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

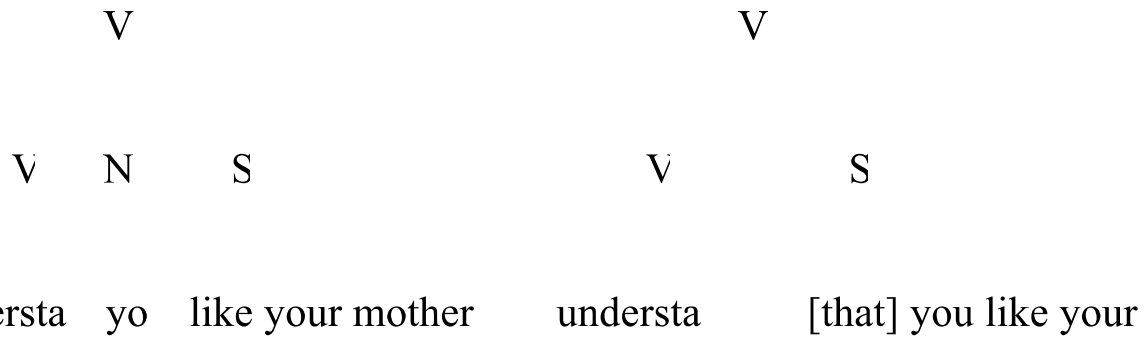
1 and 3: Does this mean well, or poorly?

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"

At the syntactic level:



Different structures lead to different interpretations.

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

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

Thisis aninstance of anaphora,whereshe co-referees to some other discourse entity

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

Case study: Determiner Placement

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

Relevant Grammar Rules

Determiner placement is largely determined by:

1. Type of noun (countable, uncountable)
2. Reference (specific, generic)
3. Information value (given, new)
4. 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 seen during training, compute its probability to take a certain determiner
  p(determiner|noun) = f req(noun,determiner)/f req(noun)
- Given a new noun, select a determiner with the highest likelihood as estimated on the training corpus