Chapter 2: From Words to Major Phrase Types

Jong-Bok Kim & Peter Sells
jongbok@khu.ac.kr, sells@stanford.edu

Kyung Hee U. & Stanford U.

Introduction

The basic units of syntax are words.
Grouping words into same kind of words (also known as parts of speech, or lexical categories, or grammatical categories)

- start from words (lexical categories)
- These lexical categories then form a larger constituent 'phrase'; and phrases go together to form a 'clause'.

(1) a. The weather is lovely today.
b. I am hoping [that the weather is lovely today].
c. [If the weather is lovely today] then we will go out.
d. The birds are singing [because the weather is lovely today].

Words can be classified into different lexical categories according to three criteria: meaning, morphological form, and syntactic function.

Determining the lexical categories: by meaning

At first glance, it seems that words can be classified depending on their meaning.

(2) a. N: referring to an individual or entity

b. V: referring to an action

c. A: referring to a property

d. Adv: referring to the manner, location, time or frequency of an action

: cf. Words like *sincerity, happiness*, and *pain*, *absence* and *loss*.

words like sincerity, happiness, and pain do not simply denote any individual or entity. Absence and loss are even harder cases.

There are also many words whose semantic properties do not match the lexical category that they belong to. For example, words like assassination and construction may refer to an action rather than an individual, but they are always nouns. Words like remain, bother, appear, and exist are verbs, but do not involve any action.

Determining the lexical categories: by form

(3) a. N: __ + plural morpheme *-(e)s*

b. N: __ + possessive *'s*

c. V: __ + past tense *-ed* or 3rd singular *-(e)s*

d. V: __ + 3rd singular *-(e)s*

e. A: __ + *-er/est* (or *more/most*)

f. A: __ + *-ly* (to create an adverb)

(4) a. N: trains, actors, rooms, man's, sister's, etc.

b. V: devoured, laughed, devours, laughs, etc.

c. A: fuller, fullest, more careful, most careful, etc.

d. Adv: fully, carefully, diligently, clearly, etc.

: think about *information, furniture, love, pain, absent*, etc.

The morphological (form-based) criterion, though reliable in many cases, is not a necessary and sufficient condition for determining the type of lexical categories.

Determining the lexical categories: by distribution

The most reliable criterion in judging the lexical category of a word is based on its syntactic function or distributional possibilities.

- Distribution tests

(5) a. They have no __ .
b. They can __ .
c. They read the __ book.
d. He treats John very __ .
e. He walked right __ the wall.

The categories that can go in the blanks are N, V, A, Adv, and P (preposition). Roughly only one lexical category can appear in each position.

Determining the lexical categories: by distribution

- N, V, Adj, P

(6) a. They have no TV/car/information/friend.
b. They have no *went/*in/*old/*very/*and.

(7) a. They can sing/run/smile/stay/cry.
b. They can *happy/*down/*door/*very.

(8) a. They read the big/new/interesting/scientific book.
b. They read the *sing/*under/*very book.

(9) a. He treats John very nicely/badly/kindly.
b. He treats John very *kind/*shame/*under.

(10) a. He walked right into/on the wall.
b. He walked right *very/*happy/*the wall.

Some examples for basic lexical categories

As shown, only a restricted set of lexical categories can occur in each position; we can then assign a specific lexical category to these elements:

(11) a. N: TV, car, information, friend, . . .

b. V: sing, run, smile, stay, cry, . . .

c. A: big, new, interesting, scientific, . . .

d. Adv: nicely, badly, kindly, . . .

e. P: in, into, on, under, over, . . .

In addition to these basic lexical categories, does English have other lexical categories?
There are a few more. Consider the following syntactic environments:

What other categories?

(12) a. __ student hits the ball.

b. John sang a song, __ Mary played the piano.

c. John thinks __ Bill is honest.

(13) We found out that __ jobs were in jeopardy.

Here we see that only words like the, my, his, some, these, those, and so forth can occur here. These articles, possessives, quantifiers, and demonstratives all 'determine' the referential properties of jobs here, and for this reason, they are called determiners.

Determiners and complementizers

(14) a. *[My these jobs] are in jeopardy.

b. *[Some my jobs] are in jeopardy.

c. *[The his jobs] are in jeopardy.

(15) a. I think __ learning English is not easy at all.

b. I doubt __ you can help me in understanding this.

c. I am anxious __ you to study English grammar hard.

Words like 'my' and 'these' or 'some' and 'my' cannot occur together, indicating that they compete with each other for just one structural position.

Complementizers

The italicized words here are different from the other lexical categories that we have seen so far. They introduce a complement clause, marked above by the square brackets, and may be sensitive to the tense of that clause. A tensed clause is known as a 'finite' clause, as opposed to an infinitive. For example, that and if introduce or combine with a tensed sentence (present or past tense), whereas for requires an infinitival clause marked with to.

consider these

(16) a. I think *that* [learning English is not all that easy].

b. I doubt *if* [you can help me in understanding this].

c. I am anxious *for* [you to study English grammar hard].

(17) a. *I think *that* [learning English to be not all that easy].

b. *I doubt *if* [you to help me in understanding this].

c. *I am anxious *for* [you should study English grammar hard].

(18) a. I think *that* [learning English is not all that easy].

b. I doubt *if* [you can help me in understanding this].

c. I am anxious *for* [you to study English grammar hard].

The term 'complement' refers to an obligatory dependent clause or phrase relative to a head.2 The italicized elements in (18) introduce a clausal complement and are consequently known as Complementizers (abbreviated as 'C'). There are only a few complementizers in English (that, for, if , and whether),

Auxiliaries

(19) a. John __ not leave.
b. John __ drink beer last night.
c. __ John leave for Seoul tomorrow?
d. John will study syntax, and Mary __ , too.

(20) a. He left.
b. He did not leave.

21: ?

(21) a. Students wanted *to* write a letter.
b. Students intended *to* surprise the teacher.

(22) a. Students objected *to* the teacher. 22. Prep
b. Students sent letters *to* the teacher.

