

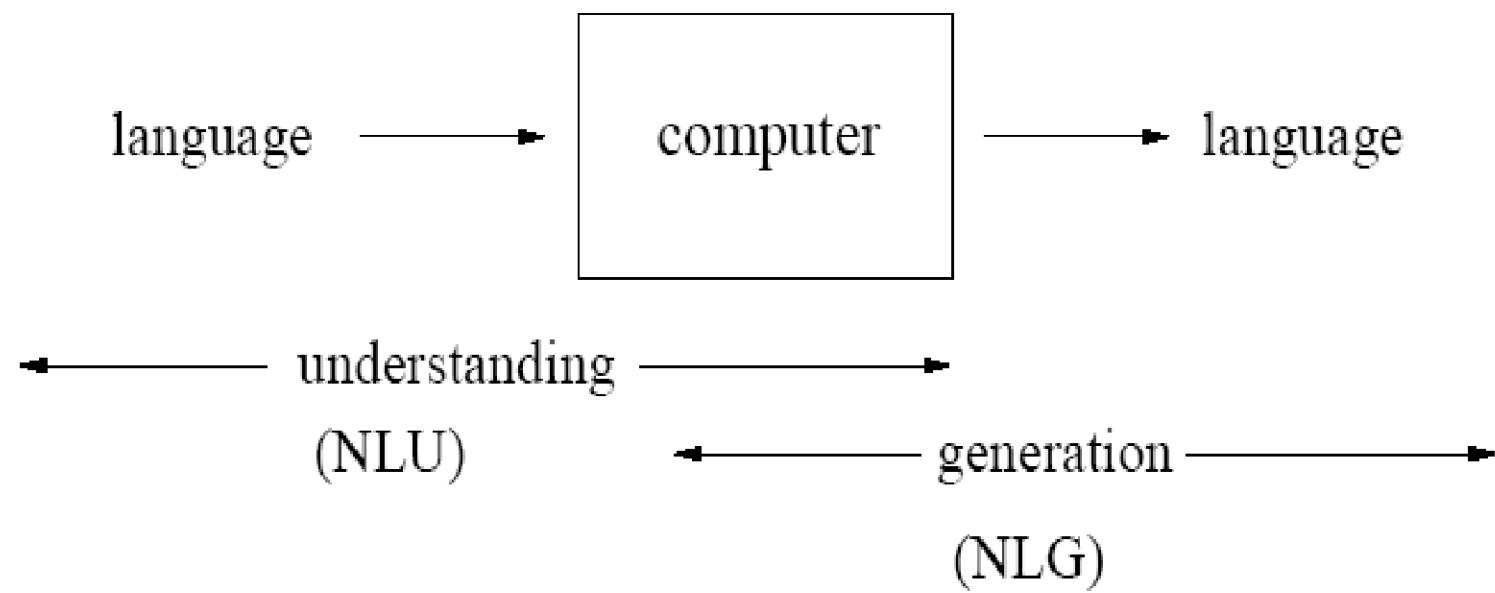
Natural Language Processing

What is Natural Language Processing (NLP)

- The process of computer analysis of input provided in a human language (natural language), and conversion of this input into a useful form of representation.
- The field of NLP is primarily concerned with getting computers to perform useful and interesting tasks with human languages.
- The field of NLP is secondarily concerned with helping us come to a better understanding of human language.

Forms of Natural Language

- The input/output of a NLP system can be:
 - **written text**
 - **speech**
- We will mostly concerned with written text (not speech).
- To process written text, we need:
 - **lexical, syntactic, semantic knowledge about the language**
 - **discourse information, real world knowledge**
- To process spoken language, we need everything required to process written text, plus the challenges of speech recognition and speech synthesis.





Okay, Ginger! I've had it!
You stay out of the garbage!
Understand, Ginger? Stay out
of the garbage, or else!



blah blah GINGER blah
blah blah blah blah blah
blah blah GINGER blah
blah blah blah blah...



Okay, Ginger! I've had it!
You stay out of the garbage!
Understand, Ginger? Stay out
of the garbage, or else!



blah blah GINGER blah
blah blah blah blah blah
blah blah GINGER blah
blah blah blah blah...



