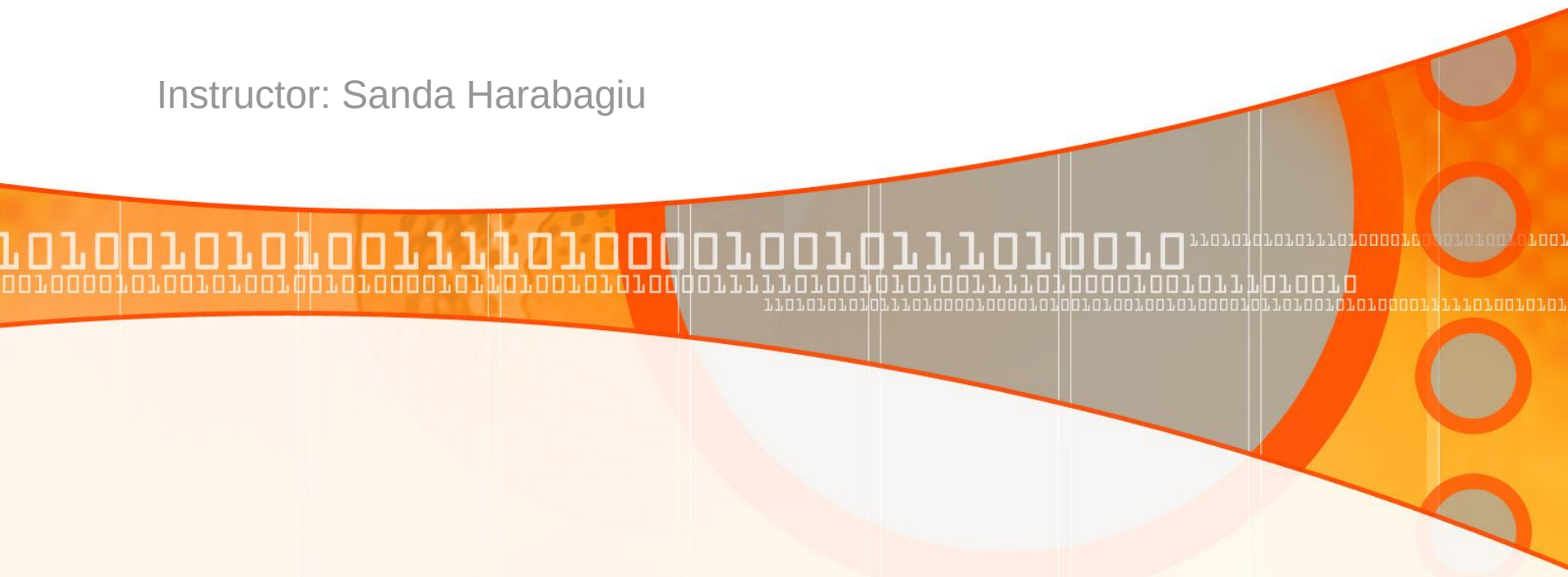
Natural Language Processing
CS 6320
Lecture 1
Introduction to NLP

Instructor: Sanda Harabagiu



Definition

- *NLP is concerned with the computational techniques used for processing human language. It creates and implements computer models for the purpose of performing various natural language tasks.*
  - *These tasks include :*
    - *Mundane applications, e.g. word counting, spell checking, automatic hyphenation*
    - *Cutting edge applications, e.g. automated question answering on the Web, building NL interfaces to databases, machine translation, and others.*
- *What distinguished these applications from other data processing applications is their use of knowledge of language.*

- *NLP is playing an increasing role in curbing the information explosion on Internet and corporate America.*

Related areas

- *NLP is a difficult, and largely unsolved problem. One reason for this is its **multidisciplinary** nature:*
  - ***Linguistics** : How words, phrases, and sentences are formed.*
  - ***Psycholinguistics** : How people understand and communicate using human language.*
  - ***Cognitive Modeling:** Deals with models and computational aspects of NL (e.g. algorithms).*

Related areas

- ***Philosophy**: relates to the semantics of language; notion of meaning, how words identify objects. NLP requires considerable knowledge about the world.*
- ***Computer science**: model formulation and implementation using modern methods.*
- ***Artificial intelligence**: issues related to knowledge representation and reasoning.*
- ***Statistics:** many NLP problems are modeled using probabilistic models.*
- ***Machine learning:** automatic learning of rules and procedures based on lexical, syntactic and semantic features.*
- ***NL Engineering**: implementation of large, realistic systems. Modern software development methods play an important role.*

