Conditions on Consistency of Probabilistic Tree Adjoining Grammars

COLING/ACL 1998

Anoop Sarkar
Dept. of Computer and Information Science
University of Pennsylvania
anoop@linc.cis.upenn.edu

Consistency of Probabilistic Grammars

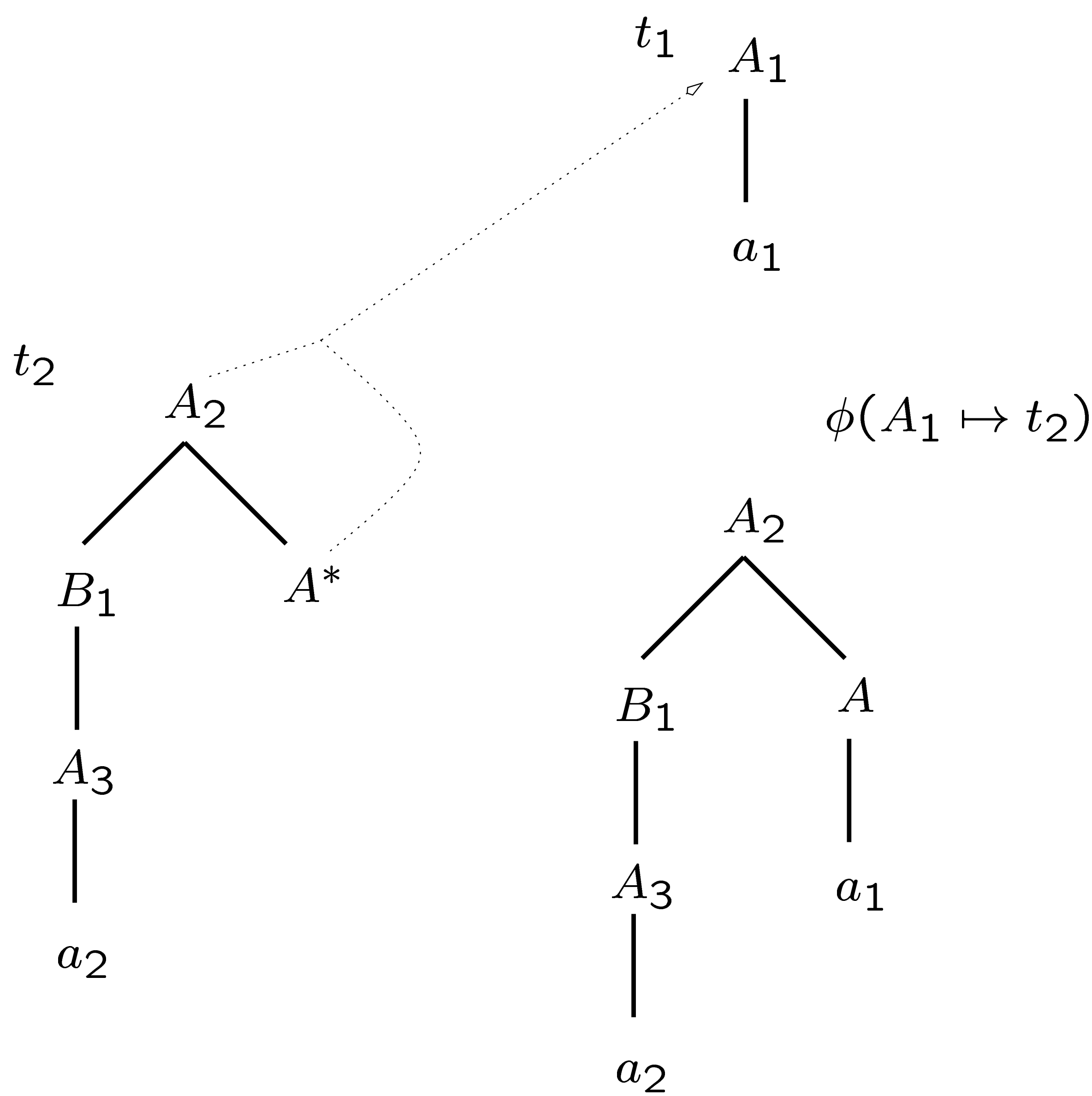
- Pr assigns a probability to each string v in the language.

