

# Computational Linguistics

CSC 485/2501  
Fall 2023

2C

## 2c. Understanding language

Gerald Penn

Department of Computer Science, University of Toronto  
(many slides taken or adapted from others)

What does it mean to “understand” language?

# The Turing Test

In the first line of your sonnet which reads  
“Shall I compare thee to a summer’s day,”  
would not “a spring day” do as well or better?

It wouldn’t scan.

How about “a winter’s day”? That would scan  
all right.

Yes, but nobody wants to be  
compared to a winter’s day.



# The Turing Test

Would you say Mr Pickwick reminded you of Christmas?

In a way.

Yet Christmas is a winter's day, and I do not think Mr Pickwick would mind the comparison.

I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.



Using language  
is not necessarily  
understanding language

## A computer psychiatrist

Men are all alike.

In what way?

They're always bugging us about something or the other.

Can you think of a specific example?

Well, my boyfriend made me come here.

Your boyfriend made you come here.

He says I'm depressed.



## A computer psychiatrist

I'm sorry to hear you are depressed.

It's true. I am unhappy.

Do you think coming here will help you not to be unhappy?

Perhaps I could learn to get along with my mother.

Tell me more about your family.



# Levels of understanding 1

- ***0. Keyword processing:***  
Limited knowledge of **particular words or phrases**, or their collocations.
  - Chatbots.
  - Information retrieval.
  - Web searching.

# Levels of understanding 2

- **1. Limited linguistic ability:**  
Appropriate response to simple, highly constrained **sentences**.
- Database queries in NL.  
*“Show all sales staff who exceeded their quota in May.”*
- Simple NL interfaces.  
*“I want to fly from Toronto to Vancouver next Sunday.”*

# Levels of understanding 2



Know your meme at [FAILKING.COM](http://FAILKING.COM)

# Levels of understanding 3

- **2. *Full text comprehension:***  
Understanding **multi-sentence** text and its relation to the “real world”.
  - Conversational dialogue.
  - Automatic knowledge acquisition
  - Machine translation?
- **3. *Emotional understanding/generation:***
  - Responding to **literature, poetry, humour**
  - Story narration.

# Levels of linguistic structure and analysis 1

- Phonology
  - The **sound system** of a language.
- Morphology
  - The **minimal meaningful units** of language (root of a word; suffixes and prefixes), and how they combine.
- Lexicon
  - The semantic and syntactic **properties of words**.

# Levels of linguistic structure and analysis 2

- Syntax
  - The means of expressing meaning: **how words can combine**, and in what order.
- Semantics
  - The **meaning** of a sentence (a logical statement?).
- Pragmatics
  - The **use** of a sentence: pronominal referents; intentions; multi-sentence structure.

# “Building blocks” of CL systems 1

- Language interpretation + language generation = machine translation?
  - Part-of-speech (PoS) tagging.
  - Parsing and grammars.
  - Reference resolution.
  - Dialogue management.
- These are better thought of as functional units now rather than as modular components of modern NLP architectures.

# Natural language interpretation

Does Flight 207 serve lunch?



YNQ (  $\exists e$  SERVING( $e$ )  $\wedge$  SERVER( $e$ , flight-207)  
 $\wedge$  SERVED( $e$ , lunch) )

