

Introduction to Natural Language Understanding

Goal: To create *computational models* of language in enough detail that you could write computer programs to perform various tasks involving natural language.

Ultimate Goal: To be able to specify models that approach human performance in the linguistic tasks of reading, writing, hearing, and speaking.

Dual purpose:

scientific: to explore the nature of linguistic communication

practical: to enable effective human-machine communication

Slide CS472 – NLP 1

NLU Draws Heavily from Other Disciplines

linguistics	emphasis on structural regularities
psycholinguistics	emphasis on the process of language comprehension and production
philosophy	emphasis on logical accounts of communicative abilities
computational linguists	emphasis on building a computational theory of language comprehension and language generation

Slide CS472 – NLP 2

Information Retrieval (IR)

Choose from a set of documents the ones that are relevant to a query.

Find me all articles on leveraged buyouts.

Find me all articles on leveraged buyouts involving more than 100 million dollars that were attempted but failed during 1986 and 1990.

Slide CS472 – NLP 3

Machine Translation

Translate from one natural language to another.

The spirit is willing, but the flesh is weak.

The extension of the coverage of the health services to the underserved or not served population of the countries of the region was the central goal of the Ten-Year Plan and probably that of greater scope and transcendence.

Welcome to Chinese Restaurant. Please try your Nice chinese Food With chopsticks. the traditional and typical of Chinese glorious history and cultural. PRODUCT OF CHINA

Slide CS472 – NLP 4

Information Extraction

Read a text and derive from it some assertions that can be put into a structured database.

BOGOTA, 9 JAN 90 (EFE) – RICARDO ALFONSO CASTELLAR, MAYOR OF ACHI, IN THE NORTHERN DEPARTMENT OF BOLIVAR, WHO WAS KIDNAPPED ON 5 JANUARY, APPARENTLY BY ARMY OF NATIONAL LIBERATION (ELN) GUERRILLAS, WAS FOUND DEAD TODAY, ACCORDING TO AUTHORITIES. CASTELLAR WAS KIDNAPPED ON 5 JANUARY ON THE OUTSKIRTS OF ACHI, ABOUT 850 KM NORTH OF BOGOTA, BY A GROUP OF ARMED MEN, WHO FORCED HIM TO ACCOMPANY THEM TO AN UNDISCLOSED LOCATION.

Slide CS472 – NLP 5

Summary

Date: 05 JAN 90

Location: COLOMBIA: BOLIVAR (DEPARTMENT): ACHI (TOWN)

Type: KIDNAPPING

Weapon: *

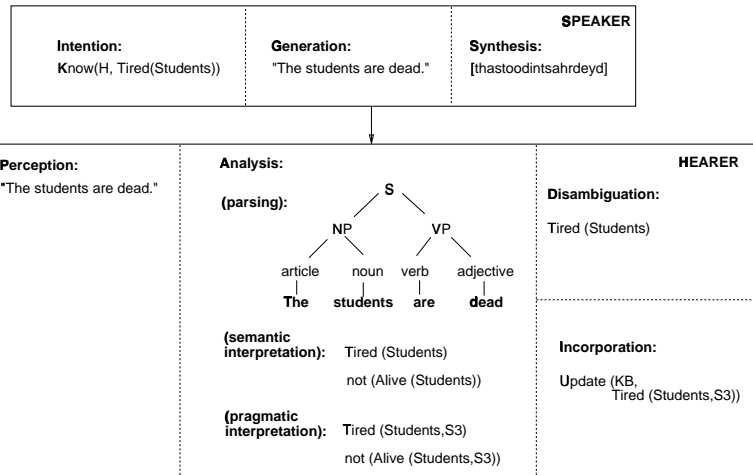
Victim: “RICARDO ALFONSO CASTELLAR” (“MAYOR OF ACHI”)

Perpetrator: “GROUP OF ARMED MEN”

Organization: “ARMY OF NATIONAL LIBERATION (ELN)”

Slide CS472 – NLP 6

Component Steps of Communication



Slide CS472 – NLP 7

Why is NLP hard?

From Professor Lee's lecture...

- massive ambiguity at all levels of analysis
- NLP is AI-complete

[Professor Lee's notes are on-line via CS150, slides 1-20.]

Case study: noun phrase coreference [separate slides]

Slide CS472 – NLP 8