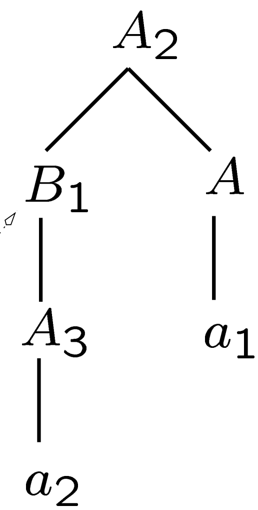
- If v is not in the language then $\Pr(v) = 0$.

- Consistency is the property that sum of probabilities assigned to all the strings in the language sum to 1.

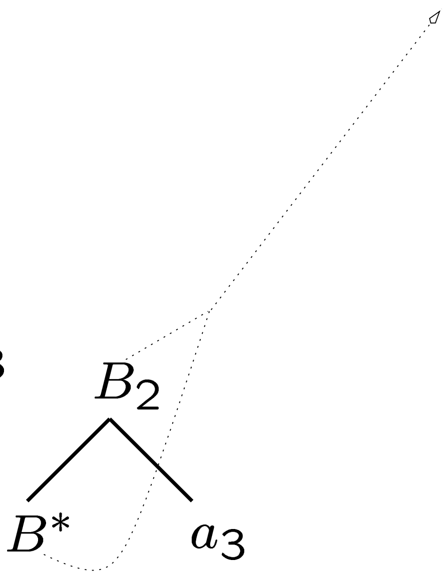
$$\sum_{v \in L(G)} \Pr(v) = 1$$

Probabilistic TAGs

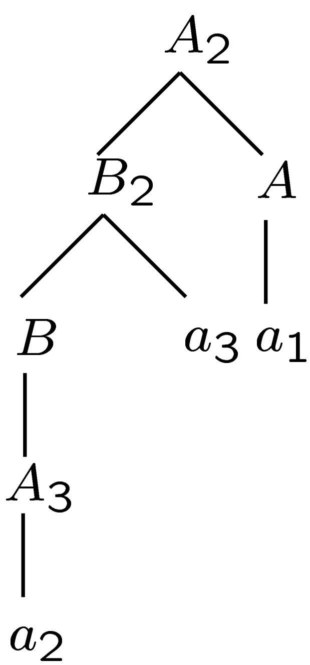


