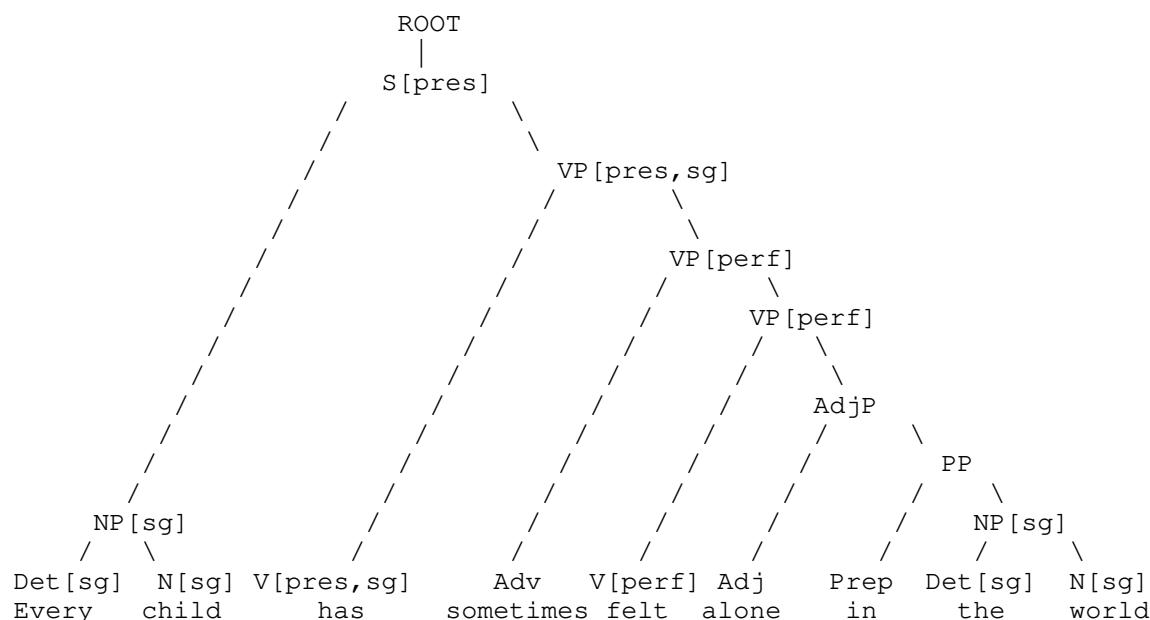


1. Here is a reasonable tree:



where "sg" is short for "number=singular"  
 "pres" is short for "tense=present"  
 "perf" is short for "tense=perfect"

SANITY CHECK: What if we changed "sg" to "pl" (plural)? Let's see what happens, leaving everything else alone as much as possible:

All children have sometimes felt alone in the world.

That is indeed still grammatical. The plural NP (with plural determiner "all" and plural noun "children") agrees with the plural VP (with plural auxiliary verb "have").

NP NUMBER FEATURES: Determiners need number features. You can see that "every" is a singular determiner:

every child  
 \*every children

And that "all" is a plural determiner:

\*all child  
 \*all children

And that "the" can be either singular or plural. You can write

Det[number=singular] -> the  
 Det[number=plural] -> the

or just compress those two rules into one rule with a variable:

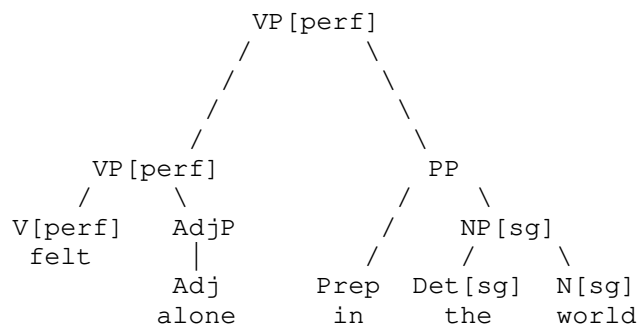
Det[number=alpha] -> the

which says that the Det -> the rule doesn't care about number.

VP NUMBER FEATURES: VP[pres, sg] realizes its singular feature on its left child -- the V[pres, sg] "have". We didn't bother to pass it down also to the right child, the VP[perf] "sometimes felt alone in the world".

Passing it down to the right child would be superfluous because it wouldn't be used there: the VP[perf] doesn't have separate singular and plural forms (at least in English). In general, the number feature on an English verb phrase only affects the leftmost verb: "every child FEELS" or "every child HAS felt."