# Natural language generation

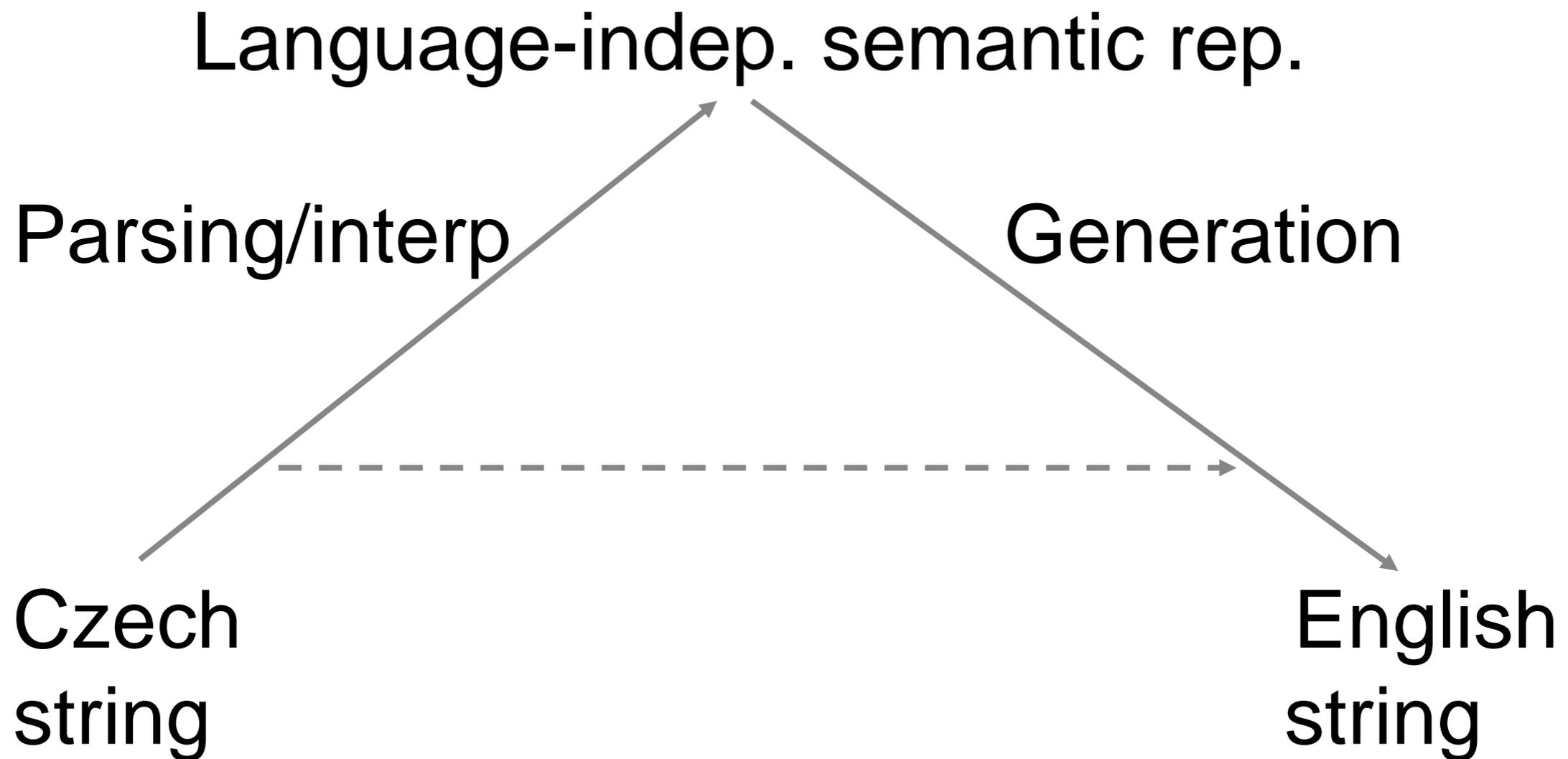
```
(spray-1 (OBJECT paint-1)
  (PATH (path-1
    (DESTINATION wall-1))))
(CAUSER sally-1)
```



Sally sprayed paint onto the wall.

# Machine translation

- History lesson: *the Vauquois triangle* (1968).



# Machine translation

- History lesson: *the Vauquois triangle* (1968).
- Current systems based purely on statistical associations and lexical semantic embeddings.
- Getting incrementally better as they learn from more data.
- Probably more emergent knowledge of linguistics in there than we give them credit for, but it's awfully difficult for us to extract it.



[Úvodní strana](#)

[Historie](#)

[Studium](#)

[Aktuální školní rok](#)

[Přijímací zkoušky](#)

[Maturity](#)

[Lidé](#)

[Předměty](#)

[Google Apps](#)



upc

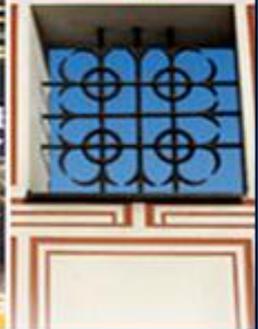
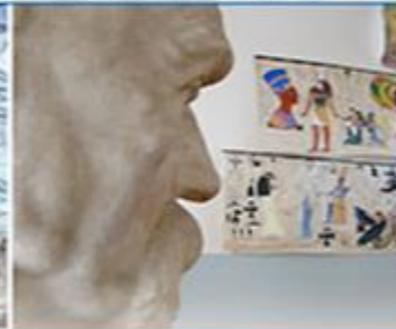


