Sangmi Yun

LING 472 A

5 / 1 / 2022

**Part 1**

**Constituency Parsing**

a. Kim adores snow in Oslo.

b. The PCFG

| Rules | | Lexicon | | Numerical indices of non-terminals* |
|---|---|---|---|---|
| S → NP VP | [.80] | NP→Waikiki | [.10] | 1 S |
| S → Aux S | [.20] | NP→Oslo | [.10] | 2 VP |
| VP → V S | [.25] | NP→Kim | [.10] | 3 NP |
| VP → V NP VP → VP | [.45] | NP→snow | [.10] | 4 PP |
| PP | [.30] | V→adores | [.70] | 5 Det |
| NP → Det N | [.40] | V→snores | [.30] | 6 N |
| NP → NP PP | [.20] | Aux→does | [.20] | 7 P |
| PP → P NP PP → P S | [.90] [.10] | Aux→can | [.25] | 8 Aux |
| | | Aux→is | [.55] | 9 V |
| | | P→in | [.20] | |
| | | P→on | [.40] | |
| | | P → before | [.40] | |
| | | Det → this | [.20] | |
| | | Det → these | [.20] | |
| | | Det → the | [.60] | |

| | 1 Kim | 2 adores | 3 snow | 4 in | 5 Oslo. |
|---|---|---|---|---|---|
| 1 | NP: .10 | | S:.80<br>p= 0.00252 | | S: .80<br>p=0.00000907<br>S: .80<br>p=0.0000136<br>S: .80<br>p=0.00252<br>NP: .20<br>p=0.0036 |
| 2 | | V: .70 | VP: .45<br>p= 0.0315 | | VP: .45<br>p=0.0001134<br>VP: .30<br>p=0.0001701<br>VP: .45<br>p=0.0315 |
| 3 | | | NP: .10 | | NP: .20<br>p=0.00036 |
| 4 | | | | P: .20 | PP: .90<br>p=0.018 |
| 5 | | | | | NP: .10 |

2 possibility if the meaning of the sentence is

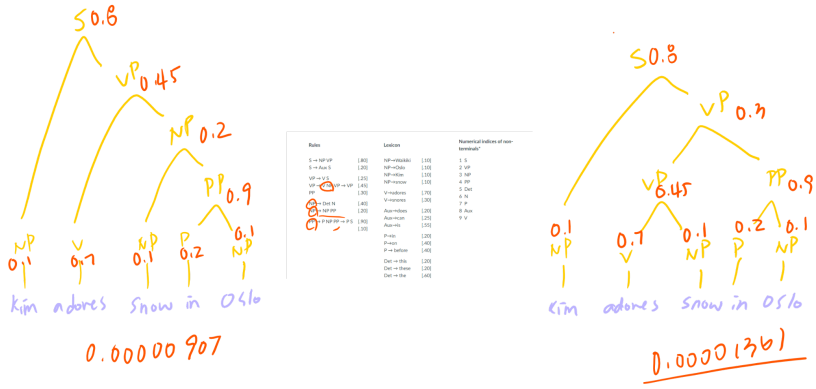
1. [[Kim] [[adores snow] [in Oslo]]] → adores snow/ in Oslo. → p = 0.0000136
2. [[Kim] [[adores] [snow in Oslo]] → adores/ snow in Oslo. → p = 0.00000907

According to the information that we've got from the chart, the 2nd one's PCFG is lower than the 1st possible meaning, so the 2nd which is [[Kim] [[adores] [snow in Oslo]] is a true possibility.

0.00000907 < 0.0000136

<u>Reflection</u>

It was an interesting activity that we can do with the syntax tree. Before I fill out the numbers in the chart I drew a syntax tree to see what can be a possible answer. Since I'm more familiar with the syntax tree it was a fun activity to do.