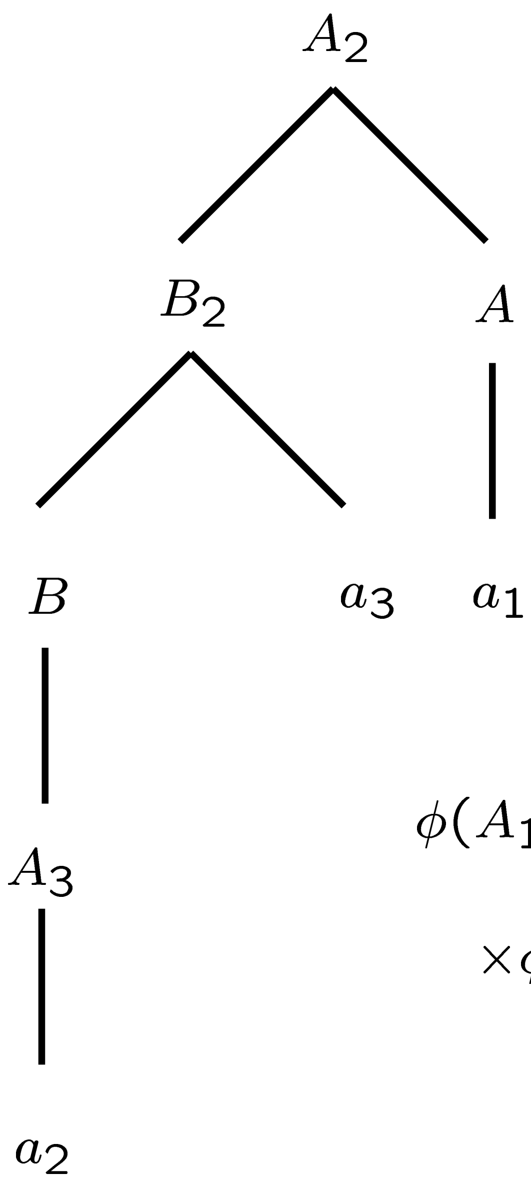


t_3

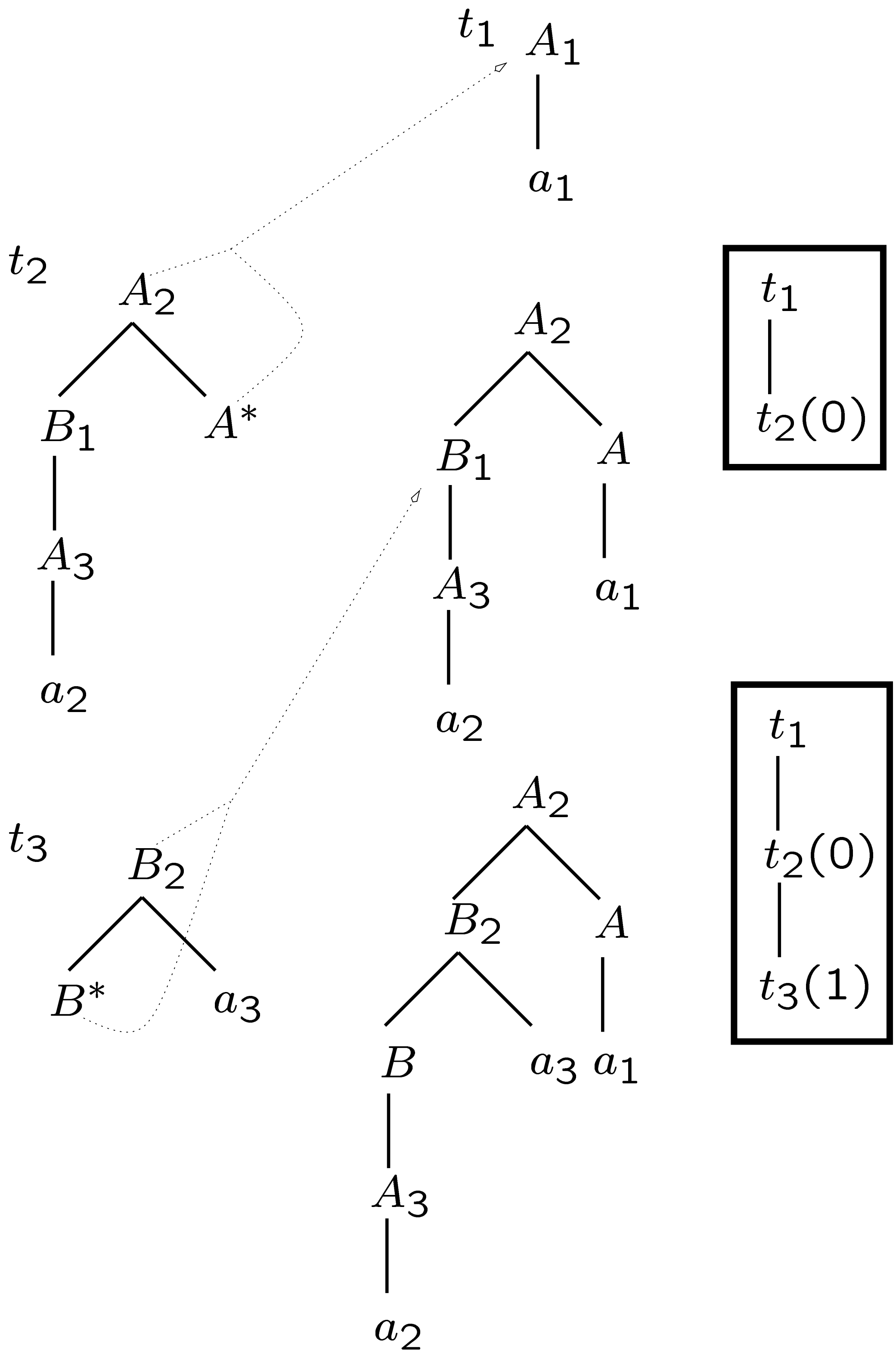


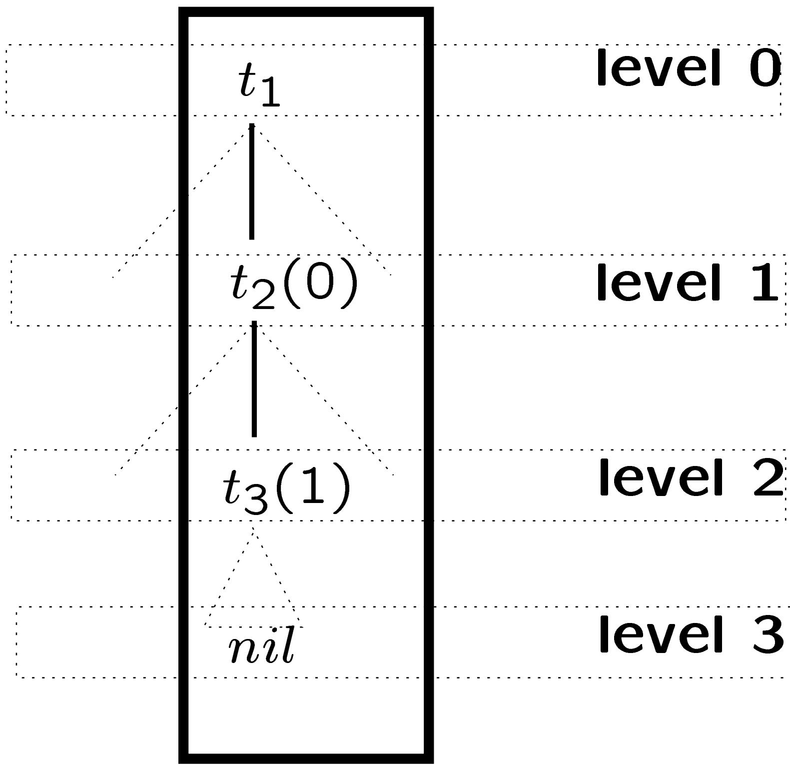
$$\phi(A_1 \mapsto t_2) \times \phi(B_1 \mapsto t_3)$$



