Natural Language Processing

2-3 Week

Natural language understanding

Raw speech signal

↓ • **Speech recognition**

Sequence of words spoken

↓ • **Syntactic analysis** using knowledge of the grammar

Structure of the sentence

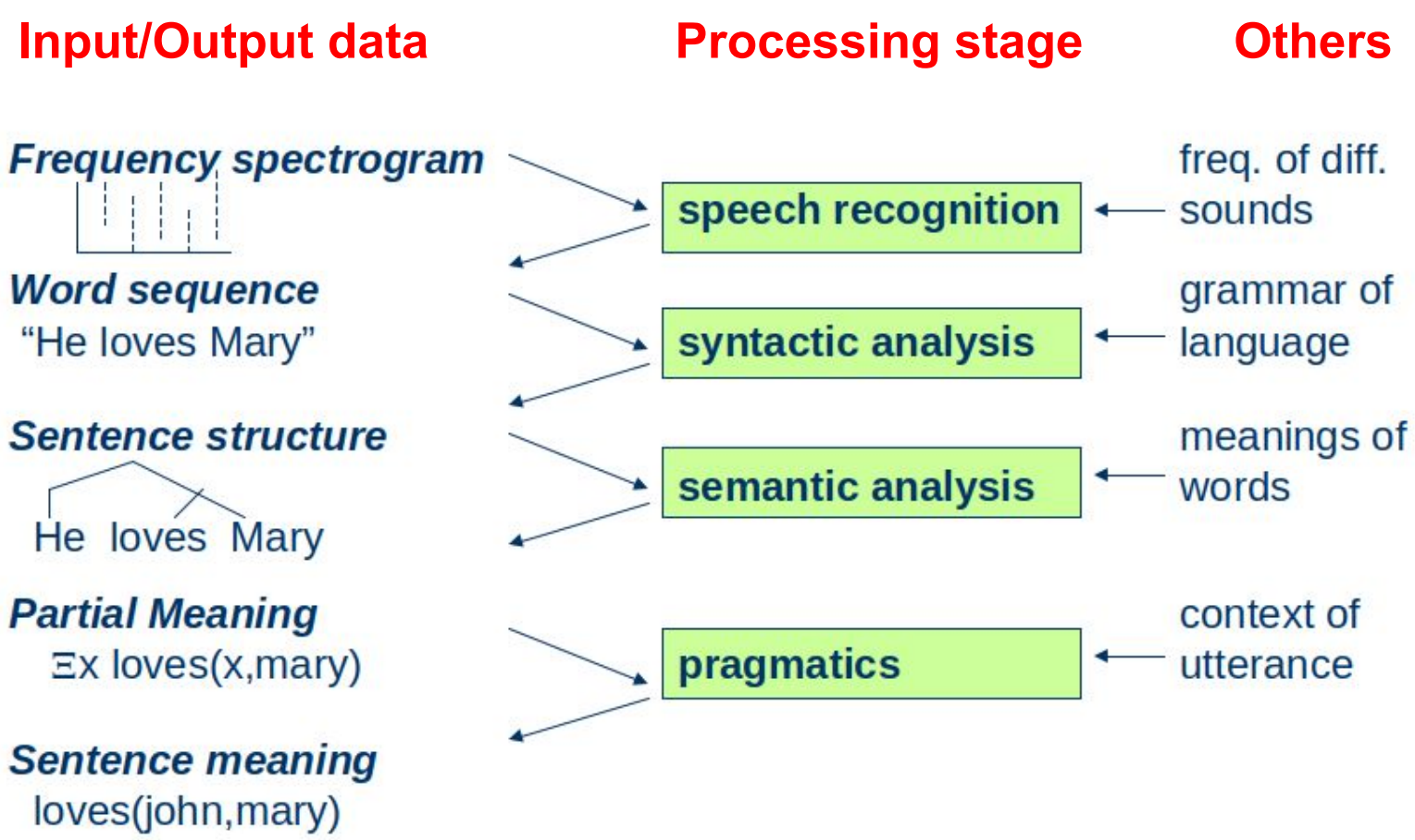
↓ • **Semantic analysis** using info. about meaning of words

Partial representation of meaning of sentence

↓ • **Pragmatic analysis** using info. about context

Final representation of meaning of sentence

Natural Language Understanding



Phonetics, Phonology

- Phonetics is the study of human sounds.

- Phonology is the classification of the sounds within the system of a particular language or languages.Phonetics is divided into three types according to the

    - production (articulatory),
    - transmission (acoustic) and
    - perception (auditive) of sounds.

-
- · Three categories of sounds must be recognised at the outset: phones (human sounds), phonemes (units which distinguish meaning in a language), allophones (non-distinctive units).

.

Phonetics, Phonology

- Production (articulatory):
  - How speech sounds are articulated. Description and classification of speech sounds.
- Transmission (acoustic):
  - How speech sounds are generated and how they are transmitted. The relationship between articulation and acoustic output.
- Perception (auditive):
  - How human ears perceive speech sounds

.