- Fundamental goal: deep understand of broad language
 - Not just string processing or keyword matching!
- End systems that we want to build:
 - Ambitious: speech recognition, machine translation, information extraction, dialog interfaces, question answering...
 - Modest: spelling correction, text categorization...

Components of NLP

- **Natural Language Understanding**
 - Mapping the given input in the natural language into a useful representation.
 - Different level of analysis required:
morphological analysis,
syntactic analysis,
semantic analysis,
discourse analysis, ...
- **Natural Language Generation**
 - Producing output in the natural language from some internal representation.
 - Different level of synthesis required:
deep planning (what to say),
syntactic generation
- NL Understanding is much harder than NL Generation.
But, still both of them are hard.

Why is NLP hard?

- Reason (1) – human language is ambiguous.
- Task: Pronoun Resolution
 - Jack saw Sam yesterday. ***He*** went back to the restaurant to get another meal.
 - Jack saw Sam yesterday. ***He*** clearly had eat too much.

Why is NLP hard?

- Reason (1) – human language is ambiguous

Why is NLP hard?

- Reason (2) – requires reasoning beyond what is explicitly mentioned (**A,B**) , and some of the reasoning requires world knowledge (**C**)
 - *I couldn't submit my homework because my horse ate it.*

Why is NLP hard?

- Reason (3) – Language is difficult even for humans.

Ambiguity

Why is Language Ambiguous?

- Having a unique linguistic expression for every possible conceptualization that could be conveyed would make language overly complex and linguistic expressions unnecessarily long.
- Allowing resolvable ambiguity permits shorter linguistic expressions, i.e. data compression.
- Language relies on people's ability to use their knowledge and inference abilities to properly resolve ambiguities.
- Infrequently, disambiguation fails, i.e. the compression is lossy.

Natural Languages vs. Computer Languages

- Ambiguity is the primary difference between natural and computer languages.
- Formal programming languages are designed to be unambiguous, i.e. they can be defined by a grammar that produces a unique parse for each sentence in the language.
- Programming languages are also designed for efficient (deterministic) parsing.

Ambiguity

- Ambiguity at multiple levels
 - Word senses: **bank** (finance or river ?)
 - Part of speech: **chair** (noun or verb ?)
 - Syntactic structure: **I can see a man with a telescope**
 - Multiple: **I made her duck**



Natural Language Tasks

- Processing natural language text involves many various syntactic, semantic and pragmatic tasks in addition to other problems.

Why NL Understanding is hard?

- Natural language is extremely rich in form and structure, and **very ambiguous**.
 - How to represent meaning,
 - Which structures map to which meaning structures.
- One input can mean many different things. Ambiguity can be at different levels.
 - Lexical (word level) ambiguity -- different meanings of words
 - Syntactic ambiguity -- different ways to parse the sentence
 - Interpreting partial information -- how to interpret pronouns
 - Contextual information -- context of the sentence may affect the meaning of that sentence.
- Many input can mean the same thing.
- Interaction among components of the input is not clear.

The Challenges of “Words”

- Segmenting text into words.
- Morphological variation
- Words with multiple meanings: bank, mean
- Domain-specific meanings: latex
- Multiword expressions: make a decision, take out, make up

