

# FLST: Linguistic Foundations

Francesca Delogu  
delogu@coli.uni-saarland.de

<http://www.coli.uni-saarland.de/courses/FLST/2014/>

# Linguistic Foundations

- Today
  - *Intro*
  - *Phonetics / Phonology (Trouvain)*
  
- Tuesday
  - *Phonetics / Phonology (Trouvain)*
  - *Morphology (Delogu)*
  
- Wednesday and Thursday
  - *Syntax (Drenhaus)*
  
- Friday
  - *Phonetics / Phonology - Tutorial (Trouvain)*
  - *Morphology-Syntax - Tutorial (Delogu-Drenhaus)*

# Linguistic Foundations

- Slides and exercises will be uploaded on the website as soon as possible after each lecture.
  - <http://www.coli.uni-saarland.de/courses/FLST/2011/page.php?id=schedule>
- Exercises are due before the tutorial session (morphology/syntax tutorial on Friday).

# Goal

- To provide you with a general overview of key concepts and main topics in linguistic research, including:
  - Competence vs. performance
  - Prescriptive vs. descriptive grammar
  - Phonetics / Phonology
  - Morphology
  - Syntax

# What is Linguistics?

- “The scientific study of human language” (Lyons, 1968).
  - What is the subject-matter of Linguistics?
    - What is human language?
  - Is linguistics a science?

# What is (a) language?

- *Language* as a capacity that all humans possess (and that distinguishes humans from other animals).
- *A language* as a specific instantiation of this uniquely human capacity (e.g., English, German, Basque, etc.).
  - Natural languages vs. Artificial languages
- Is linguistics about language or individual languages?
  - Both!

# Some definitions

- **Sapir (1921):** *“Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntarily produced symbols.”*
- **Bloch and Trager (1942):** *“A language is a system of arbitrary vocal symbols by means of which a social group co-operates”.*
- **Noam Chomsky (1957):** *“A language is a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements”.*

# Some distinctive features of human language

- **Arbitrariness:** There is no inherent relationship between form and meaning.
- **Duality:** Language has two layers of structure - a finite number of meaningless sound elements (phonemes) that can be recombined in different orders to form a large number of meaningful units (words).
- **Discreteness:** The sounds of language are perceived categorically, not continuously.
- **Productivity/Creativity:** The property that makes it possible for language users to produce and understand completely novel utterances.



# Implications of productivity

- The ability to produce and understand novel utterances cannot be explained by memorization of all possible sentences → the brain is finite.
- Productivity must be subject to systematic constraints, i.e., a *system of rules* for pairing sounds with meaning (a **grammar**).
- Language users must have a *subconscious knowledge* of the grammar of their language (**linguistic competence**).

# Linguistic competence

Any native speaker of a language knows:

- which sounds belong to the language and which don't
- which words are possible in the language and which are not
- which sentences are grammatical and which are not
- how to assign meaning to words, phrases, sentences
- how to use language in communicative contexts

# Subfields of Linguistics

<i>Rules</i>	<i>Levels of analysis</i>
Sound-formation	Phonetics/Phonology
Word-formation	Morphology
Sentence-formation	Syntax
Meaning-assignment	Semantics/Pragmatics

# Phonetics

- Phonetics is the study of the primary medium of human language, i.e., **sound**.
- Main areas:
  - **Articulatory phonetics**: The study of how the vocal organ produces speech-sounds.
  - **Acoustic phonetics**: The study of the physical properties of the sound-waves created by the vocal organ (duration, frequency, etc.).
  - **Auditory phonetics**: The examination of how speech sounds are perceived and identified by the hearer's ear and brain.

# Phonology

- **Phonology** is the study of the *sound system* of a particular language and the combination and distribution of sounds which occur on that language.
- The basic unit of analysis is the **phoneme**, i.e., the smallest unit of speech that can distinguish the meaning of words.
- Some issues:
  - What is the phoneme inventory of a particular language?
  - How phonemes get realized in different environments (allophones).
  - How speech sounds are combined (phonotactics)?
  - Prosody, intonation, etc.