Phonetics, Phonology

- Phones:
  - An individual sound of speech; an elementary sound unit.
    - [k], [b], [f], [u], [d]
- Phonemes:
  - The smallest sound unit in a language that distinguishes word meanings.
    - /p/, /b/, /e/
- Allophones:
  - is a variant of a phoneme.
    - [p^h] as in “pin” and [p] as in “spin”

IPA

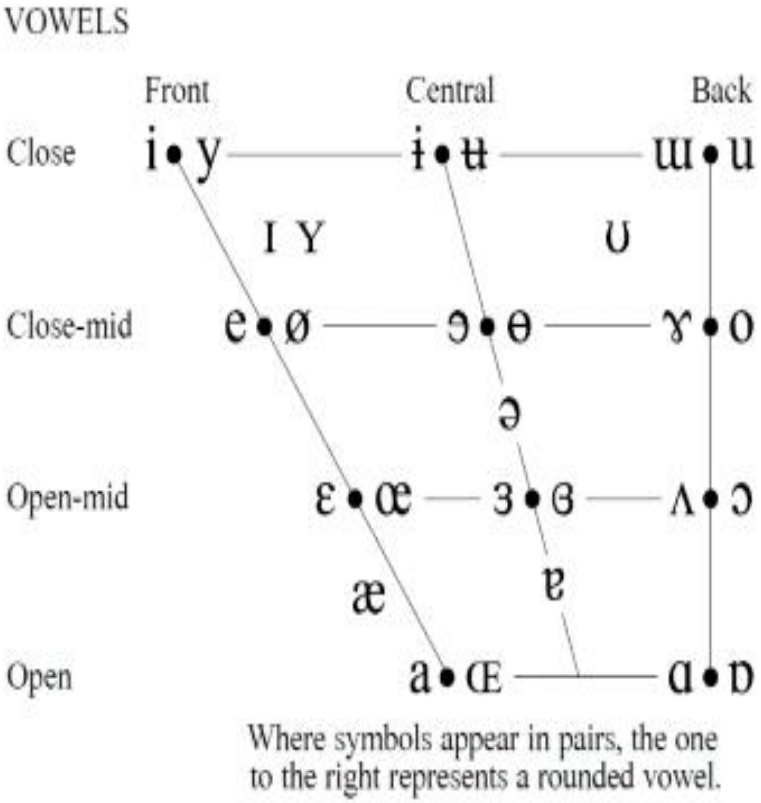
THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

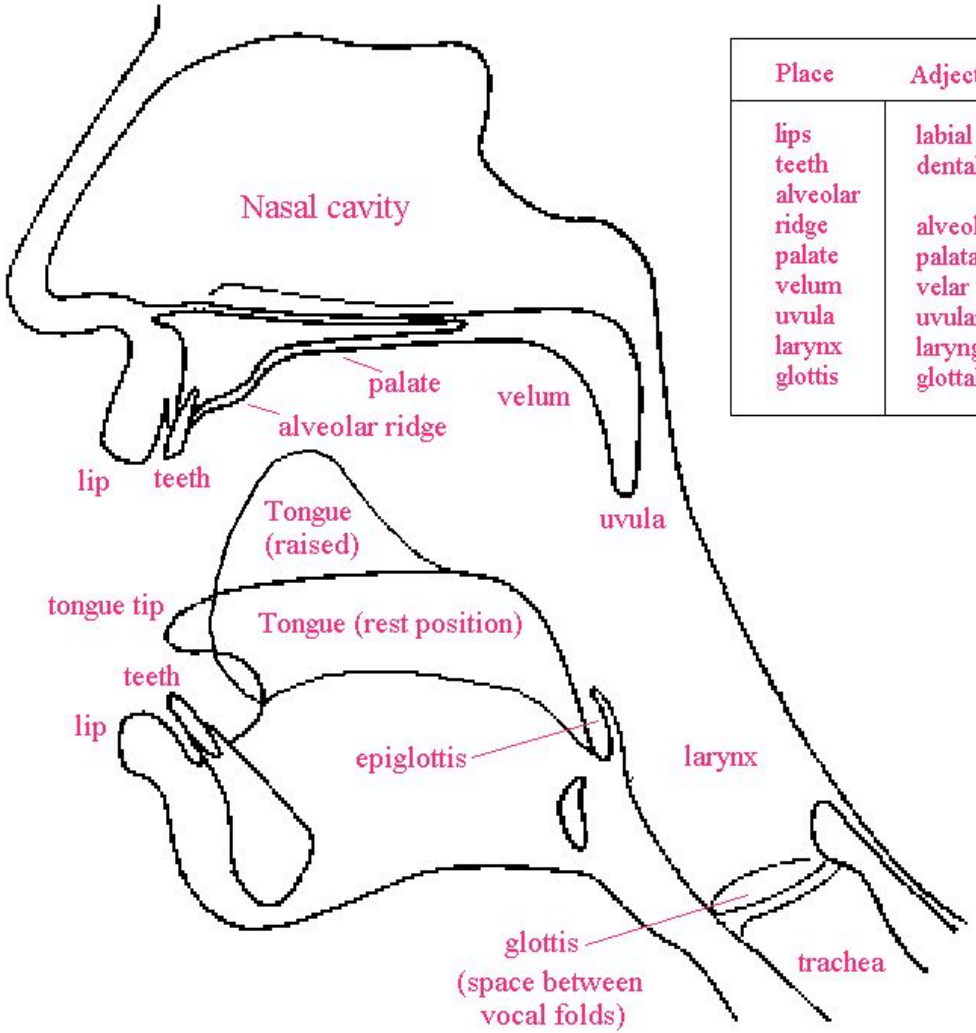
IPA


VOWELS
Front
Central
Back
Close
Close-mid
Open-mid
Open
Where symbols appear in pairs, the one to the right represents a rounded vowel.

