Overview

- Introduction
- Constituency
- Context-Free Rules and Trees
- Sentence-Level Constructions

Introduction

- Syntax – setting out together or arrangement (Greek)

- **Syntax:** the way words are arranged together

- Main ideas of syntax:

  - *Constituency*

  - *Grammatical relations*

  - *Subcategorization and dependencies*

Introduction

- *Constituency*

  – Groups of words may behave as a single unit or phrase, called **constituent**, e.g., NP

  she, Rama Krishnan, the house, Raisina Hill, a well-weathered three-storey structure

  – CFG – a formalism allowing us to model the constituency facts

- *Grammatical relations*

  – A formalization of ideas from traditional grammar: SUBJECT and OBJECT

  She ate a mammoth breakfast

Introduction

- *Subcategorization and dependencies*

  – Referring to certain kinds of relations between words and phrases

  – e.g., the verb *want* can be followed by an infinitive or a noun phrase

  I want to fly to Delhi   *I want a flight to Delhi*

  – But the verb *find* can not be followed by an infinitive

  I found to fly to Mumbai

  – Called facts about the *subcategorization* of the verb

Introduction

- All of the kinds of syntactic knowledge can be modelled by various kinds of CFG-based grammars.

  – CFGs are thus backbone of many models of the syntax of NL.

  – Integral to computer applications: grammar checking, semantic interpretation, dialogue understanding, machine translation

Constituency

- **Noun Phrase:**

  - A sequence of words surrounding at least one noun, e.g.,

    – three parties from Brooklyn *arrive*…

    – a high-class spot such as Mindy's *attracts*…

    – the Broadway coppers *love*…

    – They *sit*

  – Evidences of constituency

  - The NPs can all appear in similar syntactic environment, e.g., before, a verb – *Noun Phrases can occur before verbs*

  - **Preposed** or **postposed** constructions, e.g., the PP, *on September seventeenth*, can be placed in a number of different locations

    – On September seventeenth, I'd like to fly from Atlanta to Denver.

    – I'd like to fly on September seventeenth from Atlanta to Denver.

    – I'd like to fly from Atlanta to Denver on September seventeenth.

Context-Free Rules and Trees

- **CFG (or Phrase-Structure Grammar):**

  - The most commonly used mathematical system for modeling constituent structure in English and other NLs

  - A context-free grammar consists of:

    - a set of *rules* or productions

      » Rules expresses the ways that symbols of the language

      can be grouped and ordered together

    - and a *lexicon* of words and symbols

Context-Free Rules and Trees

- **CFG (or Phrase-Structure Grammar):**

  - The following <u>productions</u> express that:

    – a NP can be composed of either a *ProperNoun* or a

    determiner (*Det*) followed by a *Nominal*; a *Nominal* can be

    one or more *Nouns*

      $NP \rightarrow Det\ Nominal$

      $NP \rightarrow ProperNoun$

      $Nominal \rightarrow Noun \mid Noun\ Nominal$

    – <u>Lexicon:</u> $Det \rightarrow a \mid the$
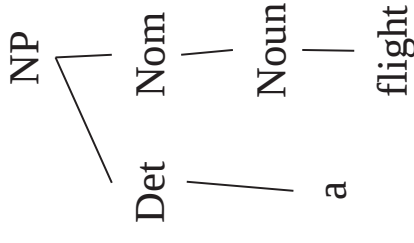
      $Noun \rightarrow flight$

Context-Free Rules and Trees

- **CFG (or Phrase-Structure Grammar):**

  - The symbols used in a CFG are divided into two classes:

    – *Terminal symbols*: the symbols that correspond to words in the language

    – *Non-terminals*: the symbols that express clusters or generalizations of these

  - In each context-free rule:

    – The item to the right of the arrow ($\rightarrow$) is ordered list of one or more terminals and non-terminals

    – Left of the arrow is a single non-terminal symbol expressing some cluster or generalization

Context-Free Rules and Trees

- **CFG (or Phrase-Structure Grammar):**

  - *NP → Det Nominal → Det Noun → a flight*

  - The string *a flight* can be derived from the non-terminal *NP*

  - This sequence of rule expansions is called a **derivation** of the string of words

  - Derivation is represented by a **parse tree**

  - The set of strings that are derivable from the **start symbol** is often called *S*

```
              NP
           /      \
         Det      Nom
          |        |
          a       Noun
                   |
                 flight
```

Context-Free Rules and Trees

Noun → flight | breeze | trip | morning | ...
Verb → is | prefer | like | need | want | fly ...
Adjective → cheapest | non-stop | first | latest | other | direct | ...
Pronoun → me | I | you | it | ...
Proper-Noun → Alaska | Baltimore | Los Angeles | Chicago | United | American | ...
Determiner → the | a | an | this | these | that | ...
Preposition → from | to | on | near | ...
Conjunction → and | or | but | ...