# Morphology

- **Morphology** is the study of word structure and word formation.
- The minimal unit of analysis is the **morpheme**, i.e., an arbitrary union of sound and meaning (or a grammatical function) that cannot be further analyzed.
- Main areas:
  - **Inflectional morphology**: The study of how words vary in order to express grammatical contrasts (e.g., singular/plural).
  - **Derivational morphology**: The study of the rules of word-formation.

# Syntax

- **Syntax** is the study of how words combine into phrases and how phrases combine into sentences; the study of sentence structure.
- A theory of syntax should account for speakers' ability to produce and understand an unlimited number of sentences never produced or heard before

# Semantics

- **Semantics** is the study of linguistic meaning of words, phrases and sentences.
- Main areas:
  - **Lexical semantics:** The study of the meaning of words and meaning relationships among words.
  - **Sentential semantics:** The study of the meaning of larger syntactic units.
  - **Pragmatics:** The study of how context affects meaning.



# Interdisciplinary areas

- **Historical linguistics:** the study of how language changes over time
- **Sociolinguistics:** the study of how language is used in society.
- **Psycholinguistics:** the study of how language is processed in the mind
- **Neurolinguistics:** the study of how language is encoded in the brain
- **Computational linguistics:** the study of how natural language can be processed automatically

# Syntactic competence

- Our syntactic knowledge includes rules that specify the correct *word order* for a language, the *grammatical relations* of a sentence, the *subcategorization* properties of verbs, the *constituent structure* of sentences.
- \**Child the drank milk the.*
- *Your dog chased my cat vs. Your cat chased my dog*
- \**The boy found.*
- You don't know how good meat tastes.

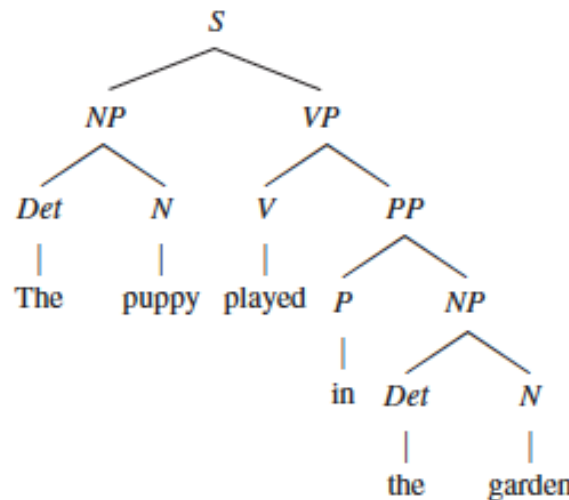
# Phrase structure (PS) rules

- PS rules make explicit this knowledge precisely and concisely

$NP \rightarrow Det N$

- A noun phrase can contain a determiner followed by a noun in that order;
- A determiner followed by a noun is a noun phrase.

- PS trees



- 1)  $S \rightarrow NP VP$
- 2)  $NP \rightarrow Det N$
- 3)  $VP \rightarrow V PP$
- 4)  $PP \rightarrow P NP$