IPA

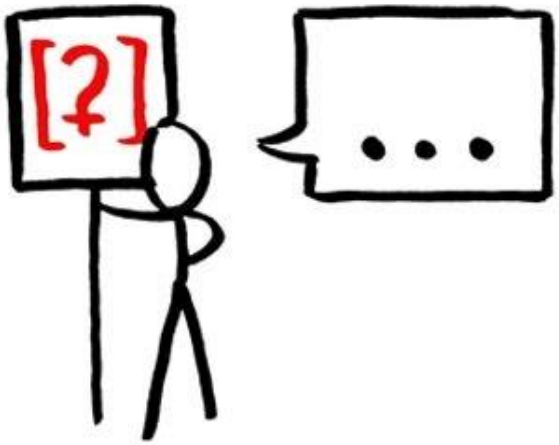
Articulation

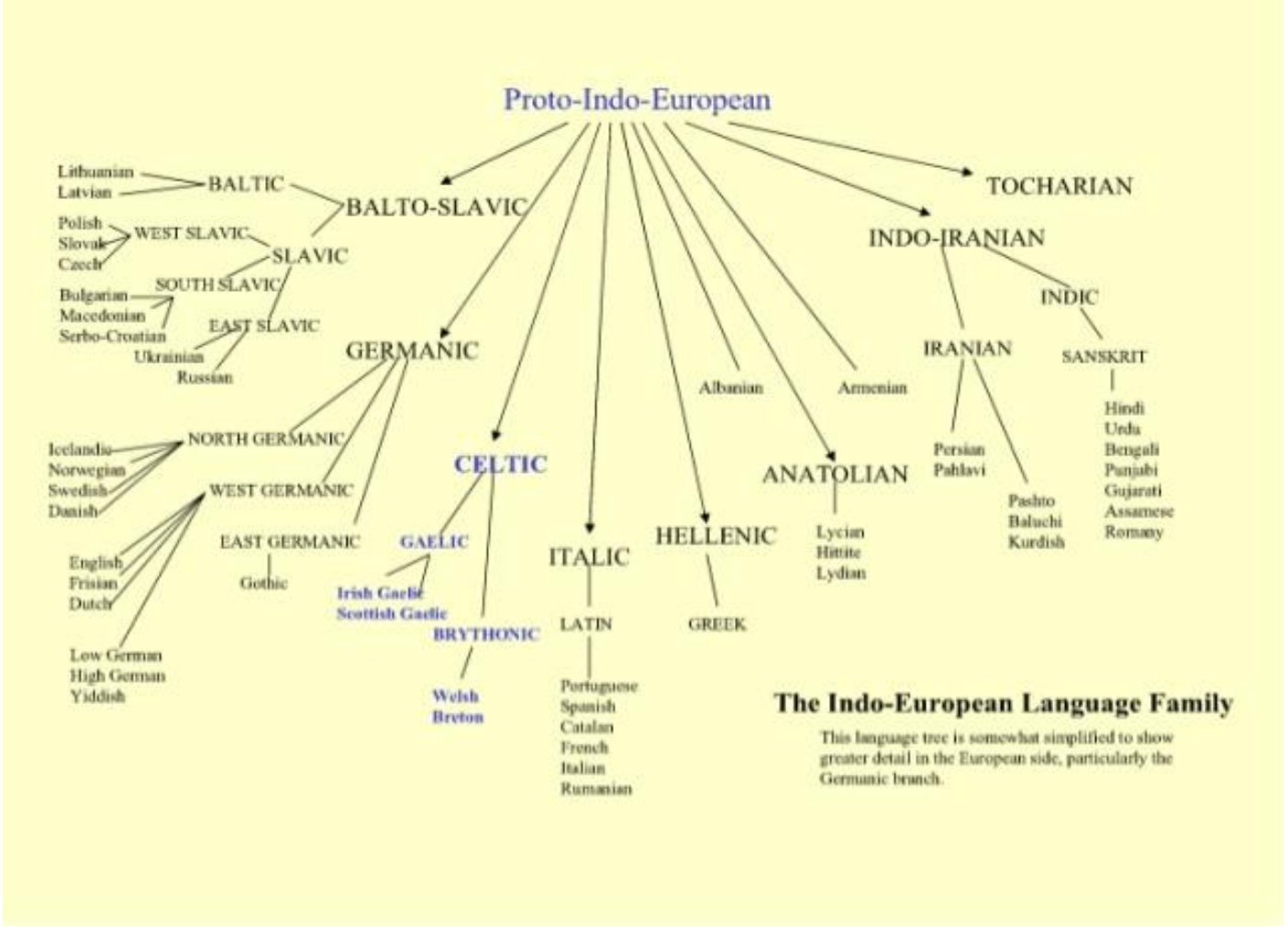
- Places
  - lips (labial)
  - teeth (dental)
  - alveolar ridge (alveolar)
  - hard palate (palatal)
  - soft palate (velar)
  - uvula (uvular)
  - pharynx (pharngeal)
  - larynx/glottis (glottal)
- Manners
  - stop/plosive
  - fricative
  - nasal
  - lateral
  - glide/approximant
  - trill
  - tap/flap