The words that can appear in the blanks are neither main verbs nor adjectives, but rather words like will', 'can', 'shall' and 'must'. In English, there is clear evidence that these verbs are different from main verbs, and we call them auxiliary verbs (Aux). The auxiliary verb appears in front of the main verb, which is typically in its citation form, which we call the 'base' form. Note the change in the main verb form in (20b) when the negation is added.

There is also one type of to which is auxiliary-like. Consider the examples in (21) and (22):

Auxiliaries

We can observe that 'to' behaves like an auxiliary verb 'should':

(23) a. It is crucial for John to show an interest.

b. It is crucial that John should show an interest.

(24) a. I know I should [go to the dentist's], but I just don't want to.

b. I don't really want to [go to the dentist's], but I know I should.

(25) a. She thought it was likely [that everyone *to/might/would fit into the car].

b. She thought it was easy [for everyone to/*might/*would fit into the car].

In (23), to and should introduce the clause and determines the tenseness of the clause.
In (24), they both can license the ellipsis of its VP complement.

Another property to shares with other auxiliary verbs like will is that it requires a base verb to follow.

Particles vs. prepositions

- Particle

(26) a. The umpire called *off* the game.
b. The two boys looked *up* the word.

Words like off and up here behave differently from prepositions, in that they can occur after the object.

(27) a. The umpire called the game *off*.
b. The two boys looked the word *up*.

- preposition

(28) a. The umpire fell *off* the deck.
b. The two boys looked *up* the high stairs (from the floor).

(29) a. *The umpire fell the deck *off*.
b. *The students looked the high stairs *up* (from the floor).

Particles vs. prepositions

(30) a. The umpire called it *off*. (particle)

b. *The umpire called *off* it.

(31) a. *The umpire fell it *off*.

b. The umpire fell *off* it. (preposition)

Content and function words

(32) a. The student will take a green apple.
b. The teachers are fond of Bill.

(33) a. *Student take green apple
b. *Teachers fond Bill.

(34) Content words:
a. N: computer, email, fax, Internet, ...
b. A: happy, new, large, grey, tall, exciting, ...
c. V: email, grow, hold, have, run, smile, make, ...
d. Adv: really, completely, also, well, quickly, ...

(35) a. P: of, at, in, without, between, ...
b. Det: the, a, that, my, more, much, ...
c. Conj: and, that, when, while, although, or, ...
d. Aux: can, must, will, should, ought, ...
e. C: for, whether, that, ...
f. Part: away, over, off, out, ...

Grammar with lexical categories

- sentences

32 (36) a. A man kicked the ball.
b. A tall boy threw the ball.
c. The cat chased the long string.
d. The happy student played the piano.

- Rule:

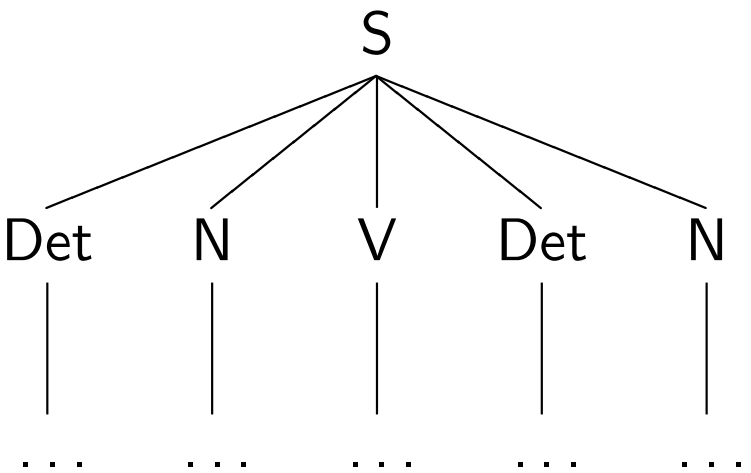
33 (37) S → Det (A) N V Det (A) N

the main goal of syntax is building a grammar that can generate an infinite set of well-formed, grammatical English sentences. Let us see what kind of grammar we can develop now that we have lexical categories.

Grammar with lexical categories

34

(38)



So this rule characterizes any sentence which consists of a Det, N, V, Det, and N, in that order, possibly with an A in front of either N. We can represent the core items in a tree structure as in (38):

Grammar with lexical categories

We assume a lexicon, a list of categorized words, to be part of the grammar along with the rule in 33

(39) a. Det: a, that, the, this, . . .

b. N: ball, man, piano, string, student, . . .

c. V: kicked, hit, played, sang, threw, . . .

d. A: handsome, happy, kind, long, tall, . . .

(40) a. That ball hit a student.

b. The piano played a song.

c. The piano kicked a student.

d. That ball sang a student.

By inserting lexical items into the appropriate pre-terminal nodes in the structure, where the labels above . . . are, we can generate grammatical examples like those (32/40) as well as those like the following, not all of which describe a possible real-world situation:

Grammar with lexical categories

Notice that even this simple grammar rule can easily extend to generate an infinite number of English sentences by allowing iteration of the A

(41) $S \rightarrow Det\ A^*\ N\ V\ Det\ A^*\ N$

(42) a. The tall man kicked the ball.

b. The tall, handsome man kicked the ball.

c. The tall, kind, handsome man kicked the ball.

(43) The happy, happy, happy, happy, happy, happy man sang a song.

The operator ( ) allows us to repeat any number of As, thereby generating sentences like (38). Note that the parentheses around 'A' in (34) are no longer necessary in this instance, for the Kleene Star operator means any number including zero.

Grammar with lexical categories: Problems

A grammar using only lexical categories can be specified to generate an infinite number of well-formed English sentences, but it nevertheless misses a great deal of basic properties that we can observe. For example, this simple grammar cannot capture the agreement facts seen in examples like the following:

(44) a. The mother of the boy and the girl **is** arriving soon.

b. The mother of the boy and the girl **are** arriving soon.

(45) a. [The mother of [the boy and the girl]] is arriving soon.

b. [The mother of the boy] and [the girl] are arriving soon.