COMPLEX AdjP: The most interesting property of this sentence is the PP "in the world". What is the function of that PP? Maybe it is modifying the VP:



Then the PP would be describing where the VP's event happened: where the child felt that emotion. But it would be strange to MENTION that the child was "in the world" when they had this feeling -- because we all know that EVERY event happens in the world! You don't have to bother to say it. So although this parse is possible, it doesn't correspond to the meaning that was probably intended.

A more likely meaning is that the child's feeling is a feeling of being "alone in the world." In other words, "in the world" modifies the type of aloneness. Being "alone in the basement" means that no one else is in the basement with you. Being "alone in the world" means that no one else is in the world with you.

In the recommended parse, to make the phrase "alone in the world," we used the rule AdjP -> Adj PP. You should be a bit worried about this rule because we haven't seen an PP combine with an adjective before! However, we can test that "alone in the world" is actually a legal AdjP by trying to use it in contexts that call for an AdjP (and only an AdjP). It seems to work!

Tintin was not so tall.

Tintin was not so alone in the world.

Whether poor or rich, you are valuable.

Whether alone in the world or surrounded by a large family,  
you are valuable.

So the AdjP -> Adj PP rule seems to be okay, at least with some adjectives like "alone." You can look for other examples.

In fact, one such example is above: "surrounded by a large family" can also be analyzed using AdjP -> Adj PP. (This particular example is more specifically a passive construction, where the Adj is a past participle and the preposition is "by". Semantically, it's related to "a large family surrounds e".)

ADVERB PLACEMENT: The other interesting property of the sentence is that adverb "sometimes" is stuck into the middle of the main VP, "has felt alone in the world." We explain this by saying that it is adjoined to one of the sub-VPs, using a rule like this:

VP[tense=alpha] -> Adv VP[tense=alpha]

(What does "adjoined" mean? Remember that adjunction involves a rule of the form X -> X Y or X -> Y X. It enhances the lower X by providing additional details in Y, something like an optional keyword argument in Python.)

When we apply this adjunction rule at VP[perf], we get

"has sometimes felt alone in the world"

But notice that we could also have applied it at VP[pres], giving

"sometimes has felt alone in the world"

which is also grammatical.

2. (a) This was a bit of a trick question, because it probably reminded you of a sentence we discussed in class:

Papa ate the caviar with a spoon.

As discussed in class, that sentence has a SYNTACTIC ambiguity, regarding where the PP attaches.

However, this syntactic PP-attachment ambiguity goes away in

Papa ate them with the spoon.

because "them with a spoon" is not a grammatical constituent: it sounds horrible! (You can't say \*"Them with paprika is what Papa ate.") Apparently, a PP can attach to a noun ("caviar"), but not to a full NP, and certainly not to a pronoun such as "them."

Similarly, the syntactic PP-attachment ambiguity is NOT present in  
What did you mix them with?

Rather, this example has is a LEXICAL ambiguity in the meaning of "with" or perhaps "mix with". Does "mix with" mean "mix together with" or "mix using"?

Remember from week 1 of the class that a "lexicon" contains information about all the words of a language, and a "lexical" ambiguity arises when we're not sure which word is present: is this "with\_1" or "with\_2", which are spelled the same but have different meanings?

Notice that the same lexical ambiguity is present in

Papa [[ate the caviar] with the spoon].

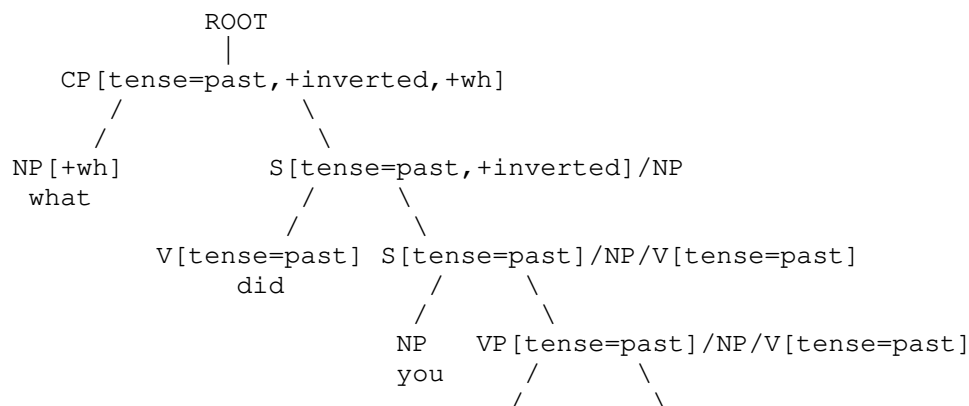
although we have not mentioned it in class: Did Papa eat the spoon along with the caviar?