Part of Speech Tagging

ikr smh he asked fir yo last name

so he can add u on fb lololol

Part of Speech Tagging

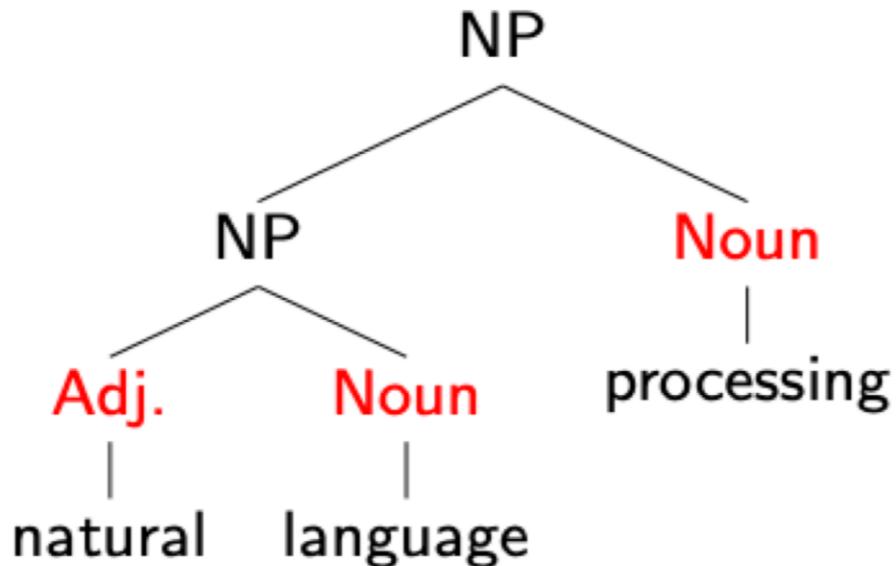
I know, right shake my head
ikr smh he asked fir yo last name

you Facebook laugh out loud
so he can add u on fb lololol

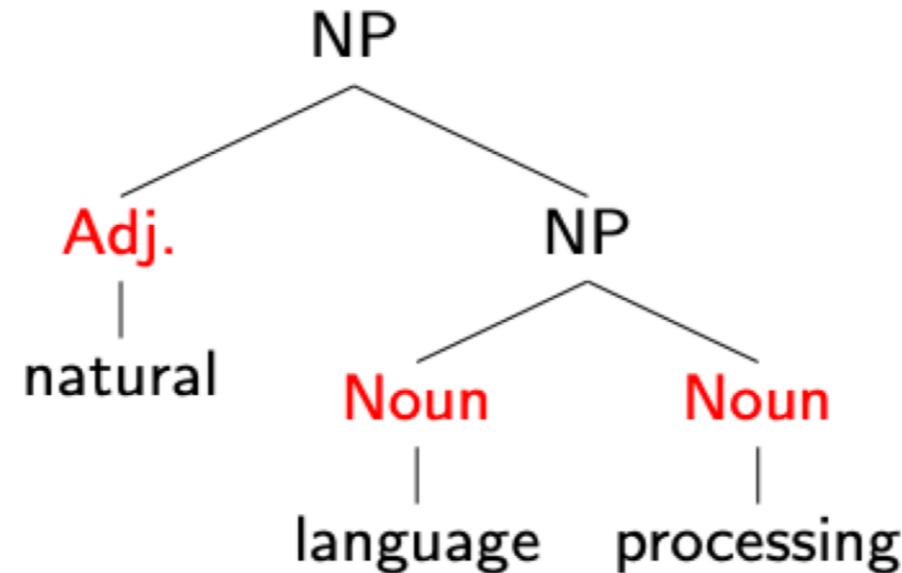
Part of Speech Tagging

I know, right	shake my head			for	your		
ikr	smh	he	asked	fir	yo	last	name
!	G	O	V	P	D	A	N
interjection	acronym	pronoun	verb	prep.	det.	adj.	noun
		you		Facebook		laugh out loud	
so	he	can	add	u	on	lololol	
P	O	V	V	O	P	^	!
preposition				proper noun			

Syntax



vs.



Morphology + Syntax



A ship-shipping
ship, shipping
shipping-ships

Syntax + Semantics

- We saw the woman with the telescope wrapped in paper.
 - Who has the telescope?
 - Who or what is wrapped in paper?

Ambiguity

- “Get the cat with the gloves.”

Ambiguity

- “Get the cat with the gloves.”





Knowledge of Language

- **Phonology** – concerns how words are related to the sounds that realize them.
- **Morphology** – concerns how words are constructed from more basic meaning units called morphemes. A morpheme is the primitive unit of meaning in a language.
- **Syntax** – concerns how can be put together to form correct sentences and determines what structural role each word plays in the sentence and what phrases are subparts of other phrases.
- **Semantics** – concerns what words mean and how these meaning combine in sentences to form sentence meaning. The study of context-independent meaning.

Knowledge of Language (cont.)

