Introduction to Natural Language Processing (NLP)

Source: Elaine Rich (UT Austin)

Spoken Language

The dream: HAL (2001 A Space Odyssey)

Going Both Ways

- Understanding

- Generation

English: Put the kid's cereal on the bottom shelves.

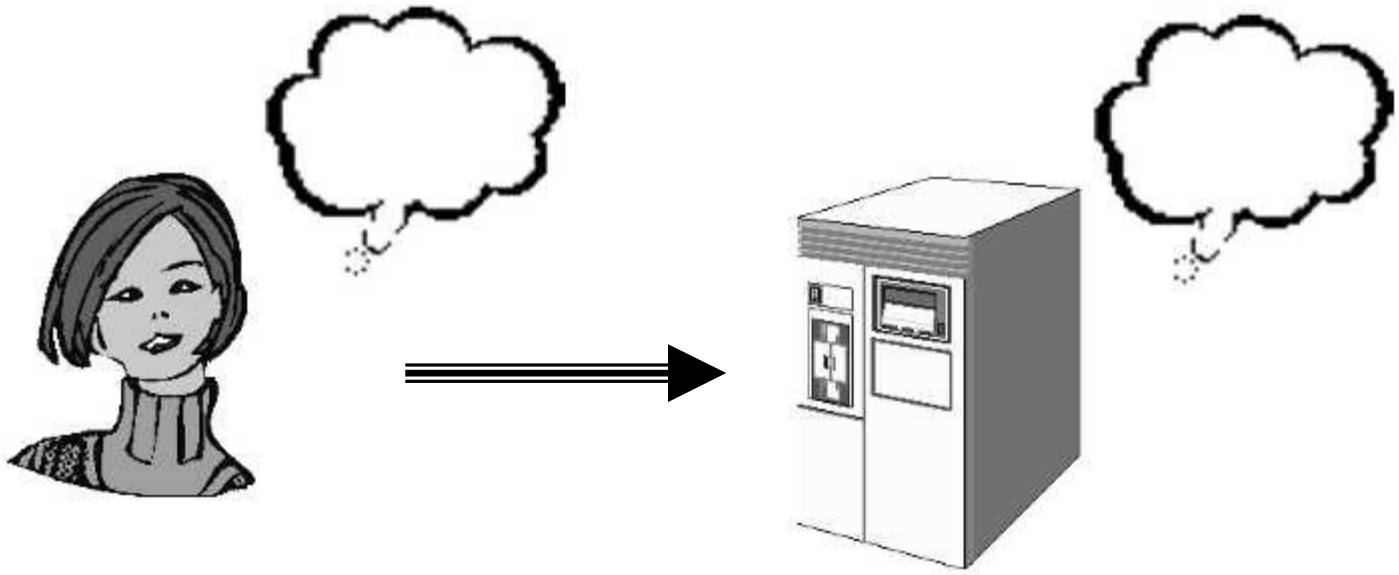
Java

```
import java.util.ArrayList;
public class GroceryStore
{
  private int[][][] shelves;
  private ArrayList products;

  public void placeProducts(String productFile)
  { FileReader r = new FileReader(productFile);
    GroceryItemFactory factory = new GroceryItemFactory();

    while(r.hasNext())
      products.add( factory.createItem(r.readNext()));

    ThreeDLoc startLoc;
    GroceryItem temp;
    for(itemNum = 0; itemNum < products.size(); itemNum++)
    { temp = (GroceryItem)(products.get(itemNum))
      startLoc = temp.getPlacement(this);
      shelves[startLoc.getX()][startLoc.getY()][startLoc.getY()]=
        tempgetIDNum();
      }
  }
}
```

Java, Continued

```
public class ChildrensCereal extends GroceryItem
{   private static final int PREFERRED_X = -1;
    private static final int PREFERRED_Y = 0;
    private static final int PREFERRED_Z = 0;

    public ThreeDLoc getPlacement(GroceryStore store)
    {   ThreeDLoc result = new ThreeDLoc();
        result.setX(store.find(this));
        result.setY(PREFERRED_Y);
        result.setZ(PREFERRED_Z);
        return result;
    }
}
```