Applications of NLP

- ***Text - based applications***:
  - *Finding documents on certain topics (document classification)*
  - *Information extraction: extract information related events, relations, concepts*
  - *Complete understanding of texts: requires a deep structure analysis,*
    - *Reading comprehension*
  - *Translation from a language to another,*
  - *Summarization,*
  - *Knowledge acquisition,*
  - *Question-Answering*
- ***Dialogue - based applications*** *(involve human - machine communication):*
  - *Conversational Agents*
  - *Tutoring systems*
  - *Problem solving.*
- ***Speech processing***

Basic levels of language processing 1/2

1. ***Phonetic*** *- how words are related to the sounds that realize them. Essential for speech processing.*
2. ***Morphological Knowledge*** *- how words are constructed : e.g friend, friendly, unfriendly, friendliness.*
3. ***Syntactic Knowledge*** *- how words can be put together to form correct sentences, and the role of each play in the sentence. e.g.:*
4. John ate the cake.
5. ***Semantic Knowledge*** *- Words and sentence meaning:*

   They saw a log.

   They saw a log yesterday.

   He saws a log.

Basic levels of language processing 2/2

5. ***Pragmatic Knowledge***- *how sentences are used in different situations(or contexts).*

   Mary grabbed her umbrella.

   a) It is a cloudy day.
   b) She was afraid of dogs.

5. ***Discourse Knowledge*** *- how the meaning of words and sentences is effected by the proceeding sentences; pronoun resolution.*

   John gave his bike to Bill.

   He didn't care much for it anyway.

5. ***World Knowledge*** *- the vast amount of knowledge necessary to understand texts. Used to identify beliefs, goals.*
6. ***Language generation*** *- have the machine generate coherent text or speech. Needs planning.*

Examples of NLP difficulties

1. ***Syntactic ambiguity****- when a word has more than one part of speech:*

   Example: Rice flies like sand.

   Note that these syntactic ambiguities lead to different parse structures. Sometimes it is possible to use grammar rules (like subject verb agreement) to disambiguate:

   Flying planes are dangerous.

   Flying planes is dangerous.

2. ***Semantic ambiguity****- when a word has more than one possible meaning (or sense):*

   John killed the wolf.

   John killed the project.

   John killed that bottle of wine.

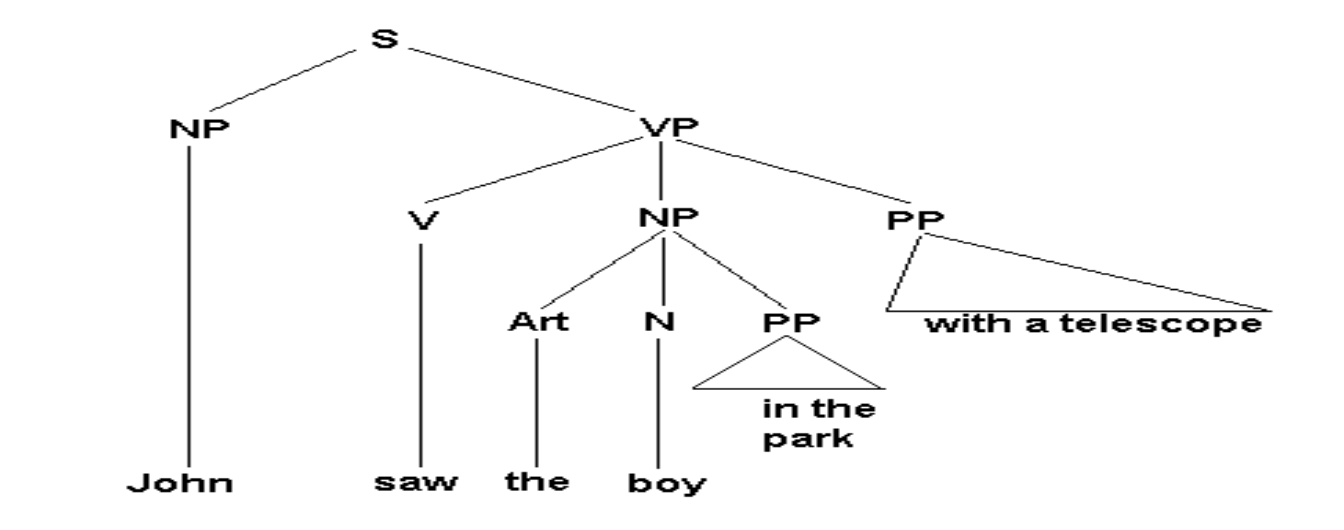
   John killed Jane. *(at tennis , or murdered her)*