- **Pragmatics** – concerns how sentences are used in different situations and how use affects the interpretation of the sentence.
- **Discourse** – concerns how the immediately preceding sentences affect the interpretation of the next sentence. For example, interpreting pronouns and interpreting the temporal aspects of the information.
- **World Knowledge** – includes general knowledge about the world. What each language user must know about the other's beliefs and goals.

Phonology

- Red and Read
- Flower and Flour
- I and Eye
- Write and Right
- Knows and Nose
- Hear and Here
- Weight and Wait
- A part and Apart
- Piece and Peace
- ate and eight

What is Morphology?

- Study of Words
 - Their **internal structure**

washing → wash + -ing

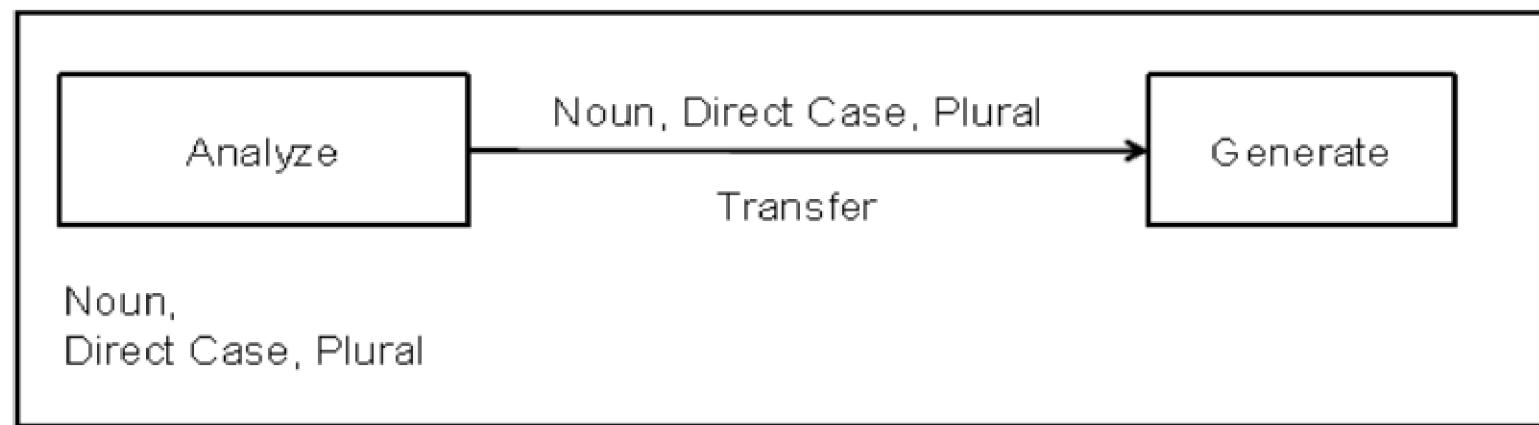
- How they are **formed?**

bat	→	bats	⋮	rat	→	rats
write	→	writer	⋮	browse	→	browser

- Morphology tries to formulate rules

Morphology for NLP

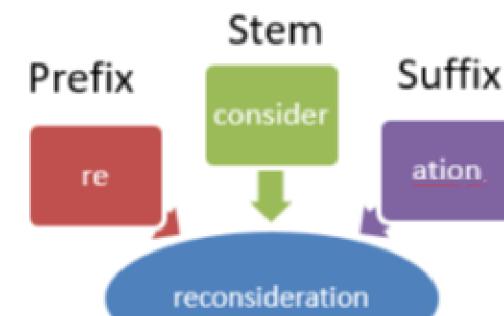
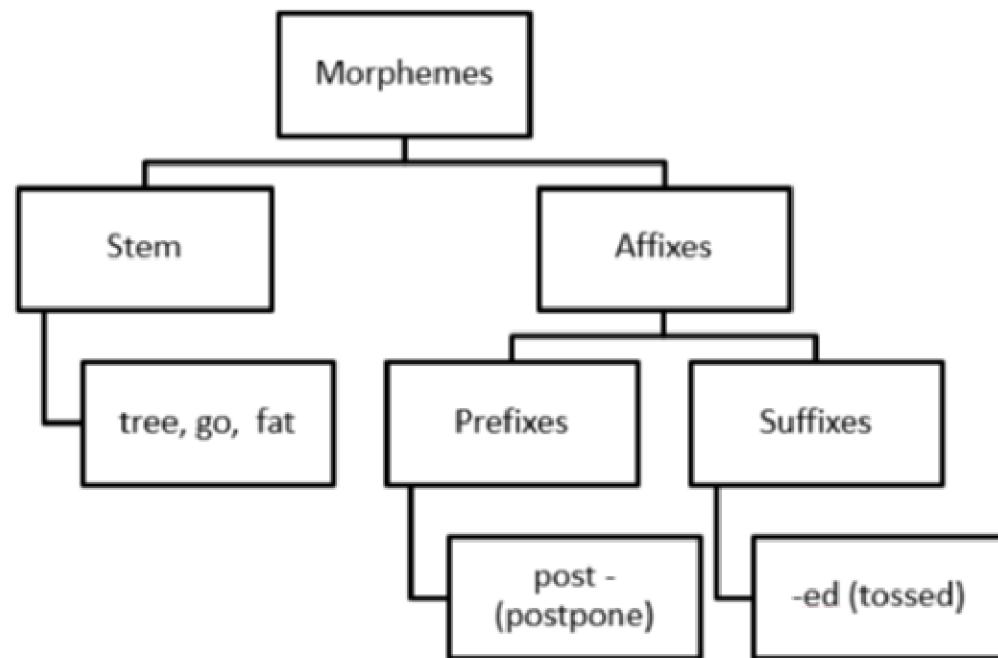
- Machine Translation



- Information Retrieval
 - **goose** and **geese** are two words referring to the same root **goose**

Morphemes

- Smallest meaning bearing units constituting a word



Inflection vs Derivation morphemes

Ambiguity

I made her duck.

- How many different interpretations does this sentence have?
- What are the reasons for the ambiguity?
- The categories of knowledge of language can be thought of as ambiguity resolving components.
- How can each ambiguous piece be resolved?
- Does speech input make the sentence even more ambiguous?
 - Yes – deciding word boundaries

Ambiguity (cont.)

- Some interpretations of : **I made her duck.**



