6.891 (Fall 2003): Machine Learning Approaches for Natural Language Processing

Michael Collins
MIT CSAIL/EECS

Lecture 1: September 3, 2003

Welcome to 6.891!

- Overview and Goals
- General Information
- Syllabus
- An overview of the course...

Course Coverage

- **NLP sub-problems:** part-of-speech tagging, parsing, word-sense disambiguation, etc.

- **Machine learning techniques:** probabilistic context-free grammars, hidden markov models, estimation/smoothing techniques, the EM algorithm, log-linear models, cotraining, etc.

- **Applications:** information extraction, machine translation, natural language interfaces...

General Information

Evaluation:

- Survey paper (30% of final grade). Read three papers on a particular research project, and write a paper that synthesizes and compares the results of the three papers.

- Project (70% of final grade). A course project on some research problem in machine learning for natural language.

General Information

Intended goals of the course:

- By the end of the course, you should have the background to read research papers in the general area of statistical NLP; and be able to start research in the area.
- Techniques and problems should be of interest to researchers in related fields: e.g., speech recognition, computational biology, linguistics, brain and cognitive sciences (psycholinguistics), machine learning, statistics, computer vision, probabilistic methods in AI.

Background required for the course:

- Course should be fairly self-contained; **but** a background in basic probability theory will be very helpful.

Readings for the course:

- I'll give pointers to relevant papers as we go along.
- No set textbook, but *Manning and Schutze: Foundations of Statistical Natural Language Processing; MIT Press* should be useful.

A Syllabus

- Introduction (1 lecture)
- Estimation techniques, and language modeling (1 lecture)
- Stochastic parsing (3 lectures)
- Stochastic tagging, and maximum entropy/log-linear models (2 lectures)
- The EM algorithm for PCFGs, HMMs, and machine translation (2 lectures)
- Information extraction (2 lectures)
- Machine translation (3 lectures)
- Vision and language (1 lecture)
- Dialogue systems (2 lectures)
- NLP issues in different languages,
  e.g., Czech, Chinese, Arabic, German (1 lecture)
- Word-sense disambiguation (1 lecture)
- Semi-supervised approaches (2 lectures)
- Discriminative/reranking approaches, and kernels over NLP structures
  (2 lectures)
- Learning in Optimality Theory (1 lecture)
- Conclusions/open problems (1 lecture)

Different Aspects of the field

- Statistical models
- Algorithms
- Systems and experiments
- Linguistics
- How do we use and acquire language?

Goals of the Field

- Build natural language systems
- Theoretical and experimental understanding of language and learning
- In the longer term: ideas may influence our understanding of cognition

NLP Technology

- Information extraction
- Machine translation
- Vision and language
- Dialogue systems

NLP Technology: Information Extraction

- Goal: Map a document collection $\Rightarrow$ Structured database
- Example: **Whizbang!** labs work on job postings

10TH DEGREE is a full service advertising agency specializing in direct and interactive marketing. Located in Irvine CA, 10TH DEGREE is looking for an Assistant Interactive Account Manager to help manage and coordinate interactive marketing initiatives for a marquee automotive account. Experience in online marketing, automotive and/or the advertising agency field is a plus.

Assistant Account Manager Responsibilities

Ensures smooth implementation of programs and initiatives Helps manage the delivery of projects and key client deliverables ...

Compensation: $50,000 – $80,000 Hiring Organization: 10TH DEGREE

Principals only. Recruiters, please don't contact this job poster. Please, no phone calls about this job! Please do not contact job poster about other services, products or commercial interests. Reposting this message elsewhere is NOT OK. this is in or around Orange County - Irvine

⇓

INDUSTRY	Advertising
POSITION	Assistant Account Manager
LOCATION	Irvine, CA
COMPANY	10th Degree
SALARY	$50,000 – $80,000