NADACE ČEZ

## Historie Gymnázia Duchcov

Všeobecná touha českých obyvatel po zřízení střední školy s českým vyučovacím jazykem byla naplněna až po první světové válce. Dne 6. října 1919 začalo české gymnázium prozatímně působit v části německého reálného gymnázia. Obrovský zájem o studium (české gymnázium v Duchcově bylo jediným pro teplický, duchcovský a bílinský okres) přiměl ředitelství ústavu otevřít několik tříd i mimo německé gymnázium. Toto a všeobecné přání získat vlastní objekt stálo u zrodu záměru postavit pro duchcovské české gymnázium účelnou a důstojnou budovu. Od přání z roku 1919 k realizaci uplynulo ještě dlouhých osm let. Novostavba byla předána do užívání v neděli 22. května 1927. V průběhu desetiletí se škola stala nejen významným vzdělávacím ústavem, ale i centrem českého menšinového života, řada vynikajících absolventů svědčí o úrovni výuky. Bylo vybudováno zázemí, vytvořeny tradice, škola se dostala do širokého povědomí.

Stejně jako celá naše země prožívala své slavné i méně slavné dny. Ty nejhorší snad za války, kdy byla výuka zcela přerušena. Z hlediska stavebního budova velmi utrpěla i v letech 1948 - 1953, v době, kdy bylo zrušeno reálné gymnázium a v objektu se střídaly různé druhy škol, které zde pro své provozní potřeby prováděly stavební úpravy.



[Home](#)

[History](#)

[Studied](#)

[Current school year](#)

[Admission](#)

[Maturity](#)

[People](#)

[Subjects](#)

[Google Apps](#)



upc



NADACE ČEZ

INFOPHARM®

## *History Gymnasium Duchcov*

General desire Czech population after the establishment of secondary schools and English language teaching has been filled up after the First World War. On 6 October 1919 became the Czech grammar school provisionally operate in the German real grammar school. Huge interest in the study (Czech grammar school in Duchcov was only for Teplice, duchcovský and bílinský District) led the Directorate of the Institute to open a few classes and outside the German grammar school. Such a universal desire to acquire its own building stood for the birth of intent to build a school for duchcovské Czech efficient and dignified building. From the desire of the year 1919 to be implemented yet expired long eight years. New was transferred to use on Sunday 22nd May 1927. Over the decades, the school became not only an important educational institution, but also the centre of minority life, a number of outstanding graduates shows the level of instruction. It was built facilities, established tradition, the school into a broad awareness.

Like our country prožívala its famous and less famous days. The worst, perhaps during the war, when it was completely interrupted instruction. From the perspective of building a building suffered in the years 1948 - 1953, when it was cancelled real school building to pieces and various types of schools, which are for their operational needs to carry out structural adjustments.

## History Grammar Duchcov

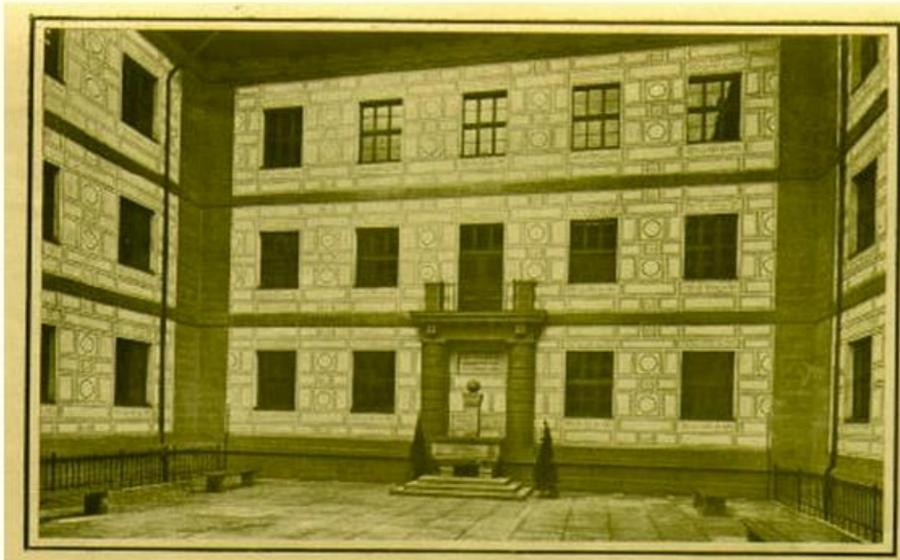


General desire of the Czech population after the establishment of secondary schools in the Czech language of instruction was filled to the First World War. 6 October 1919 Czech Grammar School began to operate provisionally in the German grammar school. The enormous interest in the study (in Czech high school was only for Duchcov Teplice, Duchcovská and Bilina district) headquarters forced the institute to open a few classes outside of the German school. This and a

general desire to get their own building was the genesis of the project to build a gymnasium for Duchcovská Czech efficient and dignified building. From the desire of the year 1919 to implement more long passed for eight years. New building was handed over for use on Sunday 22 May 1927. Over the decades, the school became not only an important educational institution, but also the center of the minority lives, many outstanding graduates shows the level of teaching. It was built facilities, created a tradition, the school turned into a broad awareness.

