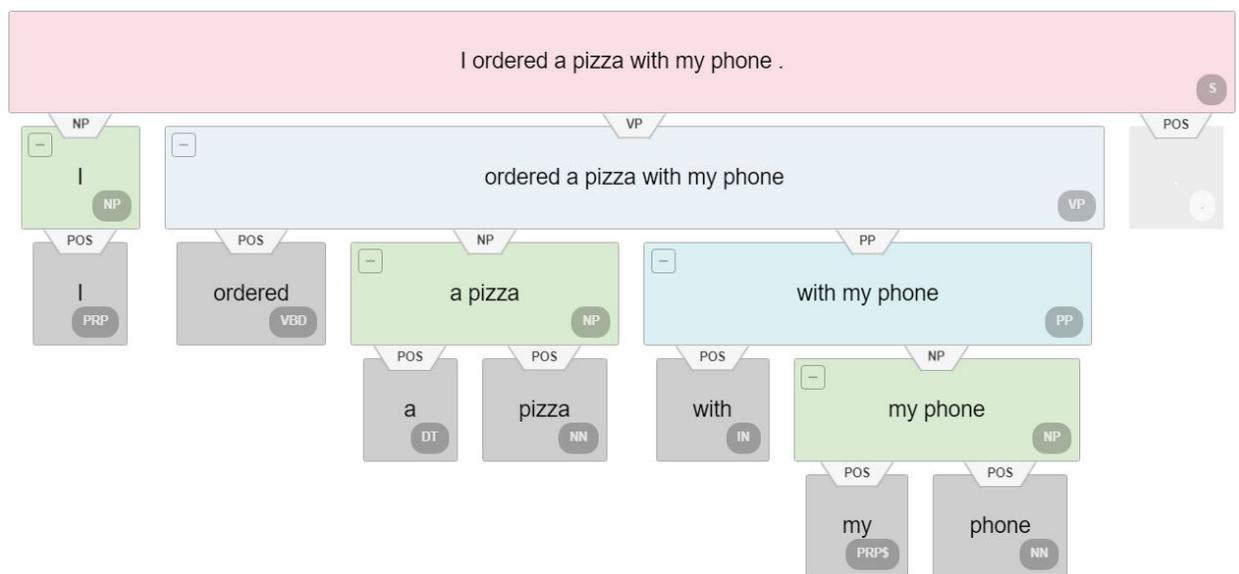


Homework 4: Parsing and Parse Trees

Question 1

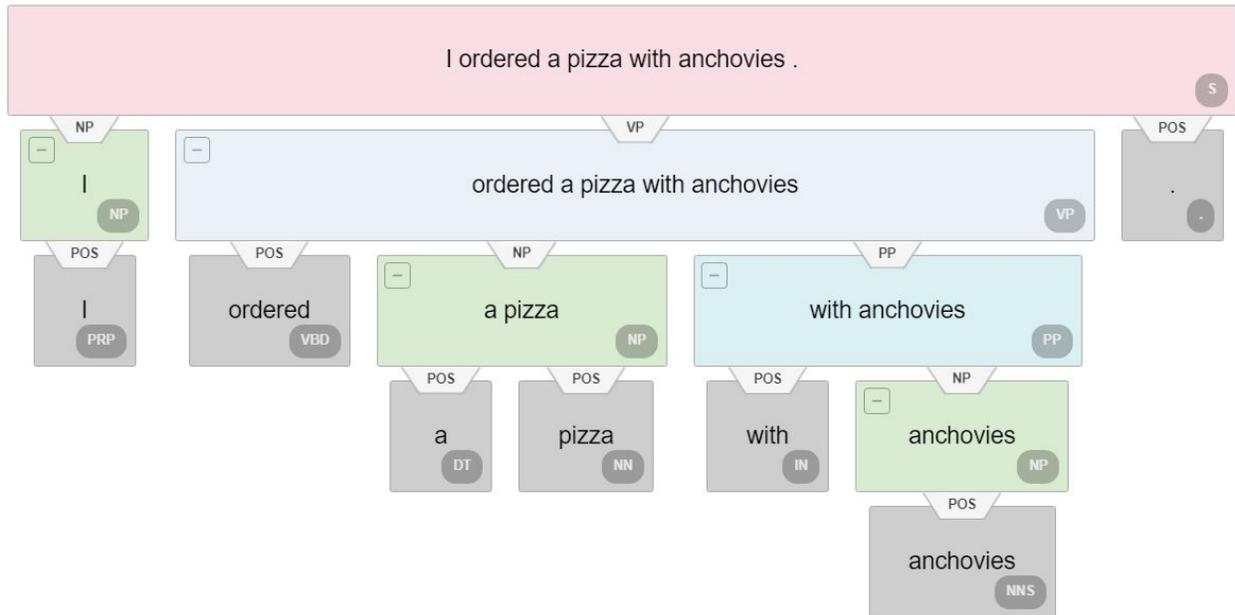
Parse the following sentences:

- a. "I ordered a pizza with my phone."



