

Analisi Sintattica Automatica

Giorgio Satta
Università di Padova

Pisa, 12 ottobre 2013

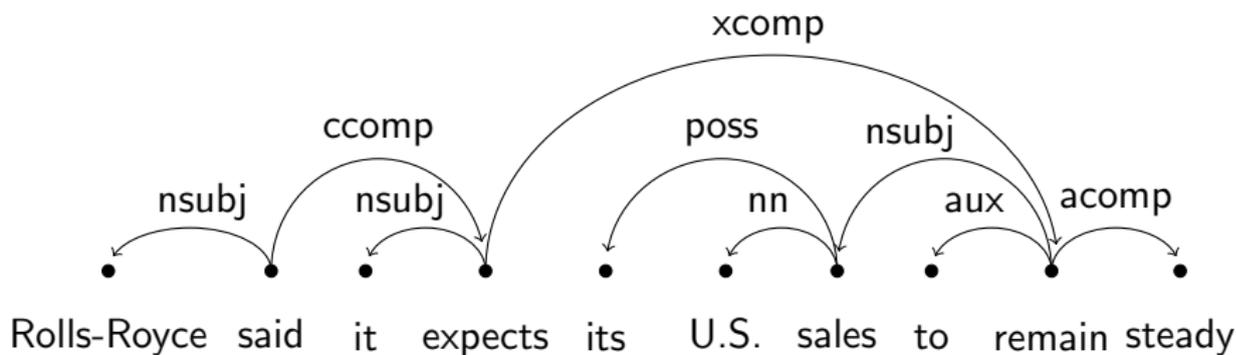
La **sintassi** è un insieme astratto di **regole** che stabilisce quali **combinazioni di parole** siano corrette in una determinata lingua

L'**analisi sintattica automatica** della frase produce una struttura che rappresenta le **relazioni grammaticali** realizzate dalla frase stessa

Gli **algoritmi** che producono l'analisi sintattica di una frase sono indicati dagli specialisti con il termine di **parser**

Albero delle Dipendenze

Esempio : La grammatica valenziale o **grammatica alle dipendenze**, sviluppata ad opera del linguista francese Lucien Tesnière (1893-1954)



Il parser è un **componente fondamentale** in tutte le applicazioni basate sulla **comprensione** del linguaggio naturale :

- **IBM Watson**, sistema di **risposta automatica** a domande in linguaggio naturale; nel 2013 ha battuto i due campioni storici del quiz televisivo statunitense **Jeopardy!**
- **Google Knowledge Graph**, sistema che **estrae informazioni** dai documenti Web, e le rappresenta come una **rete** di relazioni tra concetti ed entità
- **Facebook Graph Search**, motore di ricerca in linguaggio naturale integrato nella versione inglese del **social network**

Sino agli anni '80, i parser erano basati su **sistemi di regole** prodotte manualmente da esperti linguisti

Principali svantaggi di questa metodologia :

- **costi** di progettazione e manutenzione molto elevati
- **precisione** limitata nelle applicazioni su larga scala
- **dipendenza** del sistema da una specifica lingua

A partire dagli anni '90, i parser impiegano tecniche di **apprendimento automatico**, e vengono **addestrati** sulla base di esempi linguistici

Principali vantaggi dei moderni parser :

- **precisione** elevata: 93% relazioni grammaticali corrette sull'Inglese (parser alle dipendenze)
- **velocità**: 10000 parole al secondo (parser alle dipendenze)
- **trasportabilità** a diverse lingue

Il parser è un **componente fondamentale** in applicazioni basate sulla **comprensione** del linguaggio naturale

La ricerca sul parsing del linguaggio naturale è **multidisciplinare** e coinvolge l'informatica, la linguistica e le scienze cognitive

In Italia i **maggiori gruppi** di ricerca in questo campo operano presso l'Università di Pisa e l'Università di Torino