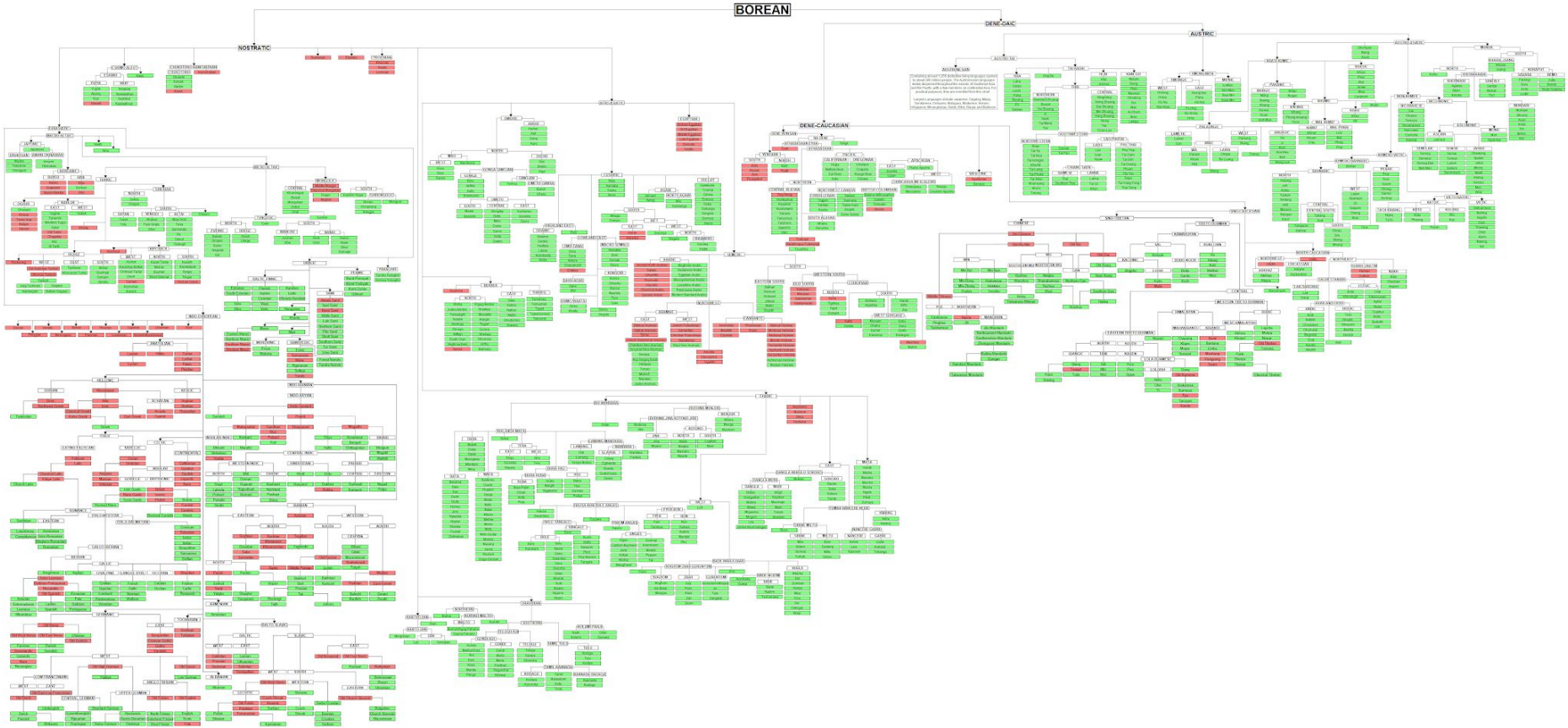
Place	Adjective
lips	labial
teeth	dental
alveolar ridge	alveolar
palate	palatal
velum	velar
uvula	uvular
larynx	laryngeal
glottis	glottal

IPA: Articulation



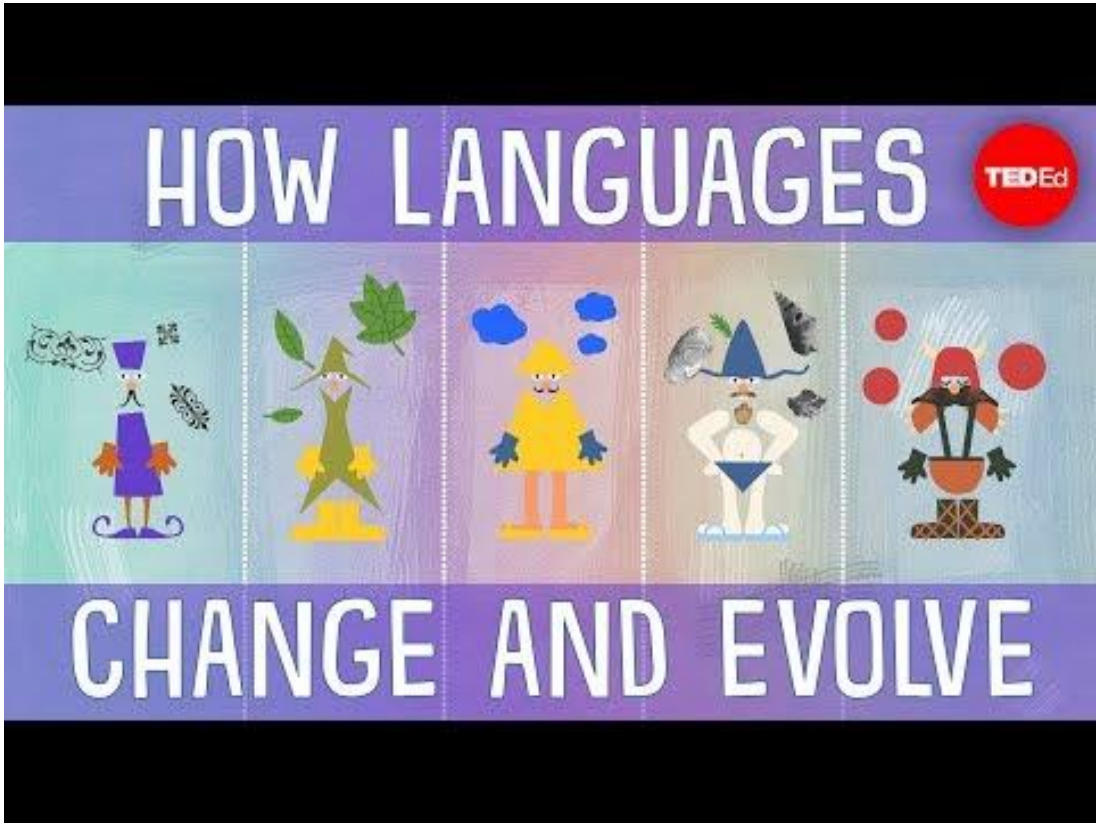


https://image.slidesharecdn.com/indoeuropean-language-tree-1192712218474033-1/95/indoeuropean-language-tree-4-638.jpg?cb=1422672641