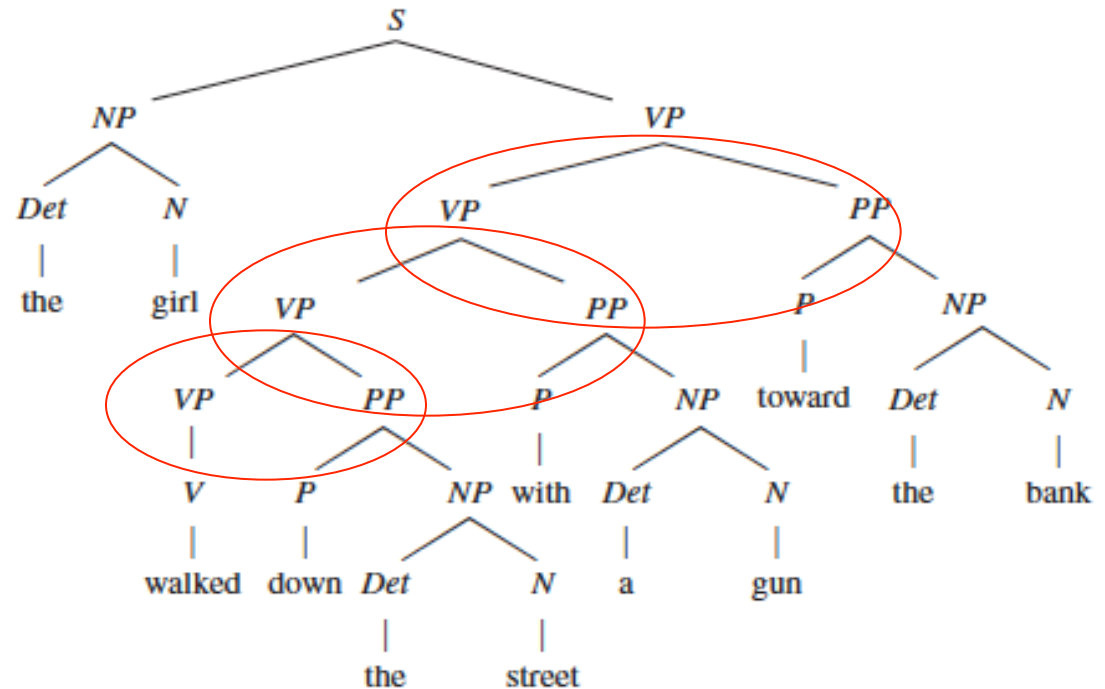
# Recursive rules

$S \rightarrow NP VP$

$VP \rightarrow VP PP$

$PP \rightarrow P NP$

$NP \rightarrow Det N$



- Recursive rules allow the grammar to produce an infinite set of sentences.

# Grammaticality

➤ Are the following sentences **grammatical**?

- *The cat sleeps.*
- *The cat that the dog chased sleeps.*
- *The cat that the dog that the boy petted chased sleeps.*
- *Colorless green ideas sleep furiously.*

# Grammaticality vs. Acceptability

- A sentence is **grammatical** if it is formed according to the rules of the language.
  - Ideally, a sentence is either grammatical or not.
  - Grammaticality should not be confused with probability of use or meaningfulness.
- A sentence is **acceptable** if it would normally occur in the language.
  - A sentence can be perceived as more or less acceptable.
- A sentence can be grammatical but unacceptable for independent reasons (e.g., hard to process).

# Competence vs Performance

## ➤ Linguistic Competence

- The implicit knowledge language users have of the grammar (the system of rules) of their language, which enables them to produce and understand any possible sentence of that language.
- Linguistics is concerned with linguistic competence.

## ➤ Linguistic Performance

- The actual use of language in real situations, which is conditioned by physiological and psychological constraints (memory limitations, shifts of attention, etc.).
- The central subject matter of Psycholinguistics.

# What is grammar

- The term *grammar* can refer to:
  - the set of phonological, morphological, syntactic and semantic rules that represent (together with a lexicon) our linguistic knowledge (***mental, internalized grammar***)
  - a model, a description, of the speaker's internalized grammar (***descriptive grammar***)
  - the set of rules that prescribe how language should be used (***prescriptive grammar***)



# Descriptive vs. prescriptive grammars

- Grammars in contemporary linguistics are *descriptive*.
  - All languages and dialects have a grammar and every grammar is equally complex, logical, capable of producing an infinite set of sentences to express any thought → Linguistics aims at describing these grammars.
- So called *traditional grammars* were *prescriptive*.
  - Prior to the birth of modern linguistics, the study of language had the primary aim of preserving the purity and perfection of the language used by great writers of the past → Grammarians aimed at prescribing grammars.

# Some prescriptive rules for English

- Don't use the double negative.
- Don't split the infinitive.
- “It's me” is wrong.
  
- These rules were introduced in “A short introduction to English Grammar with Critical Notes” (1762) written by Bishop Robert Lowth.
  
- Double negatives are standard in Roman languages.
- Many of these prescriptive rules were based on Latin grammar and made little sense for English.