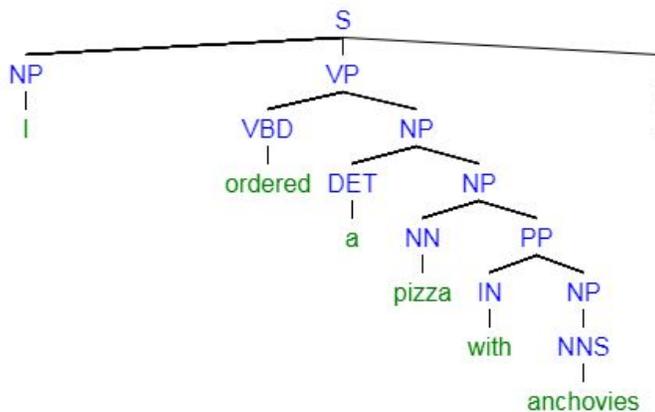
I cannot spot any errors in this parsing. The PP attachment (as a sibling of the VBD verb, "ordered", and semantically as an ARGM-MNR) is correct, and the individual parts of speech all look correct.

B. "I ordered a pizza with anchovies."



Here, we are less fortunate. The parser assumes that this sentence's PP attachment is similar to sentence a, and assumes that "anchovies" are an instrument used in the ordering of the pizza (and a sibling of the verb "ordered" and a direct child of the VP), instead of a PP that should be a child of the NP "a pizza with anchovies" and the sibling of "pizza."

A correct tree would be parsed as follows:



c. "John went to the shop yesterday night."

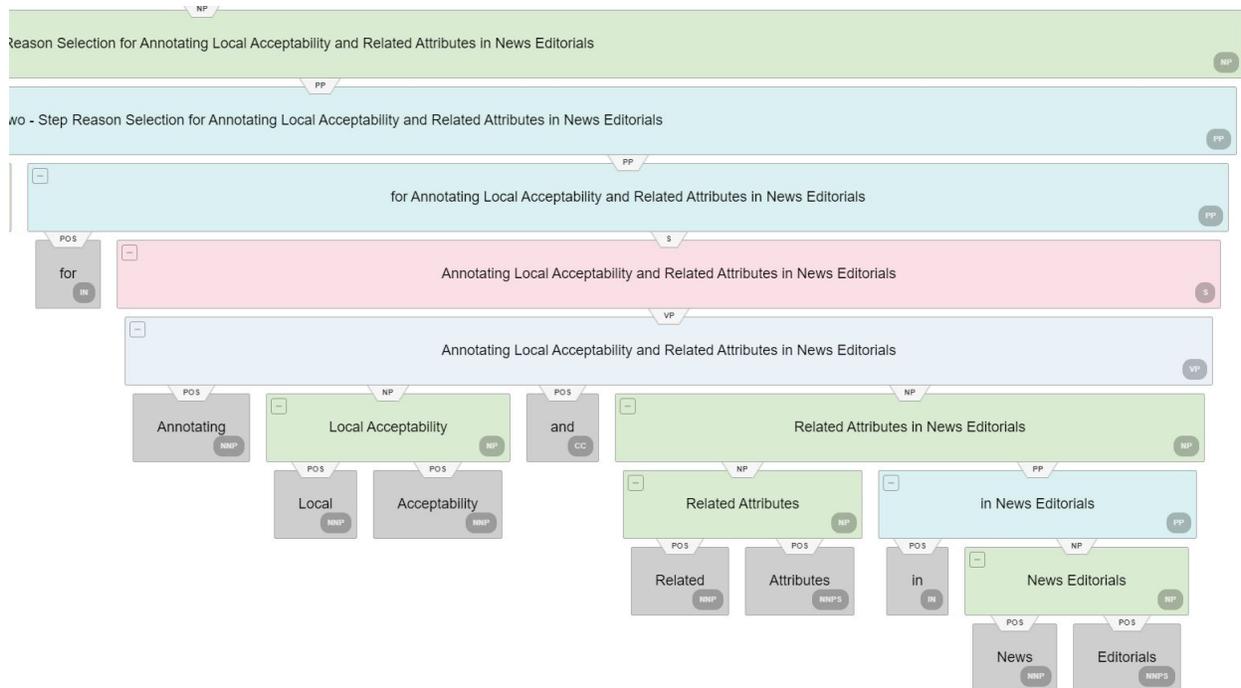


According to how the Penn Treebank usually parses sentences, this sentence seems sound. In Linguistics courses, I would usually tree an NP that is used adverbially (such as "yesterday night," here answering the question of "when") as an AdvP, not an NP, but I have noticed that POS tagging and parsing tagsets usually tag these phrases as NPs, so I'll count this as acceptable. The siblings of the past-tense verb "went" ("to the shop" and "yesterday night") answer core argument questions WHERE and WHEN, respectively.