It's All about Mapping

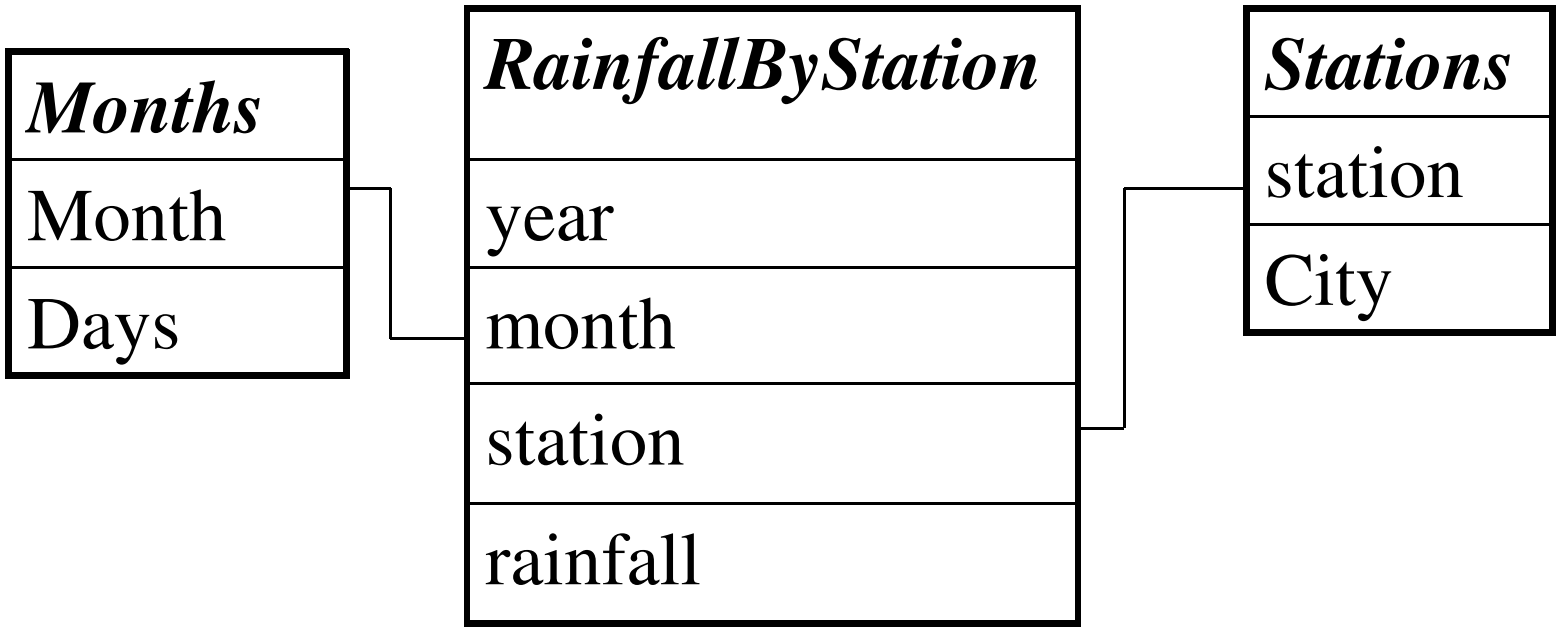


Put the kids cereal on the bottom shelves.

What Are We Going to Map to?

English: Do you know how much it rains in Austin?

The database:



English: What is the average rainfall, in Austin, in months with 30 days?

SQL:

```
SELECT Avg(RainfallByStation.rainfall) AS
   AvgOfrainfall FROM Stations INNER JOIN
   (Months INNER JOIN RainfallByStation
ON Months.Month =
                RainfallByStation.month)
ON Stations.station =
                RainfallByStation.station
HAVING (((Stations.City)="Austin")
AND ((Months.Days)=30));
```

Ambiguity – the Core Problem

- Time flies like an arrow.
- I hit the boy with the blue shirt (a bat).
- I saw the Grand Canyon (a Boeing 747)
  flying to New York.
- I know more beautiful women than Kylie.
- I only want potatoes or rice and beans.
- Is there water in the fridge?
- Who cares?
- Have you finished writing your paper?
- I've written the outline.

