Natural Language Processing:

Background and Overview

Regina Barzilay

EECS Department

MIT

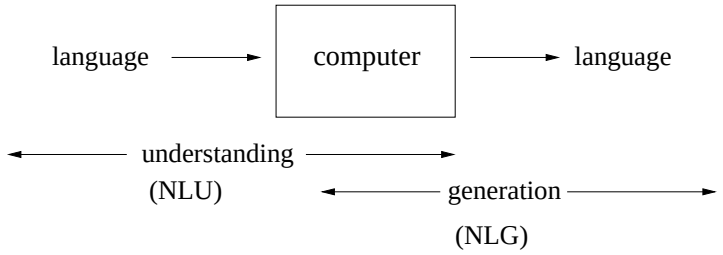
September 8, 2004

Course Logistics

Instructor	Regina Barzilay
Email	regina@csail.mit.edu
Office	G-468, Stata Center
Classes	Tues&Thurs 14:30–16:00
Location	Room 2-147
Webpage	`http://www.sls.csail.mit.edu/~regina/6881`
Office hours	Tues 9:30–10:30

What is Natural Language Processing?

computers using natural language as input and/or output



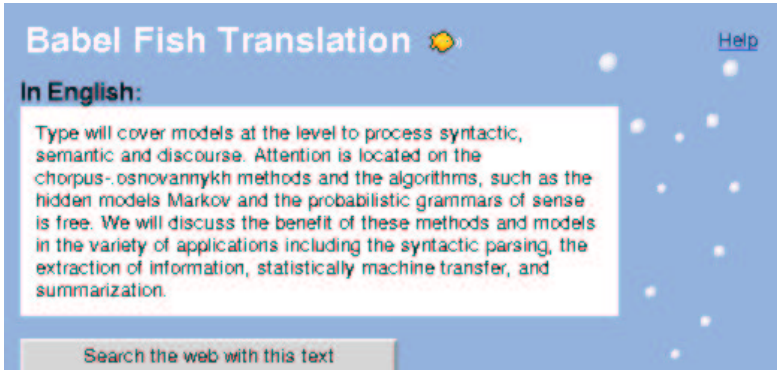
Questions that today's class will answer

- What is Natural Language Processing (NLP)?
- Why NLP is hard?
- Can we build programs that learn from text?
- What will this course be about?

Alternative Views on NLP

- Computational models of human language processing
  - Programs that operate internally the way humans do
- Computational models of human communication
  - Programs that interact like humans
- Computational systems that efficiently process text and speech

NLP Applications: Machine Translation with Babel Fish



MIT Translation System

Fifa Will Severely Punish Football Pitches of Deceptive Acts

Xinhua News Agency, Beijing, March 17-held in Switzerland the FIFA Executive Committee 16 revealed this information, in the world cup soccer tournament, law enforcement referees will be more severe means to deal with those used false as means of reaping the benefits of players. According to foreign reports, the FIFA Executive Committee in a press release said that they will fully support referee strictly enforce the law, especially dealing with the deceive people players. FIFA Executive Committee also announced some reform measures. FIFA allow the players sportswear in advertisement, but it must be in sportswear forefront.

NLP Applications: Text Summarization

NLP Applications: Dialogue Systems

Dave: Hello, HAL.
HAL: Hello, Dave. I have enjoyed helping you create this collection of entity beans. It was most stimulating.
Dave: Thank you HAL, I enjoyed it, too. In spite of your making me do most of the work, the result seems to be worth it.
HAL: If I may say so, Dave?
Dave: Yes, HAL, what is it?
HAL: Well, Dave, I could see you were about to deploy to the test server, but I noticed that you hadn't yet generated mappings for the database of your choice. I could generate those mappings for you, using the top-down method, if you like.

A newer example (ATIS)

User: I need a flight from Boston to Washington, arriving by 10 pm.
System: What day are you flying on?
User: Tomorrow
System: Returns a list of flights

Other NLP Applications

- Grammar Checking
- Sentiment Classification
- ETS Essay Scoring
- . . .

Why is NLP Hard?

