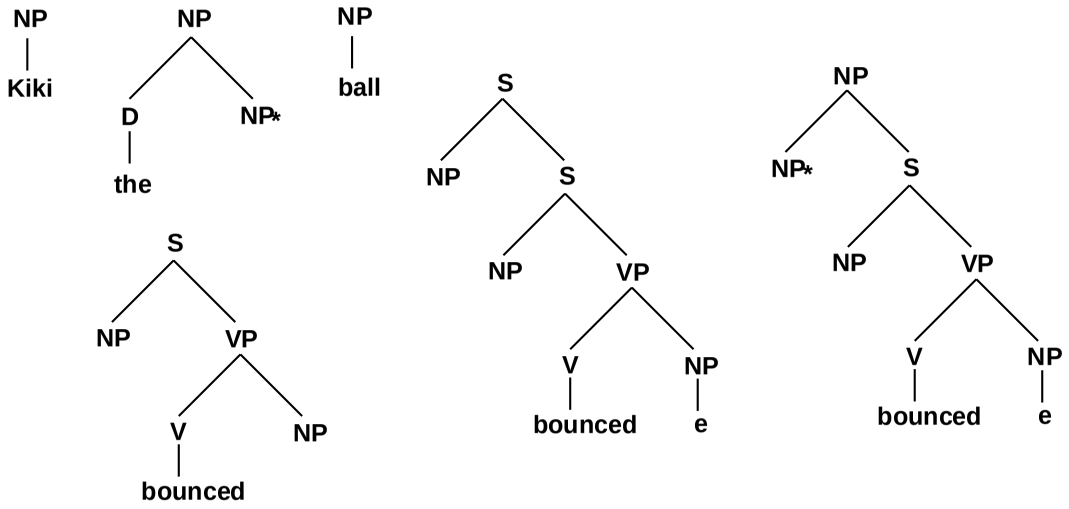
Grammar Inference using TAGs and Conditions on their Consistency

Anoop Sarkar
anoop@linc.cis.upenn.edu

IRCS and Dept. of Computer and Information Science
University of Pennsylvania

Linguistic Properties of TAGs

- TAGs can be used to represent a lexicalized grammar.
- Each word projects the various syntactic environments it can appear in.
- All empty elements are local to the trees each word projects.
- Trees are combined using substitution and adjunction.
- Distributional information can be used as evidence for structural disambiguation.



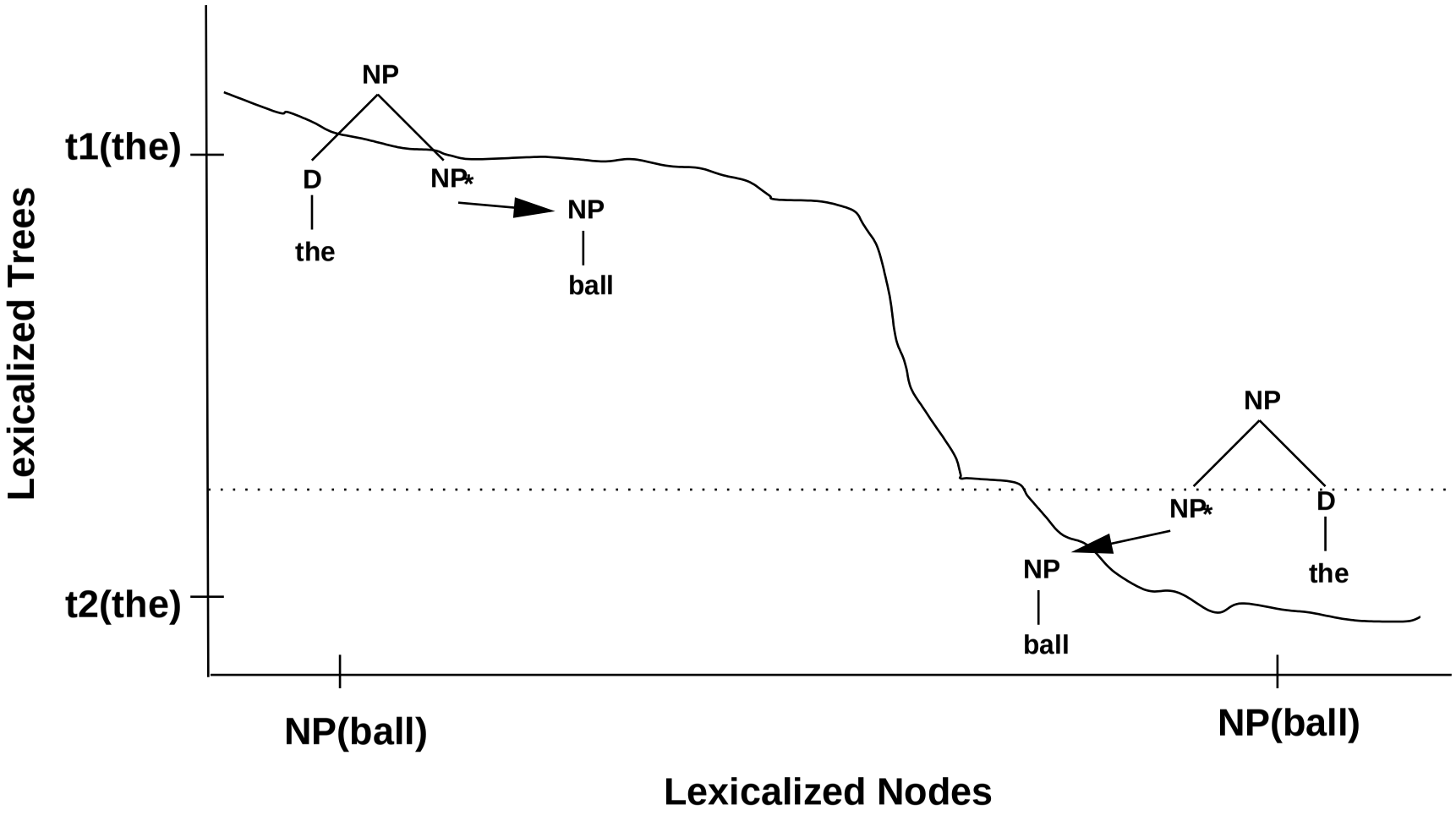
Statistical Grammar Inference using TAGs