Why do the verbs in these two sentences have different agreement patterns? Our intuitions tell us that the answer lies in two different possibilities for grouping the words:

The different groupings shown by the brackets indicate who is arriving: in (45a), the mother, while in (45b) it is both the mother and the girl. The grouping of words into larger phrasal units which we call constituents provides the first step in understanding the agreement facts in (41).

Grammar with lexical categories: Problems

These sentences have different meanings depending on how we group the words.

(46) a. John saw the man with a telescope.

b. I like chocolate cakes and pies.

c. We need more intelligent leaders.

(47) a. John saw [the man with a telescope].
(the man had the telescope)

b. John [[saw the man] with a telescope].
(John used the telescope)

Even these very cursory observations indicate that a grammar with only lexical categories is not adequate for describing syntax. In addition, we need a notion of 'constituent', and need to consider how phrases may be formed, grouping certain words together.

Grammar with phrases

In addition to these agreement and ambiguity facts, our intuitions may also lead us to hypothesize constituency. If you (the native speakers) were asked to group the words in (44/48) into phrases, what constituents would you come up with?

(48) The student enjoyed his English syntax class last semester.

(49) a. [The student] [enjoyed [his English syntax class last semester]].

b. [The] [student enjoyed] [his English syntax class] [last semester].

c. [The student] [[enjoyed his English] [syntax class last semester]].

What kind of knowledge, in addition to semantic coherence, forms the basis for our intuitions of constituency? Are there clear syntactic or distributional tests which demonstrate the appropriate grouping of words or specific constituencies? There are certain salient syntactic phenomena which refer directly to constituents or phrases.

Constituent tests

Cleft: The cleft construction, which places an emphasized or focused element in the X position in the pattern 'It is/was X that . . . ', can provide us with simple evidence for the existence of phrasal units.

(50) The policeman met several young students in the park last night.

(51) a. It was [the policeman] that met several young students in the park last night.

b. It was [several young students] that the policeman met in the park last night.

c. It was [in the park] that the policeman met several young students last night.

d. It was [last night] that the policeman met several young students in the park.

Constituent tests

(52) a. *It was [the policeman met] that several young students in the park last night.

b. *It was [several young students in] that the policeman met the park last night.

c. *It was [in the park last night] that the policeman met several young students.

Constituent tests

Constituent Questions and Stand-Alone Test: For any given *wh*-question, the answer can either be a full sentence or a fragment. This stand-alone fragment is a constituent:

(53) A: Where did the policeman meet several young students?
B: In the park.

(54) A: Who(m) did the policeman meet in the park?
B: Several young students.

Constituent tests

(55) John put old books in the box.

(56) A: What did you put in your box?
B: Old books.
B: *Old books in the box.

(57) A: Where did you put the book?
B: In the box.
B:*Old books in the box.

(58) A: What did you do?
B: *Put old books.
B: *Put in the box.
B: Put old books in the box.

Constituent tests: preposition and particles again

(59) a. John looked up the inside of the chimney.

b. John looked up the meaning of ‘chanson’.

(60) A: What did he look up?

B: The inside of the chimney.

B: The meaning of ‘chanson’.

(61) A: Where did he look?

B: Up the inside of the chimney.

B:*Up the meaning of ‘chanson’.

(62) A: Up what did he look?

B: The inside of the chimney.

B: *The meaning of ‘chanson’.

Constituent tests

Substitution by a Pronoun: English, like most languages, has a system for referring back to individuals or entities mentioned by the use of pronouns.

(63) a. What do you think the man who is standing by the door is doing now?

b. What do you think *he* is doing now?

(64) a. Have you been [to Seoul]? I have never been *there*.

b. John might [go home], *so* might Bill.

c. John might [pass the exam], and *as* might Bill.

d. If John can [speak French fluently] – *which* we all know he can – we will have no problems.

Constituent tests

(65) a. John asked me to put the clothes in the cupboard, and to annoy him I really stuffed *there* [there=in the cupboard].

b. John asked me to put the clothes in the cupboard, and to annoy him I stuffed *them there* [them=the clothes].

c. *John asked me to put the clothes in the cupboard, but I did *so* [=put the clothes] in the suitcase.

Constituent tests

- Words and phrases can be coordinated by *conjunctions*, and each conjunct is typically the same kind of constituent as the other conjuncts:

(66) a. The girls [played in the water] and [swam under the bridge].
b. The children were neither [in their rooms] nor [on the porch].
c. She was [poor] but [quite happy].
d. Many people drink [beer] or [wine].

(67) a. *Mary waited [for the bus] and [to go home].
b. *Lee went [to the store] and [crazy].

NP

The above tests confirm our assumption that there are strings of words which function as one unit. These strings have some internal structure, and are projected from lexical categories. We call these strings phrases. Again, we use distributional evidence to classify these first and then to specify rules for the distributions we observe.

- template

  (68) __ [liked ice cream].

- NPs

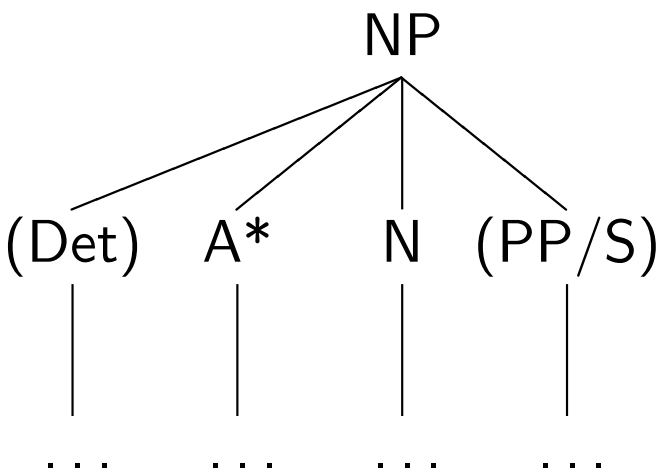
  (69) Mary, I, you, students, the students, the tall students, the students from Seoul, the students who came from Seoul, etc.