Designing a Mapping Function

- Morphological Analysis and Part Of Speech tagging
  - *The womans goed home.*
- Syntactic Analysis (Parsing)
  - *Fishing went boys older*
- Extracting Meaning
  - *Colorless green ideas sleep furiously.*
  - *Sue cooked. The potatoes cooked.*
- Putting it All in Context
  - *My cat saw a bird out the window. It batted at it.*

Morphological Analysis and POS Tagging

Morphological Analysis:

played = play + ed = play + PAST

saw = see + PAST

compute → computer → computerize → computerization

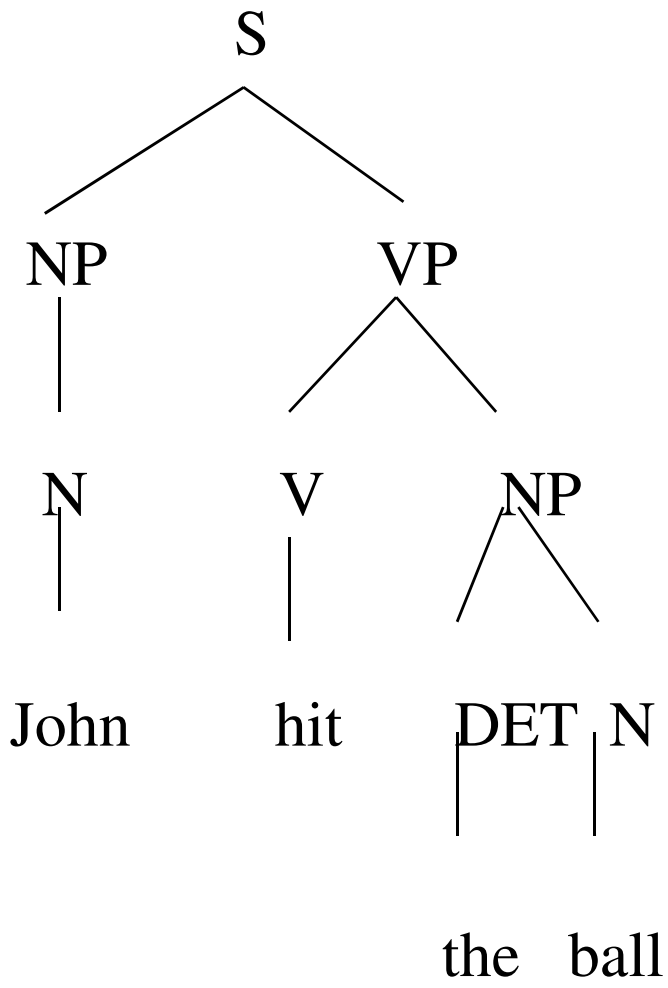
POS Tagging:

I hit the bag.

Parsing

John hit the ball.

(S (NP (N John))
(VP (V hit)
(NP (DET
the)
(N
ball))))



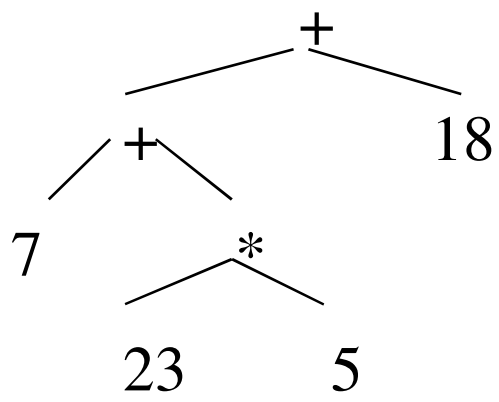
Syntax: Dealing with Ambiguity

English:

Water the flowers with the hose.

Water the flowers with brown leaves.

Java: 7 + 23 * 5 + 18



Using Domain Knowledge

(plant (isa living thing))

(flower (isa plant)

(has parts leaf))

(water (isa action)

(instrument mustbe container))

(hose (isa container))

Syntax: Gapping

English: Who did you say Mary gave the ball to?

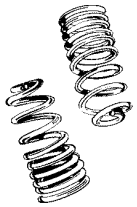
Java: 7 + 23 * 5 + 18

Semantics: The Meaning of Words

Getting it right for the target application:

"month" ⇒ RainfallByStation.month

Dealing with ambiguity:

"spring" ⇒  or  or 

"stamp" ⇒  or 

Noun Phrases Describe Objects

Corn oil

Coconut oil

Cooking oil

Baby oil

How do Modifiers Work?