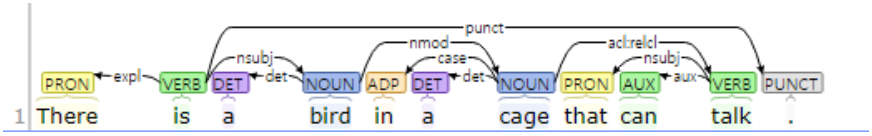



Here I also got the same answer for the PCFG number. While I was learning syntax, when we are drawing a tree of ambiguous sentences we didn't think about which one is a true possibility or not. I think the concept that we can number them and calculate to know which way of translating is true or false. During this CKY algorithm, I could learn more about the relationships between computational linguistics and syntax. I agree that the CKY process is efficient from a computational linguistic point of view and it seems easy to know the connection between the words. However, according to my experience, I had a bit confused while I was filling in the blank because there is too much information going around and it easily gets lost while tracking down. But in the syntax tree, it only has a needs information on the tree and it is easier to track down the numbers and where the PP connected with.

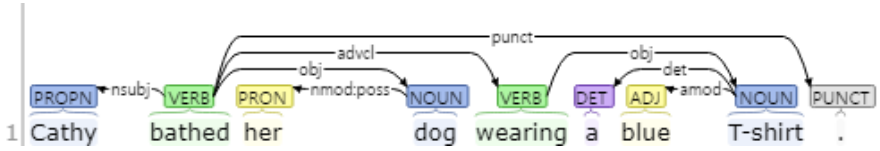
**Part 2**

1. Errors happen in the Ambiguous sentences.

   There is a bird in a cage that can talk.



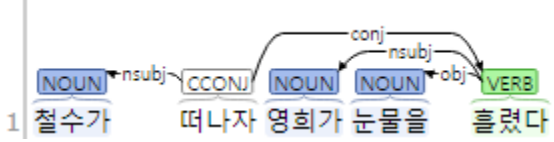

   Cathy bathed her dog wearing a blue T-shirt.




   I put an English sentence “There is a bird in a cage that can talk” to the Stanza. According to this analysis, we can see the complementizer phrase [that can talk] is modifying [a cage]. So it means the cage can talk. This sentence also can have a different meaning by grouping [a bird in a cage] [that can talk]. This gives the meaning that the bird is talking. Both translations could be possible, however, in this case, [that can talk] should modify [a bird in a cage] according to the meaning and possibility of the sentence. I think this error happened because this sentence is an ambiguous sentence. For the correct description. These types of sentences are more difficult for the parser because they don’t have the ability to recognize where complementizer phrases should modify. And distinguish which meaning is more acceptable to the people. We can conclude that understanding the context and the meaning of the language is one of its limitations of Stanza.

2. Errors in Combined sentences.

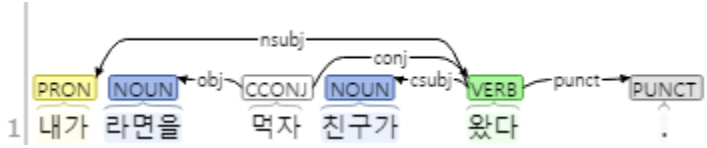
   철수가 떠나자 영희가 눈물을 흘렸다.



cheolsuga tteonaja yeonghuiga nunmul-eul heullyeossda.

When Cheolsu left, Yeonghui started to cry.

내가 라면을 먹자 친구가 왔다.




naega lamyeon-eul meogja chinguga wassda.

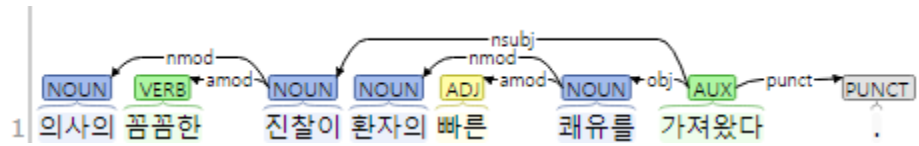
When I was eating ramen, a friend came.

I put the Korean sentence “철수가 떠나자 영희가 눈물을 흘렸다” into the Stanza. When we look at the analysis, there are some errors in the verbs. This sentence is a complex sentence that is combined of the two sentences “철수가 떠나자 (cheolsuga tteonaja)” and “영희가 눈물을 흘렸다(yeonghuiga nunmul-eul heullyeossda).” In the analysis by Stanza ‘떠나자’ is expressed as a conjunction. However, ‘떠나자’ is a verb for the noun ‘철수가.’ In this case I assume that when the word has an end suffix “-다” Stanza is regarded as a main verb. One of the rules for the conjunction is that both of the sides that are connected with the conjunct should be the same. For example, SV conj SV.