2. parse Nonsense!: Quality Control via Two-Step Reason Selection for Annotating Local Acceptability and Related Attributes in News Editorials

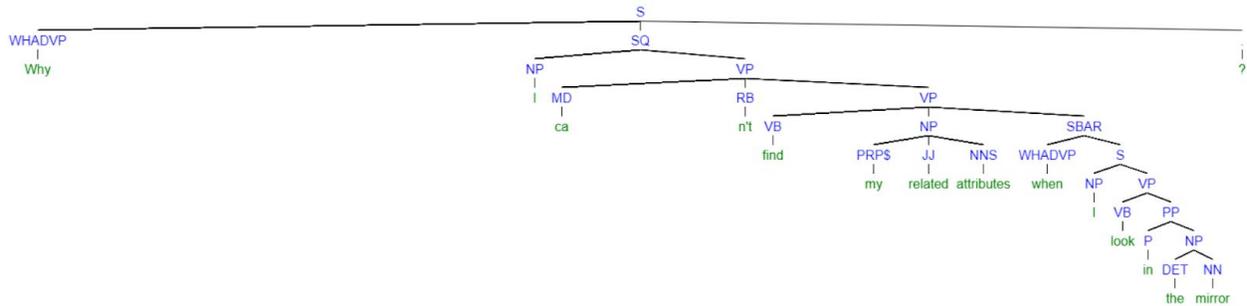
The parse tree for this article title will have to be inserted in two pieces:





The algorithm got more right than expected. The biggest error, in my opinion, was the separation of the hyphenated adjective “Two-Step” into a cardinal number (“Two”), a hyphen, and a supposed noun phrase “Step Reason Selection”; this isn’t a slight error in judgment, but is outright nonsensical, separating naturally related elements of the phrase and grouping them with other, less-related elements. There are several other errors: “Nonsense” here is clearly an interjection (“UH” in the Penn Treebank tagset), but the tagger labels it a noun (“NN”). The tagger does correctly separate “Quality Control” from the rest of the gargantuan NP it belongs to (probably taking this cue from the presence of “via,” a preposition), but tags each element of this NP as a proper noun, likely thrown off by the capital letters where they do not usually exist outside titles. In fact, EVERY noun and adjective (aside from the cardinal number “Two”) falls prey to this bias for capitalized words being NNPs and capitalized words with “s” on the end being NNPS. It is especially strange when the parser tags “Annotating Local Acceptability and Related Attributes in News Editorials” correctly as a VP, and then completely fails to recognize the verb “annotating,” despite the clear VBZ “-ing” ending, because it is capitalized. Prepositional phrases are tagged accurately, likely because the prepositions in titles are not capitalized. A correct parse of this title would look something more like this (if this is too hard to see, please paste

(I reversed the order of the modal and the pronoun so I could parse properly)



If this sentence is too small to read, please paste this code into the website given earlier: [S [WHADVP Why] [SQ [NP I] [VP [MD ca] [RB n't] [VP [VB find] [NP [PRP\$ my] [JJ related] [NNS attributes]]] [SBAR [WHADVP when] [S [NP I] [VP [VB look] [PP [P in] [NP [DET the][NN mirror]]]]]]]]]. ?]