NLP Technology: Information Extraction

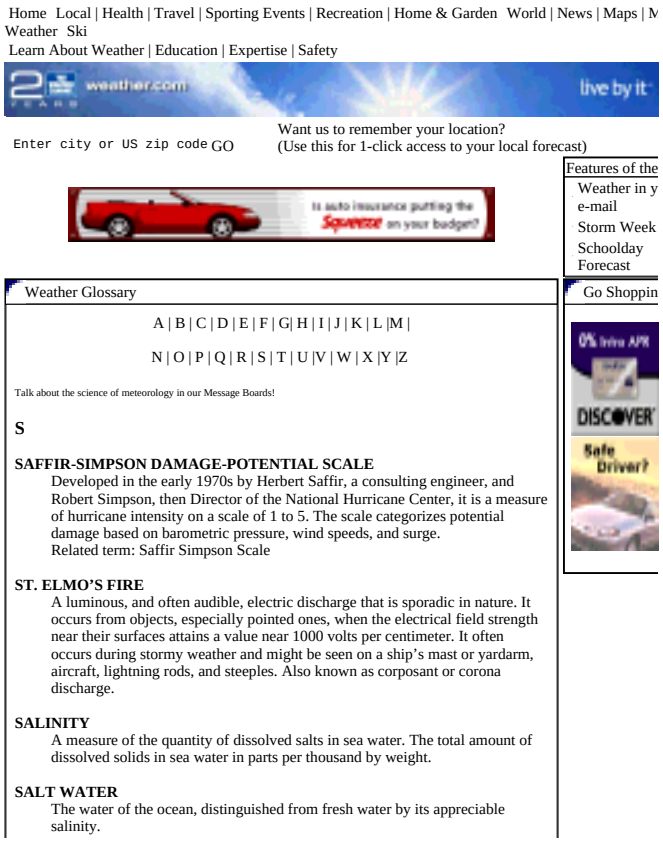
Motivation

- Complex searches (“Find me all jobs in advertising paying at least $50,000 in California”)

- Statistical queries (“How many jobs in advertising are posted in Los Angeles”)

Extracting Glossary Entries from the Web

Input:



Home Local | Health | Travel | Sporting Events | Recreation | Home & Garden World | News | Maps | M
Weather Ski
Learn About Weather | Education | Expertise | Safety

weather.com live by it

Enter city or US zip code GO

Want us to remember your location?
(Use this for 1-click access to your local forecast)

Features of the
Weather in y
e-mail
Storm Week
Schoolday
Forecast

Weather Glossary

A | B | C | D | E | F | G| H | I | J | K | L |M |

N | O | P | Q | R | S | T | U |V | W | X |Y |Z

Talk about the science of meteorology in our Message Boards!

S

SAFFIR-SIMPSON DAMAGE-POTENTIAL SCALE
Developed in the early 1970s by Herbert Saffir, a consulting engineer, and Robert Simpson, then Director of the National Hurricane Center, it is a measure of hurricane intensity on a scale of 1 to 5. The scale categorizes potential damage based on barometric pressure, wind speeds, and surge.
Related term: Saffir Simpson Scale

ST. ELMO'S FIRE
A luminous, and often audible, electric discharge that is sporadic in nature. It occurs from objects, especially pointed ones, when the electrical field strength near their surfaces attains a value near 1000 volts per centimeter. It often occurs during stormy weather and might be seen on a ship's mast or yardarm, aircraft, lightning rods, and steeples. Also known as corposant or corona discharge.

SALINITY
A measure of the quantity of dissolved salts in sea water. The total amount of dissolved solids in sea water in parts per thousand by weight.

SALT WATER
The water of the ocean, distinguished from fresh water by its appreciable salinity.

Go Shoppin

DISCOVER

Safe Driver?

Output: **St. Elmo's Fire:** A luminous, and often audible, electric discharge that is sporadic in nature. It occurs from objects, especially pointed ones, when the electrical field strength near their surfaces attains a value near 100 volts per centimeter...

Machine Translation

From Babel Fish:

Aznar ha premiado a Rodrigo Rato (vicepresidente primero), Javier Arenas (vicepresidente segundo y ministro de la Presidencia) y Eduardo Zaplana (ministro portavoz y titular de Trabajo) en la septima remodelacion de Gobierno en sus dos legislaturas. Las caras nuevas del Ejecutivo son las de Juan Costa, al frente del Ministerio de Ciencia y Tecnologia, y la de Julia Garcia Valdecasas, que ocupara la cartera de Administraciones Publicas.

