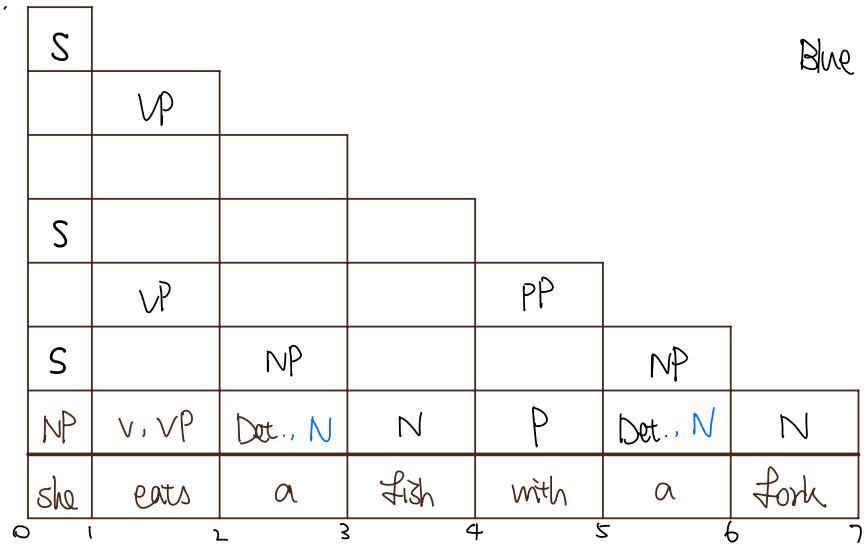
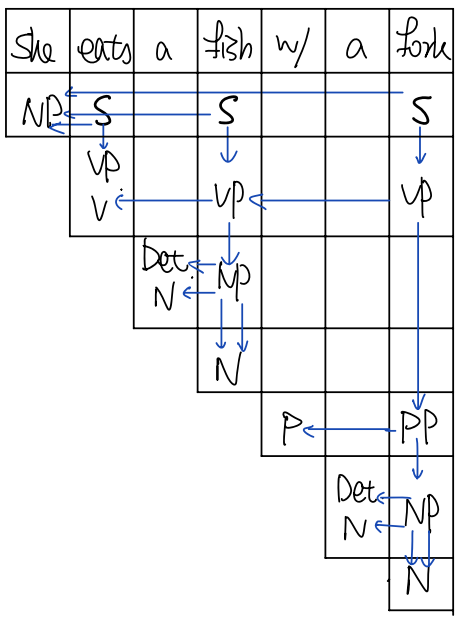
1. a. i.



Blue ones are added

ii. To make it cleaner, an upper-triangle is used.
It's the same thing, it's just cleaner.



Black parts are a(i)
Blue parts are parsing.

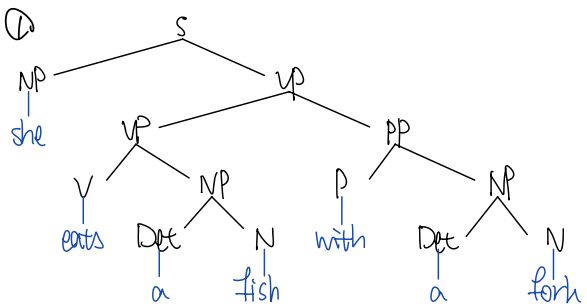
b 4.

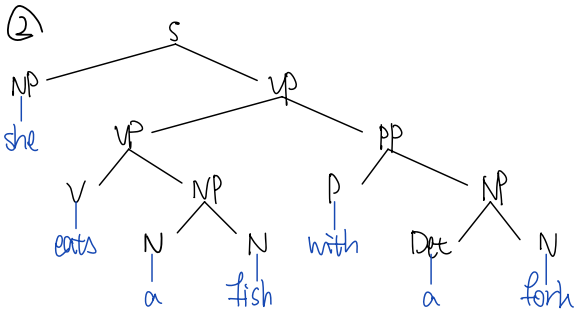
① S(NP she)(VP(VP(V eats)(NP(Det a)(N fish)))(PP(P with)(NP(Det a)(N fork))))

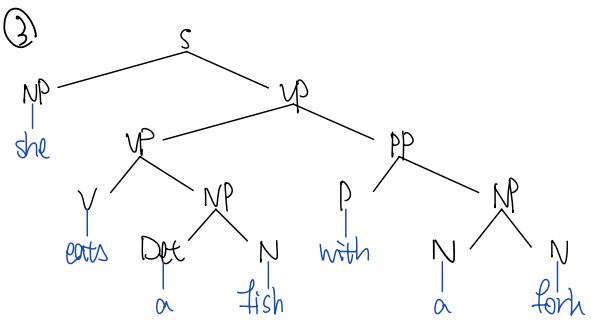
② S(NP she)(VP(VP(V eats)(NP(N a)(N fish)))(PP(P with)(NP(Det a)(N fork))))

