

Chapter 7. Natural Language Processing

The Quest for Artificial Intelligence, Nilsson, N. J., 2009.

Lecture Notes on Artificial Intelligence, Spring 2016

Summarized by Jang, Ha-Young and Lee, Chung-Yeon

Biointelligence Laboratory
School of Computer Science and Engineering
Seoul National University

Contents

7.1 Linguistic Levels

- Levels of Language Analysis

- Phrase-Structure Grammar

- Parse Tree

- Semantic Analysis

7.2 Machine Translation

- First Suggestion

- Demonstration of possibility

- Disappointed Subsequent Work

7.3 Question Answering

- Communication with Computers Using Natural Language

Overview of Chapter 7

- To distinguish languages such as English from the languages used by computers, the former are usually called “natural languages.”
- In artificial intelligence, ‘understanding’ natural language input usually means either converting it to some kind of memory model (such as the one used by Raphael in his SIR system or the semantic network used by Quillian) or the evocation of some action appropriate to the input.
- Natural languages are spoken as well as written. And, because speech sounds are not as well segmented as are the characters printed on a page, speech understanding presents additional difficulties.
- Translating from one language to another involves both understanding and generation. So does carrying on a conversation. All of these problems - understanding, generation, translation, and conversing - fall under the general heading of “natural language processing” (sometimes abbreviated as NLP).

Chapter 7. Natural Language Processing

7.1 Linguistic Levels

Levels of Language Analysis

- Basic components of language (sounds and word parts)
 - Phonetics, phonology and morphology.
- Syntax
 - “John saw the man with a telescope” has two different meanings depending on its syntactic structure.
- Semantics
 - Determines the meaning of a sentence by employing logical analysis.
- Pragmatics
 - Deals with meanings in the context of specific situations.

Phrase-Structure Grammar

■ Nonterminal symbols

- 'S', 'DET', 'NP', and so on

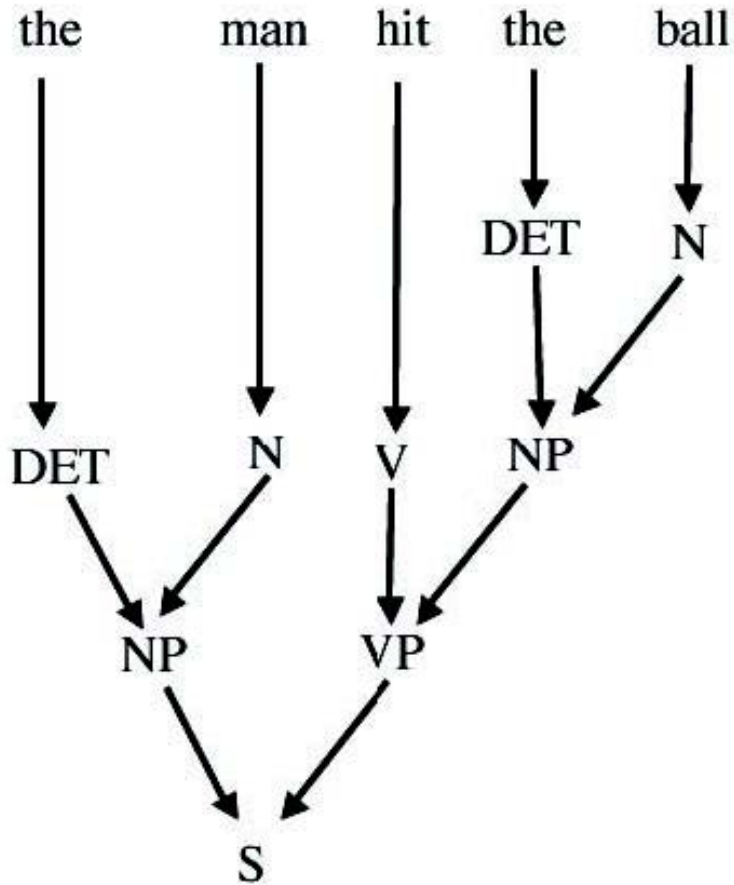
■ Terminal symbols

- Vocabulary words such as 'ball', 'john', and 'threw'