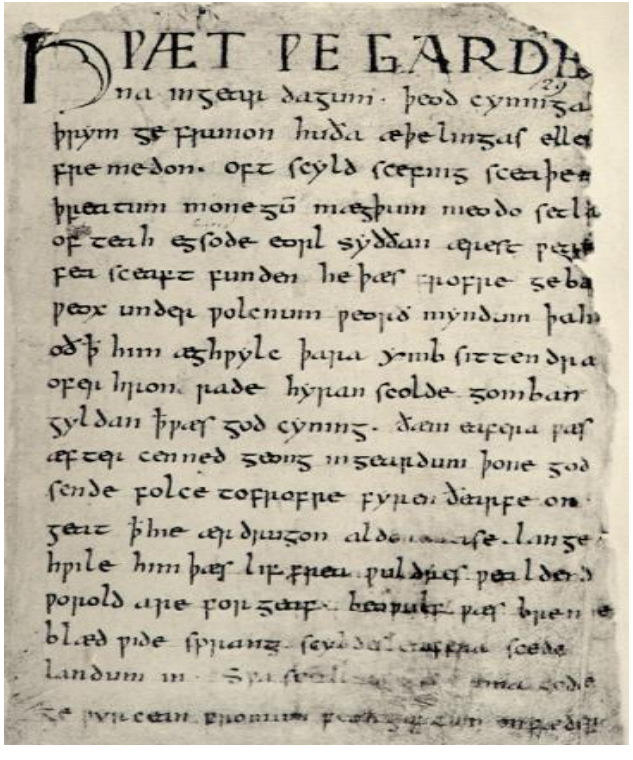
https://upload.wikimedia.org/wikipedia/commons/1/14/BoreanLanguageTree.png

Language Family



Introduction

- What language?



- Old English
- 8th-11th Century

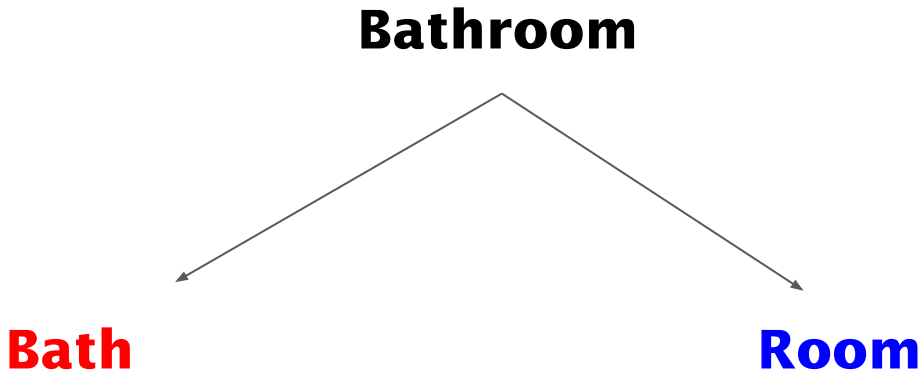
Beowulf

Hwæt! We Gardena in geardagum,
þeodcyninga, þrym gefrunon,
hu ða æþelingas ellen fremedon.
Oft Scyld Scefing sceaþena þreatum,
monegum mægþum, meodosetla ofteah,
egsode eorlas. Syððan **ærest** wearð
feasceaft funden, he þæs frofre gebad,
weox under wolcnum, weorðmyndum þah,
oðþæt him æghwylc þara ymbsittendra

Lo! the Spear-Danes' glory through splendid achievements
The folk-kings' former fame we have heard of,
How princes displayed then their prowess-in-battle.
Oft Scyld the Scefing from scathers in numbers
From many a people their mead-benches tore.
Since **first** he found him friendless and wretched,
The earl had had terror: comfort he got for it,
Waxed 'neath the welkin, world-honor gained,
Till all his neighbors o'er sea were compelled to …

Morphology

- The study of meaningful components of words.
    - It is the smallest meaningful units in a language.



Syntax

- The study of structural relationships among words

The boy is playing football in the field.

The/DT boy/NN is/VBZ playing/VBG football/NN
in/IN the/DT field/NN

Lexical semantics

- The study of word meaning
  - Open
  - Close
  - ……..

Compositional semantics

- The study of the meaning of sentences
  - The boy is playing football in the field.
  - Ravi closed the door.
  - …….

Pragmatics

- The study of the use of language to accomplish goals
  - Lift your arm
  - Where is your shoe?

Discourse

- The study of conventions of dialogue

  - Satya started an export business in 2010. After five years he started another business.

  - Satya started an export business in 2010. After **five years** **he** started **another business**.

Language Understanding

Lexical Analysis − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.