# The birth of Modern Linguistics

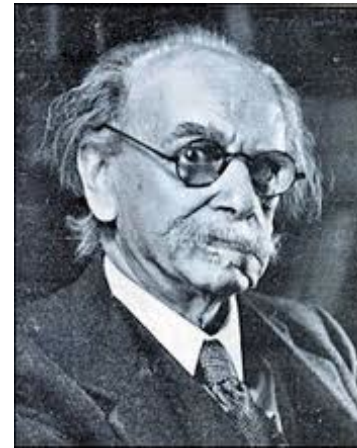
At the turn of the 20<sup>th</sup> century, a new approach to the study of language emerged.

## European Structuralism



Ferdinand De Saussure  
(1857-1913)

## American Structuralism



Franz Boas  
(1857-1913)

# Linguistics in the Structuralist Era

- According to the Structuralists, every language has its own internal structure (descriptive categories developed in traditional grammar were inappropriate to describe non European spoken languages).
- The task of Linguistics is to discover and **describe** this structure with scientific precision (i.e., using a strongly empirical and inductive approach).
- Linguistics became an autonomous discipline, whose object of investigation was human speech spoken in speech-communities (Bloomfield, 1925).

# The methodology

- Collection of transcribed corpora of spoken language produced by native speakers.
- Identification of the phonological system by segmentation and classification of the phonetic units and analysis of their arrangements → **phonetic/phonology.**
- Analysis and description of recurring groups of phonetic units identified as morphemes and of their arrangement in more complex structures → **morphology.**
- A grammar was considered to be an inventory of elements (phonemes, morphemes, morpheme sequences) and Linguistics as a classificatory type of study.

# The Chomskyan Revolution

Noam Chomsky, *Syntactic Structure* (1957) developed:

- A new scientific program - **the generative program** - arising from rejection of both traditional Structuralism and Behaviorism
- A new methodology
- A redefinition of the object of investigation

# The generative program

Developed from two key observations:

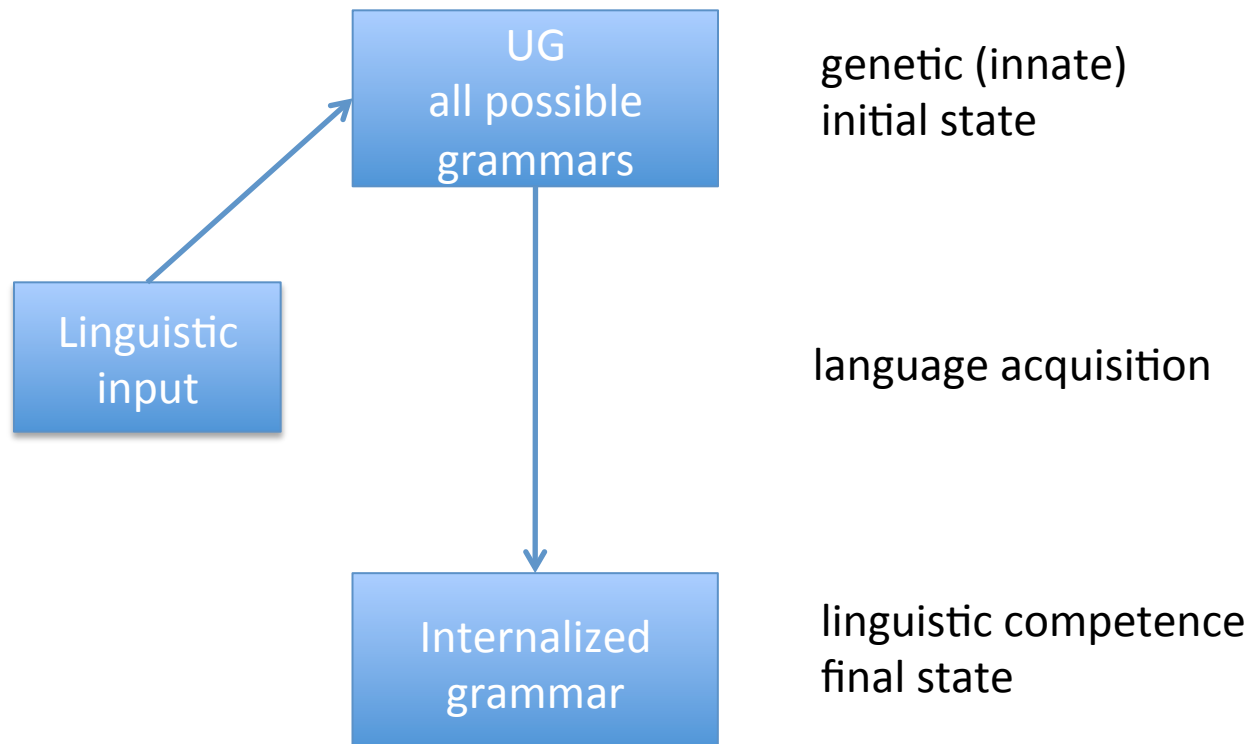
- 1) **Linguistic competence:** People are able to understand and produce an infinite number of grammatical sentences.
  - The inductive method used by the Structuralists was unable to account for this knowledge and for key linguistic phenomena like structural ambiguity.
  
