Linguistics for CS

Lecture 1

Anca Dinu

FMI master programmes

University of Bucharest, October 2025

01 What is Linguistics?

02 Linguistic Levels

03 Computers and Human Language

04 Modern Linguistics - Structuralism

What is Linguistics (not)?

IS

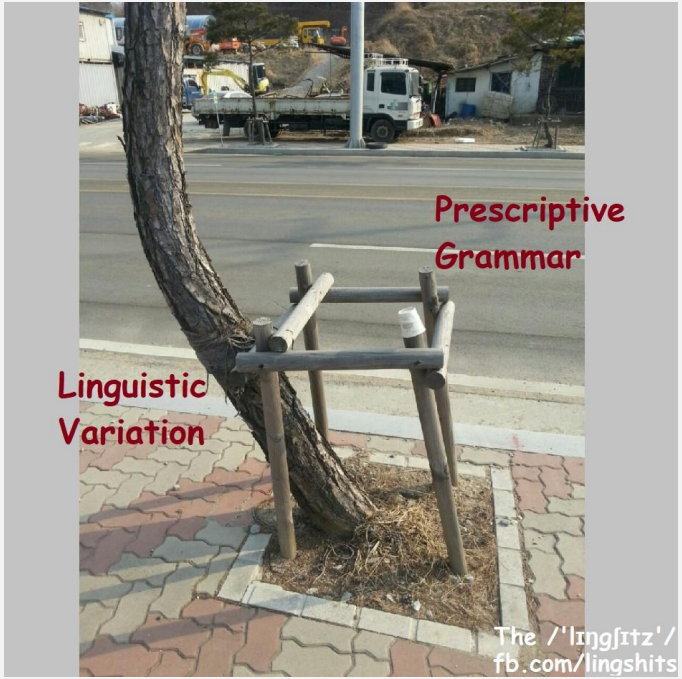
- Linguistics is the science of language.
- Like any science, it attempts to systematize and explain a domain of the empirical world - language.

IS NOT

- The learning of many languages
- Literary criticism
- Comparative philology
- Traditional study of grammar

What Linguistics is NOT

- **The learning of many languages**: few linguists are fluent in more than two languages. Speaking many languages doesn't make you a linguist but being able to speak about the language (David Crystal, 1941).
- **Literary criticism**: Linguists do not try to evaluate literary text by aesthetic or moral standards.
- **Comparative philology**: historical linguistics (how the languages appeared and evolved) is just a sub-field of Linguistics.
- **Traditional study of grammar** from school is prescriptive, while linguistic study is descriptive.



What Linguistics IS

- Linguistics is the science of language. It attempts to systematize and explain a domain of the empirical world:
  - Particular languages (accurate and complete descriptions)
  - Language in general (how language works: universal grammar, language acquisition, et.)
- Empirical = based on, concerned with, or verifiable by observation or experience rather than theory or pure logic (from empiria [gr] = experience).
- What facts of the empirical world constitute the object of linguistics?
- The observable data used in linguistics are the examples.

HIBRID DISCIPLINES (applied linguistics)

COMPUTATIONAL-LINGUISTICS

PSYCO-LINGUISTICS

CORPUS-LINGUISTICS

NEURO-LINGUISTICS

FORENSIC-LINGUISTICS

CLINICAL-LINGUISTICS

What Linguistics IS

(1) There is *language.* (the ‘factum linguae’)

- There is first *the fact that people talk* and that ability to speak is one of man's essential properties.
- The *existence* of *language* is an *axiom* for linguistics.
- If one wonders on the *conditions of possibility for language*, one would be conducting an investigation within the *philosophy of language*, not linguistics.
- From the point of view of linguistics, language is given.

What Linguistics IS

(2) There are *languages.* (the 'factum linguarum').