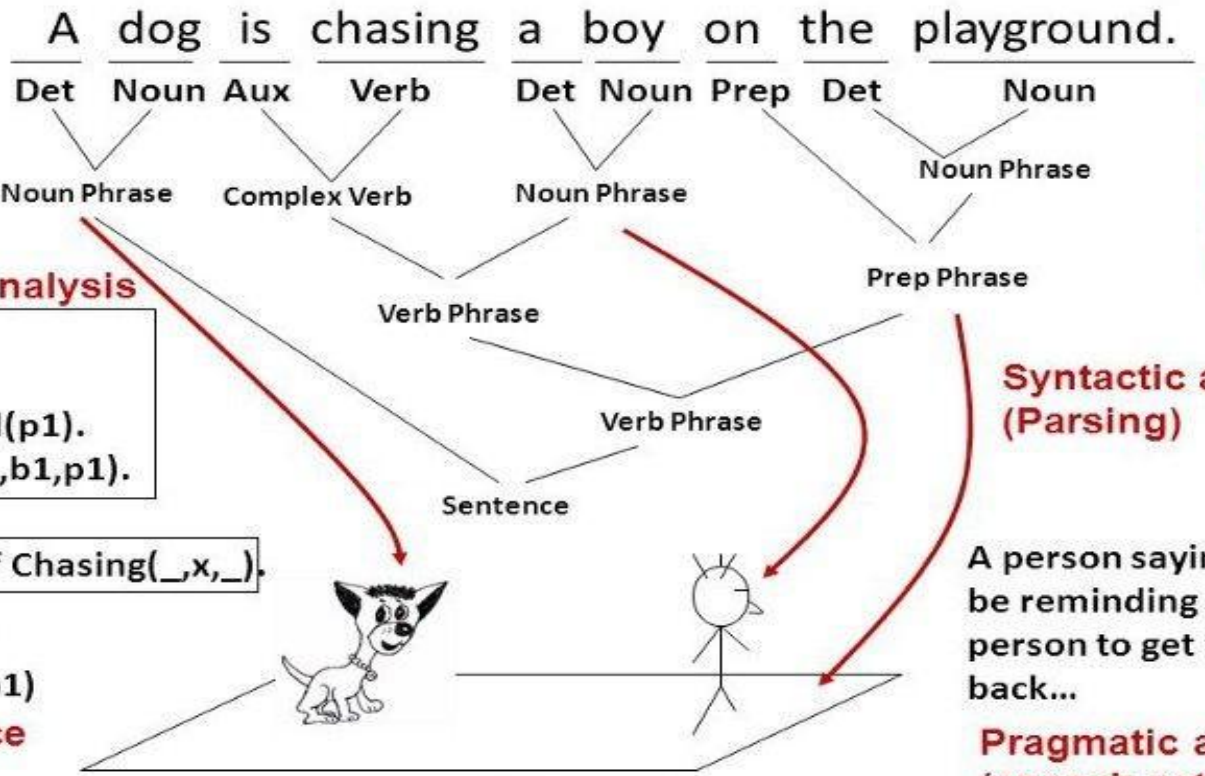
Syntactic Analysis (Parsing) − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.

Semantic Analysis − It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.

Discourse Integration − The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.

Pragmatic Analysis − During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

An example of NLP



NLP hard

- Natural Language is:
  - Ambiguous
  - complex and subtle use of context to convey meaning
  - fuzzy, probabilistic
  - involves reasoning about the world

- But NLP can also be surprisingly easy
  - sometimes rough text features can often do half the job

NLP hard

- Language is a complex social process
- Tremendous ambiguity at every level of representation
- Modeling it is AI-complete (requires first solving general AI)

NLP hard

- Speech acts (“can you pass the salt?)
- Conversational implicature (“The opera singer was amazing; she sang all of the notes”).
- Shared knowledge (“Clinton is running for election”)
- Variation/Indexicality (“This homework is wicked hard”)

Ambiguity

- Meaning
- Different ways to interpret sentence
- Interpreting pronouns
- Basing on context

Ambiguity

- *One morning I shot an elephant in my pajamas*

Ambiguity

- *One morning I shot an elephant in my pajamas*

  - *One morning I shot an elephant in my pajamas*
  - *One morning I shot an elephant in my pajamas*
  - *One morning I shot an elephant in my pajamas*
  - *One morning I shot an elephant in my pajamas*

