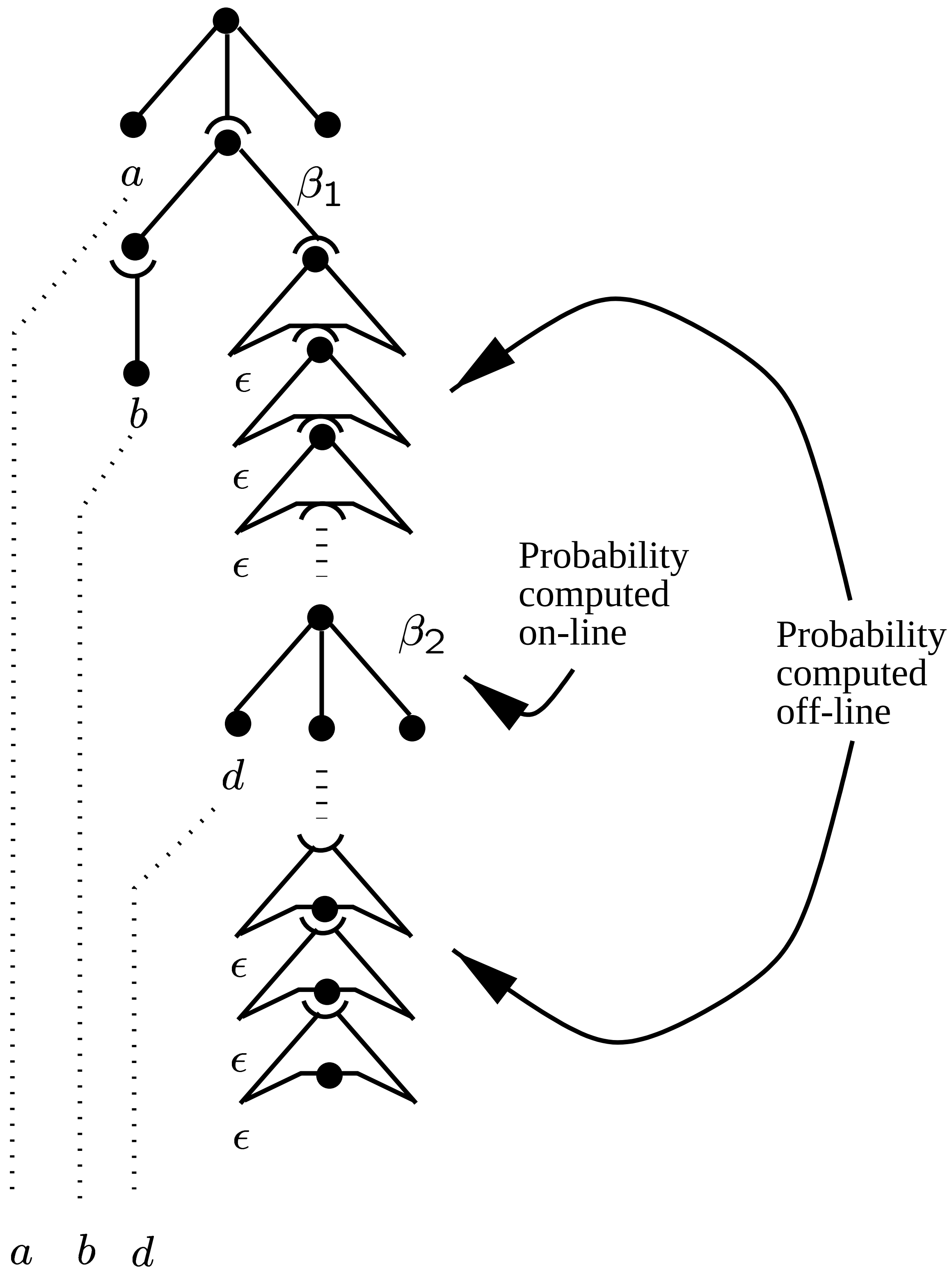
$$\sum_{w \in \Sigma^*} P(a_1, \ldots, a_n w)$$

Stochastic Tree Adjoining Grammars



Let prefix $= abd$




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

TAG Derivations and LIGs

- CFGs with stack symbols, Indexed Grammars: $A[..\eta_r] \to B[..\eta_1]C[..\eta_2]$

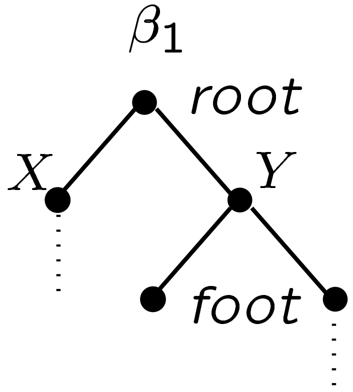
- Unique rhs gets stack = Linear Indexed Grammars: $A[..\eta_r] \to B[..\eta_1]C[\$\eta_2]$

- LIGs and TAGs are weakly equivalent.

