Syntactic Theory

Introduction

Yi Zhang & Antske Fokkens

Department of Computational Linguistics
Saarland University

October 15, 2009

Syntax: What does it mean?

We can view syntax/syntactic theory in a number of ways, two of which are the following:

- Psychological way/model: syntactic structures correspond to what is in the heads of speakers and hearers
- Computational way/model: syntactic structures are formal objects which can be mathematically treated/manipulated

Syntactic Analysis

- Focus on collection of words and rules with which we generate
  - strings of those words (weak generative power)
  - structures which license strings of those words (strong generative power)
- Syntax attempts to capture the nature of those rules:
  1. Colorless green ideas sleep furiously
  2. *Furiously sleep ideas green colorless.
  3. *Sally talk to man.
  4. Sally talks to a man.
- What generalizations are needed to capture the difference between grammatical and ungrammatical sentences?

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Why is the Study of Syntax Relevant?

[Sag, Watson and Bender, 2003]

- A window on the structure of the mind
  - Innateness of the language faculty (Chomsky)
    - Universal Grammar
- A window on the mind's activity
  - Cognitive process
    - Ambiguity management
- Natural language technologies
  - Parsing
  - Generation
  - Grammar checking

Two Conceptions of Grammar

Prescriptive grammar

Consists of admonitions not to use certain forms or constructions that are common in everyday speech:

- *Never split an infinitive.*
- *A preposition is a bad word to end a sentence with.*

Human language is a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities

Natural phenomena can NOT be legislated!

Two Conceptions of Grammar

Prescriptive grammar

Consists of admonitions not to use certain forms or constructions that are common in everyday speech:

- *Never split an infinitive.*
- *A preposition is a bad word to end a sentence with.*

Human language is a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities

Natural phenomena can NOT be legislated!

Two Conceptions of Grammar

Prescriptive grammar

Consists of admonitions not to use certain forms or constructions that are common in everyday speech:

- *Never split an infinitive.*
- *A preposition is a bad word to end a sentence with.*

Human language is a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities

Natural phenomena can NOT be legislated!

Two Conceptions of Grammar

Prescriptive grammar

Consists of admonitions not to use certain forms or constructions that are common in everyday speech:

- *Never split an infinitive.*
- *A preposition is a bad word to end a sentence with.*

Human language is a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities

Natural phenomena can NOT be legislated!

Two Conceptions of Grammar

Prescriptive grammar

Consists of admonitions not to use certain forms or constructions that are common in everyday speech:

- *Never split an infinitive.*
- *A preposition is a bad word to end a sentence with.*

Human language is a phenomenon amenable to scientific investigation, rather than something to be regulated by the decrees of authorities

Natural phenomena can NOT be legislated!

Two Conceptions of Grammar

Descriptive grammar

- Observes language and creates conceptual categories for it without establishing rules of language
- Consults intuitions of native speakers on *what sounds good*
  1. They saw Pat with Chris.
  2. They saw Pat and Chris.
  3. Who did they see Pat with?
  4. * Who did they see Pat and?

Syntax vs. Grammar

Two terms are in many cases used interchangeably, but

Syntax contrasts with semantics, morphology and phonology, and focus on the way words are put together into phrases, and phrases into sentences, although the boundaries are not always sharp.

Grammar Formalisms

Computational grammar formalisms share several properties

- Descriptive adequacy
- Precise encoding
- Constrained formalism

Descriptive Adequacy

Some researchers try to explain the underlying mechanisms, but we are most concerned with being able to describe linguistic phenomena

- Provide a structural description for every well-formed sentence
- Give us an accurate encoding of a language
- Give us broad-coverage, i.e., can (try to) describe all of a language

Precise Encoding

Mathematical Formalism: formal way to generate sets of strings or structures

Precisely define:

- elementary structures
- ways of combining those structures

Such an emphasis on mathematical precision makes these grammar formalism more easily implementable

Constrained Formalism

A formalism must be **constrained**:

- Linguistic motivation: limits the scope of the theory of grammar
- Computational motivation: allows us to define efficient processing models

Simplistic Syntactic Theory Example I

List as grammars

A grammar consists of a list of all the well-formed sentences in the language

- *Some sentences go on and on.*
- *Some sentences go on and on and on.*
- *Some sentences go on and on and on and on.*
- ...

- Grammar $\mathscr{G}_1$ is defined by the language $\mathscr{L}$ it self, as a set of strings $\mathscr{G}_1 = \{s_i | s_i \in \mathscr{L}\}$
- Weak expressive power: cannot enumerate all possible sentences in a language
- No (useful) structure
- No generalization over linguistic phenomena

Simplistic Syntactic Theory Example I

List as grammars

A grammar consists of a list of all the well-formed sentences in the language

- *Some sentences go on and on.*
- *Some sentences go on and on and on.*
- *Some sentences go on and on and on and on.*
- ...

- Grammar $\mathscr{G}_1$ is defined by the language $\mathscr{L}$ it self, as a set of strings $\mathscr{G}_1 = \{s_i | s_i \in \mathscr{L}\}$
- Weak expressive power: cannot enumerate all possible sentences in a language
- No (useful) structure
- No generalization over linguistic phenomena

Simplistic Syntactic Theory Example I

List as grammars

A grammar consists of a list of all the well-formed sentences in the language

- *Some sentences go on and on.*
- *Some sentences go on and on and on.*
- *Some sentences go on and on and on and on.*
- ...

- Grammar $\mathscr{G}_1$ is defined by the language $\mathscr{L}$ it self, as a set of strings $\mathscr{G}_1 = \{s_i | s_i \in \mathscr{L}\}$
- Weak expressive power: cannot enumerate all possible sentences in a language
- No (useful) structure
- No generalization over linguistic phenomena

Simplistic Syntactic Theory Example II

Regular Expressions

Regular Expressions, i.e. patterns making use of Kleene star (and Kleene plus), parentheses for optionality, and the vertical bar for alternatives, can be used to describe grammars

- $\mathscr{G}_2$: *Some sentences go on [and on]*$^+$ *.*
- Insufficient description power to capture generalizations

Syntactic Theories to be Reviewed

In this course, we will introduce the following linguistic frameworks

- Chomskyan Transformational Tradition & Minimalism
- Dependency Grammar
- Tree Adjoining Grammar
- Lexical Functional Grammar
- Head-Driven Phrase Structure Grammar

With particular focus on the `LFG` and `HPSG`

Organizational Matters

- Time: Tuesday 16:15 - 17:45, Thursday 14:15 - 15:45
- Location: Seminar Room, C72
- Office hours: Thursday 13:00 - 14:00 (after email contact)
- Credit Points: 6 CP
- Course Homepage:

  `http://www.coli.uni-saarland.de/courses/syntactic-theory-09/`

Lectures, Exercises, and Exam

- Regular attendance of the lectures are required
- Exercises needs to be submitted in, and will be corrected
- One must “pass” at least half of the exercises to be qualified for the final exam