Linguistics for CS

Lecture 5 - Syntax1

Anca Dinu

NLP master programme

University of Bucharest, October 2025

01 Formal vs. Traditional Grammars

02 Generative Grammar as a continuation of Structuralism

Recap

- **Modern** Linguistics begins with Ferdinand de **Saussure**' Structuralism and comprises the **fundamental concepts** of modern linguistic:
  - the **linguistic sign**
  - Language as a system of **relations and oppositions**
- This theory is founded on a few **maximally general concepts and maximally simple formalization**, in the absence of any more complex formal machinery.
- It was the **starting point** of more complex theories.

Recap

- **What facts of the empirical world constitute the object of linguistics?**
  (1) There is *language.* (the `factum linguae`)
  (2) There are *languages.* (the `factum linguarum`)
  (3) There are *grammars.* (the `factum grammaticae`)

- **Teaching grammar in school is quite an ancient activity.**
- **Linguistics, as a comparatively more recent discipline, is founded on *grammatical activity (as teached in schools)*.**

Formal vs. Traditional Grammars

Formal Grammars

synchronic
descriptive
spoken language
holistic view
stress on relations
immanent definition of lg
formal definitions

Traditional Grammars

diachronic
normative-prescriptive
written language
atomistic view
stress on independent units
instrumental definition of lg
notional definitions

Recap

Examples of notional definitions in traditional grammar:

- **“Substantivul este partea de vorbire flexibilă care denumeşte obiecte in sens larg”.**
- **"Subiectul este partea de propoziţie care arată cine înfăptuieşte acţiunea exprimată de predicatul verbal sau cui i se atribuie o însuşire exprimată prin numele predicativ.”**

Recap

Examples of formal definitions in structural grammars:
(from C. C. Fries, *The Structure of English)*

- **The noun**: The word class whose members occur in the context "The ---- is / are good."
- **The verb**: The word class whose members occur in the contexts:
  - to---(to ask);
  - ---ing (asking);
  - --- s (asks)
  - --- ed (asked)

Recap

- Compare the following two definitions of transitive verbs. Which one is formal?

1. A transitive verb is a verb whose action "passes onto an object".

2. A transitive verb is a verb that occurs immediately before a nominal phrase, i.e., in the context --- NP.

Generative Grammar as a continuation of Structuralism

- Linguistics in the XXth century is marked by the *advent of Structuralism*, with two qualitatively different stages in the evolution of linguistic structuralism:
  - the stage of ***analytical structuralism***, which prevails in Europe and the US in the first half of the century, starting with De Saussure's *Cours de linguistique générale* (1916);
  - the stage of ***synthetic structuralism***, i.e., the phase of Generative Grammar (GG), starting with Chomsky's *Syntactic Structures (1957)*.

Generative Grammar as a continuation of Structuralism

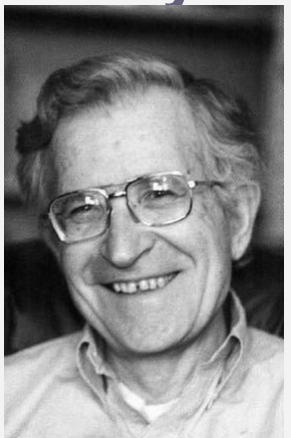
- ***Structural*** (analytic) grammars - the direction of analysis is from the (infinite) text to the (finite) invariant units and structures, and to the classification of those units.
- ***Generative*** (synthetic) grammars - start from an inventory of units (the lexicon) and a set of combinatory rules and aim at producing the language, the infinite text.
- Structural grammars are mainly paradigmatic. In fact, de Saussure defined 'langue' as a system of paradigms. Syntagmatic relations are treated as auxiliary tools for the discovery of the paradigms.
- Generative grammars are primarily syntagmatic - they offer rules for sentence construction and interpretation. The sentence itself is a syntagmatic structure. Paradigms are assumed to be given.