⇓

Aznar has awarded to Rodrigo Short while (vice-president first), Javier Sands (vice-president second and minister of the Presidency) and Eduardo Zaplana (minister spokesman and holder of Work) in the seventh remodeling of Government in its two legislatures. The new faces of the Executive are those of Juan Coast, to the front of the Ministry of Science and Technology, and the one of Julia Garci'a Valdecasas, who will occupy the portfolio of Public Administrations.

Machine Translation

An approach pioneered by researchers at IBM:

- Use the Canadian Parliamentary Proceedings ("Hansards") as a training set of French/English translation pairs

- Learn a statistical model of how sentences in French are translated to sentences in English

---

Morgen fliege ich nach Kanada zur Konferenz

Tomorrow I will fly to the conference in Canada

Vision and Language

- The Corel dataset: images paired with captions.
- Learn a model that maps images $\Rightarrow$ words, and words $\Rightarrow$ images.
- Note the similarity to machine translation...

Natural Language Interfaces

An old example:

Dave: Open the pod bay doors, Hal.
Hal: I'm sorry, Dave, I'm afraid I can't do that.
Dave: What's the problem?
Hal: I think you know what the problem is just as well as I do.
Dave: What're you talking about, Hal?
Hal: This mission is too important for me to allow you to jeopardise it.
Dave: I don't know what you're talking about, Hal.
Hal: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.
Dave: Where the hell'd you get that idea, Hal?
Hal: Dave, although you took very thorough precautions in the pod against my hearing you, I could see your lips move.
Dave: Alright, Hal. I'll go in through the emergency airlock.
Hal: Without your space helmet, Dave, you're going to find that rather difficult.
Dave: Hal, I won't argue with you any more. Open the doors.
Hal: Dave, this conversation can serve no purpose any more. Goodbye.
Dave: Hal? Hal. Hal. Hal! Hal!

Natural Language Interfaces

A newer example (ATIS):

User: I need a flight from Boston to Washington, arriving by 10 in the morning

System: What day are you flying on?

User: Tomorrow

System: (Returns a list of flights)

Natural Language Interfaces

Wizard-of-Oz data collection:

- Method: Build an interface where the "system" is a human ("Wizard") sat at a terminal... Collect interactions between a user and the "system". Goal is to mimic the Wizard's behaviour.

- Result:
  A set of **(context, user-utterance, system-action)** triples

- Problem: How to learn the mapping from
  (context, user-utterance) $\Rightarrow$ **system-action**?

Basic NLP Problems: Tagging

TAGGING: Strings to Tagged Sequences

a b e e a f h j $\Rightarrow$ a/C b/D e/C e/C a/D f/C h/D j/C

Example 1: Part-of-speech tagging

Profits/N soared/V at/P Boeing/N Co./N ,/, easily/ADV topping/V forecasts/N on/P Wall/N Street/N ,/, as/P their/POSS CEO/N Alan/N Mulally/N announced/V first/ADJ quarter/N results/N ./.