- 2) **Language acquisition:** Children are able to learn a language perfectly, even though they are exposed to defective inputs.
  - The idea that language is learnt through *imitation*, *reinforcement*, *analogy* (i.e., *stimulus-response* processes) as argued by the Behaviorists (e.g., Skinner 1957) was no longer tenable.

# The innateness hypothesis (UG)

- Children can acquire a complex grammar quickly and easily because they are born with a genetic disposition to language, called Universal Grammar (UG)
- UG provides is a set of principles and parameters that constrain all human languages, e.g.,
  - Principle: A sentence must always have a subject
  - Parameter: Whether or not the subject must be overtly realized
- Children only need to learn the values of the parameters specific to their native language



# Universal Grammar (UG)



(Chomsky, 1964)

# The new methodology

- The task of a linguist is comparable to a child's acquisition of linguistic competence
- Children do not acquire language inductively, but deductively → the new methodology must be deductive.
- The task of a linguist is to make hypotheses about possible grammars for a given language and select the one that ***generates*** all and only the grammatical sentences in that language.

# Generative grammar

- A generative grammar is a formal system (a set of rules) that predicts which combinations of words is a grammatical sentence
- If a grammar predicts a sentence which is never encountered or consistently rejected by native speakers => the grammar is not adequate
- An adequate grammar for a language must account for the knowledge system underlying the intuitions of the native speaker (***descriptive grammar***) and explain how such knowledge is acquired (***explanatory grammar***)

# Summary

- The object of investigation in generative linguistics is the linguistic competence of native speakers and, ultimately, the universal grammar.
- Linguistic competence is an internalized (mentally-represented) grammar capable of generating an infinite number of grammatical sentences and none of the ungrammatical ones.
- The goal of linguistics is to provide an adequate generative grammar that models linguistic competence and explains language acquisition.

# Some of Chomsky's controversial claims

- The superficial diversity of human languages masks their underlying similarity.
  - Several approaches emphasize on language diversity.
- All languages are fundamentally alike because linguistic knowledge is largely innate.
  - Cognitive linguistics explains competence in terms of general cognitive abilities.
- The central problem of linguistics is explaining language acquisition.
  - For many formal theories, linguistics does not and cannot aim at explanation.

# Generative theories of grammar

## ➤ Transformational grammars (TG)

- Standard theory (ST)
- Extended Standard theory (EST)
- Revised Extended Standard Theory (REST)
- Principle and Parameter Theory (P&P)
  - Government and Binding Theory (GB)
  - X bar theory
  - Minimalist Program (MP)

# Alternative theories of grammar

## Non-transformational grammars

- Arose from the need to provide computationally implementable and/or psychologically plausible grammars.
- **Some non-transformational approaches**
- Lexical-functional grammars (LFG)
  - Generalized Phrase-Structure Grammar (GPSG)
  - Head-driven Phrase-Structure Grammar (HPSG)

# Summary

- Linguistics as the scientific study of human language
- Productivity/creativity as a distinctive feature of human language
- Levels of linguistic analysis and knowledge: Phonetics, Phonology, Morphology, Syntax, Semantics
- Recursion and how it accounts for productivity
- Competence vs. performance
- Grammaticality vs. acceptability
- Descriptive vs. prescriptive grammars
- The modern era of Linguistics
- The generative program