Empirical problems addressed by GG

Problem 1 Linguistic creativity

The problem is that the normal use of language is innovative and potentially infinite, in the sense that that much of what one says in the course of common language use is not a repetition of anything one has heard or said before. Similarly, the number of sentences in one's native language that one will understand without any feeling of difficulty is astronomical.

"The central fact to which any significant linguistic theory must address itself is that the typical language learner has observed [only] a certain limited set of utterances of his language, but can, on the basis of this finite linguistic experience, produce and understand an indefinite number of new utterances." (Chomsky 1973).



Noam Chomsky

Empirical problems addressed by GG

Proposed answer

- Users produce and understand new sentences because they *possess an internalized grammar*, a device which provides a semantic and a phonetic interpretation for any sentence of the given language, L.
- Any grammar of L will project the finite and somewhat accidental set of observed utterances to a presumably infinite set of grammatical utterances the speaker can produce or understand.

Empirical problems addressed by GG

Problem 2 Language acquisition and poverty of stimulus

- Grammars are learnable. By the age of three, any normal child has already acquired his grammatical competence; he will have mastered the morpho-syntactic structures of his language; in later years, he will mostly enrich his language at the level of the vocabulary.
- The learnability of grammars indicates that *grammars are finite devices*. A grammar G can only contain a finite number of rules. A grammar G of a language L is a f*inite set of rules* which produces (and interprets) all and only the grammatical sentences of the language L. How do we learn that?

Empirical problems addressed by GG

Proposed answer

- Given the complexity of language, the fact that language is learnable can only result from the fact that the child is *genetically endowed with a language faculty.* Chomsky's strong claim is that the psychological counterpart of UG is the child's *language faculty*, *which is itself a kind of universal grammar*, a component of the child's mind, *part of his genetic endowment*. Learning cannot simply proceed by analogy, induction and generalization, as was believed by empiricists (and structuralists).
- Chomsky radically departs from empirical theories of learning, adopting the view that *the learning of a complex system like language would be impossible in the absence of some well-structured innate mental mechanism.* The child is innately equipped with a UG.

The poverty of stimulus argument

Chomsky's most powerful argument is the argument from "the poverty of stimulus".
The mastered linguistic system is of great complexity.
Consider the pairs of questions below.

a. Who do you believe ___ came ?
b. *Who do you believe that ___ came ?
c. Who do you believe Peter saw __ ?
d. Who do you believe that Peter saw ___?

Examples in each set are perfectly synonymous, yet in the first set there is a sharp contrast of grammaticality in a) vs. b). It is quite unlikely that such grammaticality judgments, mirroring "rules" of the language, can be arrived at by induction over the data. What kind of examples should one generalize over? Moreover, there is no explicit instruction given the child to prevent the occurrence of examples like b), and such errors occur seldom if ever. "Rules" like those cannot be learned on the basis of such an 'impoverished stimulus'.
Given the 'poverty of stimulus' and the complexity of knowledge attained, a reasonable hypothesis to entertain is that the child is innately equipped with the mental analogue of a UG.

GG - *Competence and Performance.*

- **Competence** is the speaker's internalized grammar the speaker's tacit knowledge of his language, which enables the speaker to use language.
- **Performance** represents the actual use of language.

GG - *Competence and Performance.*

In time, competence has become a cover-term for several different abilities:
a) grammatical competence - meaning 'tacit knowledge of grammar';
b) conceptual (lexical) competence - ability to use and understand the vocabulary of a language;
c) pragmatic (communicative) competence - ability to communicate, to write texts, etc.

GG - *Competence and Performance.*

- Members of a speech community may differ considerably in their lexical competence (in the number of words they can use appropriately), as well as in their pragmatic competence (e.g., ability to produce different text-types varies a great deal).
- Grammatical competence, in contrast, is homogeneous across a given speech community: in what follows, 'competence' is understood as 'grammatical competence'.