Like our country has endured its famous and less famous days. Perhaps the worst of the war, when teaching was completely interrupted. In terms of construction the building was seriously damaged in the years 1948 - 1953, when school was canceled and the real object is rotated different kinds of schools here for their operational needs to carry out construction work.

## History Gymnasium Duchcov



General desire Czech population after the establishment of secondary school teaching in Czech language was filled to the First World War. On 6 October 1919 began Czech grammar provisionally act in the German grammar school. The enormous interest in the study (Czech secondary school in Duchcově was only for Teplitz, and Duchcov Bílina district) headquarters induced Institute opened a few classes outside of German grammar. This and the general desire to obtain its object was the genesis of the project to build a high school for Duchcov Czech efficient and dignified building. From the desire of 1919 for implementation has passed even eight long years. The new building was handed over for use on Sunday 22 May 1927. Over the decades, the school has become not only a major educational institution, but also the center of the Czech minority, many outstanding graduates testifies to the level of instruction. It was built

facilities created tradition, the school came into widespread awareness.

Like our entire country had its famous and less famous days. The worst perhaps during the war when teaching was completely interrupted. In terms of construction building was seriously damaged in the years 1948 - 1953, when it was abolished grammar school building and alternated different types of schools, which are for its operational needs in construction work.

## History Grammar School Duchcov



The general desire Czech population after the establishment of secondary school teaching in Czech language was filled to the First World War. On October 6, 1919 began Czech gymnasium provisionally act in the German grammar school. The huge interest in the study (Czech secondary school in Duchcově was only for Teplice, Bílina Duchcov and district) caused the Directorate of the Institute opened a few classes outside the German Gymnasium. This and a general desire to get your own object was the genesis of the project to build a high school

for Duchcov Czech efficient and dignified building. From the desire of 1919 for implementation has passed even eight long years. The new building was handed over for use on Sunday 22 May 1927. During the decade, the school has not only become an important educational institution, but also the center of the minority, many outstanding graduates indicative of the level of education. It was built facilities, established tradition, the school came into widespread awareness.

Like our whole country had its famous and less famous days. The worst perhaps during the war when teaching was completely interrupted. In terms of construction and the building was seriously damaged in the years 1948 - 1953, when the grammar school was canceled in the building and alternated different types of schools, which are for its operational needs in construction work.

## History of the school

Added: 14.11.2014 | Author: Holikova | Viewed: 45142x |

The school was formed by the merger of two institutions - a grammar school and a secondary technical school. The tradition of both schools dates back to the 1920s. At present, both schools are located in the building of the grammar school, which was built in 1927. After extensive reconstruction, which took place in 1990-1996, it dominates the town of Duchcov. As one of the most beautiful buildings in the north of Bohemia is listed in the regional list of architectural monuments

### HISTORY OF INDEPENDENT INDUSTRIAL SCHOOL

#### 1919 - 1938

Soon after the proclamation of the Czechoslovak Republic on October 28, 1918, it seemed paramount to establish a Czech vocational school. Until then, only a German vocational school (Bergschule) existed in Duchcov. On **1 December 1919**, the Ministry of Education established the **State Vocational Secondary School** based in Duchcov. 46 Czech miners signed up for the first year.

The first year of teaching took place under difficult conditions. The school had only one room at the Duchcov German grammar school. The teaching staff consisted practically of two professors, the director of Dr. Ing. Alois Parma and his assistant Ing. Václav Přibáně. Later the school moved to the more spacious rooms of the Waldstein chateau in Duchcov.

#### 1938 - 1941

The time of occupation was very difficult for the school as well as for the whole republic. In 1938 the borderland was seized by the Germans. On September 23, the class was stopped and on October 10 the Ministry of Education and National Enlightenment ordered the school to be transferred to Kladno, saying that pupils enrolled in Duchcov would leave the school and new ones would not be admitted. At the end of the school year on 30 June 1941 the school was liquidated.