Ambiguity

- *One morning I shot an elephant in my pajamas*

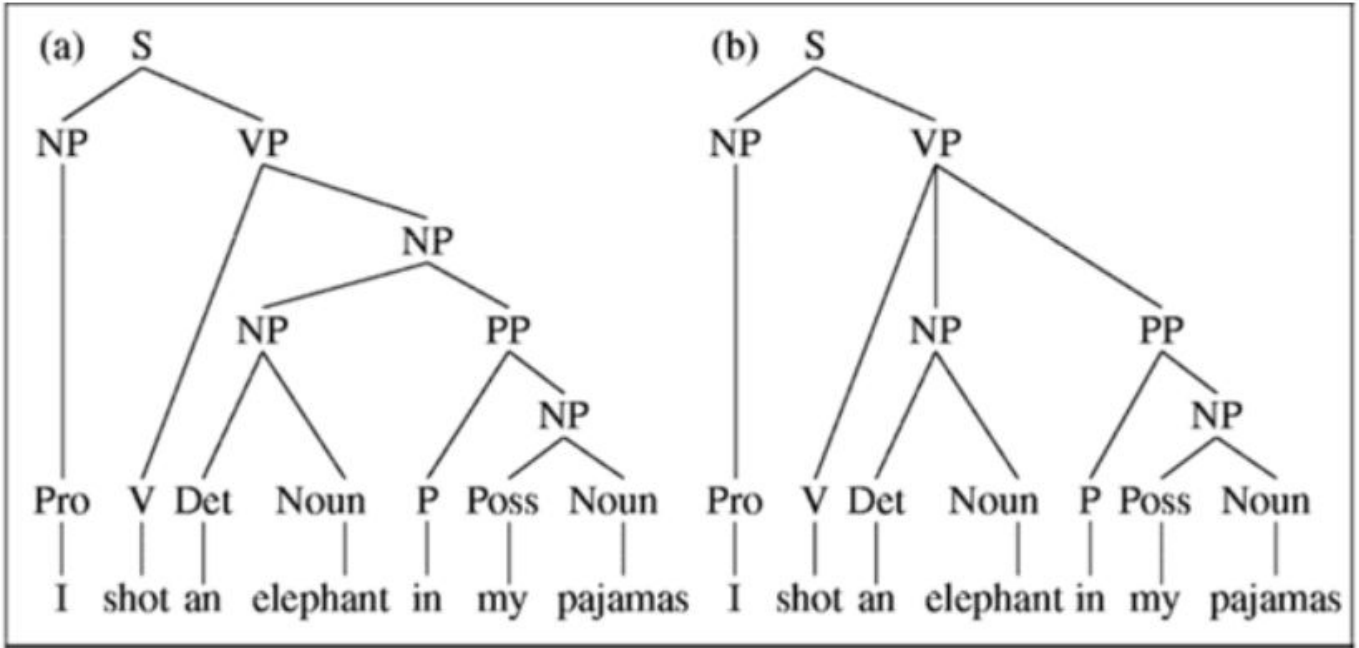


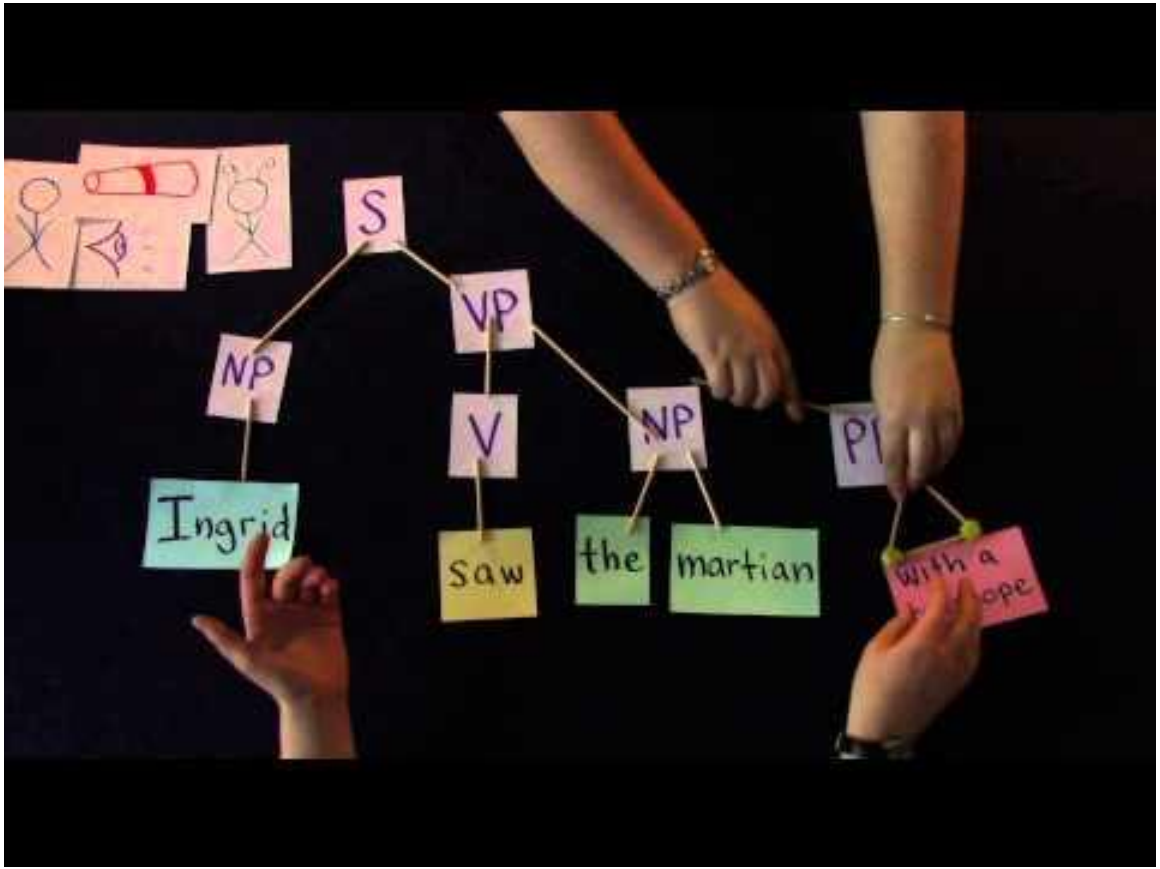
Figure 10.11 Two parse trees for an ambiguous sentence. Parse (a) corresponds to the humorous reading in which the elephant is in the pajamas, parse (b) to the reading in which Captain Spaulding did the shooting in his pajamas.

Ambiguity

“One morning I shot an elephant in my pajamas”

Communication involves recursive reasoning: how can X choose words to maximize understanding by Y?

Ambiguity



https://www.youtube.com/watch?v=djK1t1eiPnc

Ambiguity

I made her duck.

