

Context-Free Grammars for English

From: Chapter 12 of *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, by Daniel Jurafsky and James H. Martin

Overview

- **Syntax:** the way words are arranged together
- Main ideas of syntax:
 - **Constituency**
 - Groups of words may behave as a single unit or phrase, called **constituent**, e.g., NP
 - CFG, a formalism allowing us to model the constituency facts
 - **Grammatical relations**
 - A formalization of ideas from traditional grammar about SUBJECT, OBJECT and other such relations
 - **Subcategorization and dependencies**
 - Referring to certain kind of relations between words and phrases, e.g., the verb *want* can be followed by an infinitival phrase, as in *I want to fly to Detroit*.

Background

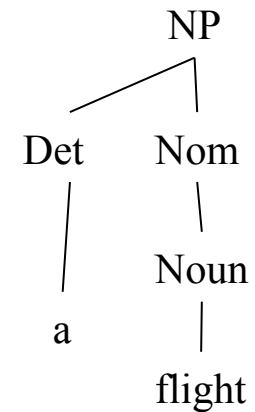
- All of the kinds of syntactic knowledge can be modeled by various kinds of CFG-based grammars.
- CFGs are thus backbone of many models of the syntax of NL.
- They are powerful enough to express sophisticated relations among the words in a sentence, yet computationally tractable enough that efficient algorithms exists for parsing sentences with them.
- Also probability version of CFG are available
- Example sentences from the Air Traffic Information System (ATIS) domain

Constituency

- **NP:**
 - 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* ...
 - They *sit*
 - Harry the Horse
 - the reason he comes into the Hot Box
- Evidences of constituency
 - The above NPs can all appear in similar syntactic environment, e.g., before, a verb.
 - **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
 - Terminals and non-terminals
 - Derivation
 - Parse tree
 - Start symbol