Ambiguity is Everywhere

- Lexical category: part of speech
 - ▶ Duck can be a Noun or Verb
 - V: Duck! I caused her to quickly lower her head or body.
 - N: I cooked waterfowl for her benefit
 - ▶ Her can be possessive (of her) or dative (for her)
 - Possessive: I cooked waterfowl belonging to her.
 - Dative: I cooked waterfowl for her benefit
- Lexical Semantics:
 - ▶ Make can mean create or cook
 - create: I made the (plaster) duck statue she owns
 - cook: I cooked waterfowl for her benefit

Ambiguity (cont.)

- Some interpretations of : **I made her duck.**
 1. I cooked *duck* for her.
 2. I cooked *duck* belonging to her.
 3. I created a toy duck which she owns.
 4. I caused her to quickly lower her head or body.
 5. I used magic and turned her into a *duck*.
- duck – morphologically and syntactically ambiguous:
noun or verb.
- her – syntactically ambiguous: dative or possessive.
- make – semantically ambiguous: cook or create.
- make – syntactically ambiguous:
 - Transitive – takes a direct object.
 - Di-transitive – takes two objects.
 - Takes a direct object and a verb.



Ambiguity (cont.)

- Some interpretations of : **I made her duck.**
 1. I cooked *duck* for her.
 2. I cooked *duck* belonging to her.
 3. I created a toy duck which she owns.
 4. I caused her to quickly lower her head or body.
 5. I used magic and turned her into a *duck*.
- duck – morphologically and syntactically ambiguous:
noun or verb.
- her – syntactically ambiguous: dative or possessive.
- make – semantically ambiguous: cook or create.
- make – syntactically ambiguous:
 - Transitive – takes a direct object. =>
 - Di-transitive – takes two objects. =>
 - Takes a direct object and a verb. =>