#### 1945 - 1959

After the end of World War II in May 1945 the building of the former German upper school (in Bezručova Street) was secured by representatives of the MNV in Duchcov. The former headmaster of the school, Ing. Hugo Malý took over the school building and inventory. 37 pupils were admitted to the first year. The period of study was set at 2 years. It was taught only in the morning. In the afternoon the pupils went down the mines.

The school year 1952 - 53 began in the new building of the former national school Hynek Holub in Kubicových Street. In the school year 1954 - 55, the study of the previous field of Coal Mining was extended to include the field of Mining Measurement. In the school year 1955 - 56, a friendship was established with the mining school of E. Thälmann in Senftenberg in the GDR, which lasted until the 1980s.

#### 1959 - 1969

The beginning of the 1960s was a milestone in the development of the school. The boom in mining in the 1960s led to an increase in the requirements for the number of graduates of the **Secondary Technical Mining School** in Duchcov, as the school was called since 1961. In order to speed up the training of graduates, a three-year field of study **Coal Mining** was introduced for trained miners. The study field **Mining Mechanization and Electrification** and the specialization **Coal Treatment** was also established.

Between 1960 and 1970, the study forms were extended by distance study, **Geology and Drilling**. At the end of the 1960s, the fields of **Power Electrical** and **Geodesy Equipment** were introduced.

#### 1969 - 1979

On September 1, 1979, the school changed its name to the **Secondary Technical School**. In the second half of the 1980s, the mining branch changed its name to **Mining and Mining Geology**.

<https://www.gspsd.cz/index.php?type=Post&id=256&ids=249>  
Translated by Google Translate, 5th September 2019.

# Information extraction

“Bridgestone Sports Co. said Friday it has set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be shipped to Japan. The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990.”



**Tie-up-1:** *Relation:* Tie-up

*Entities:* Bridgestone Sports Co.

a local concern

a Japanese trading house

*Joint venture:* Bridgestone Sports Taiwan Co.

*Activity:* Activity-1

*Amount:* NT \$ 20,000,000

**Activity-1:** *Company:* Bridgestone Sports Taiwan Co.

*Product:* golf clubs

*Start date:* January 1990

# “Building blocks” of CL systems 2

- Information extraction
  - Chunking (instead of parsing).
  - Template filling.
  - Named-entity recognition.

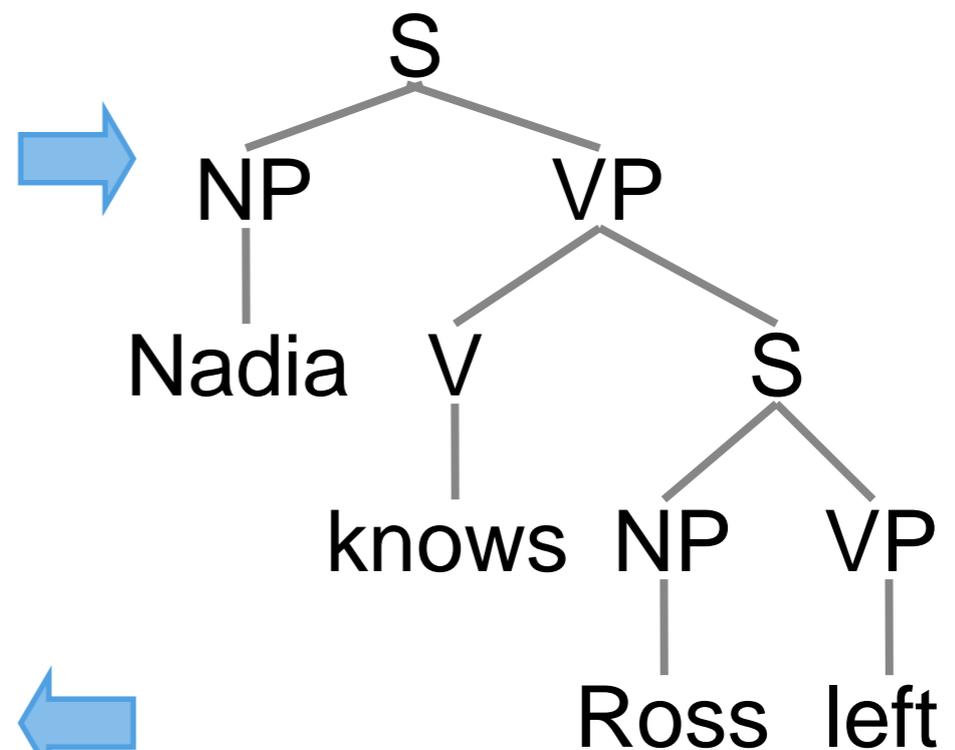
# “Building blocks” of CL systems 3

- Lexical semantics
  - Word sense disambiguation (WSD).
  - Taxonomies of word senses.
  - Analysis of verbs and other predicates
  - Embeddings of words into continuous vector space (word2vec, BERT, XLNet, etc.).
- Computational morphology

# Why is understanding hard? 1

- The structures that we are interested in are richer than strings – often *hierarchical* or *scope-bearing*.