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

$S \rightarrow NP VP$

$NP \rightarrow Pronoun$

| *Proper-Noun*

| *Det Nominal*

$Nominal \rightarrow Noun Nominal$

| *Noun*

$VP \rightarrow Verb$

| *Verb NP*

| *Verb NP PP*

| *Verb PP*

$PP \rightarrow Preposition NP$

I + want a morning flight

I

Los Angeles

a + flight

morning + flight

flights

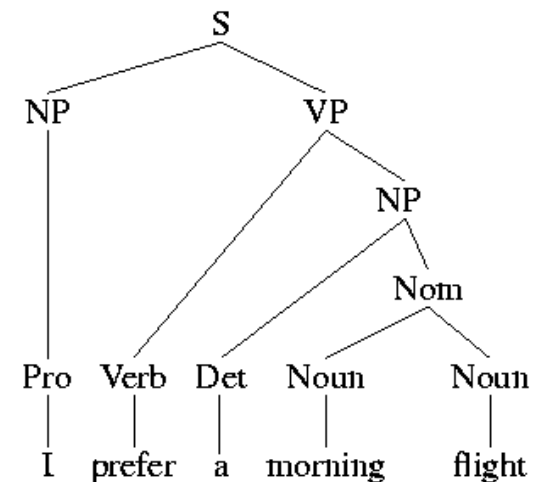
do

want + a flight

leave + Boston + in the morning

leaving + on Thursday

from + Los Angeles



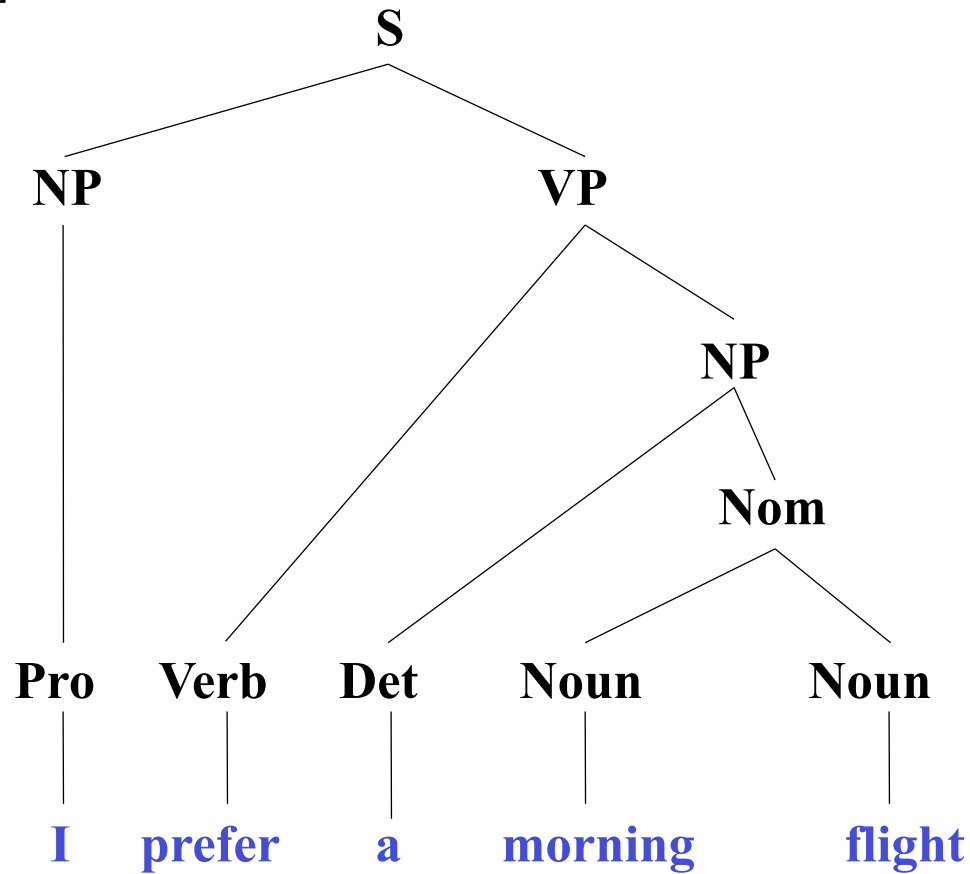
The grammar for L_0

Context-Free Rules and Trees

- **Bracket notation** of parse tree (see next page)
- Grammatical vs. ungrammatical sentences
- 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.
- The formal definition of a CFG is a 4-tuple: (A set of non-terminal symbols, a set of terminal symbols, a set of rules, a start symbol).

Bracketed Notation

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



Sentence-Level Constructions

- There are a great number of possible overall sentence structures, but four are particularly common and important:
 - *Declarative structure, imperative structure, interrogative 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.*
 - *I need a flight to Seattle leaving from Baltimore making a stop in Minneapolis.*
 - *The return flight should leave at around seven p.m.*
 - *I would like to find out the flight number for the United flight that arrives in San Jose around ten p.m.*
 - *I'd like to fly the coach discount class.*
 - *I want a flight from Ontario to Chicago.*
 - *I plan to leave on July first around six thirty in the evening.*
 - $S \rightarrow NP VP$

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 flight arriving in Las Vegas from Memphis and New York City.*
 - *List all flights between five and seven p.m.*
 - *List all flights from Burbank to Denver.*
 - *Show me all flights that depart before ten a.m. and have first class fares.*
 - *Show me all the flights leaving Baltimore.*
 - *Show me flights arriving within thirty minutes of each other.*
 - *Please list the flights from Charlotte to Long Beach arriving after lunch time.*
 - *Show me the last flight to leave.*
 - $S \rightarrow VP$

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?*
 - $S \rightarrow Aux NP VP$

Sentence-Level Constructions

- The **wh-subject-question** structure
 - 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.?*
 - *Which flights serve breakfast?*
 - *Which of these flights have the longest layover Nashville?*
 - $S \rightarrow Wh-NP VP$
- The **wh-non-subject-question** structure
 - *What flights do you have from Burbank to Tacoma Washington?*
 - $S \rightarrow Wh-NP Aux NP VP$

The Noun Phrase

- View the NP as revolving around a **head**, the central noun in the NP.
 - The syntax of English allows for both pre-nominal (pre-head) modifiers and post-nominal (post-head) modifiers.

