

# Computational Linguistics

Shuly Wintner, [shuly@cs.haifa.ac.il](mailto:shuly@cs.haifa.ac.il)

Department of Computer Science  
University of Haifa  
Haifa, Israel

Winter 2008

# What is this course about?

## Natural language processing:

- A subfield of computer science, and in particular artificial intelligence
- Concerned with computational processing of natural languages
- Emulates cognitive capabilities without being committed to a true simulation of cognitive processes
- Provides such novel products as computers that can understand everyday human speech, translate between different human languages, and otherwise interact linguistically with people in ways that suit people rather than computers.

# Natural language processing applications

- Machine translation
- Natural language interfaces to computer systems
- Speech recognition
- Text to speech generation
- Automatic summarization
- E-mail filtering
- Intelligent search engines

# What is this course about?

## Computational linguistics:

- An approach to linguistics that employs methods and techniques of computer science
- A formal, rigorous, computationally based investigation of questions that are traditionally addressed by linguistics
- What do people know when they know a natural language?
- What do they do when they use this knowledge?
- How do they acquire this knowledge in the first place?

# Example of an application: machine translation

The spirit is willing but the flesh is weak  
The vodka is excellent but the meat is lousy

# Example of an application: machine translation

From <http://babelfish.altavista.com/>, using technology developed by SYSTRAN; and from Google translation tools.

## Example of an application: machine translation

*Language is one of the fundamental aspects of human behavior and is a crucial component of our lives. In written form it serves as a long-term record of knowledge from one generation to the next. In spoken form it serves as our primary means of coordinating our day-to-day behavior with others. This book describes research about how language comprehension and production work.*

## Example of an application: machine translation

*Il linguaggio è una delle funzioni fondamentali di comportamento umano ed è un componente cruciale delle nostre vite. Nella forma scritta serve da record di lunga durata di conoscenza da una generazione al seguente. Nella forma parlata serve da nostri mezzi primari di coordinazione del nostro comportamento giornaliero con altri. Questo libro descrive la ricerca circa come la comprensione di una lingua e la produzione funzionano.*



## Example of an application: machine translation

Altavista:

*The language is one of the fundamental functions of human behavior and is a crucial member of our screw. In the written shape servants from record of long duration of acquaintance from one generation to following. In the shape speech she serves from our primary means of coordination of our every day behavior with others. This book describes the search approximately as the understanding of a language and the production work.*

# Example of an application: machine translation

Google:

*Language is a fundamental aspect of human existence behavior and is a key component of our lives. In writing that serves as a Long term record knowledge from one generation to another neighbor. In spoken form it serves as our principal means of coordinate our day to day behavior with others. This book describes Search as language comprehension and production work.*

# Comparison

*Language is one of the fundamental aspects of human behavior and is a crucial component of our lives*

*The language is one of the fundamental functions of human behavior and is a crucial member of our screw*

# Comparison

*In written form it serves as a long-term record of knowledge from one generation to the next*


*In the written shape servants from record of long duration of acquaintance from one generation to following*

# Comparison


*This book describes research about how language comprehension and production work*


*This book describes the search approximately as the understanding of a language and the production work*

# Question answering (From <http://www.ask.com/>)

 who was the second US president?  Advanced Search

Web

[John Adams](#)  
 John Adams followed George Washington as **president** of the United States, becoming the country's **second** chief executive. An early colonist agitator against the Stamp Act of 1765, John Adams helped draft the... [More »](#)  
 Go To: [Official Site](#) · [Encyclopedia](#)  
 Related Searches: [George Washington](#) · [James Madison](#) · [John Quincy Adams](#)  
[Source](#)

[Images of second us president](#)  
  
[images.ask.com](#) · [More images »](#)

[List of Presidents of the United States - Wikipedia, the free](#)  
 The **President** is indirectly elected to a four year term by an Electoral College. Since the ratification of the Twenty-**second** Amendment to the U.S.  
[en.wikipedia.org/wiki/List\\_of\\_Presidents\\_of\\_the\\_United\\_...](#)

[George W. Bush - Wikipedia, the free encyclopedia](#)  
 Main articles: **Presidency** of George W. Bush, George W. Bush's first term as **President** of the **United States**, George W. Bush's **second** term as **President** of the...  
[en.wikipedia.org/wiki/George\\_W\\_Bush](#)

[Biography of John Adams](#)  
 Biography of John Adams, **the second President** of the **United States** (1797-1801). When Adams became **President**, the war between the French and British was...  
[www.whitehouse.gov/history/presidents/ja2.html](#)