  (70) NP $\rightarrow$ (Det) A* N (PP/S)

This rule characterizes a phrase, and is one instance of a phrase structure rule (PS rule). The rule indicates that an NP can consist of an optional Det, any number of optional A, an obligatory N, and then an optional PP or a modifying S.8 The slash indicates different options for the same place in the linear order.

NP

(71)



(72) *the whistle tune, *the easily student, *the my dog, . . .

VP

- template

(73) The student __ .

(74) snored, ran, sang, loved music, walked the dog through the park, lifted 50 pounds, thought Tom is honest, warned us that storms were coming, etc.

(75) VP $\rightarrow$ V (NP) (PP/S)

(76) *leave the meeting sing, *the leave meeting, *leave on time the meeting, ...

VP

- Tensed VP

(77) a. The monkey wants to leave the meeting.
b. *The monkey eager to leave the meeting.

(78) a. The monkeys approved of their leader.
b. *The monkeys proud of their leader.

(79) a. The men practice medicine.
b. *The men doctors of medicine.

(80) $S \rightarrow NP\ VP$

VP: auxiliary

- template

(81) a. The students __ .
b. The students want __ .

(82) a. run, feel happy, study English syntax, . . .
b. can run, will feel happy, must study English syntax, . . .
c. to run, to feel happy, to study English syntax, . . .

- VP rule

(83) $VP \rightarrow V[AUX +] VP$

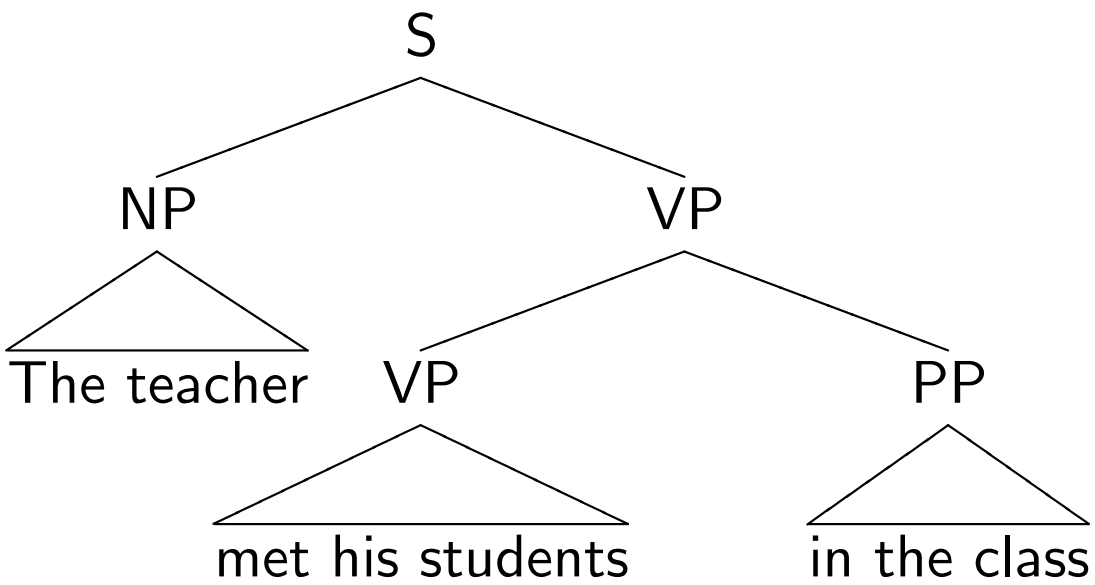
VP with a modifier

(84) a. John [[read the book] loudly].

b. The teacher [[met his students] in the class].

(85) VP → VP Adv/PP

(86)



AP

(87) John feels __ .

(88) happy, uncomfortable, terrified, sad, proud of her, proud to be his student, proud that he passed the exam, etc.

(89) AP → A (PP/VP/S)

AP

(90) a. John sounded happy/uncomfortable/terrified/proud of her.

b. John sounded *happily/*very/*the student/*in the park.

(91) a. *The monkeys seem [want to leave the meeting].

b. The monkeys seem [eager to leave the meeting].

(92) a. *John seems [know about the bananas].

b. John seems [certain about the bananas].

AdvP

(93) soundly, well, clearly, extremely, carefully, very soundly, almost certainly, very slowly, etc.

(94) a. He behaved very __ .

b. They worded the sentence very __ .

c. He treated her very __ .

(95) AdvP $\rightarrow$ (AdvP) Adv

PP

(96) from Seoul, in the box, in the hotel, into the soup, with John and his dog, under the table, etc.

(97) a. John came from Seoul.

b. They put the book in the box.

c. They stayed in the hotel.

d. The fly fell into the soup.

(98) The squirrel ran straight/right __ .

(99) a. The squirrel ran straight/right up the tree.

b. *The squirrel is straight/right angry.

c. *The squirrel ran straight/right quickly.

(100) PP $\rightarrow$ P NP

(101) *in angry, *into sing a song, *with happily, . . .

PS rules

We have seen earlier that the grammar with just lexical categories is not adequate for capturing the basic properties of the language. How much further do we get with a grammar which includes phrases?