Cat

French cat

Siamese cat

House cat

Toy cat

Toy poodle

Putting Phrases Together

Bill cooked the potatoes.

The potatoes cooked in about an hour.

The heat from the fire cooked the potatoes in 30 minutes.

(cooking-event (agent)
(object)
(instrument)
(time-frame)

What About Applications Where Almost is OK?

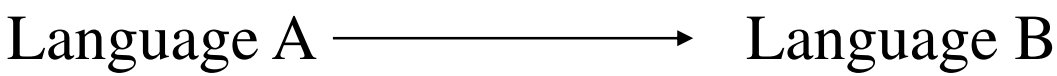
- Searching the web
  - Leaving some of the work for people
  - Retrieval failures are ok
- Snooping

Going the Other Way: Generation

```
(c (isa cooking-event)
        (agent x )
        (object y)
        (instrument  z)
        (time-frame   ))
(x (isa man)
        (name Bill)
        (height 6')
        (attire (head-
            covering h))
        (born-location b))
(y (some-of  potatoes)
        (type-of Idaho))
        (maturity new))
(z (isa microwave)
        (brand Sharp))
(h (isa gimme)
        (color red))
(gimme (subclass hat))
(b (isa city)
    (name Austin))
```

Machine Translation

Direct mapping:



Using an intermediate form:

MT Examples

English: Do you know how much it rains in Austin?

Spanish: ¿Usted sabe cuánto llueve en Austin?

English: You know how much you rain in Austin?

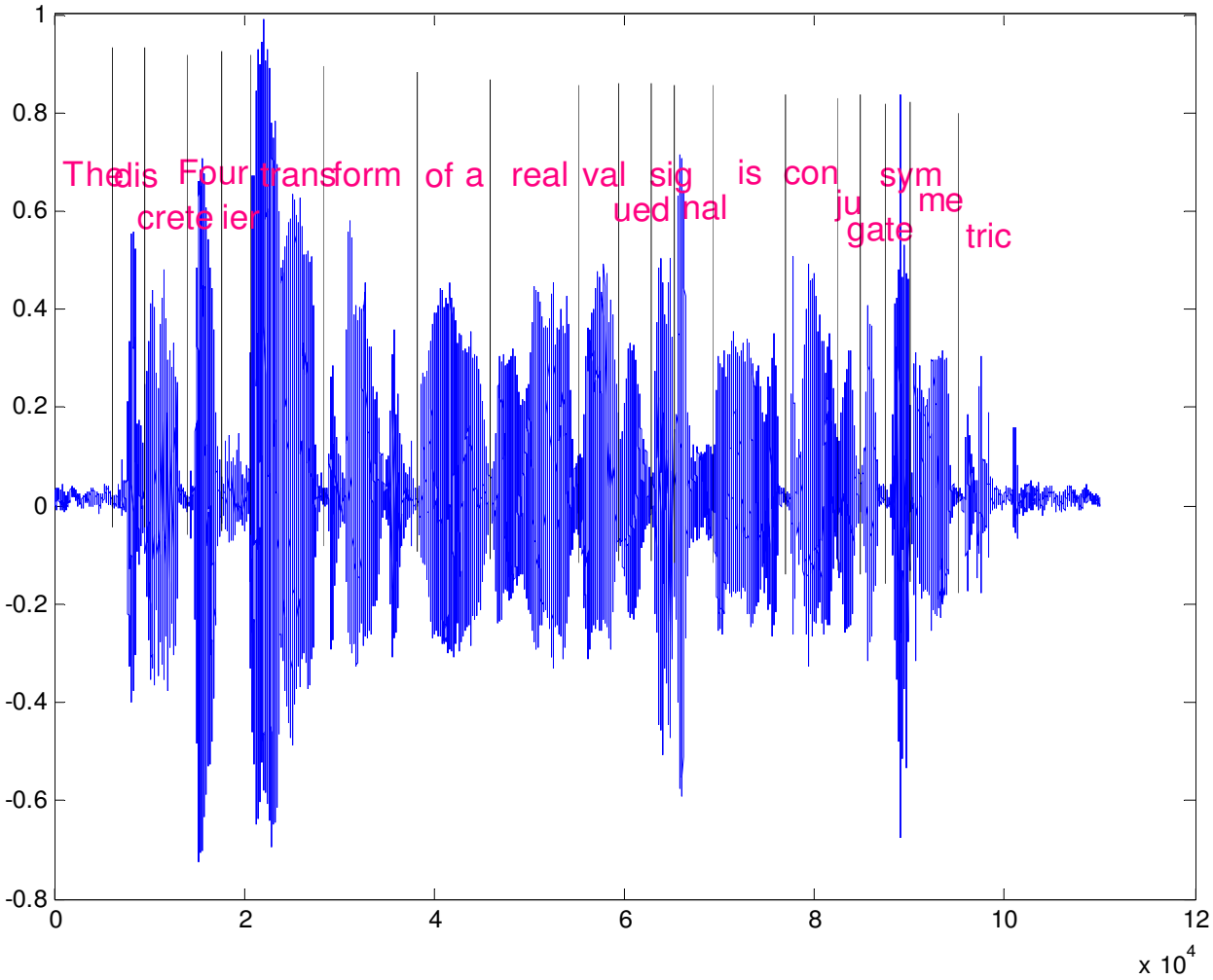
English: Please go buy some baby oil.