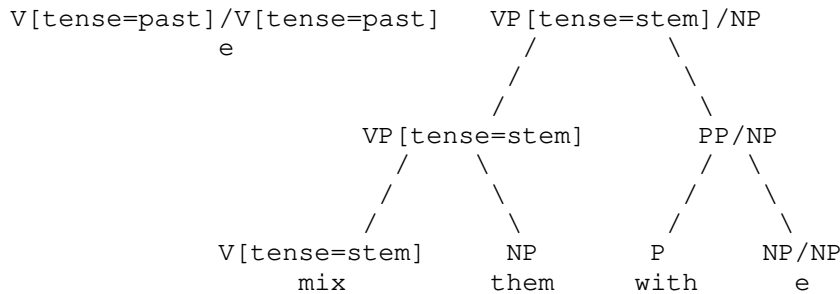
That is, "ate with" could mean "ate together with" or "ate using."

So actually, that sentence is at least 3 ways ambiguous, because "with" could modify either the verb or the noun, and if it modifies the verb, it could have either of two meanings.

- (b) This is a challenging question because we have to figure out how to "move" the auxiliary verb to the start of the sentence, and then "move" the wh-word to the start of the sentence. We'll use gaps /V and /NP to handle these two "movements".

I'll use a common notation where "+wh" is short for "wh=true" (or "wh=yes") and "-wh" is short for "wh=false" (or "wh=no").





Discussion of the above solution:

"you" and "them" are not nouns; they are NPs.

It is not grammatical to say *"Every them in the room is the you."* :-)

The tense of "mix" is stem, not present.

(You might be able to ask "Did you be good?"

but you certainly cannot ask "Did you are good?")

"what" is appropriately labeled as an NP[+wh]. It fills an NP gap, and its position in the sentence could grammatically be filled by any questioning noun phrase ("who," "what," "which shoes from your closet"), but not by questioning phrases that are not noun phrases ("where," "when," "why," "how").

For the slashed categories, just "/NP" will suffice. It is not necessary to specialize that to "/NP[+wh]". For example, "mix them with e" is just "VP[tense=stem]/NP" -- its gap must be filled by an NP, but the NP could be either +wh or -wh:

what did you [mix them with e]	-- filled by NP[+wh]
the salt that you should [mix them with e]	-- filled by NP[-wh]

We use a "+inverted" feature to mark an S that has the main tensed verb preceding the subject

("did you mix them," "are you mixing them," "will you mix them,"  
"have you been mixing them," ...)

rather than following the subject as usual

("you did mix them," "you are mixing them," "you will mix them,"  
"you have been mixing them," ...).

For an inverted sentence, we need to move the main tensed verb of the sentence ("did," "have," "are," "will," ...) to the start of the sentence, turning ordinary sentences like

you did mix them with salt  
you have mixed them with salt  
you are mixing them with salt  
you will mix them with salt  
you should mix them with salt

into inverted versions, which can be used as yes/no questions:

did you e mix them with salt  
have you e mixed them with salt  
are you e mixing them with salt  
will you e mix them with salt  
should you e mix them with salt

We handle this by a rule like

S[tense=alpha,+inverted] -> V[tense=alpha] S/V[tense=alpha]

which effectively moves the tensed V to the front.

An S[+inverted]/NP can be combined with NP[+wh] to make a wh-question.

what [did you e mix them with e]?

REMARK: Non-inverted questions can also be built, both yes-no and wh-. They show up in EMBEDDED sentences:

I wonder [what you have mixed them with].  
 I wonder [whether you have mixed them with salt].  
 instead of the inverted versions  
 \*I wonder [what have you mixed them with].  
 \*I wonder [whether have you mixed them with salt].  
 We saw examples of such embedded questions in class.  
 We label them as CP by analogy with  
 I believe [that you have mixed them with salt].  
 but they must have a +wh feature to indicate that they are  
 appropriate complements for "wonder" rather than for "believe."

CAVEAT: The S[+inverted] story above does have a flaw.  
 It predicts you can always move the tensed verb to the front  
 to get an inverted sentence. But in fact, this is only  
 possibly with auxiliary verbs. You can't do it with  
 ordinary "lexical" verbs:  
     you mixed them with salt  
 cannot be inverted as  
     mixed you e them with salt  
 That kind of movement was allowed in an older form of English,  
 but is not possible in modern English.

How would we improve the story to ensure that only auxiliary  
 verbs can move?