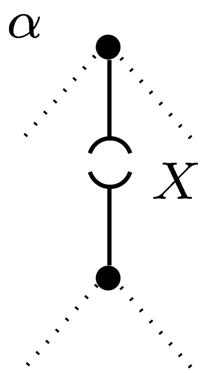
- Given a TAG, a strongly equivalent LIG can be built.

- LIGs offer a convenient way to denote TAG derivations.

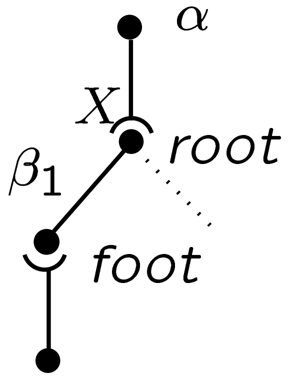
TAG Derivations and LIGs



$b[..\mathit{root}] \to t[\$X]t[..Y]$

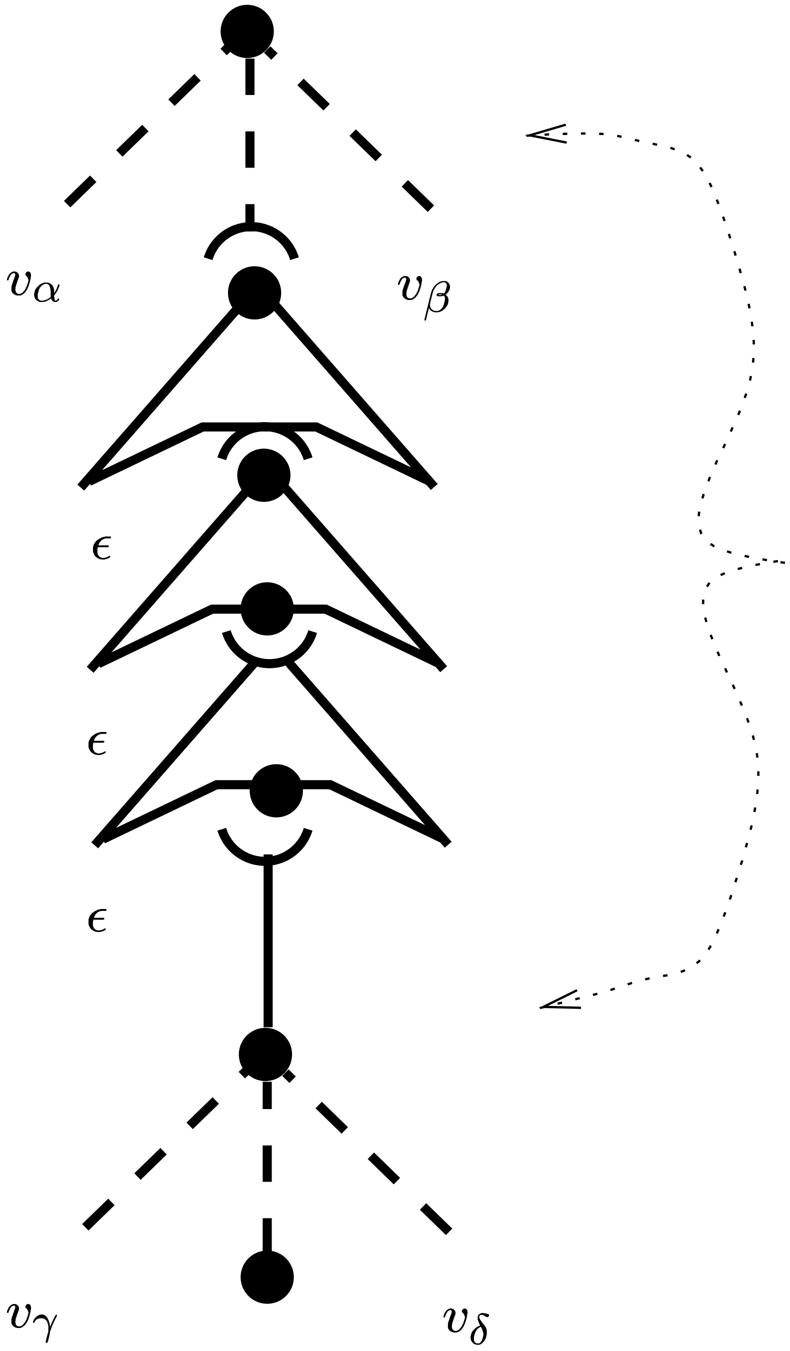