The lexicon for L_0

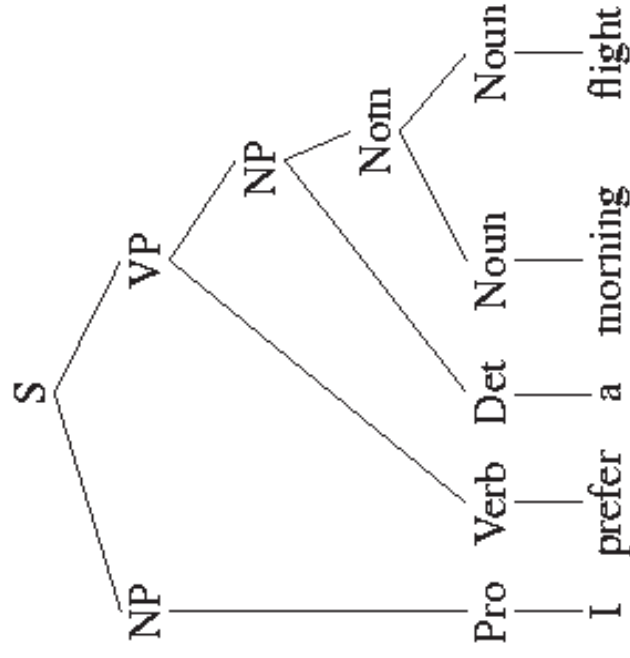|	
S → NP VP	I + want a morning flight
NP → Pronoun	I
| Proper-Noun	Los Angeles
| Det Nominal	a + flight
Nominal → Noun Nominal	morning + flight
| Noun	flights
VP → Verb	do
| Verb NP	want + a flight
| Verb NP PP	leave + Boston + in the morning
| Verb PP	leaving + on Thursday
PP → Preposition NP	from + Los Angeles

The grammar for L_0

Context-Free Rules and Trees

The parse tree for "I prefer a morning flight" according to grammar L_0

Context-Free Rules and Trees

- Bracket notation of parse tree

[$_S$ [$_{NP}$ [$_{Pro}$ I]] [$_{VP}$[$_V$ prefer] [$_{NP}$[$_{Det}$ a] [$_{Nom}$ [$_N$ morning][$_{Nom}$ [$_N$ flight]]]]]]

- Grammatical vs. ungrammatical sentences

  – Sentences that can be derived by a grammar are in the formal language defined by that grammar are called grammatical sentences

  – Sentences that cannot be derived by a given formal grammar are not in the language defined by that grammar are referred to as ungrammatical

- The use of formal languages to model NLs is called **generative grammar**, since the language is defined by the set of possible sentences "generated" by the grammar.

Context-Free Rules

- The formal definition of a CFG is a 4-tuple.

  - a set of non-terminal symbols N

  - a set of terminal symbols Σ

  - a set of productions P, each of form $A \rightarrow \alpha$, where A is a non-terminal
    and α is a string of symbols from the infinite set of strings

  - a designated start symbol S

Context-Free Rules

$S \rightarrow NP\ VP$	$Det \rightarrow that \mid this \mid a$
$S \rightarrow Aux\ NP\ VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper\text{-}Noun$	$Proper\text{-}Noun \rightarrow Houston \mid TWA$
$NP \rightarrow Det\ Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal\ Noun$	
$Nominal \rightarrow Nominal\ PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb\ NP$	
$VP \rightarrow Verb\ NP\ PP$	
$VP \rightarrow Verb\ PP$	
$VP \rightarrow VP\ PP$	
$PP \rightarrow Preposition\ NP$	

Figure 13.1 The $\mathcal{L}_1$ miniature English grammar and lexicon.

Sentence-Level Constructions

- There are a great number of possible overall sentences structures

- Four are particularly common and important:

    - *Declarative structure, imperative structure, yes-no-question structure, and wh-question structure.*

    - Sentences with **declarative** structure:

        - A subject NP followed by a VP

            - *The flight should be eleven a.m. tomorrow.*

            - *The return flight should leave at around seven p.m.*

            - *I want a flight from Ontario to Chicago.*

            - *I plan to leave on July first around six thirty in the evening.*

Sentence-Level Constructions

- Sentence with **imperative** structure

  - Begin with a VP and have no subject

  - Always used for commands and suggestions

    – *Show the lowest fare.*

    – *Show me the cheapest fare that has lunch.*

    – *Give me Sunday's flights arriving in Las Vegas from Memphis and New York City.*

    – *List all flights between five and seven p.m.*

  - *S → VP*

Sentence-Level Constructions

- Sentences with **yes-no-question** structure

  - Begin with auxiliary, followed by a subject *NP*, followed by a *VP*.

    - *Do any of these flights have stops?*

    - *Does American's flight eighteen twenty five serve dinner?*

    - *Can you give me the same information for United?*

  - *S → Aux NP VP*

Sentence-Level Constructions

- The **wh-subject-question** structure

  - Wh-word – *who, where, what, which, how, why*

  - Identical to the declarative structure, except that the first NP

    contains some wh-word.

    – *What airlines fly from Burbank to Denver?*

    – *Which flights depart Burbank after noon and arrive in Denver*

      by six p.m.?

    – *Whose flights serve breakfast?*

  - *S → Wh-NP VP*

Sentence-Level Constructions

- The **wh-non-subject-question** structure

  - Auxiliary appears before the subject *NP*, just as in the yes-no-question

    – *What flights do you have from Burbank to Tacoma Washington?*

  - *S → Wh-NP Aux NP VP*

References

- Speech and Language Processing, Daniel Jurafsky and James H. Martin

<Thank You />