Nadia knows Ross left.



KNOWS(Nadia, LEFT(Ross))

# Why is understanding hard? 2

- Mapping from **surface-form** to meaning is many-to-one: Expressiveness.

Nadia **kisses** Ross. Ross **is kissed by** Nadia.



KISS (Nadia, Ross)



Nadia **gave** Ross **a kiss**. Nadia **gave a kiss to** Ross.

# Why is understanding hard? 3

- Mapping is one-to-many:  
Ambiguity at all levels.
  - Lexical
  - Syntactic
  - Semantic
  - Pragmatic

# Lexical ambiguity

*The lawyer walked to the **bar** and addressed the jury.*

*The lawyer walked to the **bar** and ordered a beer.*

*What time is it? said the judge to Joey when they met / **Five to ten**, said Joey, the judge says, that's exactly what you get.* (Jacques Levy)

- Computational issues
  - Representing the possible **meanings** of words, and their **frequencies** and their **indications**.
  - Representing **semantic relations** between words.
  - Maintaining adequate **context**.

used to strain microscopic **plant life** from the  
zonal distribution of **plant life** .  
close-up studies of **plant life** and natural  
too rapid growth of aquatic **plant life** in water  
the proliferation of **plant** and **animal life**  
establishment phase of the **plant virus life** cycle  
that divide **life** into **plant** and **animal** kingdom  
many dangers to **plant** and **animal life**  
mammals . **Animal** and **plant life** are delicately  
automated **manufacturing plant** in Fremont  
vast **manufacturing plant** and distribution  
chemical **manufacturing plant** , producing viscose  
keep a **manufacturing plant** profitable without  
computer **manufacturing plant** and adjacent  
discovered at a St. Louis **plant manufacturing**  
copper **manufacturing plant** found that they  
copper wire **manufacturing plant** , for example  
's cement **manufacturing plant** in Alpena

vinyl chloride monomer **plant** , which is  
molecules found in **plant** and **animal** tissue  
Nissan car and truck **plant** in Japan is  
and Golgi apparatus of **plant** and **animal** cells  
union responses to **plant** closures .  
cell types found in the **plant** kingdom are  
company said the **plant** is still operating  
Although thousands of **plant** and **animal** species  
**animal** rather than **plant** tissues can be