I cooked duck for her.

I cooked duck belonging to her.

I created a toy duck which she owns.

I caused her to quickly lower her head or body.

I used magic and turned her into a duck.

Ambiguity

Ambiguity: Word(Lexical Ambiguity)

- ball, board, plant, tank, bank, pen, fast, bat, cricket etc.
  - The tank was full of water.
  - I saw a military tank.

Ambiguity: Word(Lexical Ambiguity)

- She bagged two **silver** medals. [Noun]
- She made a **silver** speech. [Adjective]
- The man's hair **silvered** very attractively. [Verb]

Ambiguity

Light

Bring a light for me

Ambiguity: Word Sense Disambiguation

Understanding the sense of a word based on it's context:
E.g. I am going to withdraw money from the bank.

The fisherman is sleeping at the bank.

The security is sleeping at the bank.

Knowledge Based Approaches

Rely on knowledge resources like WordNet, Thesaurus etc. May use grammar rules for disambiguation.

May use hand coded rules for disambiguation.

Ambiguity: Word Sense Disambiguation

Machine Learning Based Approaches

Rely on corpus evidence.
Train a model using tagged or untagged corpus.

Probabilistic/Statistical models.

Hybrid Approaches

Use corpus evidence as well as semantic relations form WordNet.

Ambiguity: Referential Ambiguity

Very often a text mentions an entity (someone/something), and then refers to it again, possibly in a different sentence, using another word.
Pronoun causing ambiguity when it is not clear which noun it is referring to.

Examples:

Navin met Hari and Kusum. They went to restaurant [Hari and Kusum or all of them?]

Jay met Venu before he went to store [is he Jay or Venu?]

Ambiguity: Incomplete sentence

Incomplete sentence where missing item is not clear

Example:
"Navin worked hard and passed the exam. Nayan too" Three possible interpretations of

this example are

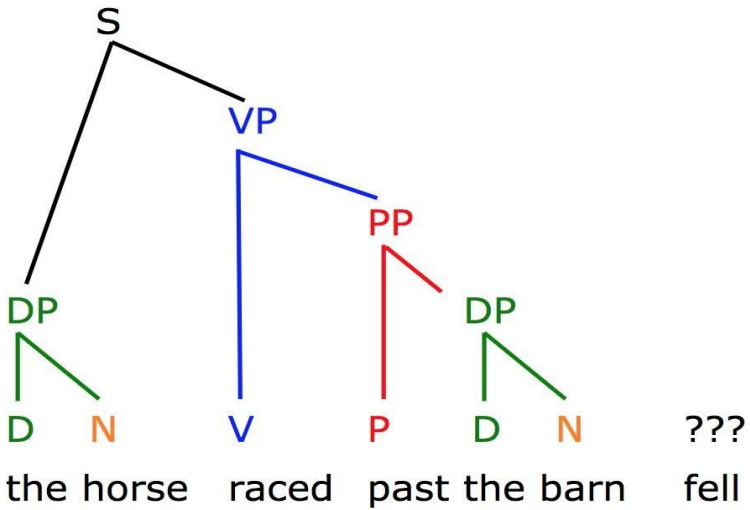
- Nayan worked hard
- Nayan passed the exam
- Nayan did both

Structural Problems

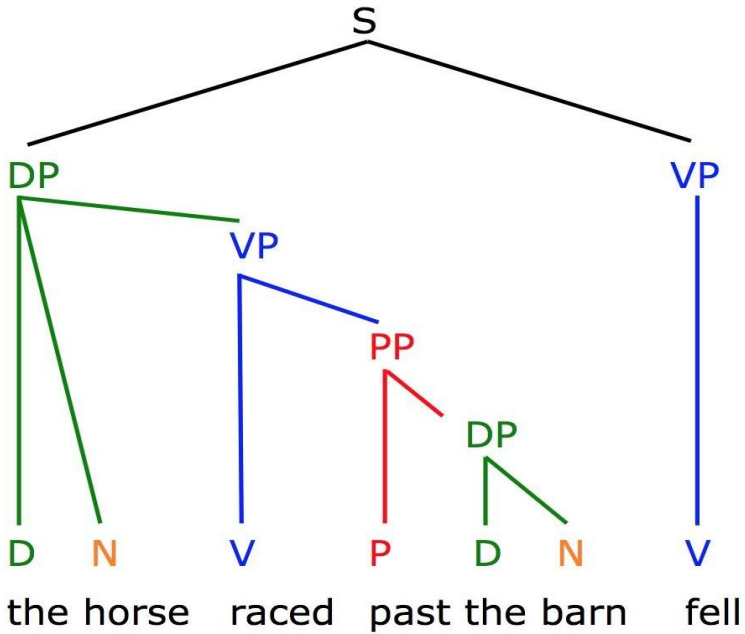
- Beverly Hills
- Beverly Sills
- The box is in the pen
- The pen is in the box
- Mary and Sue are mothers
- Mary and Sue are sisters
- Every American has a mother
- Every American has a president
- We gave the monkeys the bananas because they were hungry
- We gave the monkeys the bananas because they were over-ripe

Grammatical Structure

The ungrammatical structure



The grammatical structure



Structural Problems

- Beverly Hills
- Beverly Sills
- The box is in the pen
- The pen is in the box
- Mary and Sue are mothers
- Mary and Sue are sisters
- Every American has a mother
- Every American has a president
- We gave the monkeys the bananas because they were hungry
- We gave the monkeys the bananas because they were over-ripe

Ambiguity

https://www.youtube.com/watch?v=ldT2g2qDQNQ