

# SYNTAX

## CONTEXT-FREE GRAMMARS FOR ENGLISH

### SYNTAX

Descriptive vs. Normative

Applications: Most NLP applications

Machine Translation

Q/A

Information extraction/summarization

Grammar checking

### CONSTITUENCY

constituent (e.g., NP) behaves as a unit  
(discovery methods)

similar syntactic environments (e.g., for NPs)  
before a verb (as subject or agent of verb)  
(but individual words may not)

preposed and post posed constructions

"move" constituents in sentences as a whole

*On 9/17 I'd like to fly from Atlanta to Denver.*

Individual words can't "move"

psychological evidence

## CONTEXT-FREE RULES AND TREES

### Chomsky Hierarchy

3 Regular/Finite Grammars	simple TNs
<u>2. CFG - Phrase Structure - BNF</u>	recursive TNs
1. Context Sensitive Grammars	augmented <sub>tf</sub> TNs
0. Transformational Grammars	augmented <sub>m</sub> TNs

NP → Det Nominal (9.2)

NP → ProperNoun (9.3)

Nominal → Noun | Noun Nominal (9.4)

Det → a (9.5)

Det → the (9.6)

Noun → flight (9.7)

S → NP VP I prefer a morning flight

VP → Verb NP prefer a morning flight

VP → Verb NP PP leave Boston in the morning

VP → Verb PP leaving on Thursday

PP → Preposition NP from Los Angeles

### Terminal vs. Non-terminal symbols

Rules are not exclusive:

in general many ways to rewrite a given symbol

Formal language: all the sentences generated (or accepted) by a formal grammar (which is a set of rewrite rules)

grammatical vs. ungrammatical sentences

formal vs. natural languages

context

variations of usages/dialects/idiolects

\* two meanings of word "context"

context sensitive  $c_1$  NT  $c_2 \rightarrow c_1 \dots c_2$

context free NT  $\rightarrow \dots$

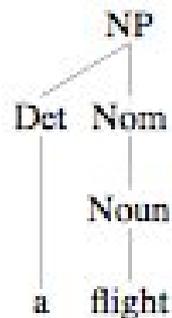
vs. "What is that?" "I don't think so." "the red one."

Synthesis vs. Analysis: (with or without an interpretation)

Generation vs. Parsing or Accepting  
Parsing vs. Accepting

Derivation:

Parse Tree



---

Lexicon

*Noun* → *flights* | *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* | ...

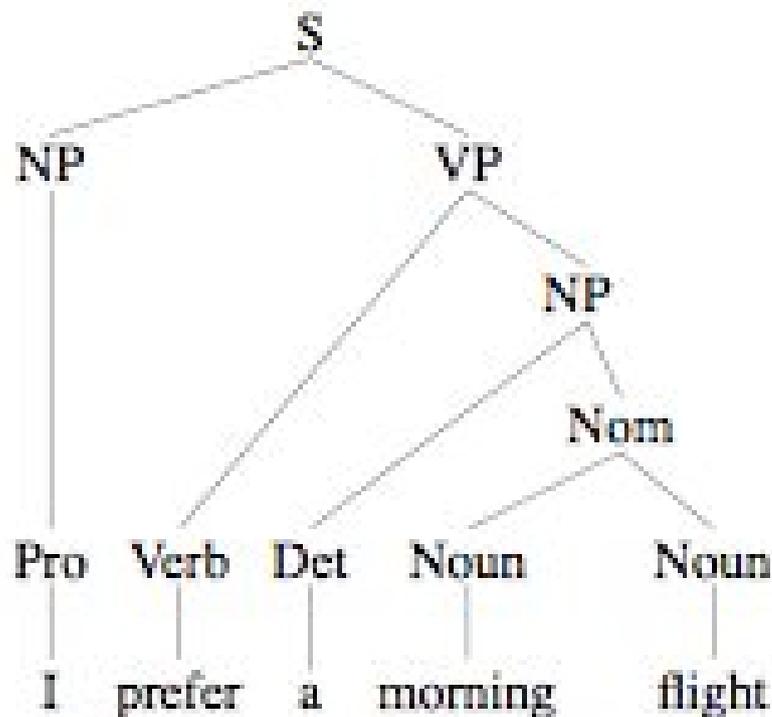
*Preposition* → *from* | *to* | *on* | *near* | ...