- People speak (different) languages (6000-7000). In saying that there are languages, there are certain implicit presuppositions:
  - that one knows what counts as a language;
  - that one can reliably distinguish between one language and another;
  - that, although languages are quite diverse in their structure, they are sufficiently alike to make up one natural kind, the language kind.
- In other words, Katz (1985):

  "Language is effable, as opposed to ineffable"

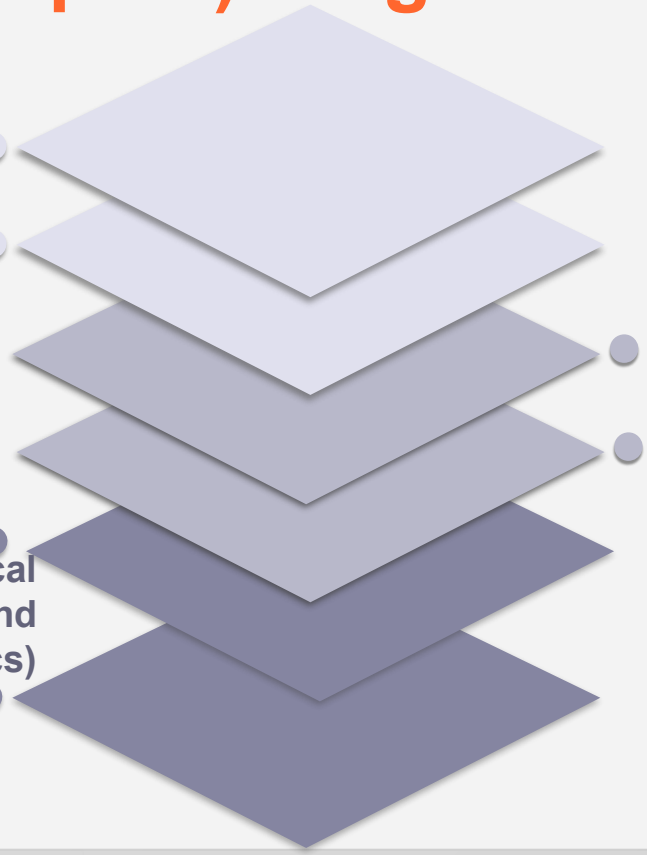
  (effable=describable in words).

(Isomorphic) Linguistic Levels

SOUND

PHONETICS
Properties of sounds

PHONOLOGY
Patterns of sounds



MORPHOLOGY
Structure of words

SYNTAX
Structure of sentence

STRUCTURE

MEANING

SEMANTICS
Meaning of words (lexical semantics), sentences and discourse (formal semantics)

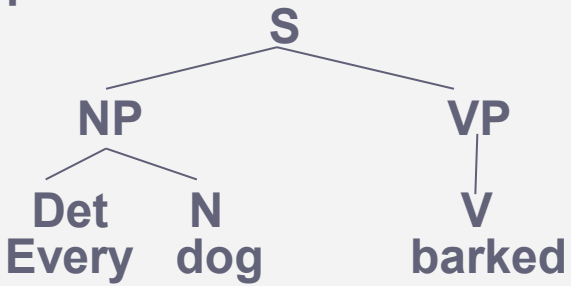
PRAGMATICS
Use of language (in extra-linguistic context, intentions)

(Isomorphic) Linguistic Levels

- Let's sketch the representation of the following utterance at different linguistic levels:
    - *Every dog barked.*
- phonemic representation
    - e^v^r^i^d^o^g^b^a:^k^t
- morphemic representation
    - every^dog^ba:k^d

(Isomorphic) Linguistic Levels

- syntactic representation



- semantic representation

$$\forall x.(dog(x) \rightarrow barks(x))$$

COMPUTERS AND HUMAN LANGUAGE

- It so happens that what computers are best at is precisely representing and working with quantification and structures, two of the most prominent features of natural language, so they seem a perfect fit.
- However, people struggled for 70 years to make computers understand and produce language, until very recently with lame results.
- Why computers did not reach the standard of human language (speech and written text), for such a long period of time, although they can process data very fast and store huge amounts of information?
- Possible answers:
  - Computers are dumb.
  - Machines work differently than human brain.
  - Humans do not fully understand the way language works so they cannot programme the computers accordingly.