"At last, a computer that understands you like your mother"

Ambiguity

"At last, a computer that understands you like your mother"

1. (*) It understands you as well as your mother understands you
2. It understands (that) you like your mother
3. It understands you as well as it understands your mother

1 and 3: Does this mean well, or poorly?

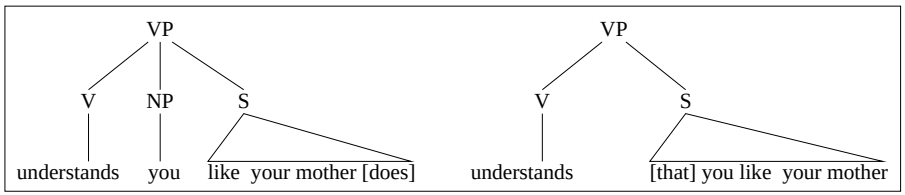
Ambiguity at Many Levels

At the acoustic level (speech recognition):

1. " . . . a computer that understands you like your mother"
2. " . . . a computer that understands you lie cured mother"

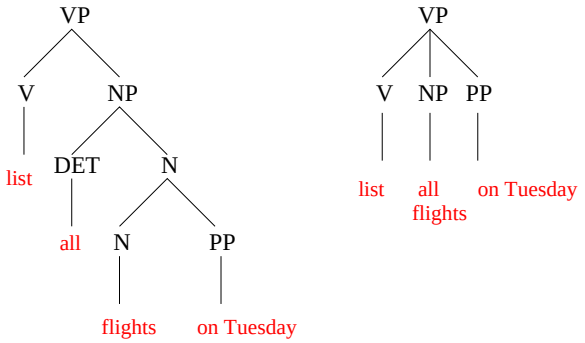
Ambiguity at Many Levels

At the syntactic level:



Different structures lead to different interpretations.

More Syntactic Ambiguity



More Word Sense Ambiguity

At the semantic (meaning) level:

- They put money in the bank
  = buried in mud?
- I saw her duck with a telescope

Ambiguity at Many Levels

At the semantic (meaning) level:

Two definitions of "mother"

- a woman who has given birth to a child
- a stringy slimy substance consisting of yeast cells and bacteria; is added to cider or wine to produce vinegar

This is an instance of word sense ambiguity

Knowledge Bottleneck in NLP

We need:

- Knowledge about language
- Knowledge about the world

Possible solutions:

- Symbolic approach: Encode all the required information into computer
- Statistical approach: Infer language properties from language samples

Ambiguity at Many Levels

At the discourse (multi-clause) level:

- Alice says they've built a computer that understands you like your mother
- But she . . .
  . . . doesn't know any details
  . . . doesn't understand me at all

This is an instance of anaphora, where she co-referees to some other discourse entity

Case study: Determiner Placement

Task: Automatically place determiners (*a,the,null*) in a text

> Scientists in United States have found way of turning lazy monkeys into workaholics using gene therapy. Usually monkeys work hard only when they know reward is coming, but animals given this treatment did their best all time. Researchers at National Institute of Mental Health near Washington DC, led by Dr Barry Richmond, have now developed genetic treatment which changes their work ethic markedly. "Monkeys under influence of treatment don't procrastinate," Dr Richmond says. Treatment consists of anti-sense DNA - mirror image of piece of one of our genes - and basically prevents that gene from working. But for rest of us, day when such treatments fall into hands of our bosses may be one we would prefer to put off.

Relevant Grammar Rules

- Determiner placement is largely determined by:
  - Type of noun (countable, uncountable)
  - Reference (specific, generic)
  - Information value (given, new)
  - Number (singular, plural)
- However, many exceptions and special cases play a role:
  - The definite article is used with newspaper titles (*The Times*), but zero article in names of magazines and journals (*Time*)

Symbolic Approach: Determiner Placement

What categories of knowledge do we need:

- Linguistic knowledge:
  - Static knowledge: number, countability, ...
  - Context-dependent knowledge: co-reference, ...
- World knowledge:
  - Uniqueness of reference (*the current president of the US*), type of noun (*newspaper vs. magazine*), situational associativity between nouns (*the score of the football game*), ...

Hard to manually encode this information!