(102) a. $S \rightarrow NP\ VP$

b. $NP \rightarrow (Det)\ A^*\ N\ (PP/S)$

c. $VP \rightarrow V\ (NP)\ (PP/S/VP)$

d. $AP \rightarrow A\ (PP/S)$

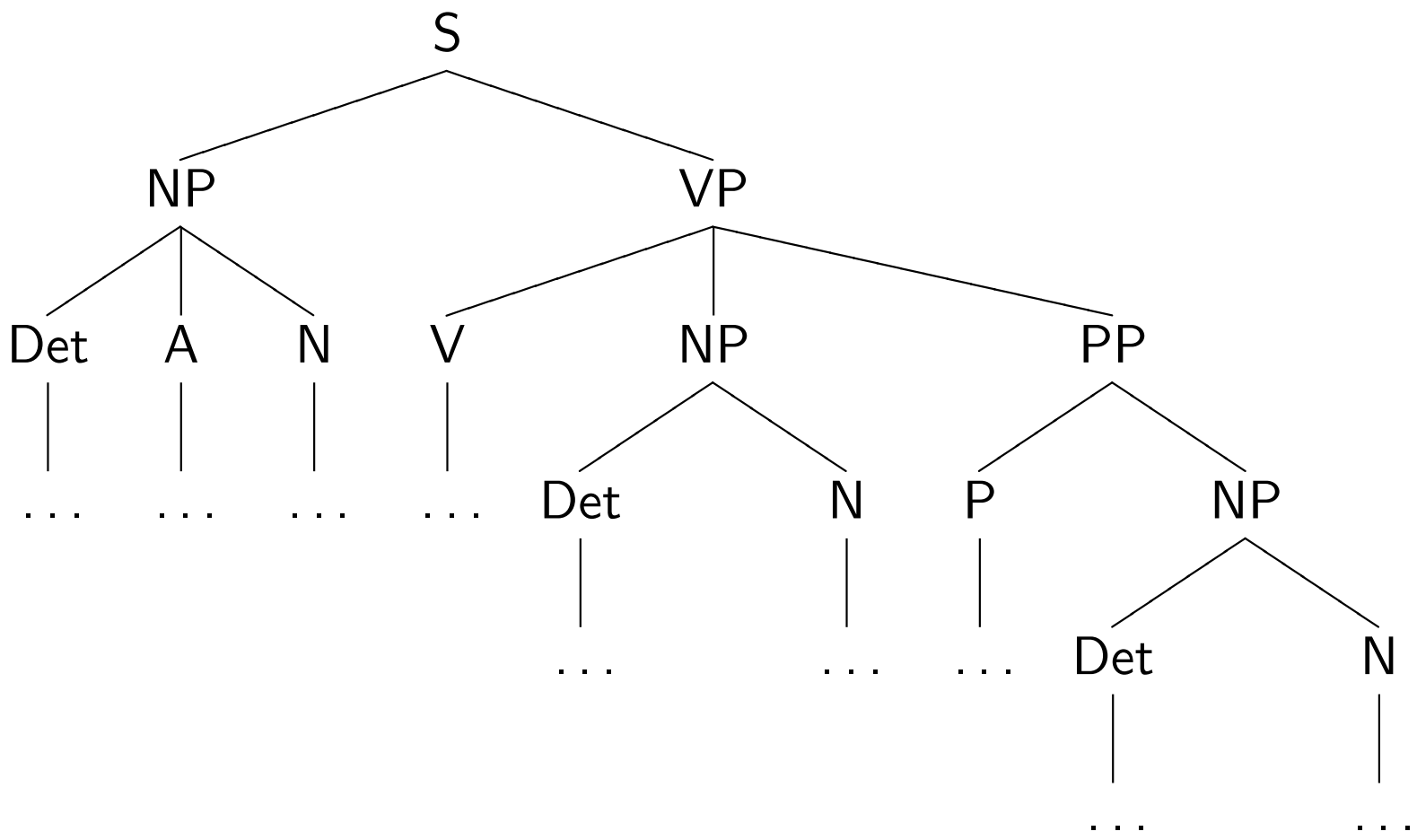
e. $AdvP \rightarrow (AdvP)\ Adv$

f. $PP \rightarrow P\ NP$

12The grammar consisting of such form of rules is often called a 'Context Free Grammar', as each rule may apply any time its environment is satisfied, regardless of any other contextual restrictions. (It's kind of 'mathematical')

From PS rules to tree structures

(103)



Lexicon and Generated outputs

let us assume that we have the following lexical entries:

(104) a. Det: a, an, this, that, any, some, which, his, her, no, etc.

b. A: handsome, tall, fat, large, dirty, big, yellow, etc.

c. N: book, ball, hat, friend, dog, cat, man, woman, John, etc.

d. V: kicked, chased, sang, met, believed, thinks, imagines, assumes etc.

(105) a. This handsome man chased a dog.

b. A man kicked that ball.

c. That tall woman chased a cat.

d. His friend kicked a ball.

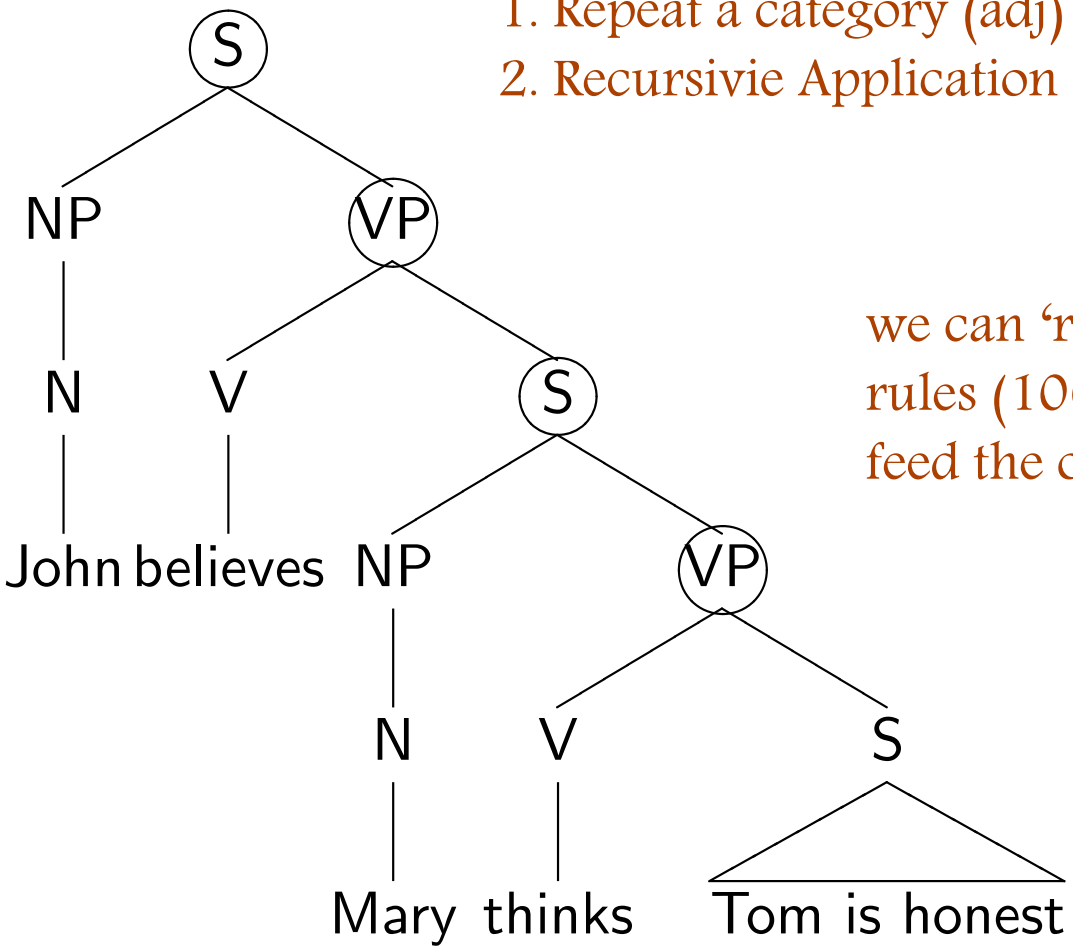
Inserting these elements in the appropriate pre-terminal nodes (the places with dots) in (99/103), we are able to generate various sentences like those in (101/105):