$t[..X] \to b[..X]$

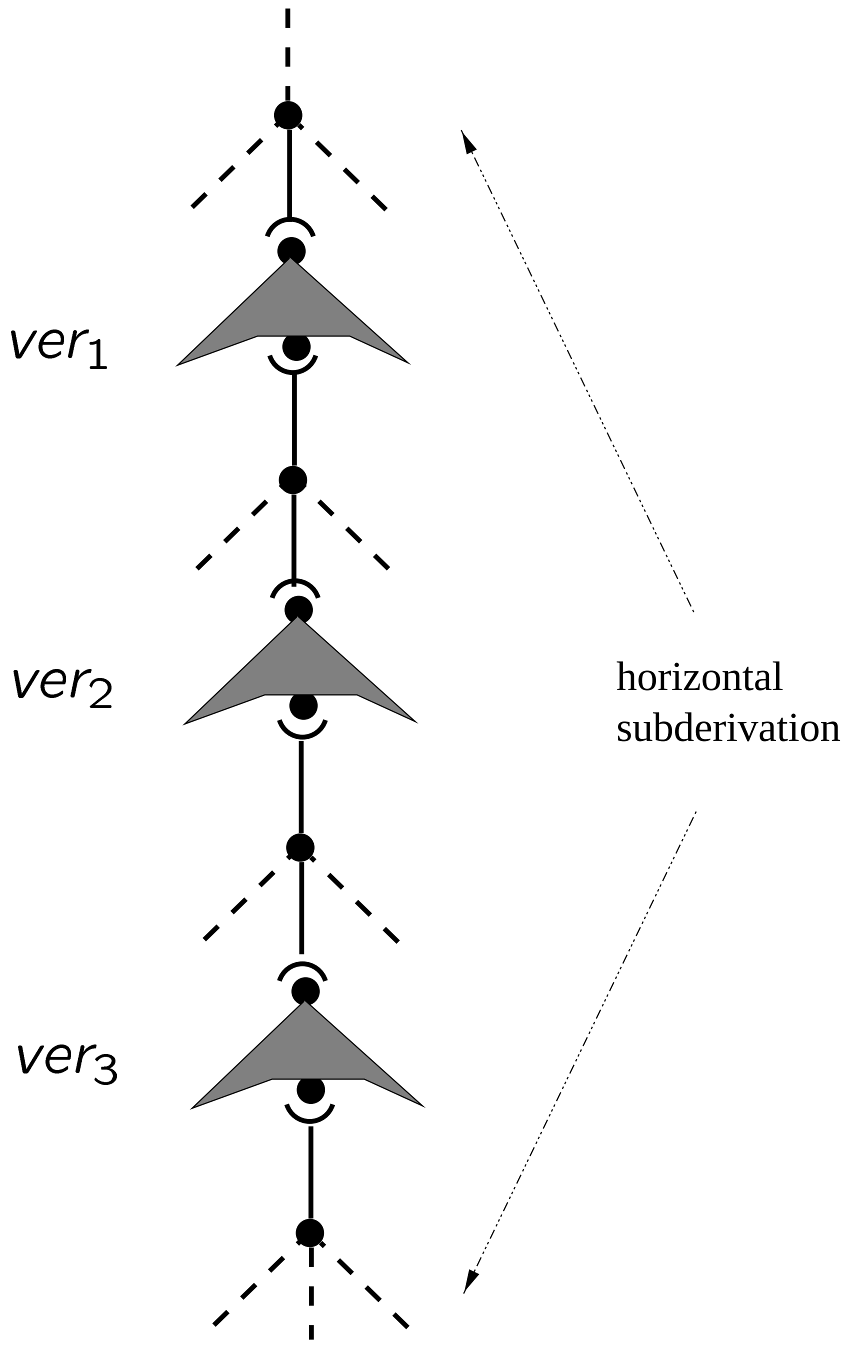


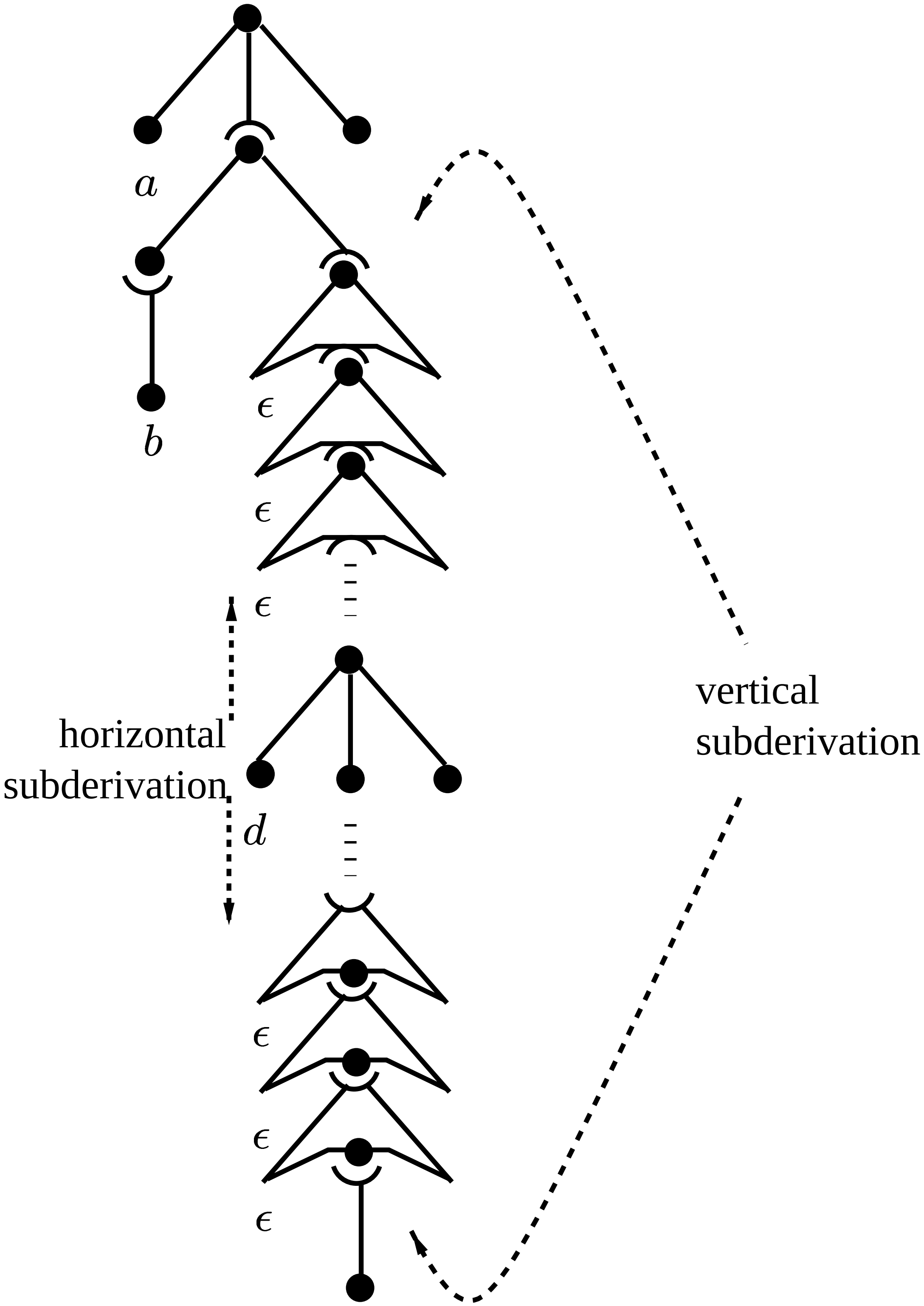
$t[..X] \to t[..X\,\mathit{root}]$

$b[..\mathit{foot}] \to b[..]$



vertical
subderivation


ver_1
ver_2
horizontal subderivation
ver_3


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

Summary

- LIGs offer a convenient way to denote TAG derivations.

- Potentially unbounded subderivations, independent of input. Probability is computed off-line.

- Bounded subderivations, depends on input symbols. Inside probability computed as prefix is recognized.

- Algorithm partitions derivations uniquely into subderivations.