[WikiAnswers - Who was the second US President](#)  
**United States** History question: **Who was the second US President?** **Second US President** John Adams from 1797-1801. Born October 30, 1735, Braintree (Quincy) MA. Died July 4, 1826, Quincy, MA. Education: Graduated...  
[wiki.answers.com/Q/Who\\_was\\_the\\_second\\_US\\_President](#) · [Cached](#)

# Question answering (From <http://www.ask.com/>)

 who was the US president after Washington?  [Advanced Search](#)

Web Images News Q&A beta More ▾

## [Biography of George Washington](#)



Biography of **George Washington**, the first **President of the United States**. (1789-1797) the Electoral College unanimously elected **Washington President**

[www.whitehouse.gov/history/presidents/gw1.html](http://www.whitehouse.gov/history/presidents/gw1.html)

## [US Presidents - George Washington](#)

Art and information provided by fifth-graders from Abingdon Elementary School in Arlington, Virginia.

[www.whitehouse.gov/kids/presidents/georgewashington.htm](http://www.whitehouse.gov/kids/presidents/georgewashington.htm)...

## [George Washington - Wikipedia, the free encyclopedia](#)

**Washington** became **President of the United States** in 1789 and established many of ..... This attempt failed **after Washington's** supporters rallied behind him.

[en.wikipedia.org/wiki/George\\_Washington](http://en.wikipedia.org/wiki/George_Washington)

## [NATO Speech: Remarks by NATO Secretary General, Lord Robertson and](#)

by NATO Secretary General, Lord Robertson and **US President** George W. Bush **after** their Bilateral Meeting...

[www.nato.int/docu/speech/2002/s020409a.htm](http://www.nato.int/docu/speech/2002/s020409a.htm)

## [Ford was accidental president | The Sun | News](#)

**FORD** — who died on Boxing Day — saved the office of **US president after** Watergate HE was the man who saved the office of **US president after** the Watergate scandal.

[www.thesun.co.uk/sol/homepage/news/article77115.ece](http://www.thesun.co.uk/sol/homepage/news/article77115.ece)

## [Jerhad!com: Boeing CEO to Run for US President After Ousted for](#)

**SEATTLE, Washington** — Boeing CEO, Harry Stonecipher, was ousted by the board of directors **after** it came to their attention that the married CEO was having an affair with a female executive.

[www.jerhad.com/jerhad/2005/03/boeing\\_ceo\\_to\\_r.html](http://www.jerhad.com/jerhad/2005/03/boeing_ceo_to_r.html)

# Question answering (From <http://www.ask.com/>)

 who was the US president following Washington?  [Advanced Search](#)

## Following

Buy **Following** here Fast, secure, reliable  
[www.auravita.com](http://www.auravita.com)

Sponsored Results

## Washington

Find Deals, Read Reviews, Compare Prices and Make Reservations.  
[www.TripAdvisor.com](http://www.TripAdvisor.com)

## Biography of George Washington



Biography of George **Washington**, the first **President of the United States**. (1789-1797) the Electoral College unanimously elected **Washington President**  
[www.whitehouse.gov/history/presidents/gw1.html](http://www.whitehouse.gov/history/presidents/gw1.html)

## President George W. Bush

U.S. Economy **President Bush** has outlined decisive government action to preserve and If you **follow** these principles, your careers will take you far,  
[www.whitehouse.gov/](http://www.whitehouse.gov/)

## George Washington - Wikipedia, the free encyclopedia

**Following** the end of the war in 1783, **Washington** retired to his **Washington** became **President of the United States** in 1789 and established many of the...  
[en.wikipedia.org/wiki/George\\_Washington](http://en.wikipedia.org/wiki/George_Washington)

## Middle East Online

Syrian FM says Damascus always favored dialogue with **Washington** following **US President's** statement.  
[www.middle-east-online.com/english/?id=5227](http://www.middle-east-online.com/english/?id=5227)



# Why are the results so poor?

- Language understanding is complicated
- The necessary knowledge is enormous
- Most stages of the process involve *ambiguity*
- Many of the algorithms are computationally intractable

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# What kind of knowledge is required?