When we translate this sentence into English, it's going to be “When Cheolsu left, Yeonghui started to cry.” With this information, we can assume ‘떠나자’ has a combining meaning of ‘when’ and ‘left,’ Thus, in my opinion, this error happened because Stanza didn’t consider the meaning of the main verb ‘떠나다(left)’ and confused when the sentences are connected each other. Eventually, it created an error in the parser that ignored the other features of the word which are important.

3. Errors because of confusion with the AUX verb with the main verb.

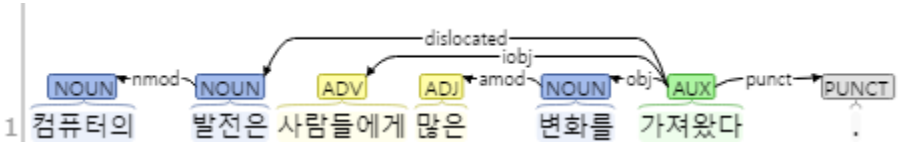
의사의 꼼꼼한 진찰이 환자의 빠른 쾌유를 가져왔다.



uisaui jinchal-i deo ppalli neujchwojinda.

The doctor’s careful examination brought a quick cure to the patient

컴퓨터의 발전은 사람들에게 많은 변화를 가져왔다.



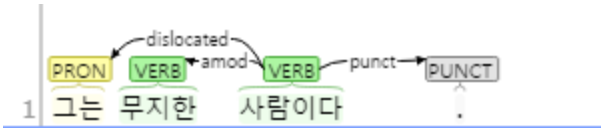

keompyuteoui baljeon-eun salamdeul-ege manh-eun byeonhwaleul gajyeowassda.

The development of computers has brought many changes to people.

In the Stanza, I put the Korean sentence “의사의 꼼꼼한 진찰이 환자의 빠른 쾌유를 가져왔다(uisaui jinchal-i deo ppalli neujchwojinda.).” When we translate this sentence into English, it becomes “The doctor’s careful examination brought a quick cure to the patient.” With the translated version, we can assume the main verb is “brought.” However, in the analysis above, the main verb is “careful” not “brought.” “꼼꼼한” is an adjective, a word merged with the derivational suffix “-한.” This suffix changes the verb to the adjective, which in this case is “꼼꼼한.” This adjective modifies the noun “진찰이(examination).” According to these reasons, we now know the errors were made because they considered the main verb “가져왔다” as an auxiliary verb. Normally “가져왔다” in Korean can be translated into various meanings, but one of the forms is “have”. In this case, it contains the meaning “bring.” However, Stanza regarded “가져왔다” as an auxiliary verb. In my opinion, this error happened because they are analyzing the meaning of the word, structured sentence formation. So, even though words are delivering different meanings to the context of the sentence it is hard for the parser to realize the error. Also, with this example, I assume that they are more reliant on the English meaning of the word.

4. Errors on the suffix that changes the verb to the adjective. (derivational suffix)

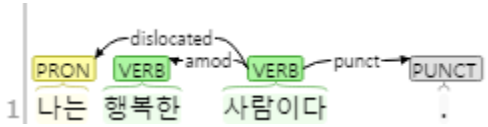
그는 무지한 사람이다.




geuneun mujihan salam-ida.

He is ignorant.

나는 행복한 사람이다.




naneun haengboghan salam-ida.

I am a happy person.

I put the Korean sentences “그는 무지한 사람이다” and “나는 행복한 사람이다.” into Stanza. According to Stanza’s analysis of sentences, they have two verbs. When we look at the sentence “그는 행복한 사람이다.” It identified “행복한” and “사람이다” as a verb. However, “행복한” is a combined word made with “행복하다(happy)” and “-한” is a derivational suffix that changes the verbs to adjectives. I think this error happens because when we are separating the word “행복한” as a part it's going to be “행복 하 ㄴ.” The verb form of happy is “행복하다” and we add “-ㄴ” to the verb to make an adjective. Because of this formation structure Stanza might be thought “행복하다” is a verb, not an adjective. This type of error can easily happen in the parser cause when they are writing the relationship between the sentence, they started with identifying the categories of the words. So when they already think “행복한” as a verb even though there is a main verb next to it, they might not be able to realize the error.