Spanish: Va por favor la compra un poco de aceite de bebé.

English: In order to please buy a little baby oil.

What if we have tons of data?

Spoken Language - Understanding



Spoken Language - Generation

The issues:

- Figuring out what to say
- Pronouncing words
- Linking them together
- Getting the prosody (rhythm / pattern) right

Evolution of NLP

1964 STUDENT solves algebra word problems

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.

1965 ELIZA models a Rogerian therapist ("chat bot")

Use at http://nlp-addiction.com/eliza/

young woman: *Men are all alike.*

eliza: *In what way?*

young woman: *They're always bugging us about something specific or other.*

eliza: *Can you think of a specific example?*

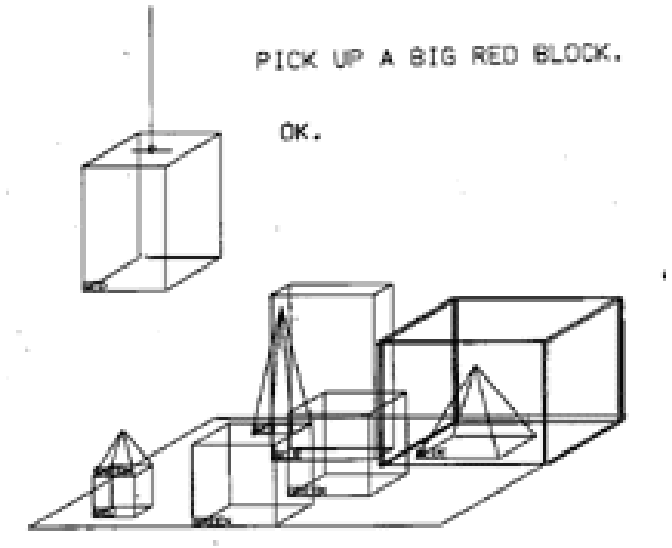
young woman: *Well, my boyfriend made me come here.*

eliza: *Your boyfriend made you come here?*

Evolution of NLP, continued

1966 Alpac (Automatic Language Processing Advisory Committee) report kills work on MT

1971 SHRDLU (Winograd)



Evolution of NLP, continued

1973 Schank – a richer limited domain: children's stories

1977 Schank – scripts add a knowledge layer – restaurant stories

1970's and 80's sophisticated grammars and parsers

But suppose we want generality? One approach is "shallow" systems that punt the complexities of meaning.

Evolution of NLP, continued – MT

1949 Warren Weaver's memo suggesting MT

1966 Alpac report kills government funding

Early 70s SYSTRAN develops direct Russian/English system

Early 80s knowledge based MT systems

Late 80s statistical MT systems

Today widely available but not very good MT engines

Is MT an "AI complete" problem?

John saw the bicycle in the store window. He wanted it.