- **Phonetic and phonological knowledge**
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# Phonetics and phonology

**Phonetics** studies the sounds produced by the vocal tract and used in language, including the physical properties of speech sounds, their perception and their production

**Phonology** studies the module of the linguistic capability that relates to sound, abstracting away from their physical properties. Defines an inventory of basic units (*phonemes*), constraints on their combination and rules of pronunciation

# Problems in phonological processing

**Homophones (homonyms):** words that are pronounced alike but are different in meaning or derivation or spelling:

weak — week

to — too — two

הקלה — הקלה+ה — הכלה+ה

**Free variation:** alternation of sounds with no change in meaning: the different pronunciations of the guttural sounds in Hebrew

# Problems in phonological processing

**Allophones:** variants of phonemes that are in complementary distribution:  
litt**l**e

**Phonotactic constraints:** restrictions on the distribution  
(occurrence) of phonemes with respect to one another:  
הצטלם — התעלם

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# Morphology

Morphology studies the structure of words.

**Morpheme:** a minimal sound-meaning unit. Can either be *bound* (not a word) or *free* (word).

Free morphemes: book, ספר

Bound morphemes: books, ספרים

**Affix:** a morpheme which is added to other morphemes, especially roots or stems.

**suffixes** follow the root/stem

**prefixes** precedes the root/stem

**infixes** are inserted into the root/stem



# Morphology

**Derivational morphology:** words are constructed from roots (or stems) and derivational affixes:

inter+national → international

international+ize → internationalize

internationalize+ation → internationalization

שלם → שלמות

**Inflectional morphology:** inflected forms are constructed from base forms and inflectional affixes: ספר+ים → ספרים

# Problems in morphological processing

*Ambiguity:* The various analyses of the word שבתה:

שבחה: [+verb] [+base] הבש [+root] הבש [+binyan]+Pa'al [+person/gender/number]+3p/F/Sg [+tense]+past

שבתה: [+verb] [+base] תבש [+root] תבש [+binyan]+Pa'al [+person/gender/number]+3p/F/Sg [+tense]+past

ה+שבת: [+noun] [+base] תבש [+gender]+fem [+number]+sing [+possessiveSuffix]+3p/F/Sg

שבתה+ש: [+subord] ש [+noun] [+base] התב [+gender]+fem [+number]+sing

שבתה+בש: [+subord] ש [+preposition] ב [+noun] [+base] הת [+gender]+masc [+number]+sing

שבתה+(ה)+בש: [+subord] ש [+preposition] ב [+def] [+noun] [+base] הת [+gender]+masc [+number]+sing

שבתה+ש: [+subord] ש [+noun] [+base] תב [+gender]+fem [+number]+sing [+possessiveSuffix]+3p/F/Sg

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- **Syntactic knowledge**
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# Syntax

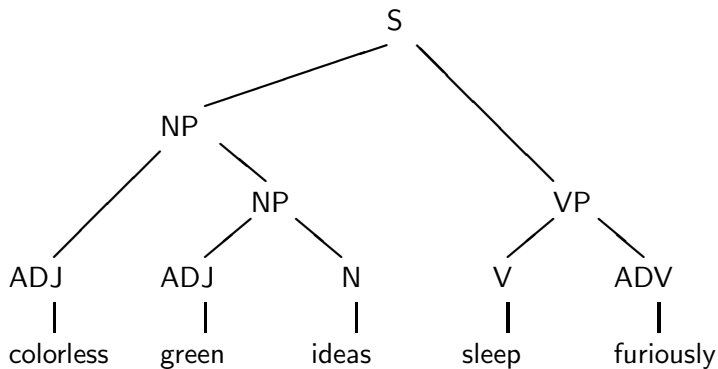
Natural language sentences have *structure*.

*Young green frogs sleep quietly*

*Colorless green ideas sleep furiously*

*Furiously sleep ideas green colorless*

## Syntax



# Problems of syntactic processing

**Expressiveness:** what formalism is required for describing natural languages?

**Parsing:** assigning structure to grammatical strings, rejecting ungrammatical ones.

- top–down vs. bottom–up
- right to left vs. left to right
- chart based vs. backtracking

# Problems of syntactic processing

## Ambiguity:

*I saw the spy with the brown hat*  
*I saw the bird with the telescope*  
*I saw the spy with the telescope*

## Control:

*Kim asked Sandy to call the plumber*  
*Kim promised Sandy to call the plumber*

## Coordination:

*This book describes research about how language  
comprehension and production work*

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- **Semantic knowledge**
- Pragmatic knowledge
- Discourse knowledge
- World knowledge



# Semantics

- Semantics assigns *meanings* to natural language utterances
- A semantic representation must be precise and unambiguous
- A good semantics is *compositional*: the meaning of a phrase is obtained from the meanings of its subphrases