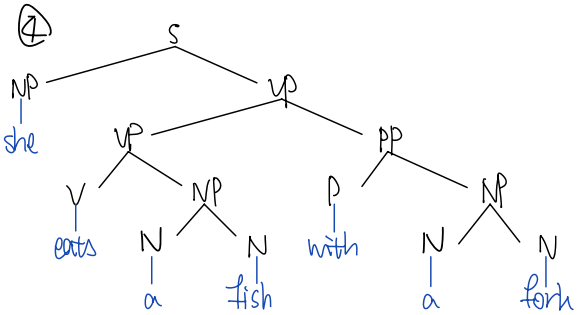
③ S(NP she)(VP(VP(V eats)(NP(Det a)(N fish)))(PP(P with)(NP(N a)(N fork))))

④ S(NP she)(VP(VP(V eats)(NP(N a)(N fish)))(PP(P with)(NP(N a)(N fork))))

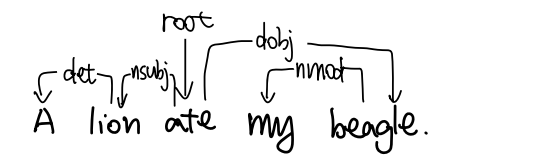


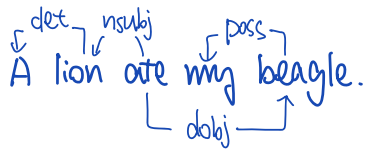


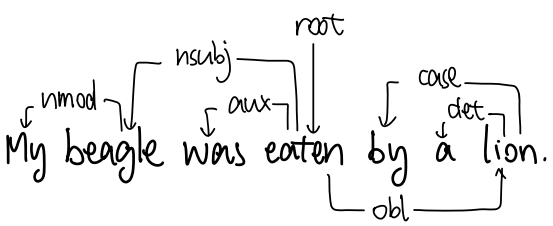


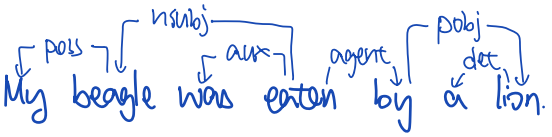


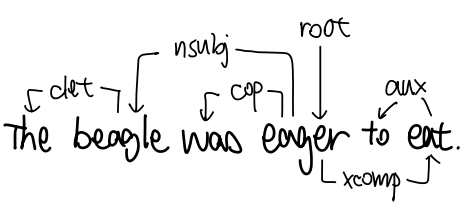
2. a.

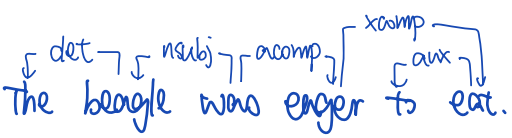


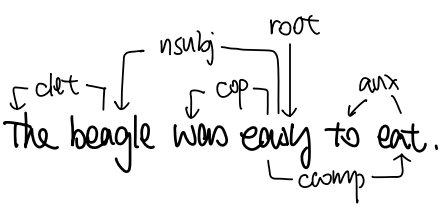


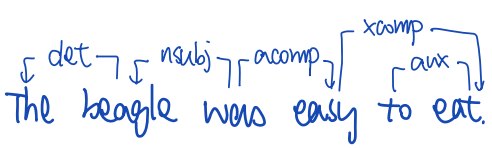












Black part is manual marking, blue part is spaCy.

b. Parser from spaCy isn't able to parse the sentence based on its meaning. By looking at the last two sentences, it can be notice that there is arbitrary explanation of "easy", which can lead to misunderstanding by the parser. Personally, I prefer the meaning of "not hard", which make the last sentence "It is easy to eat a beagle". However, it can also be explained as "The beagle is relaxed when eating", and that led to the parsing by the spaCy. In general, under such circumstance, the sentences would be parsed into the same structure, even though their specific meanings allow them to be parsed in different ways.

In addition, the choice of root is different. (But I think it's me that didn't choose the right one.)

c.

Issues:

1. Figure out a way to filter out irrelevant news articles.
2. SpaCy might not be able to understand the context. This problem would be more severe when dealing with the figurative languages used in news reports.
3. It's hard to filter out news like "attempted to eat the pet, but it got out".

Solutions:

1. Develop a filter to keep only the relevant ones, and then exclude those that "escaped"
2. Add context-aware feature into parsing model