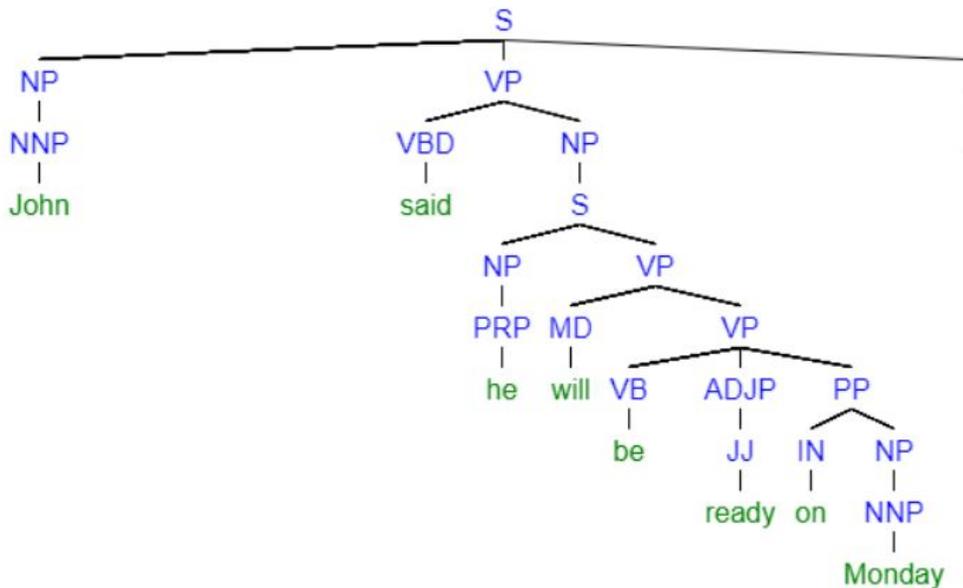
Question 2

Consider the following sentence:

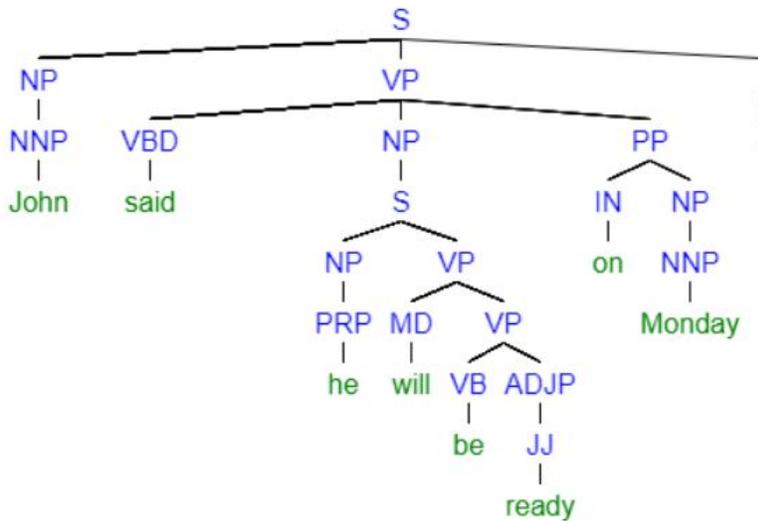
John said he will be ready on Monday.

How many interpretations can you find? Draw the parse tree and spell out what they mean.

Interpretation 1: John stated (in the present) that he (John) will be prepared on Monday (in the future).



Interpretation 2: John stated on Monday (in the past) that he (John) will be ready (at some unknown time).



Interpretation 3: John said the words “he will be ready on Monday.” (Same parse tree structure as Interpretation 1)

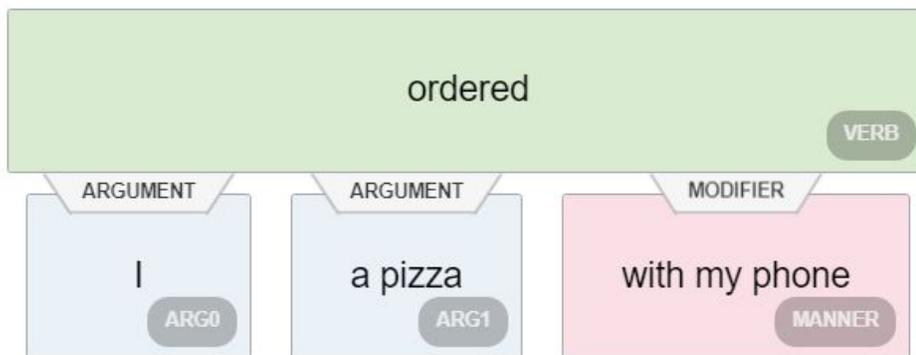
Interpretation 4: John stated (in the present) that he (unknown male referent) will be ready on Monday (in the future). (Same parse tree structure as Interpretation 1)

Interpretation 5: John stated on Monday (in the past) that he (unknown male referent) will be ready (at some unknown time). (Same parse tree structure as Interpretation 2)

Question 3

1. extract the semantic roles of the following sentences (with the tool):

(a) *I ordered a pizza with my phone.*



This seems right! The verb is ordered; “I” is the agent (the orderer of the pizza), “a pizza” is the “what” (the direct object, and a seemingly logical choice for ARG1), and “with my phone” is the way in

which the pizza was ordered. I cannot say whether “ordered” had ARG2 slated for “how” and therefore this role labeling violates the entry for this verb, but “ARGM-MNR” is a perfectly logical and sound choice for this prepositional phrase.