COMPUTERS AND HUMAN LANGUAGE

- **Human language is at the core of the 4th Revolution:**

1st: Industrial

Cause: Steam power
Effect: Mass production

2nd: Technological

Cause: electricity
Effect: globalization

3rd: Digital

Cause: Calculus power, www
Effect: Instant connection

4th: Smart?

Cause: *Big Data in Human Language*
Effect: Social?

COMPUTERS AND HUMAN LANGUAGE

Characteristics of the fourth industrial revolution :

- **Training data**: is obtained mostly through crowdsourcing of social networks, which reflect language, culture and social norms.
- **Speed**: 4th I.R. happens too fast; we don't have time to adjust and regulate;
- **Issues**: safety, distinguishing between fact and fiction/fake news, ethics.
- The new "digital minds" are increasingly powerful and no one, not even their creators, can reliably understand, predict or control them. Now we have **two problems**: we do not completely understand how **human language** works and we do not understand exactly how **large language models work**.

COMPUTERS AND HUMAN LANGUAGE

- It is no longer enough to do ordinary coding (GPT4 codes reasonably in any programming language; recent firings by big companies like Google)
- The trend in Humanities is to create user-friendly analysis tools and interfaces, that everyone can use without having a technical background.
- Rather, in this dynamic context, there is a need for smart, critical, creative, adaptive and perhaps even visionary computer scientists, who really understand the nature and the principles of the problem they are trying to solve.
- And since natural language is such a big part of the 4th industrial revolution, knowing the basic linguistic theories is a must for CSs.

COMPUTERS AND HUMAN LANGUAGE

- **Artificial Intelligence** is a reality, but so is **Artificial Stupidity**, manifesting itself in all areas of NLP (speech, chatbots and digital assistants), even in medical engineering.
- Take a look at the series of commercials that mock AI and militate for simplicity “smart house", "smart boat”, etc. at

https://www.youtube.com/watch?v=nwPtcqcqz00

MODERN LINGUISTICS

- Four main schools of linguistics:
  - FUNCTIONALISM
  - STRUCTURALISM
  - GENERATIVISM
  - COGNITIVISM

MODERN LINGUISTICS - STRUCTURALISM

- Some form of Linguistics (grammar) has always been with Humanity since Antiquity.
- In the Middle Ages the **Humanism** consisted of the standard seven liberal arts – the trivium (grammar, rhetoric and logic) plus the quadrivium (arithmetic, geometry, music and astronomy).
- But modern Linguistics begins with **Ferdinand de Saussure**' Structuralism.
- **Cours de linguistique générale** (1916) has been published after his death by his students.

MODERN LINGUISTICS - STRUCTURALISM

- The fundamental idea of structuralism is that an object is not defined by its intrinsic properties, but by the larger **structure** it is part of.
- Examples:
- Phonemes: pow / bow
- The meaning of words is defined in each language by the relation with other words and their meanings. Thinking is determined by Language! (Wittgenstein).
- It comprises the **fundamental concepts** of modern linguistic:
  - the **linguistic sign**
  - Language as a (static) system of interconnected units: **relations and oppositions**
- This theory shifts from diachronic to synchronic approach.
- It was the starting point of more complex theories.

THE LINGUISTIC SIGN

“The linguistic sign is the association of an acoustic image (the signifier) with a concept (the signified).”

Hat

/hæt/
PĂLĂRIE

F. de Saussure

OPPOSITIONS

- *"Language is entirely based on oppositions."*
- *"The units of language only exist by virtue of their relations, each one of them being what the others are not."*
- **Starting from these basic statements, linguistic structuralism develops a theory of oppositions and a theory of relations.**
- **The concept of opposition can be defined in set theoretic terms: an opposition is an ordered pair of sets of elements A: B where**
  - **The elements common to sets A, B form the base of the opposition.**
  - **The elements of A which are not elements of B, and those elements of B which are not elements of A form the differential set of A: B, or the characteristic of the opposition A: B.**

EXAMPLE OF OPPOSITIONS

- An opposition presupposes a basis of comparison (a common element, shared by the opposed terms) and the characteristic of the opposition (a differentiating feature).

p	b	
+consonant	+consonant	base of the
+bilabial	+bilabial	opposition
+plosive	+plosive	
---	---	
-voice	+voice	characteristic of the opposition

EXAMPLE OF OPPOSITIONS

- Consider the opposition om: oamenilor, analysable as:

om	: oamenilor	
+ noun	+ noun	base
+ masc	+ masc	
---	---	
+ Nom-Acc	+Gen-Dat	differential set
+ singular	+ plural	

TYPES OF OPPOSITIONS

1. **Privative** oppositions - One term is positively marked for a feature absent in the other term (inclusion).