Generating an infinite number of sentences: recursiveness

(106) a. $S \rightarrow NP\ VP$

b. $VP \rightarrow V\ S$

(107)



There are several ways to generate an infinite number of sentences with this kind of grammar.

1. Repeat a category (adj) infinitely.
2. Recursivie Application

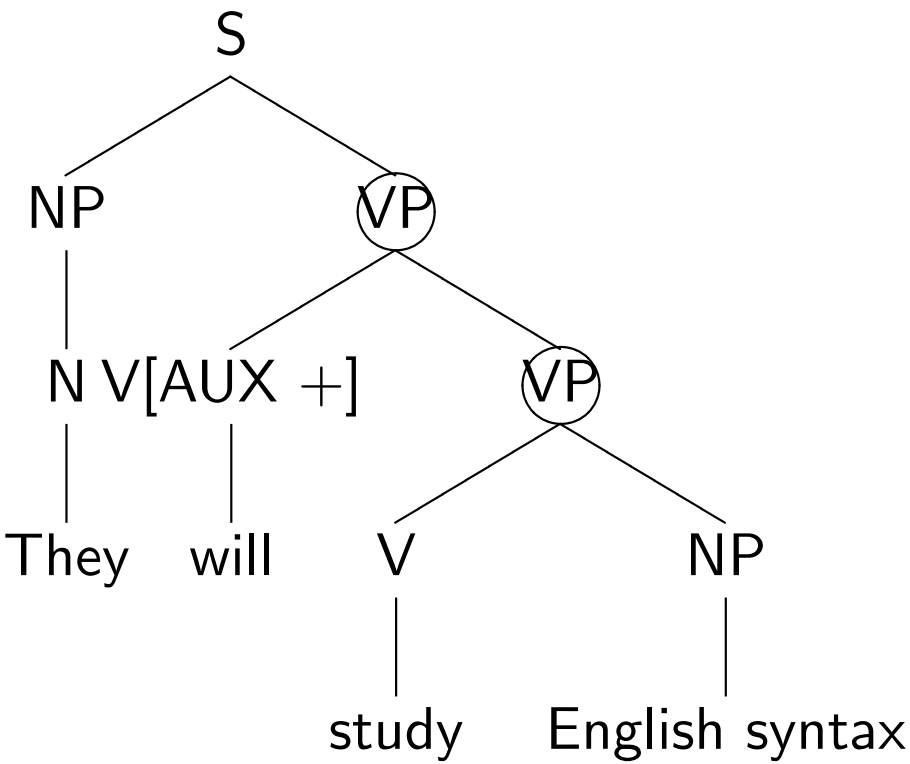
we can 'recursively' apply the two rules (106), in the sense that one can feed the other, and then vice versa:

Recursive exmaples

(108) a. Bill claims John believes Mary thinks Tom is honest.