# Problems of semantic processing

Word sense ambiguity: book; round; about; פנישה

Scope ambiguity:

*every student hates at least two courses*

*every student doesn't like math*

*all doors will not open*

# Problems of semantic processing

## Co-reference and anaphora:

*Kim went home after she robbed the bank  
After she robbed the bank, Kim went home  
In the next few paragraphs, some preliminary  
constraints are suggested and problems with them  
are discussed.*

*Language is one of the fundamental aspects of  
human behavior. In written form it serves as a  
long-term record of knowledge.*

VP anaphora: Kim loves his wife and so does Sandy.

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- **Pragmatic knowledge**
- Discourse knowledge
- World knowledge

# Pragmatics

Pragmatics is the study of how more gets communicated than is said.

# Pragmatics

**Presupposition:** the presuppositions of a sentence determine the class of contexts in which the sentence can be felicitously uttered:

*The current king of France is bald*

*Kim regrets that he voted for Obama*

*Sandy's sister is a ballet dancer*

# Pragmatics

**Implicature:** what is conveyed by an utterance that was not explicitly uttered:

*– How old are you? – Closer to 30 than to 20.*

*I have two children.*

*Could you pass the salt?*

# Pragmatics

**Speech acts:** the illocutionary force, the communicative force of utterances, resulting from the function associated with them:

*I'll see you later*

- prediction: I predict that I'll see you later
- promise: I promise that I'll see you later
- warning: I warn you that I'll see you later

*I sentence you to six months in prison*

*I swear that I didn't do it*

*I'm really sorry!*

**Non-literal use of language:** metaphor, irony etc.



# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# Discourse

- A discourse is a sequence of sentences
- Discourse has structure much like sentences do
- Understanding discourse structure is extremely important for dialog systems
- An example dialog:

*When does the train to Haifa leave?*

*There is one at 2:00 and one at 2:30.*

*Give me two tickets for the earlier one, please.*

# Problems of discourse processing

Non-sentential utterances: *aha; to Haifa; the last one*

Cross-sentential anaphora

Reference to non-NPs:

*Kim visited the University of Haifa.*

*It* changed her life.

She does *it* every year.

*It* really surprised Sandy.

It was summer *then*.

# What kind of knowledge is required?

- Phonetic and phonological knowledge
- Morphological knowledge
- Syntactic knowledge
- Semantic knowledge
- Pragmatic knowledge
- Discourse knowledge
- World knowledge

# World knowledge

*– Is the train to Haifa late? – It left Tel Aviv at 8:30.  
George W. Bush left for Vietnam today. This is the last foreign  
visit of the American president.*

# Processing Hebrew

- The script
- Writing direction
- Deficiencies of the Hebrew writing system
- Richness of the morphology
- Root-and-pattern word formation
- Paucity of linguistic resources

# Infrastructure for processing language

- Lexicons
- Dictionaries
- Morphological analyzers and generators
- Part-of-speech taggers
- Shallow parsers
- Syntactic analyzers
- Computational grammars

# Hebrew processing: the state of the art

- Lexicons
- Dictionaries
- Morphological analyzers and generators
- Part-of-speech taggers
- Shallow parsers
- Syntactic analyzers
- Computational grammars



# Conclusions

- Natural languages are complex
- Applications which require deep linguistic knowledge still do not perform well
- Applications which can rely on shallow knowledge or on statistical approaches perform better
- Hebrew poses additional problems for language processing
- To build Hebrew language applications, essential linguistic resources must be developed

# Structure of the course

## Morphology

- introduction to morphology: word structure
- inflections and derivations
- finite-state automata
- finite-state transducers

# Structure of the course

## Syntax

- introduction to syntax: the structure of natural languages
- context-free grammars: grammars, forms, derivations, trees, languages
- parsing: top-down, CYK algorithm, Earley algorithm, bottom-up chart parsing
- the complexity of natural languages
- the limitations of CFGs
- unification grammars: feature structures and unification

# Structure of the course

## Empirical methods in NLP

- POS tagging
- Shallow parsing
- Machine learning and classification
- Applications

## Other topics

- As time permits

# Practicalities

**Textbook:** Nothing mandatory or even recommended. Some of the material can be found in Daniel Jurafsky and James H. Martin, *Speech and Language Processing*, Prentice-Hall, 2000

**Grading:** Take-home exam

**Attendance:** Optional but highly recommended.