The Noun Phrase

Before the Head Noun

- NPs can begin with a determiner,
 - *a stop, the flights, that fare, this flight, those flights, any flights, some flights* (can be complex: *some but not all flights*)
- Determiners can be optional,
 - *Show me **flights** from San Francisco to Denver on weekdays.*
- **Mass nouns** don't require determiners.
 - Substances, like *water* and *snow*
 - Abstract nouns, *music, homework,*
 - In the ATIS domain, *breakfast, lunch, dinner*
 - *Does this flight server **dinner**?*

The Noun Phrase

Before the Head Noun

- **Predeterminers:**
 - Word classes appearing in the NP before the determiner
 - *all the flights, all flights*
- **Postdeterminers:**
 - Word classes appearing in the NP between the determiner and the head noun
 - **Cardinal numbers:** *two friends, one stop*
 - **Ordinal numbers:** *the first one, the next day, the second leg, the last flight, the other American flight, and other fares*
 - **Quantifiers:** *many fares*
 - The quantifiers, *much* and *a little* occur only with noncount nouns.

The Noun Phrase

Before the Head Noun

- Adjectives occur after quantifiers but before nouns.
 - *a **first-class** fare, a **nonstop** flight, the **longest** layover, the **earliest** lunch flight*
- Adjectives can be grouped into a phrase called an **adjective phrase** or **AP**.
 - AP can have an adverb before the adjective
 - *the **least** expensive fare*
- *NP* → (*Det*) (*Card*) (*Ord*) (*Quant*) (*AP*) *Nominal*
 - *The first few non-stop flights*
 - *The two first non-stop flights*

The Noun Phrase

After the Head Noun

- A head noun can be followed by **postmodifiers**.
 - Prepositional phrases
 - *All flights from Cleveland*
 - Non-finite clauses
 - *Any flights arriving after eleven a.m.*
 - Relative clauses
 - *A flight that serves breakfast*

The Noun Phrase

After the Head Noun

- PP postmodifiers
 - *any stopovers [for Delta seven fifty one]*
 - *all flight [from Cleveland] [to Newark]*
 - *arrival [in San Jose] [before seven a.m.]*
 - *a reservation [on flight six oh six] [from Tampa] [to Montreal]*
 - *Nominal → Nominal PP (PP) (PP)*

The Noun Phrase

After the Head Noun

- The three most common kinds of **non-finite** postmodifiers are the gerundive (*-ing*), *-ed*, and infinitive form.
 - A gerundive consists of a VP beginning with the gerundive (*-ing*)
 - *any of those [leaving on Thursday]*
 - *any flights [arriving after eleven a.m.]*
 - *flights [arriving within thirty minutes of each other]*

Nominal → Nominal GerundVP

GerundVP → GerundV NP | GerundV PP | GerundV | GerundV NP PP

GerundV → being | preferring | arriving | leaving | ...

- Examples of two other common kinds
 - *the last flight **to arrive** in Boston*
 - *I need to have dinner **served***
 - *Which is the aircraft **used by this flight?***

The Noun Phrase

After the Head Noun

- A postnominal relative clause
 - is a clause that often begins with a **relative pronoun** (*that* and *who* are the most common).
 - The relative pronoun functions as the subject of the embedded verb,
 - *a flight that serves breakfast*
 - *flights that leave in the morning*
 - *the United flight that arrives in San Jose around ten p.m.*
 - *the one that leaves at ten thirty five*