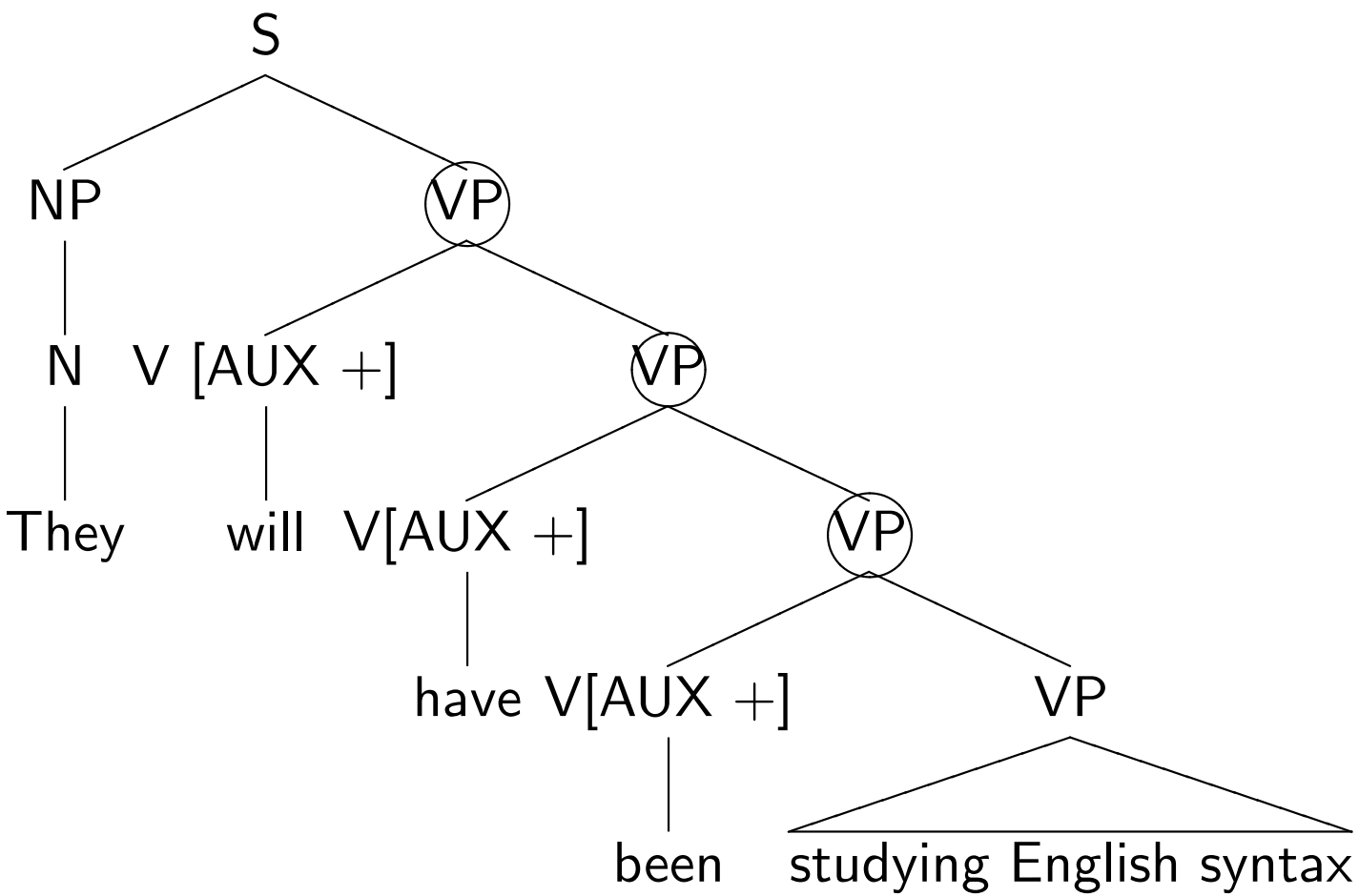
b. Jane imagines Bill claims John believes Mary thinks Tom is honest.

(109)



Recursive Aux

(110)

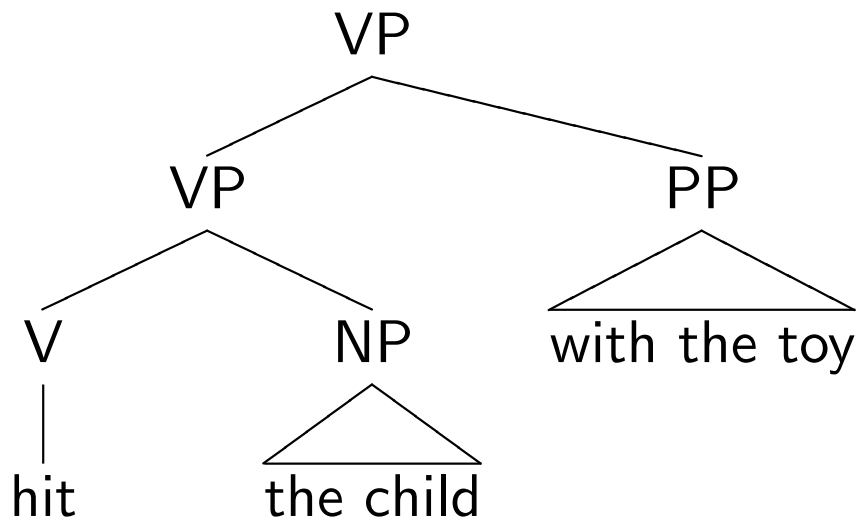


hierarchical structure: solving ambiguity

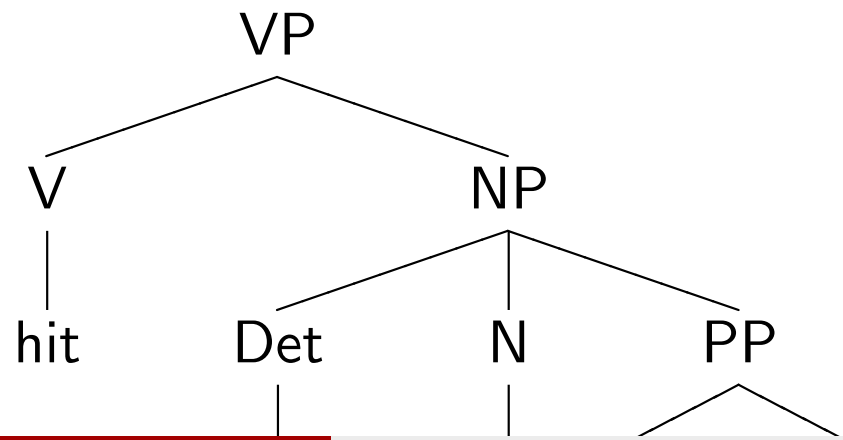
(111) a. The little boy hit the child with a toy.

b. Chocolate cakes and pies are my favorite desserts.

(112) a.



b.



Another important property that PS rules bring us is the ability to make reference to hierarchical structures within given sentences, where parts are assembled into sub-structures of the whole. One merit of such hierarchical structural properties is that they enable us to represent the structural ambiguities of sentences we have seen earlier in (42).

John saw the man with a telescope.

