

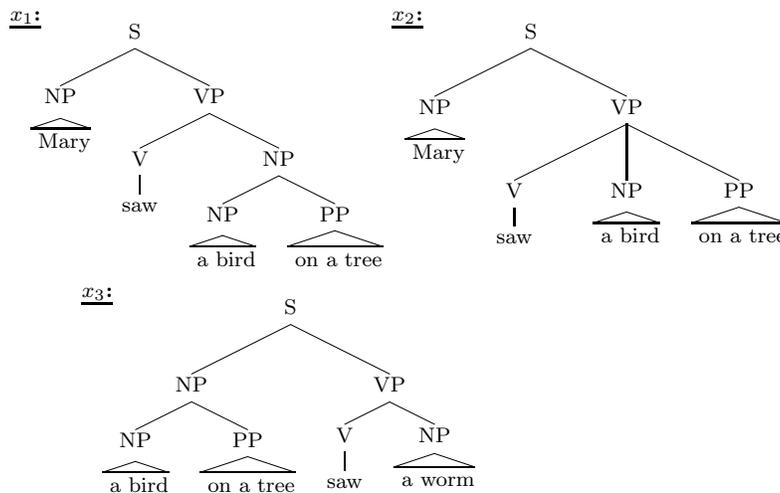
Einführung in die statistische Sprachverarbeitung

Detlef Prescher

18. Juni 2007

Hausaufgabe 9

(i) Gib eine möglichst kleine kontextfreie Grammatik an, die die folgenden Syntaxbäume generiert:



Initialisiere diese Grammatik so, dass alle Regeln mit derselben linken Seite dieselben Wahrscheinlichkeiten besitzen. Welches ist dann die wahrscheinlichste Analyse des Satzes 'Mary saw a bird on a tree'? Bist Du damit einverstanden?

(ii) Angenommen, Du bekommst die Grammatik in (i) vorgelegt, sowie einen Korpus in dem der Satz 'Mary saw a bird on a tree' 5-mal und der Satz 'a bird on a tree saw a worm' 10-mal vorkommt. Führe damit den ersten EM-Schritt durch. Welches ist nun die wahrscheinlichste Analyse des Satzes 'Mary saw a bird on a tree'? Bist Du damit einverstanden?

(iii) Führe nun den zweiten EM-Schritt durch! Erhältst Du bessere Disambiguierungsergebnisse?

(iv) Beschreibe nun in Deinen eigenen Worten (1/4-Seite), wie der generelle EM-Algorithmus abläuft. Nenne ein, zwei Gründe, warum diese Vorgehensweise zu guten Ergebnissen führen könnte...

(v) In welchen Fällen wendet man eigentlich den EM-Algorithmus an? Warum macht man in diesen Fällen keine ML-Estimierung?

Abgabetermin: Montag, 25. Juni

Besprechung: Mittwoch, 27. Juni