Nominal → *Nominal RelClause*
RelClause → (*who* | *that*) *VP*

The Noun Phrase

After the Head Noun

- Various postnominal modifiers can be combined,
 - *a flight [from Phoenix to Detroit] [leaving Monday evening]*
 - *I need a flight [to Seattle] [leaving from Baltimore] [making a stop in Minneapolis]*
 - *evening flights [from Nashville to Houston] [that serve dinner]*
 - *a friend [living in Denver] [that would like to visit me here in Washington DC]*

Coordination

- NPs and other units can be **conjoined** with **coordinations** like *and*, *or*, and *but*.
 - *Please repeat* [_{NP} [_{NP} *the flight*] *and* [_{NP} *the departure time*]]
 - *I need to know* [_{NP} [_{NP} *the aircraft*] *and* [_{NP} *flight number*]]
 - *I would like to fly from Denver stopping in* [_{NP} [_{NP} *Pittsburgh*] *and* [_{NP} *Atlanta*]]
 - *NP* → *NP and NP*
 - *VP* → *VP and VP*
 - *S* → *S and S*

Agreement

- Most verbs in English can appear in two forms in the present tense:
 - 3sg, or non-3sg

Do [_{NP} any flights] stop in Chicago?

Do [_{NP} all of these flights] offer first class service?

Do [_{NP} I] get dinner on this flight?

Do [_{NP} you] have a flight from Boston to Forth Worth?

Does [_{NP} this flight] stop in Dallas?

Does [_{NP} that flight] serve dinner?

Does [_{NP} Delta] fly from Atlanta to San Francisco?

What flight *leave* in the morning?

What flight *leaves* from Pittsburgh?

*[What flight] *leave* in the morning?

*Does [_{NP} you] have a flight from Boston to Fort Worth?

*Do [_{NP} this flight] stop in Dallas?

$S \rightarrow Aux NP VP$

$S \rightarrow 3sgAux\ 3sgNP\ VP$

$S \rightarrow Non3sgAux\ Non3sgNP\ VP$

$3sgAux \rightarrow does\ |\ has\ |\ can\ |\ \dots$

$Non3sgAux \rightarrow do\ |\ have\ |\ can\ |\ \dots$

$3sgNP \rightarrow (Det)\ (Card)\ (Ord)\ (Quant)\ (AP)\ SgNominal$

$Non3sgNP \rightarrow (Det)\ (Card)\ (Ord)\ (Quant)\ (AP)\ PlNominal$

$SgNominal \rightarrow SgNoun\ |\ SgNoun\ SgNoun$

$PlNominal \rightarrow PlNoun\ |\ SgNoun\ PlNoun$

$SgNoun \rightarrow flight\ |\ fare\ |\ dollar\ |\ reservation\ |\ \dots$

$PlNoun \rightarrow flights\ |\ fares\ |\ dollars\ |\ reservation\ |\ \dots$

Agreement

- Problem for dealing with number agreement:
 - it doubles the size of the grammar.
- The rule proliferation also happens for the noun's **case**:
 - For example, English pronouns have **nominative** (*I, she, he, they*) and **accusative** (*me, her, him, them*) versions.
- A more significant problem occurs in languages like German or French
 - Not only N-V agreement, but also **gender agreement**.
- In Sanskrit, there are three numbers: single, dual and plural, that require agreement.
- A way to deal with these agreement problems without exploding the size of the grammar:
 - By effectively **parameterizing** each non-terminal of the grammar with **feature-structures**.

The Verb Phrase and Subcategorization

- The VP consists of the verb and a number of other constituents.

VP → *Verb* disappear
VP → *Verb NP* prefer a morning flight
VP → *Verb NP PP* leave Boston in the morning
VP → *Verb PP* leaving on Thursday

- An entire embedded sentence, called **sentential complement**, can follow the verb.