- Assign each word every tree that is permissible regardless of variation across languages.
- Each substitution and adjunction is assigned a probability. e.g. $\phi(NP_{ball} \rightarrow t_{the}) = 0.03$
- $\phi(\cdot)$ form the parameters of the model.
- For each node N a $\mathbf{proper}$ assignment would be:

$$\phi(N \rightarrow nil) + \sum_{t} \phi(N \rightarrow t) = 1$$

- By receiving positive examples as input we use statistical estimation to set the values of these parameters.

Statistical Grammar Inference using TAGs



Maximum-Likelihood Estimation

- Takes an input of W sentences which are positive examples of a language.
- Computes the *expected usefulness* of each parameter in the derivation of each sentence w.
- Uses the expected value to compute the new value of the parameter by using the maximum-likelihood estimate.
- For example,

$$\hat{\phi}(NP_{ball} \rightarrow t_{the}) = \frac{\Sigma_{w \in W} \frac{1}{P(w)} \times E_{\phi}(NP_{ball} \rightarrow t_{the})}{\Sigma_{w \in W} \frac{1}{P(w)} \times E_{\phi}(NP_{ball} \rightarrow nil) + \Sigma_t E_{\phi}(NP_{ball} \rightarrow t)}$$

- Starting point: a random assignment of parameter values.
- Iterate until convergence to a local minima.

Consistency

- The probabilistic grammar assigns each sentence a probability based on the parameter values.
- This means that for each sentence w:

$$\sum_{w} \Pr(w) = 1$$

- It is not true that any assignment of probabilities will satisfy this condition.

The Main Result: Conditions on Consistency

- The conditions on consistency have to be defined over all possible derivations.
- The TAG derivations are modelled as a branching process.
- Each level in the derivation *reproduces* to form a new level or *generation*.
- Starting from the 0-th generation if the derivation does not *reproduce* indefinitely then the probabilistic TAG is consistent.
- In other words, the probability of extinction is the same as the probability that a TAG is consistent.

The Main Result: Conditions on Consistency

References

- T. L. Booth and R. A. Thompson. 1973. Applying probability measures to abstract languages. *IEEE Transactions on Computers*, C-22(5):442-450, May.
- T. E. Harris. 1963. *The Theory of Branching Processes*. Springer-Verlag, Berlin.
- Y. Schabes. 1992. Stochastic lexicalized tree-adjoining grammars. In *Proc. of COLING '92*, volume 2, pages 426-432, Nantes, France.