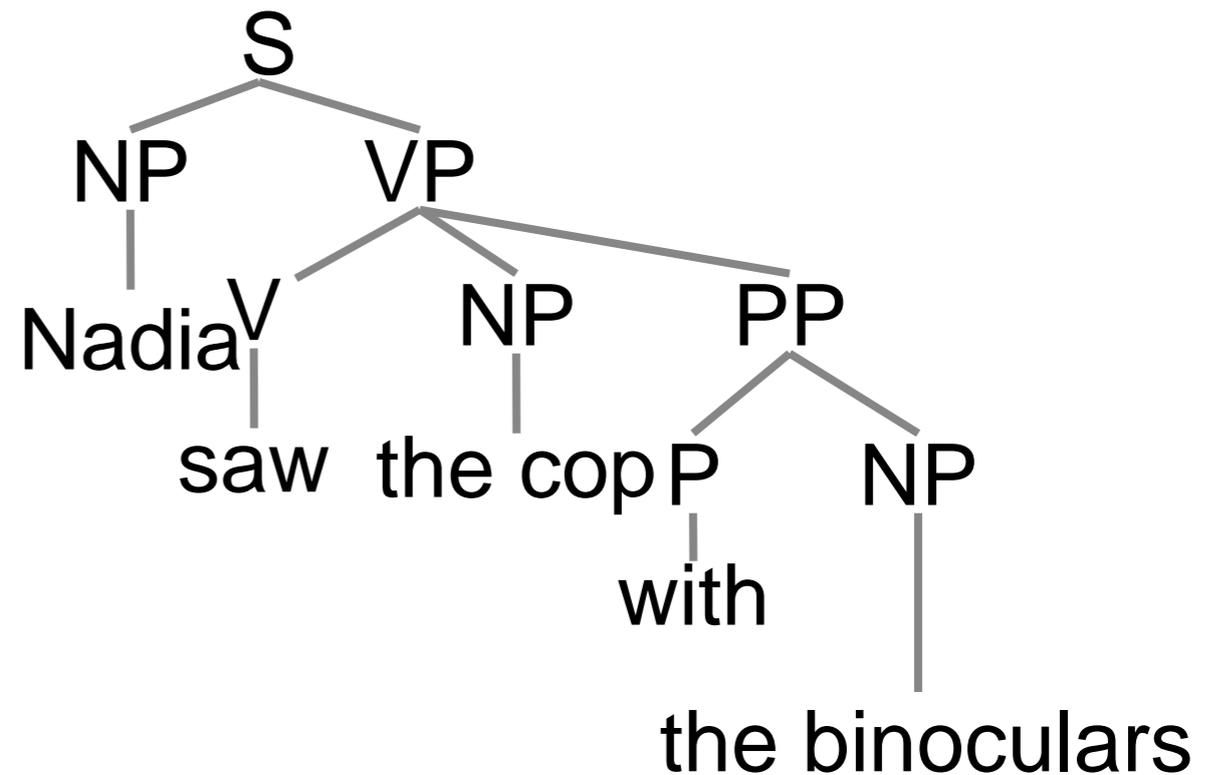
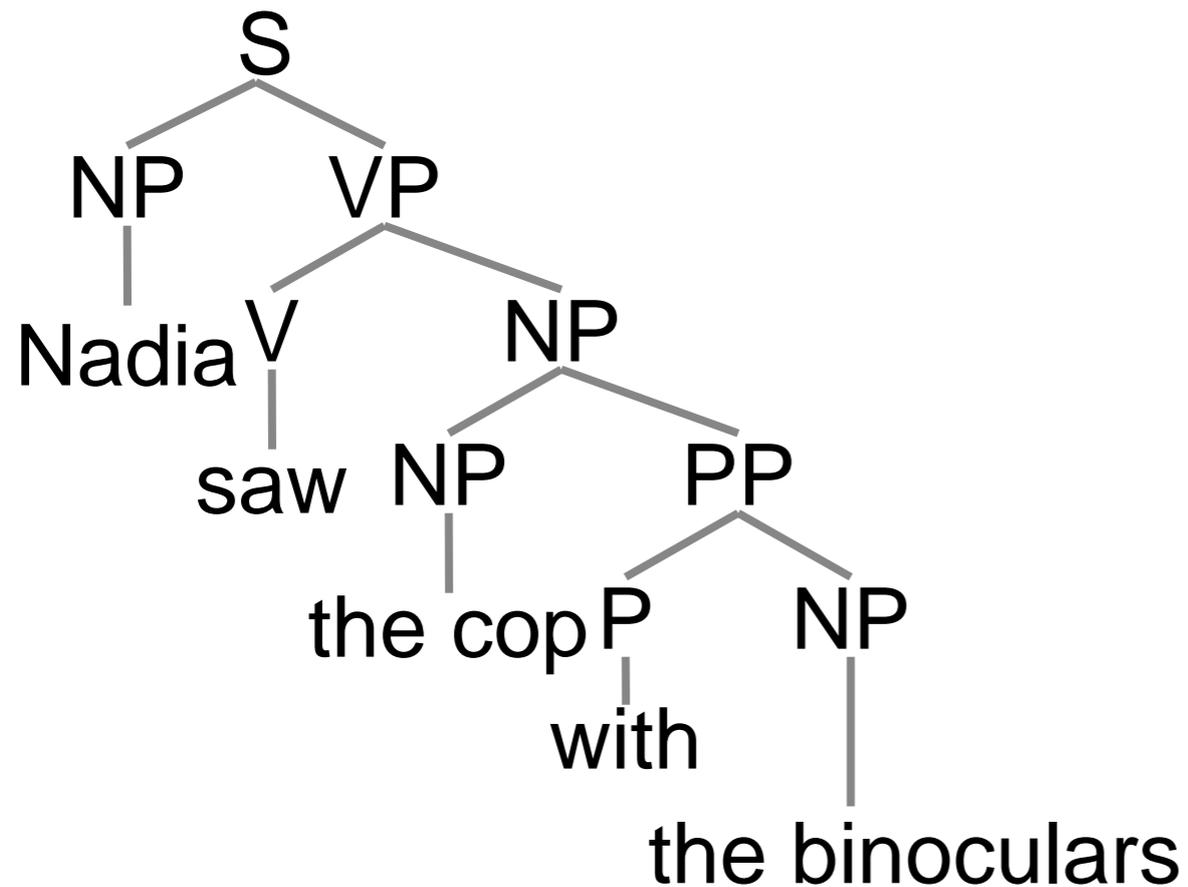
## Decision list for *plant*