You [_{VP} [_V said [_S there were two flights that were the cheapest]]]
You [_{VP} [_V said [_S you had a two hundred sixty six dollar fare]]]
[_{VP} [_V Tell] [_{NP} me] [_S how to get from the airport in Philadelphia to downtown]]
I [_{VP} [_V think [_S I would like to take the nine thirty flight]]

VP → *Verb S*

The Verb Phrase and Subcategorization

- Another potential constituent of the VP is another VP
 - Often the case for verbs like *want*, *would like*, *try*, *intent*, *need*

I want [_{VP} to fly from Milwaukee to Orlando]

Hi, I want [_{VP} to arrange three flights]

Hello, I'm trying [_{VP} to find a flight that goes from Pittsburgh to Denver after two p.m.]

- Recall that verbs can also be followed by *particles*, word that resemble a preposition but that combine with the verb to form a *phrasal verb*, like *take off*.
 - These particles are generally considered to be an integral part of the verb in a way that other post-verbal elements are not;
 - Phrasal verbs are treated as individual verbs composed of two words.

The Verb Phrase and Subcategorization

- A VP can have many possible kinds of constituents, not every verb is compatible with every VP.
 - *I want a flight ...*
 - *I want to fly to ...*
 - **I found to fly to Dallas.*
- The idea that verbs are compatible with different kinds of complements
 - Traditional grammar **subcategorize** verbs into two categories (transitive and intransitive).
 - Modern grammars distinguish as many as 100 subcategories

Frame	Verb	Example
ϕ	eat, sleep	I want to eat
<i>NP</i>	prefer, find leave	Find [<i>NP</i> the flight from Pittsburgh to Boston]
<i>NP NP</i>	show, give, find	Show [<i>NP</i> me] [<i>NP</i> airlines with flights from Pittsburgh]
<i>PP_{from} PP_{to}</i>	fly, travel	I would like to fly [<i>PP</i> from Boston] [<i>PP</i> to Philadelphia]
<i>NP PP_{with}</i>	help, load	Can you help [<i>NP</i> me] [<i>PP</i> with a flight]
<i>V_{pto}</i>	prefer, want, need	I would prefer [<i>V_{pto}</i> to go by United airlines]
<i>S</i>	mean	Does this mean [<i>S</i> AA has a hub in Boston?]

The Verb Phrase and Subcategorization

Verb-with-NP-complement → *find* | *leave* | *repeat* | ...

Verb-with-S-complement → *think* | *believe* | *say* | ...

Verb-with-Inf-VP-complement → *want* | *try* | *need* | ...

VP → *Verb-with-no-complement* *disappear*

VP → *Verb-with-NP-complement NP* *prefer a morning flight*

VP → *Verb-with-S-complement S* *said there were two flights*

Auxiliaries

- **Auxiliaries or helping verbs**
 - A subclass of verbs
 - Having particular syntactic constraints which can be viewed as a kind of subcategorization
 - Including the **modal** verb, *can, could many, might, must, will, would, shall, and should*
 - The **perfect** auxiliary *have*,
 - The **progressive** auxiliary *be*, and
 - The **passive** auxiliary *be*.

Auxiliaries

- Modal verbs subcategorize for a *VP* whose head verb is a bare stem.
 - *can go in the morning, will try to find a flight*
- The perfect verb *have* subcategorizes for a *VP* whose head verb is the past participle form:
 - *have booked 3 flights*
- The progressive verb *be* subcategorizes for a *VP* whose head verb is the gerundive participle:
 - *am going from Atlanta*
- The passive verb *be* subcategorizes for a *VP* whose head verb is the past participle:
 - *was delayed by inclement weather*

Auxiliaries

- A sentence may have multiple auxiliary verbs, but they must occur in a particular order.
 - modal < perfect < progressive < passive

<i>modal perfect</i>	could have been a contender
<i>modal passive</i>	will be married
<i>perfect progressive</i>	have been feasting
<i>modal perfect passive</i>	might have been prevented