Example 2: Named Entity Recognition

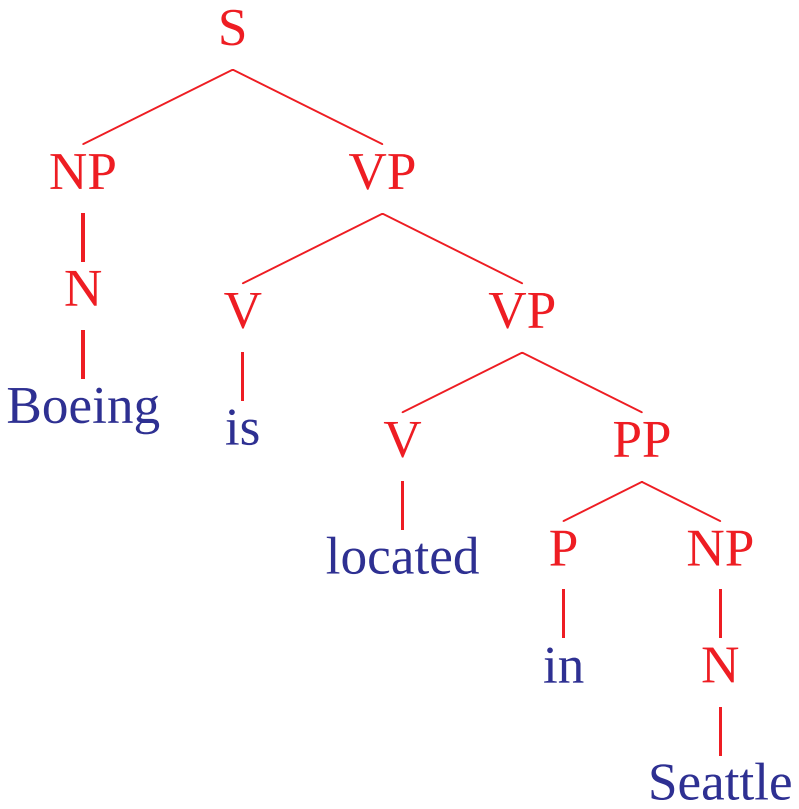
Profits/NA soared/NA at/NA Boeing/SC Co./CC ,/NA easily/NA topping/NA forecasts/NA on/NA Wall/SL Street/CL ,/NA as/NA their/NA CEO/NA Alan/SP Mulally/CP announced/NA first/NA quarter/NA results/NA ./NA

Basic NLP Problems: Parsing

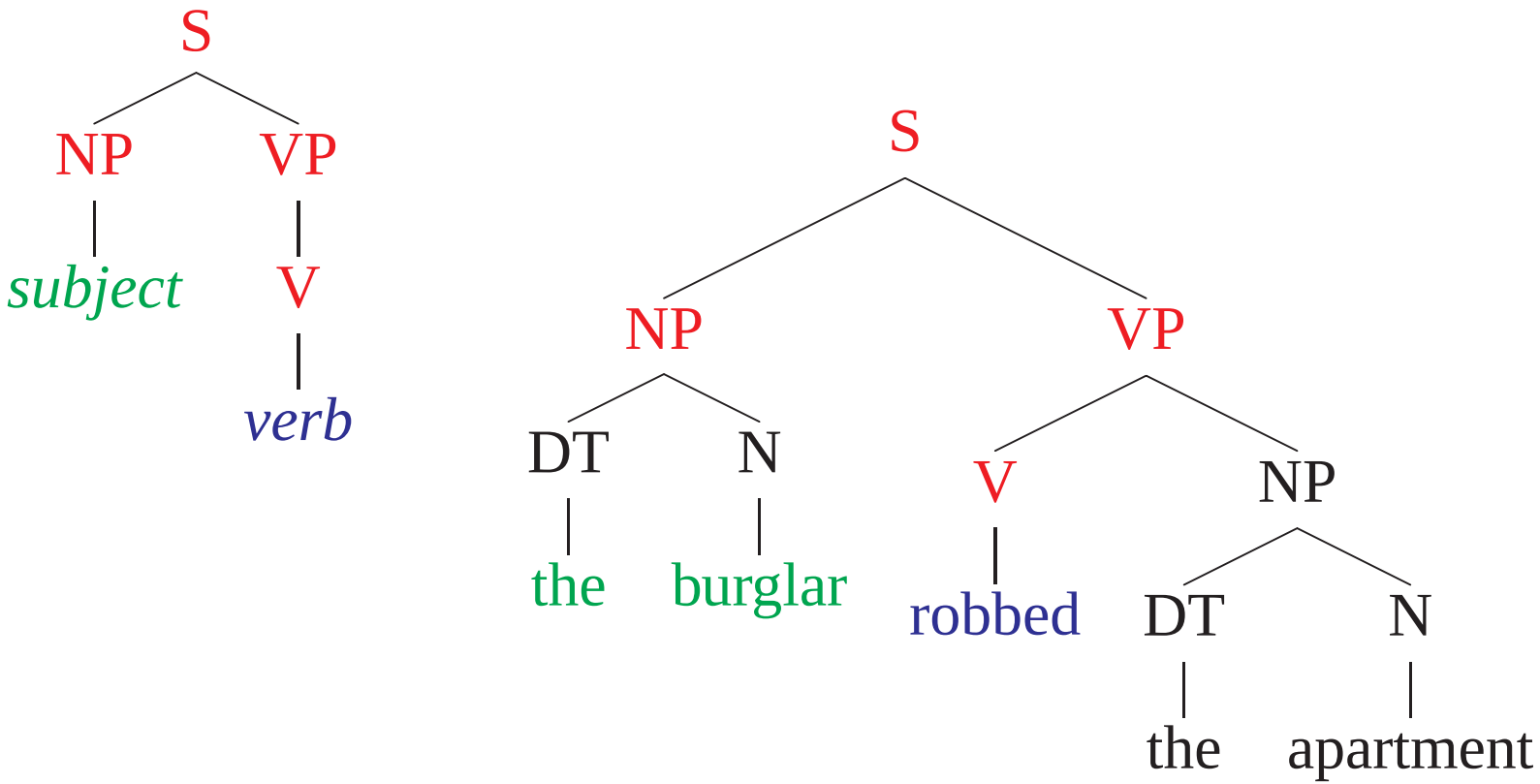
INPUT:

Boeing is located in Seattle.

OUTPUT:



Useful Relationships in parse trees



$\Rightarrow$ “the burglar” is the subject of “robbed”

An Example Application: Machine Translation

- English word order is *subject – verb – object*

- Japanese word order is *subject – object – verb*

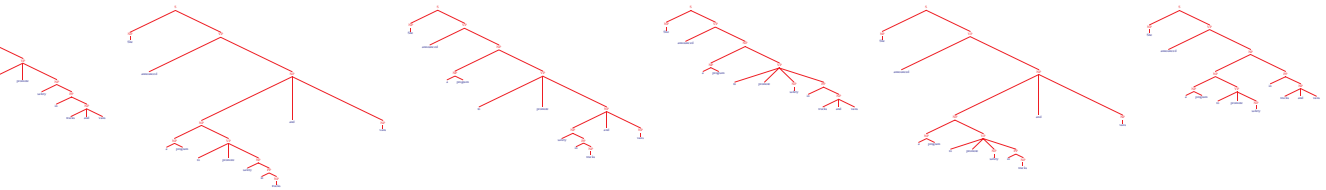
English: IBM bought Lotus
Japanese: *IBM Lotus bought*

English: Sources said that IBM bought Lotus yesterday
Japanese: *Sources yesterday IBM Lotus bought that said*

The Problem with Parsing: Ambiguity

INPUT:

She announced a program to promote safety in trucks and vans

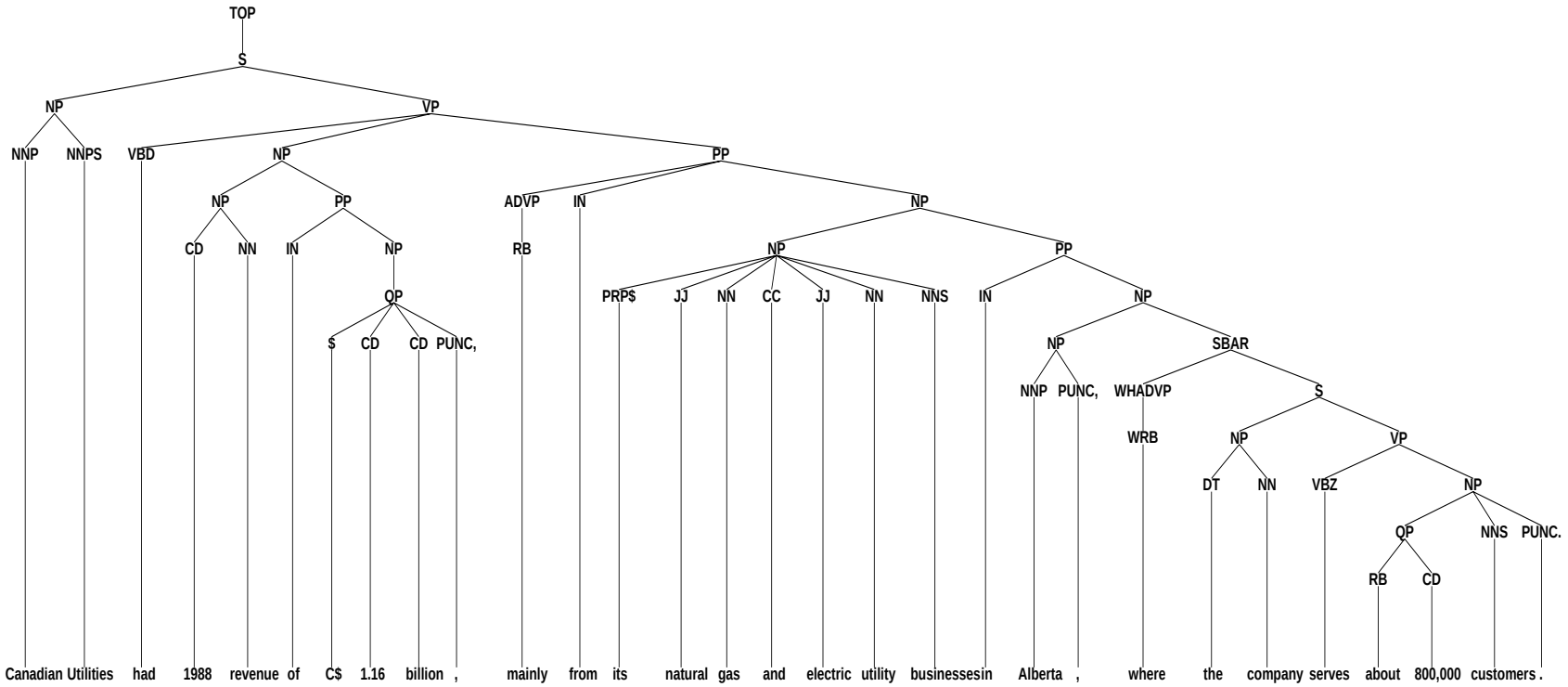


POSSIBLE OUTPUTS:

And there are more...

An Example Tree

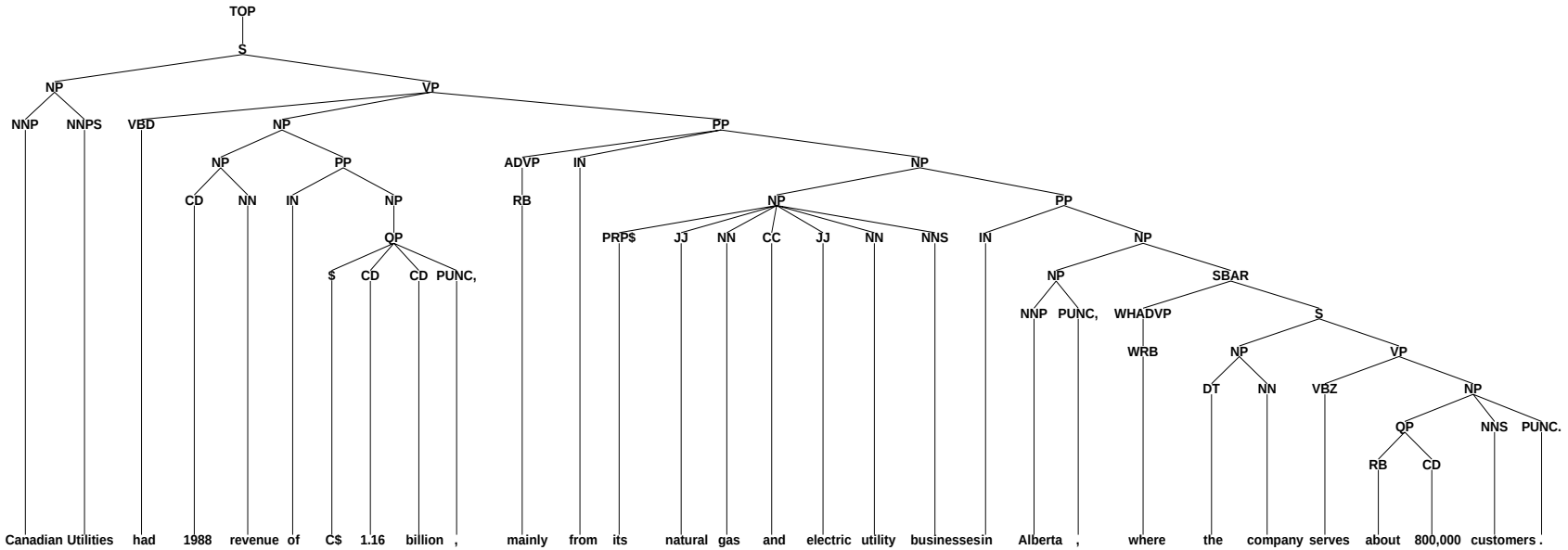
Canadian Utilities had 1988 revenue of C$ 1.16 billion , mainly from its natural gas and electric utility businesses in Alberta , where the company serves about 800,000 customers .