<u>LogL</u>	<u>Collocation</u>	<u>Sense</u>
8.10	<i>plant life</i>	→ A
7.58	<i>manufacturing plant</i>	→ B
7.39	<i>life</i> (within ±2-10 words)	→ A
7.20	<i>manufacturing</i> (in ±2-10 words)	→ B
6.27	<i>animal</i> (within ±2-10 words)	→ A
4.70	<i>equipment</i> (within ±2-10 words)	→ B
4.39	<i>employee</i> (within ±2-10 words)	→ B
4.30	<i>assembly plant</i>	→ B
4.10	<i>plant closure</i>	→ B
3.52	<i>plant species</i>	→ A
3.48	<i>automate</i> (within ±2-10 words)	→ B
3.45	<i>microscopic plant</i>	→ A

...

# Syntactic ambiguity 1

Nadia saw the cop with the binoculars.



# Syntactic ambiguity 2

*Put the book in the box on the table.*

*Put the book in the red book box.*

*Noun phrase*

*Adj Noun*

*Visiting relatives can be trying.*

*Verb Noun*

*Verb phrase*

# Syntactic ambiguity 3

- These are absolutely everywhere. Some real headlines:  
*Juvenile Court to Try Shooting Defendant*  
*Teacher Strikes Idle Kids*  
*Stolen Painting Found by Tree*  
*Clinton Wins on Budget, but More Lies Ahead*  
*Hospitals are Sued by 7 Foot Doctors*  
*Ban on Nude Dancing on Governor's Desk*
- Usually we don't even notice – we're that good at this kind of resolution.



# Syntactic ambiguity 5

- Computational issues
  - Representing the possible **combinatorial structure** of words.
  - Capturing syntactic **preferences** and frequencies.
  - Devising **incremental parsing** algorithms.

# Semantic ambiguity

- Sentence can have more than one meaning, even when the words and structure are agreed on.

*Nadia wants a dog like Ross's.*

*Everyone here speaks two languages.*

*Iraqi Head Seeks Arms.*

*DCS Undergrads Make Nutritious Snacks.*

# Pragmatic ambiguity

- A sample dialogue

Nadia: *Do you know who's going to the party?*

Emily: *Who?*

Nadia: *I don't know.*

Emily: *Oh ... I think Carol and Amy will be there.*

- Computational issues
  - Representing **intentions** and **beliefs**.
  - **Planning** and **plan recognition**.
  - **Inferencing** and diagnosis.

# Need for domain knowledge 1

Derivatization of the carboxyl function of retinoic acid by fluorescent or electroactive reagents prior to liquid chromatography was studied. Ferrocenylethylamine was synthesized and could be coupled to retinoic acid. The coupling reaction involved activation by diphenylphosphinyl chloride. The reaction was carried out at ambient temperature in 50 min with a yield of ca. 95%. The derivative can be detected by coulometric reduction (+100 mV) after on-line coulometric oxidation (+400 mV). The limit of detection was 1 pmol of derivative on-column, injected in a volume of 10 $\mu$ l, but the limit of quantification was 10 pmol of retinoic acid.

S. El Mansouri, M. Tod, M. Leclercq, M. Porthault, J. Chalom, "Precolumn derivatization of retinoic acid for liquid chromatography with fluorescence and coulometric detection." *Analytica Chimica Acta*, 293(3), 29 July 1994, 245–250.

# Need for domain knowledge 2

In doing sociology, lay and professional, every reference to the “real world”, even where the reference is to physical or biological events, is a reference to the organized activities of everyday life. Thereby, in contrast to certain versions of Durkheim that teach that the objective reality of social facts is sociology’s fundamental principle, the lesson is taken instead, and used as a study policy, that the objective reality of social facts *as* an ongoing accomplishment of the concerted activities of daily life, with the ordinary, artful ways of that accomplishment being by members known, used, and taken for granted is, for members doing sociology, a fundamental phenomenon.

Harold Garfinkel, Preface, *Studies in Ethnomethodology*, Prentice-Hall, 1967, page vii.