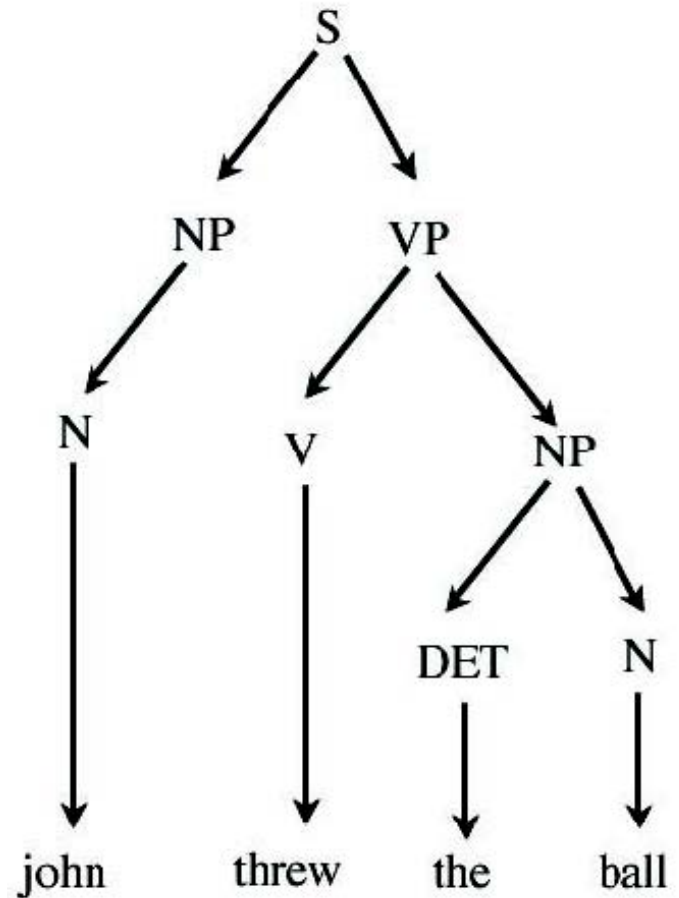
■ Rules

- the | a \rightarrow DET
- man | ball | john \rightarrow N
- hit | took | threw \rightarrow V
- DET N \rightarrow NP
- V NP \rightarrow VP
- NP VP \rightarrow S

Parse Tree



A parse tree for analyzing a sentence



A parse tree for generating a sentence

Semantic Analysis

- Sentences can have more than one parse tree
 - ‘the man hit the ball in the park’
 - ‘ball in the park’
 - ‘hit in the park’
- Deciding which parse tree is appropriate is a job for the semantics or pragmatics level
- Transformation of the input sentence into an expression in some well-defined “meaning representation language”

$$(\exists x, y, z)[\text{Past}(z) \wedge \text{Man}(x, z) \wedge \text{Ball}(y, z) \wedge \text{Event}(z) \wedge \text{Throws}(x, y, z)]$$

Chapter 7. Natural Language Processing

7.2 Machine Translation

Machine Translation: First Suggestion

- Envisioned possibility of machine translation
 - Harry D. Huskey at the National Bureau of Standards.
 - *New York Times* reported on May 31, 1949.
- Initiation of machine translation
 - Warren Weaver, a mathematician and science administrator mentioned the possibility of machine translation in a 1947, letter to Norbert Wiener.
 - Weaver has elaborated his ideas into a memorandum, titled 'Translation' that he sent to some 200 of his acquaintances in various fields.

Machine Translation:

Demonstration of possibility

- **First conference on machine translation**
 - Yehoshua Bar-Hillel, an Israeli logician organized the first conference, in June 1952 at MIT.
- **Demonstration of automatic translation**
 - Russian text to English using a small vocabulary and limited grammar, in January 1954 at IBM World Headquarters.
 - Great deal of excitement and increased funding.

Machine Translation: Disappointed Subsequent Work

- Fully automatic, high quality translation was not feasible
 - Bar-Hillel, Evaluating MT work in a 1959.
 - Because of the apparent difficulty of giving computers the ‘world knowledge’.
- No immediate or predictable prospect of useful machine translation
 - Automatic Language Processing Advisory Committee was formed in April 1964.
 - ALPAC recommended support for basic linguistics science, but not for further support of fully automatic translation in its report in August 1965
- Machine Translation survived
 - The Association for machine Translation and Computational Linguistics held its first meeting in 1962.
 - It changed its name to the Association for Computational Linguistics in 1968.

Chapter 7. Natural Language Processing

7.3 Question Answering

Communication with Computers Using Natural Language

■ BASEBALL

- Written in IPL-V, a special list-processing programming language.
- Answer simple English questions about baseball.
- ‘Where did the Red Sox play on July 7?’ would first be converted to the list:
 - Place = ?
 - Team = Red Sox
 - Month = July
 - Day = 7

■ SAD SAM

- Written in IPL-V.
- Analyze English sentences about family relationships and encode these relationships in a family tree.
- Answer English questions about relationships using this tree.
- Sentence Appraiser and Diagrammer.
 - Parse the input sentences.
- Semantic Analyzing Machine.
 - Extract the semantic information.

Communication with Computers Using Natural Language

■ Conversation with a book

- Robert F. Simmons, a psychologist and linguist.
- Synthex project and Protosynthex system.
 - Protosynthex could answer the simple questions with loaded child's encyclopedia.
 - Pioneering effort in the use of natural language for text retrieval.

■ STUDENT

- Solve algebra 'story problems' given to in in a restricted subset of English.
- Example:

The distance from New York to Los Angeles is 3000 miles. If the average speed of a jet plane is 600 miles per hour, find the time it takes to travel from New York to Los Angeles by jet.