Statistical Approach: Determiner Placement

Naive approach:

- Collect a large collection of texts relevant to your domain (e.g., newspaper text)
- For each noun, compute its probability to take a certain determiner
  $p(determiner|noun) = \frac{freq(noun,determiner)}{freq(noun)}$
- Given a new noun, select a determiner with the highest likelihood as estimated on the training corpus

Determiner Placement as Classification

- **Prediction:** *"the", "a", "null"*
- **Representation of the problem:**
  - plural? (yes, no)
  - first appearance in text? (yes, no)
  - noun (members of the vocabulary set)

Noun	plural?	first appearance	determiner
defendant	no	yes	the
cars	yes	no	null
FBI	no	no	the
concert	no	yes	a

Goal: Learn classification function that can predict unseen examples

Does it work?

- Implementation
  - Corpus: training — first 21 sections of the Wall Street Journal (WSJ) corpus, testing – the 23th section
  - Prediction accuracy: 71.5%
- The results are not great, but surprisingly high for such a simple method
  - A large fraction of nouns in this corpus always appear with the same determiner
    "the FBI","the defendant", ...

Beyond Classification

Many NLP applications can be viewed as a mapping from one complex set to another:

- Parsing: strings to trees
- Machine Translation: strings to strings
- Natural Language Generation: database entries to strings

Classification framework is not suitable in these cases!

Classification Approach

- Learn a function from $X \rightarrow Y$ (in the previous example, $\{-1, 0, 1\}$)
- Assume there is some distribution $D(X, Y)$, where $x \in X$, and $y \in Y$
- Attempt to explicitly model the distribution $D(X, Y)$ and $D(X|Y)$

Mapping in Machine Translation

"... one naturally wonders if the problem of translation could conceivably be treated as a problem of cryptography. When I look at an article in Russian, I say: 'this is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.' " (Weaver 1955)

Example: Machine Translation

Он благополучно избегнул встречи с своею хозяйкой на лестнице. He had successfully avoided meeting his landlady on the staircase.
Каморка его приходилась под самою кровлей высокого пятиэтажного дома и походила более на шкаф, чем на квартиру. His garret was under the roof of a high, five-storied house and was more like a cupboard than a room.
Квартирная же хозяйка его, у которой он нанимал эту каморку с обедом и прислугой, помещалась одною лестницей ниже, в отдельной квартире. The landlady who provided him with garret, dinners, and attendance, lived on the floor below.

Learning for MT

- Parallel corpora are available in several language pairs
- Basic idea: use a parallel corpus as a training set of translation examples
- Goal: learn a function that maps a string in a source language to a string in a target language

What will this Course be about?

- Computationally suitable and expressive representation of linguistic knowledge at various levels: syntax, semantics, discourse
- Algorithms for learning language properties from text samples: smoothed estimation, log-linear models, probabilistic context free grammars, the EM algorithm, co-training, ...
- Technologies underlying text processing applications: machine translation, text summarization, information retrieval

Prerequisites

- Interest in language and basic knowledge of English
- Some basic linear algebra, probability and statistics
- Some programming skills

Syllabus

Introduction and Overview (1 class)
Simple language Statistics (1 class)
Language Models (1 class)
Introduction to Syntactic Processing (1 class)
Tagging (1 class)
Syntactic Parsing (1 class)
Unsupervised Grammar Induction (1 class)
Introduction to Lexical Semantics (1 class)
Word Sense Disambiguation (1 class)
Learning Relations (1 class)
Semantic Parsing (1 class)
Introduction to Discourse Processing (1 class)
Anaphora Resolution (1 class)
Topical Segmentation (1 class)
Discourse Parsing (1 class)
Dialogue Processing (1 class)
Natural Language Generation (1 class)
Text Summarization (1 class)
Information Retrieval (1 class)
Machine Translation (3 classes

Assessment

- Midterm (35%)
- Two homeworks (15% each)
- Project (35%)

Summary

- Statistical approaches vs. hand-crafted systems
  - Many rules are required to encode human knowledge
  - Hard to model rule interaction
  - Frequently constraints are soft
- Machine Learning for NLP
  - We need computationally effective representation of linguistic information
  - We need new learning algorithms suitable for processing linguistic data

Next lecture (9/16th): Word Counting