Coordination

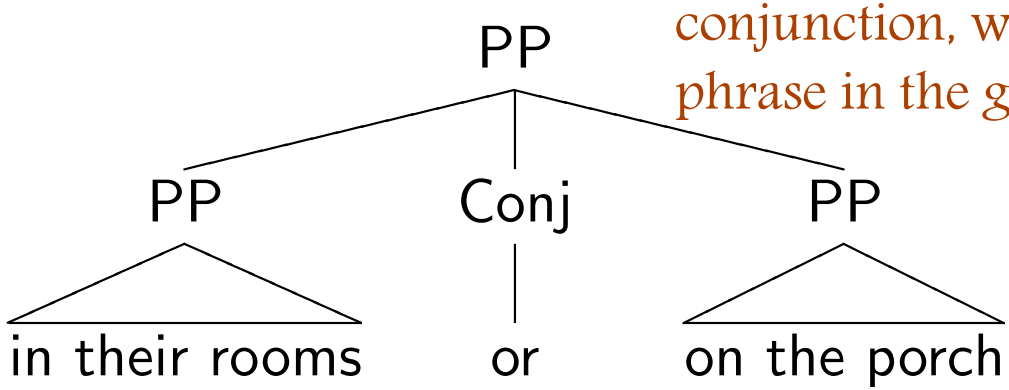
(113) a. *The children were in their rooms or happy.

b. *Lee went to the store and crazy.

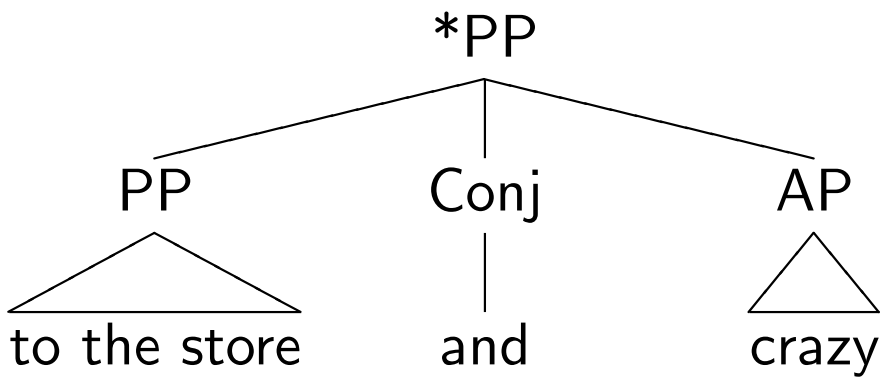
(114) XP → XP* Conj XP

English allows two alike categories to be coordinated. This can be written as a PS rule, for phrasal conjunction, where XP is any phrase in the grammar.

(115) a.



b.



The 'coordination' rule says two identical XP categories can be coordinated and form the same category XP. Applying this PS rule, we will then allow (a) but not (b).

Preposition vs. particle

(116) a. John suddenly got *off* the bus.

b. John suddenly put *off* the customers.

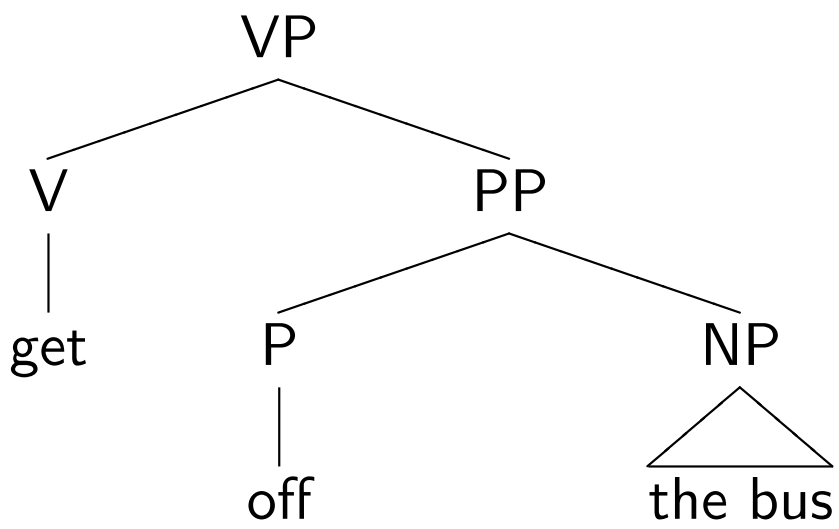
(117) a. *John suddenly got the bus off.

b. John suddenly put the customers off.

(118) VP $\rightarrow$ V (Part) (NP) (Part) PP

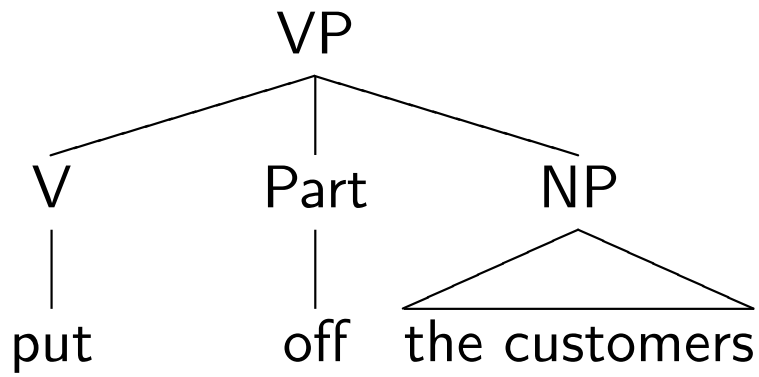
Particle vs. Preposition

(119)



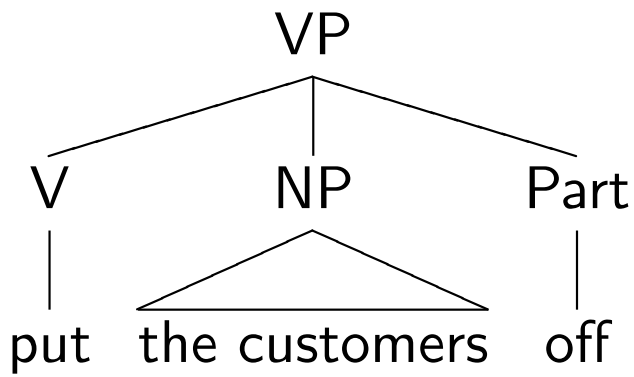
Particle vs. Preposition

(120) a.



the particle does not form a constituent with the following or preceding NP whereas the preposition does form a constituent with it.

b.



a grammar with lexical categories can not only generate an infinite number of grammatical English sentences, but also account for some fundamental properties, such as agreement and constituency. This motivates the introduction of phrases into the grammar.