More Examples of NLP difficulties

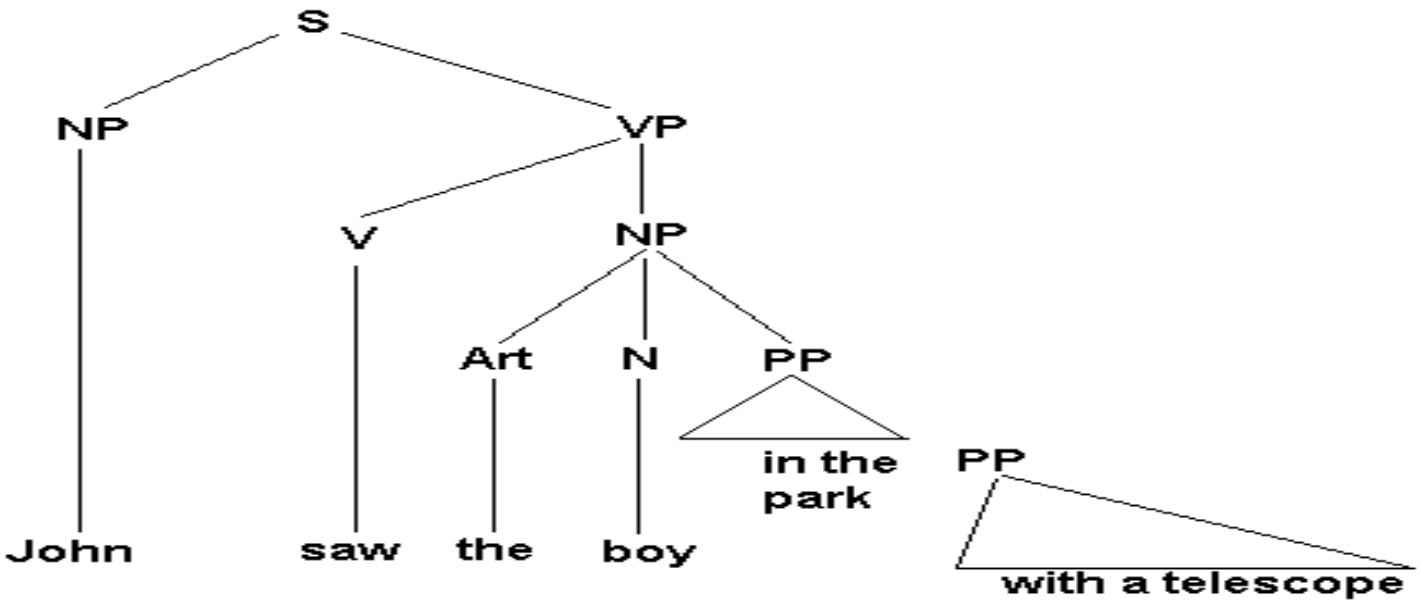
3. Structural ambiguity *- when a sentence has more than one possible parse structure; e.g. prepositional attachment .*

- *Example:*

  John saw the boy in the park with a telescope.



Another syntactic parse



Additional NLP difficulties

***Ambiguities of a sentence**:*

Example:
I made her duck.

Possible interpretations:

1. *I cooked waterfowl for her.*
2. *I cooked waterfowl belonging to her.*
3. *I created the (plaster ?) duck she owns.*
4. *I caused her to quickly lower her head or body*
5. *I wave my magic wand and turned her into undifferentiated waterfowl.*

State of the art in NLP Research 1/2

- ***NLP Publications*** *:*
  - *Association of Computational Linguistics (ACL):*
    - *Conferences: ACL, HLT-NAACL, EACL, EMNLP*
    - *Journals: Computational Linguistics, TACL*
  - *AAAI - every year proceedings.*
  - *IJCAI - every year proceedings.*
  - *The Web Conference*
- *On the WWWeb: http://aclweb.org*

- ***Natural Language Engineering*** *(journal).*

State of the art in NLP Research 2/2

- ***Machine Readable Dictionaries*** *(MRD) WordNet, LDOCE.*
- ***Large corpora****:*
  - *Penn Treebank—contains 2-3 months of Wall Street Journal articles (~ .5 million words of English, POS tagged and parsed),*
  - *Brown corpus,*
  - *SemCor.*
  - *Google GiGaword*
- ***Neural Language Processing***