(b) *I ordered a pizza with anchovies.*



Interesting that when the parser didn't get the PP-attachment of this sentence correct, the semantic role labeler does get it correct. The verb, “ordered,” is the same, and we can expect to see the agent/orderer in the ARG0 position (and indeed we do). In the previous sentence, we saw the “what” (in that case, “a pizza”) in the ARG1 position, and since “a pizza with anchovies” is now the “what,” this role labeling is entirely consistent with the last one. I do not detect any errors.

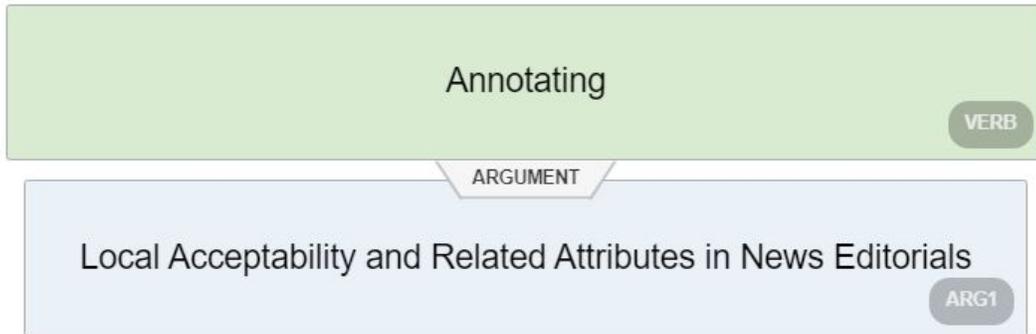
© *John went to the shop yesterday night.*



This is an interesting role labeling. It makes sense that John, the agent, would be ARG0 (this seems to be the case for most verbs - that an agent is always or almost always ARG0). I do not know what the individual arguments are for the verb “went,” but it does seem a little strange that “where,” which seems a very germane argument to a verb that describes movement in the physical plane, would

be as far down as ARG4 (or wouldn't be ARGM-LOC). "Yesterday night" is correctly labeled as ARGM-TEMP.

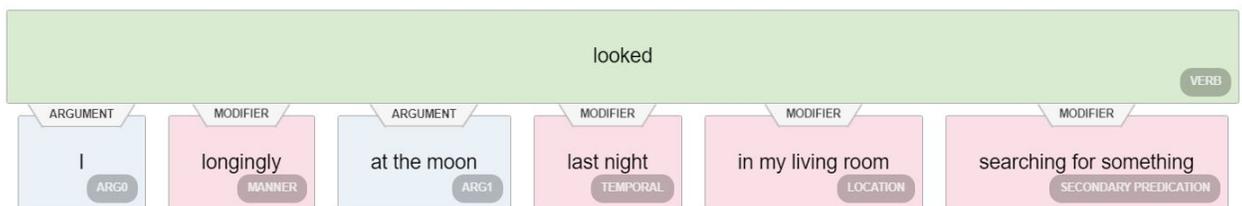
(d) *Annotating Local Acceptability and Related Attributes in News Editorials.*



This is also a fairly straightforward labeling instance. The verb, correctly identified as transitive verb "annotating," would likely take an (absent) agent as an ARG0, and a direct object as an ARG1. The direct object here (describing the "what") is "local acceptability and related attributes in news editorials," and it is especially impressive that the role labeler is able to capture information that the parser does not. It is not "thrown" at all by the titlecase nature of this VP.

2. come up with a sentence that has an argm-loc, argm-tmp and argm-mnr (and the role labeler gets all of them right).

The sentence I chose was *"I looked longingly at the moon last night in my living room, searching for something."*



It might be difficult to read, but the semantic role labeler correctly identifies "longingly" as the manner (ARGM-MNR), "last night," as the temporal argument (ARGM-TMP), and "in my living room" as the location (ARGM-LOC). It also identifies "at the moon" as the ARG1 ("what") and tags the nested verb "searching" as another verb for which to identify roles (those roles, when checked, are also correct: "I" is the ARG0, or agent, and "something" is the ARG1, or direct object - the "what"). I tailored this sentence to be easy for the labeler to label: an adverb is likely to be an ARGM-MNR argument, and I gave this one ("longingly") standard adverbial morphology. "Last night" is a very common temporal expression, and the LOC argument is

initialized by a very standard preposition of location, "in." I would have been surprised if this sentence were not correctly labeled.