- Example:

	A	: B
a.	p	: b

'b' has the feature [+ voice], missing in 'p‘.

b. horse : mare

'The semantic features of horse are included in those of mare, but mare contains the additional feature [+female]

c.

- Privative oppositions are the most characteristic, since they are binary and oriented; the term positively marked for the 'additional feature' is the 'marked' term of the opposition.

TYPES OF OPPOSITIONS

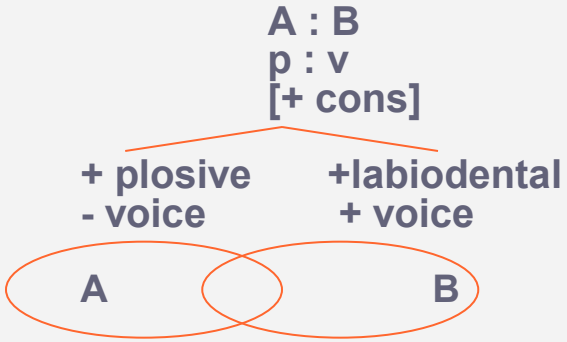
2. **Equipolent** oppositions - Both terms A, B have specific differential features; in set-theoretic terms, some of the elements in A also belong to B, and some elements of B also belong to A (intersection).

- Example:

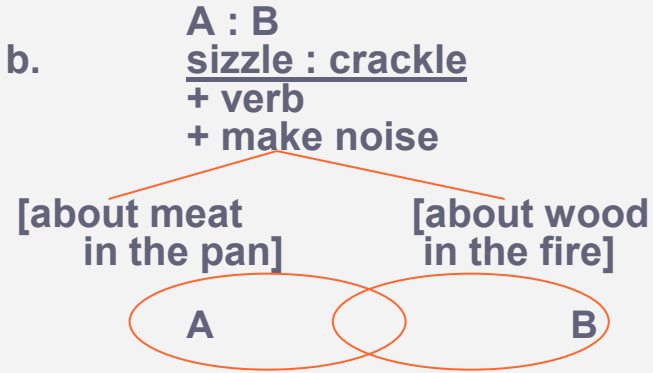
a. p:v; p is plosive and voiceless; v is labiodental and voiced; both share the fact of being consonants.

b. the lexical opposition sizzle : crackle.

a.



b.



TYPES OF RELATIONS

- Linguistic units entertain mutual relations on:
  - the vertical, paradigmatic axis, where units stand in opposition and are mutually exclusive, contracting a relation of *substitution*, an *either* ... *or* relation (cf. Hjelmslev).
  - the horizontal, syntagmatic axis, where units co-occur forming a syntagm, contracting a syntagmatic relation, called *both..... and* relation.
- To describe a linguistic system is to specify both the membership of the paradigmatic sets and the possibilities of combination of one set with another one, in well-formed syntagms.

TYPES OF RELATIONS

- Morpho-syntax example:

	Syntagmatic axis					
Paradigmatic axis	The	cat	sat	on	the	mat
	His	dog	slept	under	the	table
	Our	parrot	perched	in	its	cage

TYPES OF RELATIONS

- **Phonology example:**

	Syntagmatic axis			
Paradigmatic axis	c	l	a	y
	p	l	a	y
	t	l	a	y

THANKS

You should have some questions now!

anca.dinu@lls.unibuc.ro
ancaddinu@gmail.com
+0785641041

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik

Content contributions: some of the slides by Alexandra Cornilescu