Option 1: Constrain the rule to look at the head of the V that  
 moves, to make sure it's one of the auxiliary verbs {do, be,  
 have, will, should, ...}.

Option 2: Constrain the rules to look at the remaining VP/V,  
 to ensure that it moves an auxiliary verb but leaves behind  
 a VP with another verb in it. This could be accomplished by  
 passing an additional feature down through the S[+inverted]  
 to the VP/V, forcing it to rewrite as V/V VP and not as V/V  
 NP.

### 3a. Real -> Int

Confused? Remember that -> represents the direction of GENERATION  
 (top-down). When you are PARSING (bottom-up), the direction is  
 reversed. So this rule allows you to parse an integer as a real  
 number.

Still confused? Here's a little more of the grammar:

Real -> sqrt ( Real )  
 Real -> 3.0  
 Real -> Int  
 Int -> 3

Do you see now how sqrt(3.0) is generated, and how the Real -> Int  
 rule allows sqrt(3) to be generated as well?

### 3b. Example: "Jimmy is proud, but Mom is taller."

These rules handle cases where a phrase consists of a single word.  
 Other examples of such rules are NP -> Pronoun and  
 VP -> V[intransitive].

### 3c. Example: "Red is a nice color, but cold makes me feel sad."

The point is that adjectives can sometimes be coerced into noun  
 phrases. This changes their meaning so that they can be used  
 in a context that is expecting a noun phrase. (Just as coercing  
 3 into 3.0 changes its bitwise representation.)

For example, in the sentence above, "red" is used to mean "the

color of red things" (a noun phrase), and "cold" is used to mean "coldness" (a noun phrase).

Here are some other good examples that some of you wrote on exams:

"I put the heat on high." (high = "the high setting")  
"Hello, beautiful!" (beautiful = "beautiful person")  
"Would you like cheesy? Yes, cheesy is good."  
(cheesy = "cheesy preparation of this dish")  
"Smart is the new sexy." (smart=smartness, sexy=sexiness)  
"I prefer scary to boring." (scary=scary things, boring=boring things)  
"Billy is a particularly bad shade of grumpy today."

Some of you used the NP → Adj rule to explain predicate adjectives, such as

The boy is happy.

In other words, you analyzed "happy" as an NP direct object -- one that happened to be built from an adjective.

This answer seems weaker to me. First, if "happy" is an NP in "The boy is happy," then it should refer to some \*object\* that is equal to the boy (compare "The boy is my son"). Second, if you can use "happy" as an NP direct object, why can you only use it this way with a few verbs, like "is"/"looks"/"seems"/"feels"/"becomes"? (Compare "The boy likes happy.")

### 3d. There are several possible objections.

Probably it should be Noun → Adj rather than NP → Adj. This would let us use adjectives as nouns within a larger noun phrase: "The extreme cold in Greenland is extremely unpleasant."

In a richer grammar, one might want to use syntactic attributes: Noun[num=mass] → Adj. This indicates that the resulting Noun acts like a mass noun such as "dirt" or "weather." (It agrees with a singular verb, but uses plural determiners, including the option of no determiner at all: "(Most) extreme cold is caused by refrigeration.")

Most important, it is not clear that the rule applies to all adjectives. Some of them turn into nouns more readily than others. "I hate cold" and "Cold is dangerous" sound fine. But can you say "I hate sleepy" and "Sleepy is dangerous"? Or would you have to use the word "sleepiness" instead?

So perhaps "cold" and "red" are each in the lexicon twice, once as a noun and once as an adjective. And "sleepy" is only in the lexicon as an adjective. In this case, there is not a general grammatical ability to rewrite NP → Adj.

### 3e. This rule permits a noun to be used in ANY context that calls for an adjective. But that is not okay. Consider some other contexts that allow only adjectives, not nouns:

Roses are red.  
\*Roses are flower.

Roses are very red.  
\*Roses are very flower.

Roses are the most beautiful plant.  
\*Roses are the most flower plant.

Also notice that adjectives can trade places with each other:  
tall sweet tree

sweet tall tree  
But they can't trade places with nouns:  
tall apple tree  
\*apple tall tree

This is why we usually analyze "flower garden" as a noun-noun compound, not an adjective-noun compound. That is, English allows nouns to modify other nouns directly without being converted to adjectives first. The relevant rule was discussed in class:  
N -> N N

Note: Several of you gave a different objection: "If you could use nouns as adjectives, then you could form nonsensical phrases like 'door chair'."

But how is that an argument against having the rule in the grammar? Saying the phrase is nonsense is an objection to its semantics, not its syntax. "Colorless green ideas sleep furiously" and "Rocks eat honesty" may be nonsense semantically, but they're GRAMMATICAL nonsense and the grammar allows them, just as in Homework 1.