Data for Parsing Experiments

- Penn WSJ Treebank = 50,000 sentences with associated trees
- Usual set-up: 40,000 training sentences, 2400 test sentences

An example tree:



Canadian Utilities had 1988 revenue of C$ 1.16 billion , mainly from its natural gas and electric utility businesses in Alberta , where the company serves about 800,000 customers .

NLP Issues in Languages Other than English

Some characteristics of English:

- Relatively fixed word order
- Relatively little morphology

Some characteristics of Czech:

- Relatively free word order
- A great deal of morphology

$\Rightarrow$ Czech poses a new set of technical challenges
(and so will many other similar languages)

Machine Learning

- Classification problems
- Problems in NLP
- Motivation for learning approaches

Machine Learning: Classification Problems

- Prediction problem: will it rain tomorrow?
- **Representation of the problem**:
  - Is it raining today? (YES, NO)
  - What is the temperature today? (a value in the reals)
  - Am I in England? (YES, NO)
- **Training data:**

Raining today?	Temperature?	England?	Rain tomorrow?
Y	70	N	Y
N	65	N	N
N	70	Y	Y
N	35	Y	Y
Y	35	Y	Y
N	95	N	Y

- **Objective:** Learn a classification rule that performs well on *future* examples.

Machine Learning: Classification Problems

Formalizing the problem:

- Need to learn a function from $\mathcal{X} \to \{-1, +1\}$
  In the weather example, $\mathcal{X} = \mathbb{R}^3$

- Assume there is some distribution $D(x, y)$ where $x \in \mathcal{X}$, $y \in \{-1, +1\}$. Our training sample is drawn from $D(x, y)$.

- **Parametric approaches:** attempt to explicitly model the distribution $D(x, y)$ or $D(y \mid x)$.

- **Distribution-Free Approaches:** find a rule that performs well on the training sample, but is "simple". Statistical theory describes what is meant by a "simple" rule.

Machine Learning for NLP

The general approach:

- Annotate examples of the mapping you're interested in
- Apply some machinary to learn (and generalize) from these examples

The difference from classification:

- Need to induce a mapping from one complex set to another
  (e.g., strings to trees in parsing, strings to strings in machine translation, strings to database entries in information extraction)

Motivation for learning approaches (as opposed to "hand-built" systems):

- Often, a very large number of rules is required.
- Rules interact in complex and subtle ways.
- Constraints are often not "categorical", but instead are "soft" or violable.
- A classic example: *Speech Recognition*

A Syllabus

- Introduction (1 lecture)
- Estimation techniques, and language modeling (1 lecture)
- Stochastic parsing (3 lectures)
- Stochastic tagging, and maximum entropy/log-linear models (2 lectures)
- The EM algorithm for PCFGs, HMMs, and machine translation (2 lectures)
- Information extraction (2 lectures)
- Machine translation (3 lectures)
- Vision and language (1 lecture)
- Dialogue systems (2 lectures)
- NLP issues in different languages,
  e.g., Czech, Chinese, Arabic, German (1 lecture)
- Word-sense disambiguation (1 lecture)
- Semi-supervised approaches (2 lectures)
- Discriminative/reranking approaches, and kernels over NLP structures
  (2 lectures)
- Learning in Optimality Theory (1 lecture)
- Conclusions/open problems (1 lecture)