

Chapter NLP:V

V. Syntax

- ❑ Introduction
- ❑ Phrase Structure Grammars
- ❑ Dependency Grammars
- ❑ Features and Unification

Introduction

Problem: Given a set of symbols, how do they incur meaning?

Sun, Leipzig, the, shine, warm, in

- ❑ Leipzig shone warm in the sun.
- ❑ In Leipzig warm the sun is shining.
- ❑ Warm is the shining sun.
- ❑ The sun shines in Leipzig.

Introduction

Problem: Given a set of symbols, how do they incur meaning?

Sun, Leipzig, the, shine, warm, in

- ❑ Leipzig shone warm in the sun.
- ❑ In Leipzig warm the sun is shining.
- ❑ Warm is the shining sun.
- ❑ The sun shines in Leipzig.

Definition 1 (Grammar [Saussure])

A grammar is a system that describes the relationship between concepts (signified, *langue*) and expressions (signifier, *parole*).

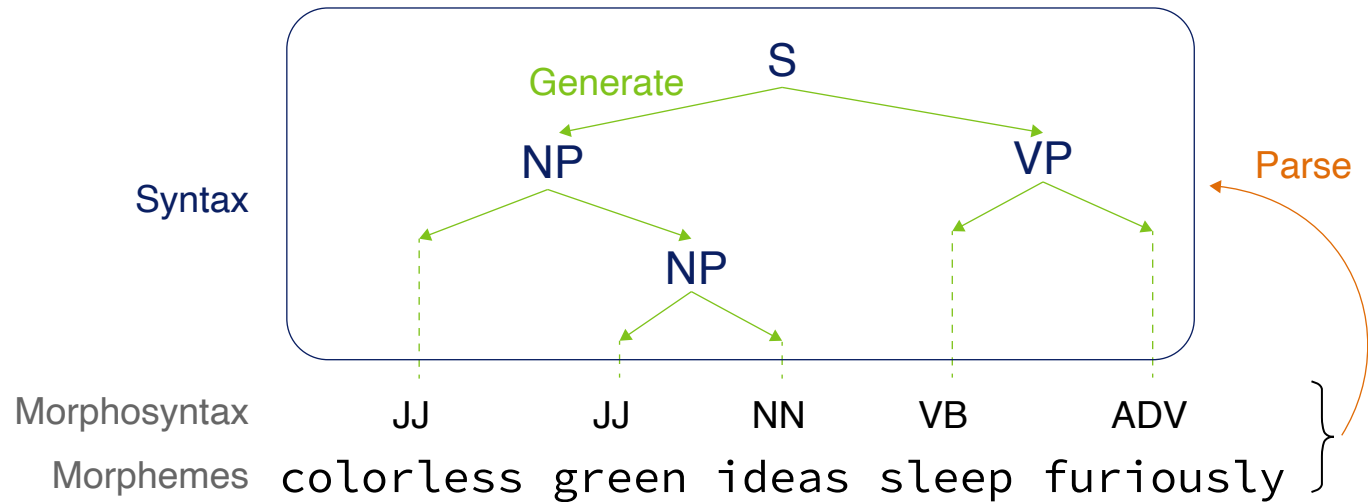
Introduction

Grammars

Definition 2 (Grammar [Chomsky])

A (generative) grammar of a language is a device that **parses** and **generates** all grammatical sentences of a language and rejects ungrammatical ones.

1. Parse: Determine the syntactic structure of a given sequence of symbols.
2. Generate: Produce valid sequence of symbols given a set of structural rules.



