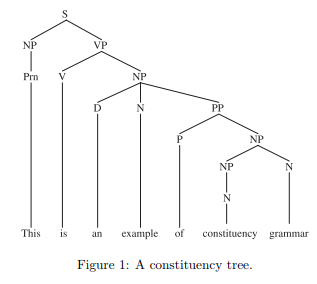
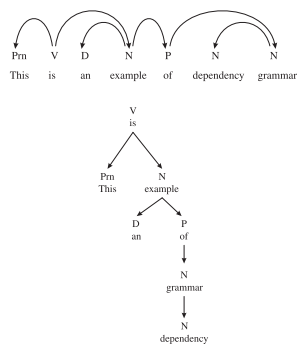
**STANDFORD PARSER**

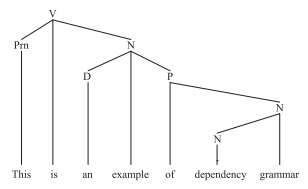
1. Tổng quan

* Phân tích cấu trúc của câu, ví dụ như những từ nào đi cùng nhau, từ nào là chủ ngữ, từ nào là vị ngữ.
* Có 2 cách mô tả cấu trúc câu trong ngôn ngữ tự nhiên
  + Constituency grammar: Tách câu thành các phần (cụm từ), sau đó tiếp tục chia thành các phần nhỏ hơn.



* + Dependency grammar: Vẽ các liên kết nối các từ riêng lẻ.





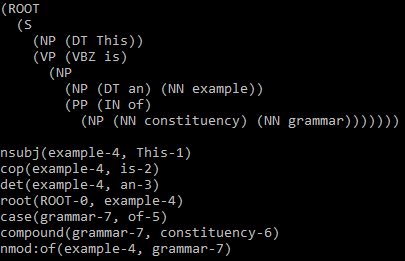
* Xây dựng model với cấu trúc từ PCFG tách biệt và sự phụ thuộc giữa các từ, mối liên hệ này được kết nối bằng suy luận chính xác và hiệu quả, sử dụng thuật toán A\*.

1. Input và output:

* Input: Có thể đọc nhiều dạng của văn bản thô.
* Output: Có thể xuất ra nhiều dạng phân tích khác nhau, bao gồm part-of-speech tagged text, phrase structure trees, and a grammatical relations (typed dependency).
* Ví dụ ta có một câu:

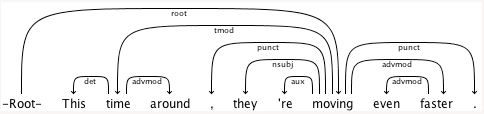
“This is an example of constituency grammar”

Kết quả thể hiện cấu trúc ngữ pháp không có ngữ cảnh và cấu trúc có ngữ cảnh.



1. Một số loại parser:
   1. **Shift – Reduce Constituent Parser**

* Version trước đây của Constituent Parser sử dụng thuật toán dựa trên biểu đồ để tìm ra điểm phân tích cao nhất theo PCFG, cách làm này đúng nhưng chậm.
* Shift – Reduce Constituent Parser sử dụng thuật toán shift-reduce để xây dựng constituent tree -> cải thiện hiệu suất.
* Cách thức hoạt động:
  + Duy trì trạng thái của cây đã được phân tích, với các từ của câu trong hàng đợi và một phần cây đã hoàn thành trong ngăn xếp, áp dụng sự chuyển đổi cho đến khi hàng đợi rổng và ngăn xếp hiện tại chỉ chứa một ngăn xếp đã hoàn thành.
  + Trạng thái ban đầu có tất cả các từ theo thứ tự trong hàng đợi. Các quá trình chuyển đổi có thể được áp dụng:
    - Shift: Chuyển một từ từ hàng đợi sang ngăn xếp.
    - Unary reduce: Thay đổi nhãn của thành phần đầu tiên trong ngăn xếp.
    - Binary reduce: Hai nút đầu tiên trong ngăn xếp được kết hợp với một nhãn mới. Có sự chuyển đổi nhị phân khác nhau cho mỗi nút nhị phân có thể
    - Finalize: Hoàn thành khi trình phân tích cú pháp chọn sự chuyển đổi hoàn tất.
    - Idle:
  1. **Neural Network Dependency Parser**
* Phân tích cấu trúc ngữ pháp của câu có ngữ cảnh, thiết lập mối quan hệ giữa một từ và những từ làm thay đổi những từ đó.



* + - Mũi tên di chuyển từ từ “moving” sang từ “faster” chỉ ra rằng “faster” làm thay đổi “moving” và từ ghi kem theo dấu mũi tên mô tả bản chất của sự phụ thuộc.

Ở trạng thái ban đầu, tất cả các từ ở trong bộ đệm, với một nút gốc trong ngăn xếp. Sau đó có thể áp dụng quá trình chuyển đổi sau:

* + LEFT – ARC: đánh dấu item thứ hai của ngăn xếp như là một sự phụ thuộc với item đầu tiên, loại bỏ item thứ 2 khỏi ngăn xếp (nếu ngăn xếp có chứa ít nhất 2 item)
  + RIGHT – ARC: đánh dấu item thứ nhất của ngăn xếp như là một sự phụ thuộc với item thứ 2, loại bỏ item thứ nhất khỏi ngăn xếp (nếu ngăn xếp có chứa ít nhất 2 item)
  + SHIFT: loại bỏ một từ trong bộ đệm và đẩy nó vào ngăn xếp (nếu ngăn xếp không rổng)

Với 3 quá trình chuyển đổi này, một parser có thể tạo ra bất kỳ dependency parse nào. Trong mỗi lần chuyển đổi ta phải xác định loại quan hệ giữa từ đó và từ phụ thuộc nó.

* Ưu điểm: hiệu suất cao.