Perhaps you also think that "door chair" is ungrammatical. But isn't it a perfectly ordinary, grammatical, noun-noun compound? (Or is your objection strong enough that you want to get rid of the rule N -> N N, as well?)

And I don't think it's nonsense, by the way. I've never heard of a door chair, but presumably it's a chair that is meant to prop a door open, or a chair that was obtained by converting the wood from an old door, or something. Noun-noun compounds can usually be interpreted if you try hard enough.

3f. Example: "I believe the tomato is a fruit."

The usual rule  
CP -> that S  
would require "I believe that the tomato is a fruit."  
The rule  
CP -> S  
allows the word "that" to be omitted.

3g. The objection is similar to the objection in 1e.  
The word "that" cannot be omitted in all contexts that call for a CP.  
For example, it cannot be omitted in

"That the tomato is a fruit perplexed the president."

since

\*The tomato is a fruit perplexed the president.

is ungrammatical.

Some of you gave a grammatical sentence that involved a CP of a different sort. (Or in some cases, a relative clause such as "that likes me" or "which likes me" -- these are actually CP/NP, not CP.) Those sentences did not happen to use the rule CP -> S. But that doesn't prove that CP -> S is bad. To show that CP -> S is bad, you need to show that including it can give rise to ungrammatical sentences.

4. (a) This is sometimes called "diary speech" or the "diary register," since it is typical of people writing diaries.

The data suggest that Twitter users can omit the subject of a statement. In the examples on the exam, the missing subject is understood semantically to be the person speaking.

Although more generally, the missing subject could be anyone or anything salient -- the kind of salient entity that one would refer to with a pronoun in standard English:

"Took midterm today. Made me happy." (i.e., IT made me happy)  
"Boyfriend came over tonight. Said he was going crazy."  
(i.e., HE said)

The data also suggest that Twitter users can utter just an NP ("Fun problems") or AdjP ("Too exhausted"). This seems very similar to dropping the subject! For in all these cases, the speaker is uttering something that is semantically a predicate, when it is obvious from context what the predicate applies to.

So let's handle all these cases in the same way. The simplest approach is to just add new rules

ROOT -> VP (this is the missing subject case)  
ROOT -> NP (other cases are handled similarly)  
ROOT -> AdjP

We should probably also add

ROOT -> ROOT and ROOT  
ROOT -> ROOT but ROOT

if it wasn't already there, in order to handle "Fun problems, but almost fell asleep."

Common mistakes:

- It's really got to be the subject that's omitted. So please don't write NP -> (e), which would overgenerate by allowing "\*Sleep is good for (e)."
- And it's really got to be omitted from the main (top-level) S. So please don't write S -> VP, which would overgenerate by allowing "\*I think need more sleep next time" (this drops the subject of the embedded S "I need more sleep next time").
- Finally, it's really got to be the WHOLE subject that's omitted. So if you're going to use an empty category for the subject, write it as (e), not as NP/NP. (Remember, an NP/NP wouldn't have to be empty. It can be something like "the crust of" (obtained by removing the embedded NP "the sandwich" from the larger NP "the crust of the sandwich"). But the subject in the Twitter utterances is dropped completely.)

The other phenomenon we see in the examples is that determiners can be dropped. The simplest approach is to just add the rule

DET -> (e)

which says that the DET can be dropped from any kind of noun phrase. An alternative would be to try

NP -> N

which is a modification of the standard rule NP -> Det N that omits the determiner. The problem with this alternative approach is that if there are other NP rules in the grammar, you would have to add determinerless modifications of them, too.

- (b) bias: This measures how wrong the model's predictions are on average. L is biased because it assumes a 5-gram model, whereas actual text comes from a more complex grammatical model that is more like G. But G is also biased because was trained on newspaper text, and expects even Twitter text to look like that.



variance: This measures how sensitive the model's answers are to the particular training sample that it saw. G has only seen a reasonable fraction of what goes on in newspaper text, and it might behave differently given a different million words. But L was trained on so much data that it is likely to have low variance. Once you've trained on a billion words, it doesn't much matter WHICH billion words you saw, particularly if you smooth. By the way, Google has publicly released 5-gram counts based on that much English web text!

runtime: n-gram language models are faster to work with than CFG language models.

- (c) By Bayes' Theorem, choose  $x$  to maximize  $p(x) p(w | x)$ . Here  $p(x)$  is the large language model L, and  $p(w | x)$  is the simplistic noisy channel model.