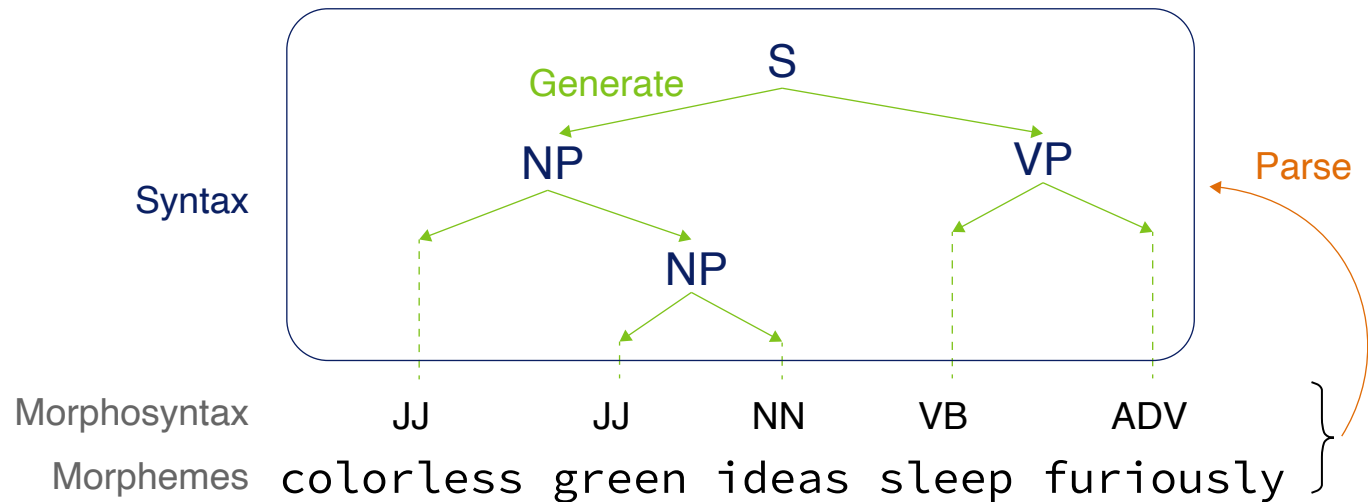
Introduction

Grammars

Grammaticality is usually understood as **intuitively acceptable to a native speaker**.

A grammatical sentence does **not** need to:

- ❑ have ever been observed in a corpus,
- ❑ be statistically probable, or
- ❑ be meaningful.



Remarks:

- ❑ There are different ways of thinking about grammar, based on the underlying theory of language. Computer scientists prefer the formalist perspective.
 - *Formalist*: Grammar is a system of rules (a formal, generative model) that describe if a sentence is grammatical.
 - *Functional*: Grammar describes how utterances are constructed to serve a function in discourse.

Discourse Act (Intent, Speaker, Addressee) → Semantics → Morphosyntax → Phonology

- ❑ Chomsky's proposed model of grammar is called 'generative transformational grammar'. It subsumes phrase structure, transformational, and morphophonemic rules. The first is determines the syntactic structure of a string, the latter two transform a string, given his phrase structure, into phonetic morphemes, a representation of the spoken sentence.
- ❑ The generative aspect of Chomsky's theories is widely adapted and different models develop the idea further. The transformational aspect is controversial.

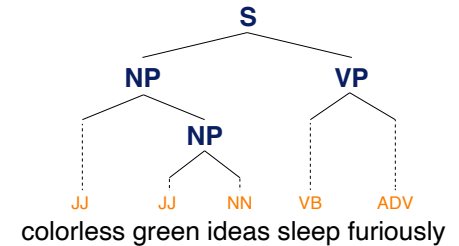
Introduction

Syntax Structures

The syntax structure of a clause is hierarchial and modeled as either:

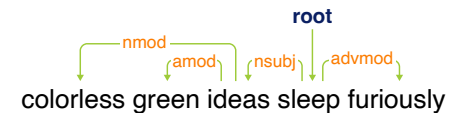
Phrase Structure Grammar (top-down):

- ❑ Clauses and phrases are divided into **one or many constituents**.
- ❑ Top-level constituents are often *Subject/Noun Phrase* (NP) and *Predicate/Verb Phrase* (VP).
- ❑ The leaves are always words.



Dependency Grammar (bottom-up):

- ❑ Each morpheme has **one head** node in the syntax structure and **zero or many dependents**.
- ❑ The root node is usually the main verb.



Introduction

Syntax Parsing

Determining the syntax structure automatically is used for:

- ❑ Grammar checkers
- ❑ Complex named entity recognition (e.g., in biological or legal domains).
- ❑ Entity relation extraction
- ❑ Syntax-based sentence compression
- ❑ Mining of opinions on aspects of products
- ❑ Source-sentence analysis for machine translation
- ❑ High precision question answering

Introduction

Ambiguity

Every sentence can have several grammatical syntax structures due to structural ambiguity. Common forms are **attachment ambiguity** and coordination ambiguity.

Attachment ambiguity:

- The attachment of many constituents (prepositional phrases, adverbial phrases, infinitives, ...) is ambiguous:

The board approved	
its acquisition	→ attaches to “approved”
by Royal Trustco Ltd.	→ attaches to “its acquisition”
of Toronto	→ attaches to “by Royal Trustco Ltd.”
for \$27 a share	→ attaches to “its acquisition”
at its monthly meeting.	→ attaches to “approved ... for \$27 a share”

- Number of potential attachments grows exponentially with the number n of constituents according to the Catalan numbers:

$$C_n = \frac{(2n)!}{(n+1)! \cdot n!}$$

Introduction

Ambiguity

Every sentence can have several grammatical syntax structures due to structural ambiguity. Common forms are attachment ambiguity and **coordination ambiguity**.

Coordination ambiguity:

- It is often unclear which phrase is coordinated by a conjunction:

[[old] [man and woman]] vs. [old man] and [woman]

Introduction

Ambiguity

Every sentence can have several grammatical syntax structures due to structural ambiguity. Common forms are attachment ambiguity and coordination ambiguity.

Syntactic disambiguation:

- ❑ Many parses are grammatical.
- ❑ Few parses are semantically plausible:

Scientists observe [whales] [from space].

vs.

Scientists observe [whales from space].

- ❑ Parsers should choose one, the most probable parse. This is called syntactic disambiguation.