GG - Grammar as a model of competence

- The linguist's grammar is a hypothesis on the structure of the speaker's internalized grammar.
- More technically, a (linguist's) grammar is said to be a ***model*** *of the speaker's competence*, a model of the speaker's internalized grammar.
- The term 'model' is understood as in the theory of modelling:

GG - *Competence and Performance.*

- *Definition of 'model'* - A model is some object or phenomenon A, which is subject to investigation as a substitute for some other object or phenomenon B, with which A is in a relation of correspondence.
- Through the study of the model A, and through the established correspondence B-A, one obtains information about the (less accessible) object B.

GG - *Competence and Performance.*

- ***The linguist's grammar is a model of competence*** **in the sense that it attempts to outline the kind of knowledge the speaker possesses, which enables him use language creatively.**
- **One thus reaches the conception of a grammar as a device capable of producing and interpreting any (therefore all) the (well-formed) sentences of a language.**

Linguistic Theory (LT) and Universal Grammar (UG)

- The idea of constructing a grammar as a system of rules leads to the fundamental problem of **justifying** this construction, of **evaluating** rival descriptions, **selecting** among them on a principled basis.
- This is why there is a need for a linguistic theory. The concept of **linguistic theory** (LT) is proposed by Chomsky in "Logical Structure of a Linguistic Theory" (= LSLT,1955, 1973).
- LT is an abstract theory which presents the basic principles and concepts of grammar, by means of which particular grammars can be written and evaluated. LT is in fact a Universal Grammar (UG).
- LT and UG should be viewed as equivalent.

Criteria of evaluation for grammars

1. External or descriptive adequacy

A grammar is externally or descriptively adequate if it is "faithful to the data", that is, if it produces all and only the correct sentences of the language.

- Exercise. We have the following data:

  Jane is here
  *The Jane is here.
  London is in England.
  *Some London is in England.

  We have the following grammars:

  G1 NP → Det + Noun
  G2 NP → (Det) + Noun

  Which of the two grammars is externally adequate with respect to these data?

Criteria of evaluation for grammars

2. Internal or explanatory adequacy

- A grammar should conform to LT/ UG, should obey the formal requirements stipulated there, should meet conditions of simplicity, and should offer intuitively satisfactory explanations, etc.
- Exercise: Suppose that G1 and G2 both generate the structuraly ambigous sentence:

Women students are tempting new subjects.

1) G1 works only with lexical categories (parts of speech): Noun (N), Verb (V), Adjective (A), Preposition (P), Copulative Verb (V cop), etc.
2) G2 possesses not only lexical categories (N, V, A, P,...), but also the corresponding phrasal categories: NPs, VPs, APs, PPs, etc.

Criteria of evaluation for grammars

- G1 can only label the constituents as to their parts of speech.

  Women students are tempting new subjects.

  N N V V_{ing} A N

- G2 can account for the syntactic homonymy:

Which grammar has better explanatory power?

Sentence constituency.
'String' versus 'Constituent'

- Sentences are hierarchically-structured strings of words.
- There is a tacit distinction between 'string' and ‘constituent’:

1. A string is defined as any sequence of two or more than two adjacent elements.
2. A constituent is a string which has formal properties, i.e., which has internal cohesion.

Sentence constituency.
'String' versus 'Constituent'

- **In the sentence**

  Women students are tempting now subjects.

- **sentence, *women, women students, women students are, students are, students are tempting new, are tempting new subjects*, etc. are all strings, but only *women, women students* and *are tempting new subjects* are constituents.**

Sentence constituency.
'String' versus 'Constituent'

- **(Sentence) constituency is the central concept of syntax. It is the Grammar's task to assign an analysis to any sentence, i.e. to find its proper constituent structure.**
- **To show the constituency of a sentence, one must indicate:**

  a) what strings (of the analyzed sentence) are constituents;

  b) what kind of constituent each one is, i.e., to what category each constituent belongs.
-

THANKS

Questions?

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 credits: slides from Alexandra Cornilescu