*Conjunction* → *and* | *or* | *but* | ...

$L_0$

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

---



[S [NP [Pro I]] [VP [V Prefer] [NP[Det A] [nom [N morning] [N flight]]]]]

[S  
  [NP [Pro I]]  
  [VP [V Prefer] [NP[Det A] [nom [N morning] [N flight]]]]]

[S  
  [NP  
    [Pro I]]  
  [VP  
    [V Prefer]  
    [NP[Det A] [nom [N morning] [N flight]]]]]

[S  
  [NP  
    [Pro I]]  
  [VP  
    [V Prefer]  
    [NP  
      [Det A]  
      [nom  
        [N morning]  
        [N flight]]]]]

## Formal description of CFG and CF Language

A CFG has four parameters (technically “is a 4-tuple”):

1. a set of non-terminal symbols (or “variables”)  $N$
2. a set of terminal symbols  $\Sigma$  (disjoint from  $N$ )
3. a set of productions  $P$ , each of the form  $A \rightarrow \alpha$ , where  $A$  is a non-terminal and  $\alpha$  is a string of symbols from the infinite set of strings  $(\Sigma \cup N)^*$
4. a designated start symbol  $S$

A Language is defined via concept of derivation

direct derivation

$\alpha \beta \gamma \delta \rightarrow \alpha \beta \gamma \delta$

if  $\alpha \beta \gamma \delta$  is a production (rewrite rule) and

$\alpha, \beta$  any strings in  $(\Sigma \cup N)^*$

derivation

$$\alpha_1 \Rightarrow \alpha_2, \alpha_2 \Rightarrow \alpha_3, \dots, \alpha_{m-1} \Rightarrow \alpha_m \quad (9.8)$$

$\alpha_1$  derives  $\alpha_m$ , or  $\alpha_1 \overset{*}{\Rightarrow} \alpha_m$

$$L_G = \{ w \mid w \text{ is in } \Sigma^* \text{ and } S \overset{*}{\Rightarrow} w \} \quad (9.9)$$

Main English Constituents:

Sentences

NPs

VPs

PPs

## SENTENCE-LEVEL CONSTRUCTIONS

main Sentence Types:

**Declaratives:** **A plane left.**

*S -> NP VP*

**Imperatives:** **Leave! Show the lowest fare.**

*S -> VP*

**Yes-No Questions:** **Did the plane leave?**

*S -> Aux NP VP*

**WH Questions:** **When did the plane leave?**

*S -> WH Aux NP VP*

wh-subject-questions:

like declarative structure except first NP contains  
a wh-word (which may be the whole NP)

$S \rightarrow Wh-NP VP$

wh-non-subject-questions:

similar to yes-no regarding aux, but a wh-phrase  
(sometimes just a wh word) appears before the aux  
and replaces a non-subject NP

$S \rightarrow Wh-NP Aux NP VP$

## THE NOUN PHRASE

### Prenominal modifiers

#### determiners

may be omitted if noun is plural

*Show me flights from Pittsburgh to Denver.*

mass nouns don't require determiners & can't take indefinite *a*.

*Water is wet.*

\* *A water is wet.*

mass vs. count ambiguity

*Does this flight serve dinner?*

#### predeterminers

*all the flights*

#### postdeterminers

cardinal numbers

*the two friends*

ordinal numbers

*the first day*

*the first three days*

quantifiers

some appear only with plural count nouns

*many fares, a few things, several ideas*

*much* and *a little* appear only with mass nouns

adjectives

after quantifiers, before nouns

adjective (adjectival) phrases

can take adverbial modification

*the least expensive fare*

*the least expensive reasonably flexible fare*

NP → (Det) (Card) (Ord) (Quant) (AP)\* Nominal (9.10)

(noun-noun modification)

*the garage key, the wing-nut spanner,*

*the head function main loop index*

Postnominal modification

Common nominal postmodifiers:

PP

*all flights from Cleveland*

non-finite clauses

*any flights arriving after eleven a.m.*

relative clauses

*a flight that serves breakfast*

PP

Nominal → Nominal PP (PP) (PP)

(Nominal PP\* is better)

Non-finite clauses (untensed)

gerundive (-ing)

past participle (-ed)

infinitive

Gerundive:

*any of those leaving on Thursday*  
*any flights arriving after eleven a.m.*

Nominal → Nominal GerundVP

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

GerundV → being | preferring | arriving | leaving

Infinitives:

*the last flight to arrive in Boston*

Past participles

*Which is the aircraft used by this flight*

Relative Clauses

*a flight that serves breakfast*

Nominal → Nominal RelClause (9.11)

RelClause → (who|that) VP (9.12)

## COORDINATION

NP → NP and NP (9.13)  
*Please repeat [NP [NP the flights] and [NP the costs]]*

VP → VP and VP (9.14)  
*What flights do you have [VP [vp leaving Denver]  
and  
[arriving in Pittsburgh]]*

S → S and S (9.15)  
*[S  
[S I'm interested in a flight from Dallas to Washington]  
and  
[S I'm also interested in going to Baltimore]]*

## AGREEMENT

subject-verb number agreement?

expand the grammar:

$S \rightarrow \text{Aux NP VP}$

$S \rightarrow \text{3sgAux 3sgNP VP}$

$S \rightarrow \text{Non3sgAux Non3sgNP VP}$

and the lexicon:

$\text{3sgAux} \rightarrow \text{does} \mid \text{has} \mid \text{can} \mid \dots$

$\text{Non3sgAux} \rightarrow \text{do} \mid \text{have} \mid \text{can} \mid \dots$

and continue to expand the grammar:

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

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

$\text{SgNominal} \rightarrow \text{SgNoun} \mid \text{SgNoun SgNoun}$

$\text{PlNominal} \rightarrow \text{PlNoun} \mid \text{SgNoun PlNoun}$

and the lexicon:

$\text{SgNoun} \rightarrow \text{flight} \mid \text{fare} \mid \text{dollar} \mid \text{reservation} \mid \dots$

$\text{PlNoun} \rightarrow \text{flights} \mid \text{fares} \mid \text{dollars} \mid \text{reservations} \mid \dots$

other languages: gender agreement

feature structures - (ATNs)

## THE VERB PHRASE AND SUBCATEGORIZATION

### simple VPs

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

### sentential complements

VP → Verb S

*You [vp [v said] [s there were two flights that were cheaper]]*  
*[vp [v Tell] [np me][s how to get downtown]]*  
*I [vp [v think] [s I would like to take the train]]*

### VP complements

(verbs like *want, would like, try, intend, need*)

*I want [vp to fly to Orlando]*  
*I'm trying [vp to find a flight from Pittsburgh to Denver]*

### *particles- phrasal verbs*

*take off, look up, check out*

### Subcategorization terminology:

complement  
subcategorizes for....  
subcategorization frame

### The issue:

not every verb is compatible with every complement  
(e.g., transitive vs. intransitive verbs)  
ignoring this would mean overgeneration

## subcategorization frames

	eat	She eats
NP	prefer, <u>find</u> , leave,	I found the book
NP, NP	show, give, <u>find</u>	Show me the money
PP <sub>from</sub> , PP <sub>to</sub>	fly, travel	I flew from Boston to NYC
NP PP <sub>with</sub>	help, load	Sam helped me with the job
VP <sub>to</sub>	prefer, want, need	I prefer to go home
VP <sub>brst</sub>	can, would, might	I can [vp <sub>brst</sub> go from Boston]
S	mean	You mean [s I can go]?

Could treat this with separate types of verbs:

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

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

Verb-with-inf-VP-complement → want | try | need | ...

and separate types of rules for those verb types:

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

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

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

but explosion of rules .... so feature structures

## AUXILIARIES and Subcategorization

auxiliary

modals:

*can, could, may, might, must, will, would, shall, should*

perfect auxiliary

*have*

progressive auxiliary

*be*

passive auxiliary

*be*

	subcat for:	e.g.
modals:	bare stem	<u>can go</u> in the morning <u>will try</u> to find a flight
perfect aux	VP <sub>past participle</sub>	<u>have booked</u> 3 flights
progressive aux	VP <sub>gerundive participle</sub>	<u>am going</u> from Atlanta
passive aux	VP <sub>past participle</sub>	<u>was delayed</u> by weather

ordered:

modal < perfect < progressive < passive

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

# SPOKEN LANGUAGE SYNTAX

## utterances

the . [exhale] . . . [inhale] . . [uh] does American airlines . offer any . one way flights . [uh] one way fares, for one hundred and sixty one dollars
[mm] i'd like to leave i guess between [um] . [smack] . five o'clock no, five o'clock and [uh], seven o'clock . P M
around, four, P M
all right, [throat_clear] . . i'd like to know the . give me the flight . times . in the morning . for September twentieth . nineteen ninety one
[uh] one way
[uh] seven fifteen, please
on United airlines . . give me, the . . time . . from New York . [smack] . to Boise-, to . I'm sorry . on United airlines . [uh] give me the flight, numbers, the flight times from . [uh] Boston . to Dallas

differ in lexical statistics

subject is usually a pronoun

disfluencies

fragment utterances

prosody

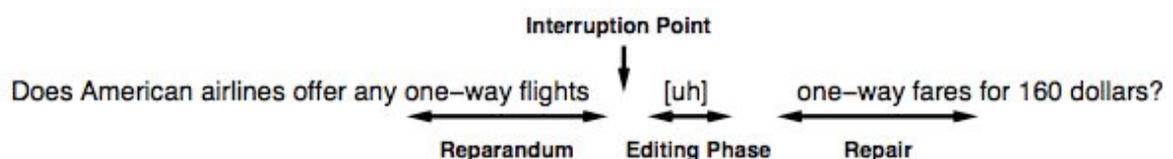
pitch contour

stress pattern

rhythm

rate, volume, tone

## Disfluencies



filled pauses

constituent structure of reparandum and repair are parallel  
(similar to errors, slips of speech)

## GRAMMAR EQUIVALENCE AND NORMAL FORM

Chomsky adequacy criteria:

- observational
- descriptive
- explanatory

equivalence

- weak - observational
- strong - descriptive

Chomsky normal form CNF

$A \rightarrow BC$

$A \rightarrow \square$

conversion to weakly equivalent CNF, e.g.,

$A \rightarrow BC D$

$A \rightarrow B X$

$X \rightarrow C D$

## FINITE-STATE AND CONTEXT-FREE GRAMMARS

### Recursion

direct:

NP → NP PP

indirect:

NP → NP PP

PP → P NP

*flights from Denver*

*Flights from Denver to Miami*

*Flights from Denver to Miami in February*

*Flights from Denver to Miami in February on a Friday*

*Flights from Denver to Miami in February on a Friday under  
\$300*

*Flights from Denver to Miami in February on a Friday under  
\$300 with lunch*

*[[flights] [from Denver]]*

*[[[Flights] [from Denver]] [to Miami]]*

*[[[[Flights] [from Denver]] [to Miami]] [in February]]*

*[[[[[Flights] [from Denver]] [to Miami]] [in February]] [on  
a Friday]]*

In a CFG rule, e.g.,

A → .....C.....

A constituent C can be used:

without regard for the internal structure of C

or the context preceding or following the NT A

recursion in finite state-grammars?

NP to head:

(Det) (Card) (Ord) (Quant) (AP) Nominal

add PP

(Det) (Card) (Ord) (Quant) (AP) Nominal (PP) \*

expand definition of PP

(Det) (Card) (Ord) (Quant) (AP) Nominal (P NP) \*

expand NP

(Det) (Card) (Ord) (Quant) (AP) Nominal (P  
(Det) (Card) (Ord) (Quant) (AP) Nominal (P NP)) \*

NP still there.....

and recursion appears with RelClause and GerundVP as well...

(Det) (Card) (Ord) (Quant) (AP) Nominal  
(RelClause | GerundVP | PP) \*

Center Embedded recursion

A □ □ A □

Limiting depth of recursion

makes FSG approximating CFG possible (not clean though)

RTNs

## GRAMMARS AND HUMAN PROCESSING

Priming studies:

ditransitive alternation:

*The wealthy widow gave [NP the church] [NP her Mercedes] (9.16)*

*The wealthy widow gave [NP her Mercedes] [PP to the church](9.17)*

prime had different semantics (e.g., locative vs. dative)

*IBM moved [NP a bigger computer] [NP to the Sears store] (9.18)*

CFG issues:

need semantic, pragmatic, social/interactional/prosodic ?

anti-modularist vs. modularist views