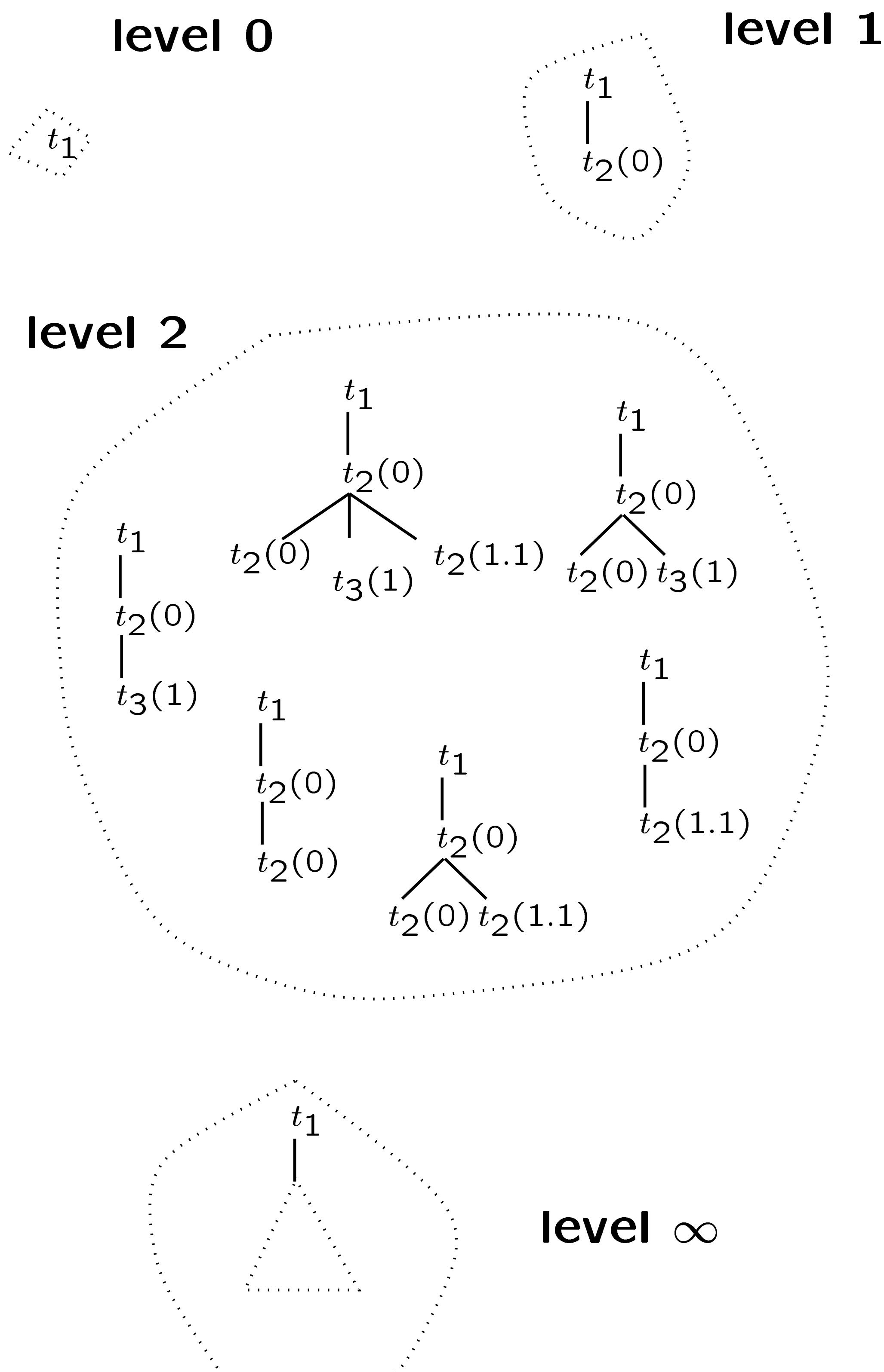


$$\phi(A_1 \mapsto t_2) \times \phi(B_1 \mapsto t_3)$$
$$\times \phi(A_2 \mapsto nil) \times \phi(A_3 \mapsto nil)$$
$$\times \phi(B_2 \mapsto nil)$$

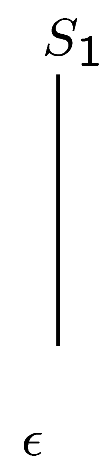

t1
A1
a1
t2
A2
B1
A*
A3
a2
A2
B1
A
A3
a1
a2
t1
t2(0)
t3
B2
B*
a3
A2
B2
A
B
a3
a1
A3
a2
t1
t2(0)
t3(1)



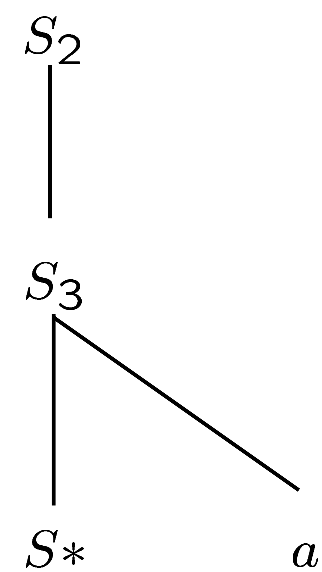
t_1
level 0
$t_2(0)$
level 1
$t_3(1)$
level 2
nil
level 3


level 0
t_1
level 1
t_1
$t_2(0)$
level 2
t_1
$t_2(0)$
$t_3(1)$
t_1
$t_2(0)$
$t_2(0)$
$t_3(1)$
$t_2(1.1)$
t_1
$t_2(0)$
$t_2(0)$
$t_3(1)$
t_1
$t_2(0)$
$t_2(0)$
t_1
$t_2(0)$
$t_2(0)$
$t_2(1.1)$
t_1
$t_2(0)$
$t_2(1.1)$
t_1
level ∞

t_1



t_2



$\phi(S_1 \mapsto t_2) = 1.0$

$\phi(S_2 \mapsto t_2) = 0.99$

$\phi(S_2 \mapsto nil) = 0.01$

$\phi(S_3 \mapsto t_2) = 0.98$

$\phi(S_3 \mapsto nil) = 0.02$

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, ... 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.

