

Natural Language Processing

Objectives of NLP

- To study the nature of language (Linguistics)
- Window into cognition (Psychology)
- Human interface technology
- Text translation
- Information management

The history of NLP

- 1948- 1st NLP application
 - dictionary look-up system
 - developed at Birkbeck College, London
- 1949- American interest
 - WWII code breaker Warren Weaver.
 - He viewed German as English in code.
- 1950- Machine Translation (Russian to English)
- 1966- Over-promised under-delivered
 - Machine Translation worked only word by word
 - NLP brought the first hostility of research funding agencies.
 - NLP gave AI a bad name before AI had a name.

NLP (backup and focus)

- NLP looked to Linguistics
 - Linguistics is language described, not prescribed.
 - Linguistics had few applicable theories for Machine Translation
- 1957- Noam Chomsky's *Syntactic Structures* revolutionized Linguistics as it applies to Machine Translation.
 - Rule based system of syntactic structures.
 - Believed there are features common to all languages that enable people to speak creatively and freely.
 - Hypothesized all children go through the same stages of language development regardless of the language they are learning.

Results of NLP Refocus

- 1957- NLP community decided that Sentences could not be processed without preformatting.
- 1958- Bar-Hillel report
 - Concluded Fully-Automatic High-Quality Translation (FAHQT) could not be accomplished without knowledge.
- 1965- All funding of NLP came to a grinding halt thanks to ALPAC report.
 - Public spent 20 million and did not like the reports generated by NLP's refocus.

1966-1980 Results

- 1967-Procedural Semantics (Woods)
 - Theory of the “meaning” of sentence.
- 1970-Augmented Transition Networks (Thorne, Woods)
 - piece of searching software that is capable of using very powerful grammars to process syntax
- Case Grammar (linking prepositions)
 - I saw a man on the hill with a telescope.

Major Issues in NLP

- **Ambiguity in Language**
- **Meaning of Language is Context Sensitive.**
- **Micro World delusion.**

Ambiguity Makes NLP difficult

- Syntactic
 - I saw the Grand Canyon flying to New York.
- Word Sense
 - The man went to the bank to get some cash.
 - The man went to the bank and jumped in the river.

Meaning of Language is Context Sensitive.

- ```
.for(int j = 0; j < 10; j++)
{
 //System.out.println("Hello NLP");
}
```
- I am going to kill you! (angry big sister or ???)
- That was bad/cool. Etc...

## Micro World delusion.

- Hope
  - A Machine Translator that can translate small words will soon be able to translate United Nation Speeches on the fly.

## Exam Question from Cambridge Dept. of NLP.

- Describe the problem Worst case (exponential) syntactic ambiguity causes and its significance for natural language processing.
  1. Answer...
    - “A computer that understands syntax, must know the semantics as well (what it means).
    - A sentence can produce a number of different analyses.
      - I am going to throw a ball today. (football player)
      - I am going to throw a ball today. (Prince Charming)
      - John saw the woman in the park with a telescope.

# Successful NLP Systems

- 1973- Shrdlu (Terry Winograd)
- 1973- Lunar (Woods)
  - Both (above) Applied Procedural Semantics
- TAUM-METEO (University of Montreal.)
- SYSTRAN (Xerox)
- Google language tools
- Microsoft Spell check.

## Shrdlu

“Put the red pyramid ontop of the green cube”

- **Good**
  - One could tell Shrdlu what to do in English
  - You could ask Shrdlu questions.
  - The System really seemed to understand what you were talking about.
- **Bad**
  - Reference ambiguity problem (scene w/ 3 pyramids)
    - Question: Take your pyramid to the left corner! Answer: “I don’t understand which pyramid you mean.”
  - Micro world success.

# Lunar

What is the average modal plagioclase concentration for lunar samples that contain rubidium?"

- **Good**
  - Allowed geologists to ask questions about the chemical analysis data of lunar rock and soil samples.
  - 78% success rate
- **But...**
  - Was never put into production because of limitations.

# TAUM-METEO

- Translates weather reports from English to French.
- Works because language used in reports is stylized and regular.
- Still used in Canada!

## SYSTRAN, developed by Xerox.

- Translated Xerox manuals into all languages that Xerox deals with.
- Utilizes pre-edited texts.