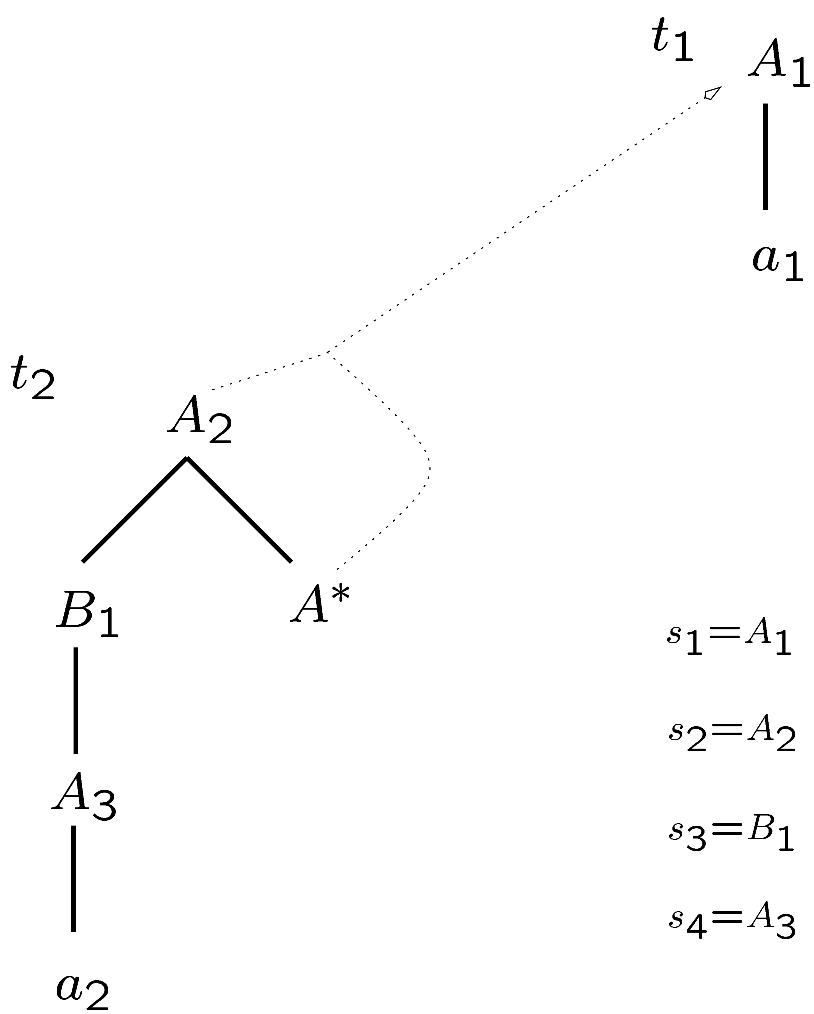
- $Z_0, Z_1, Z_2, \ldots$ form a Markov chain.

- 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

$$g_1(s_1, \ldots, s_5) = \phi(A_1 \mapsto t_2) \cdot s_2 \cdot s_3 \cdot s_4 + \phi(A_1 \mapsto nil)$$



$s_1 = A_1$

$s_2 = A_2$

$s_3 = B_1$

$s_4 = A_3$

Level generating functions

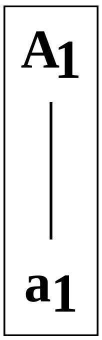
$$
\begin{aligned}
G_0(s_1, \ldots, s_k) &= s_1 \\
G_1(s_1, \ldots, s_k) &= g_1(s_1, \ldots, s_k) \\
G_n(s_1, \ldots, s_k) &= G_{n-1}[g_1(s_1, \ldots, s_k), \ldots, \\
&\quad g_k(s_1, \ldots, s_k)]
\end{aligned}
$$

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

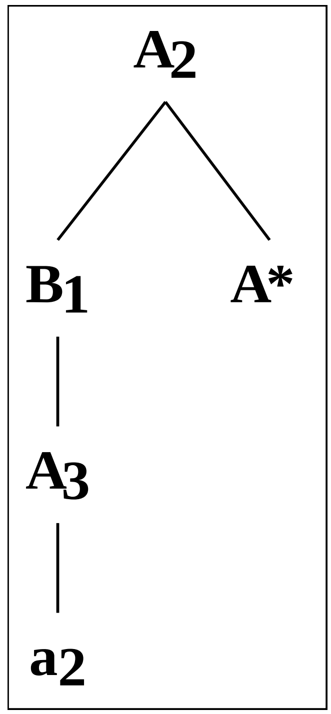
- A probabilistic TAG will be consistent if these recursive equations terminate, i.e. iff

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

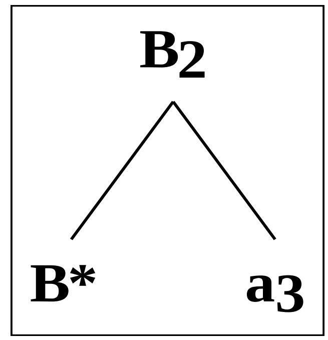
t_1



t_2



t_3



$$\phi(A_1 \mapsto t_2) = 0.8$$
$$\phi(A_1 \mapsto nil) = 0.2$$

$$\phi(A_2 \mapsto t_2) = 0.2$$
$$\phi(A_2 \mapsto nil) = 0.8$$
$$\phi(B_1 \mapsto t_3) = 0.2$$
$$\phi(B_1 \mapsto nil) = 0.8$$
$$\phi(A_3 \mapsto t_2) = 0.4$$
$$\phi(A_3 \mapsto nil) = 0.6$$

$$\phi(B_2 \mapsto t_3) = 0.1$$
$$\phi(B_2 \mapsto nil) = 0.9$$

$$\mathbf{P} = \begin{array}{c} \\ A_1 \\ A_2 \\ B_1 \\ A_3 \\ B_2 \end{array} \begin{array}{c} \begin{array}{ccc} t_1 & t_2 & t_3 \end{array} \\ \begin{bmatrix} 0 & 0.8 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.2 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.1 \end{bmatrix} \end{array}$$

$$\mathbf{N} = \begin{array}{c} \\ t_1 \\ t_2 \\ t_3 \end{array} \begin{array}{c} \begin{array}{ccccc} A_1 & A_2 & B_1 & A_3 & B_2 \end{array} \\ \begin{bmatrix} 1.0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 1.0 & 1.0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 \end{bmatrix} \end{array}$$

$$
\mathcal{M} = \mathbf{P} \cdot \mathbf{N} = \begin{array}{c} \\ A_1 \\ A_2 \\ B_1 \\ A_3 \\ B_2 \end{array} \begin{array}{c} \begin{array}{ccccc} A_1 & A_2 & B_1 & A_3 & B_2 \end{array} \\ \left[\begin{array}{ccccc} 0 & 0.8 & 0.8 & 0.8 & 0 \\ 0 & 0.2 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0.2 \\ 0 & 0.4 & 0.4 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.1 \end{array} \right] \end{array}
$$

- By representing the TAG derivations as a (Markovian) branching process we obtain a convergence result for $\mathcal{M}$.

- This allows us to test for consistency of the probabilistic TAG by computing *eig*$(\mathcal{M})$.

- In our example, the eigenvalues are $0, 0, 0, 6, 0,$ and 0.1. Since all are less than 1 the grammar is consistent.

Summary

- Derived conditions under which given probabilistic TAG can be shown to be consistent.

- Gave a simple algorithm for checking consistency.

- Gave formal justification for correctness of the algorithm.

- Useful for checking *deficiency* in a probabilistic TAG.