

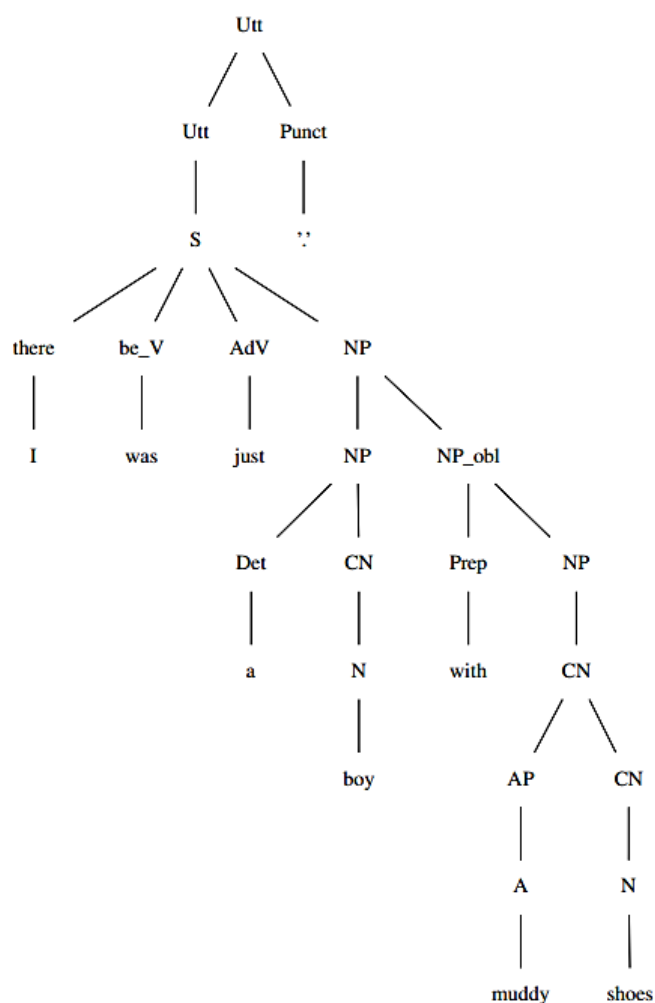
Phrase Structure Trees; Hand-drawn vs Generated

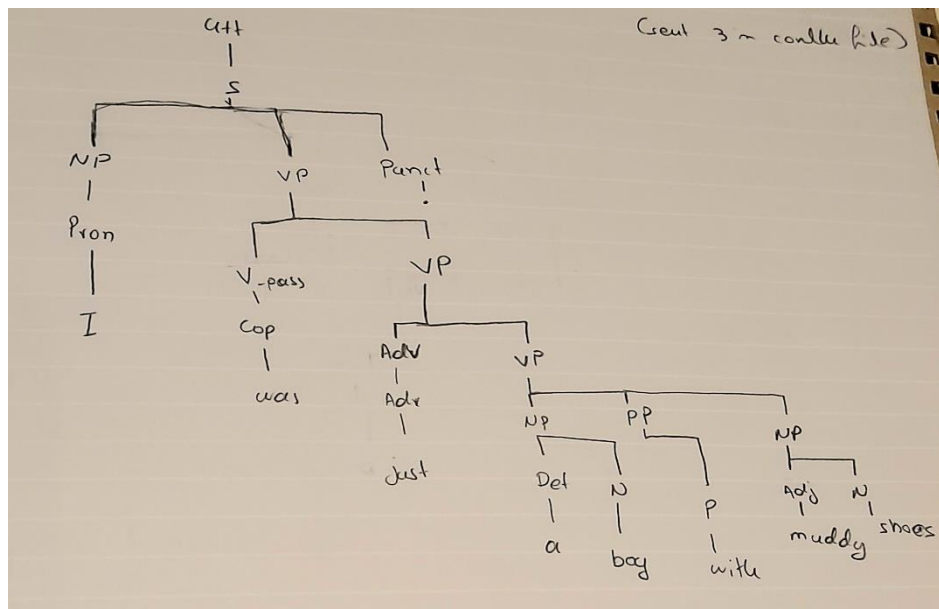
For this part of the assignment, I decided to go with 4 sentences for both English and Spanish. After drawing the phrase trees by hand, I generated them using the command line – and more specifically the command: `cat EnglishSentences.conllu | gfud extract-pos-words | gfud dbnf English.dbnf Utt` for English and `cat SpanishSentences.conllu | gfud extract-pos-words | gfud dbnf Spanish.dbnf Utt` for Spanish. Below, I am presenting the 4 trees for both languages, as well as comments on how the generated trees differ from the handmade ones.

🚩 English:

Sentence 1: I was just a boy with muddy shoes.

(sentence 3 in Conllu file and in EnglishSentences.pdf)

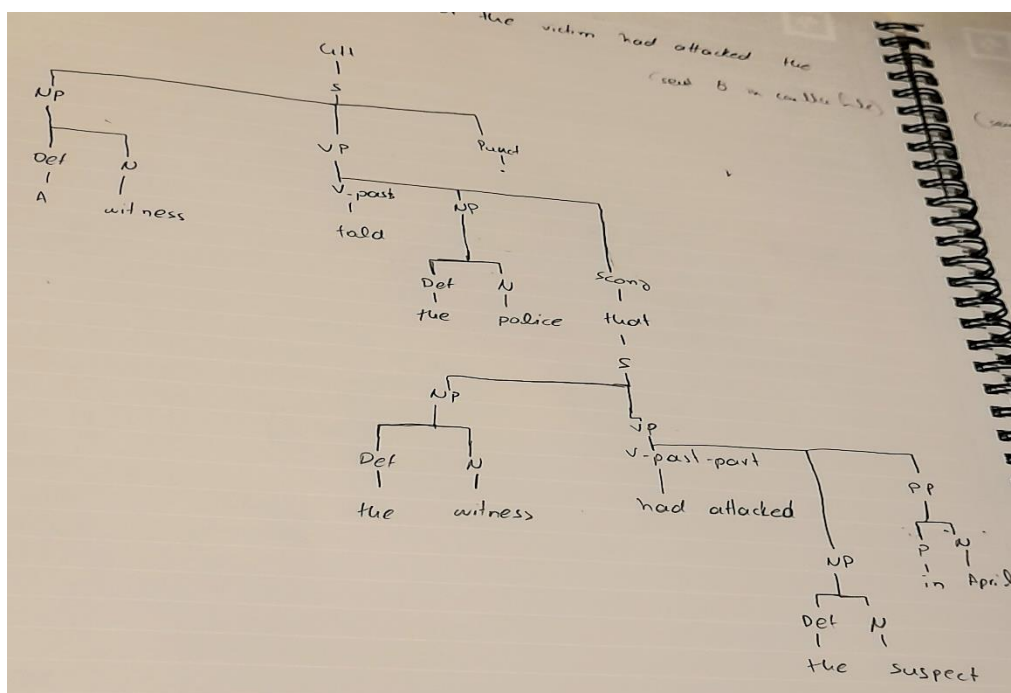


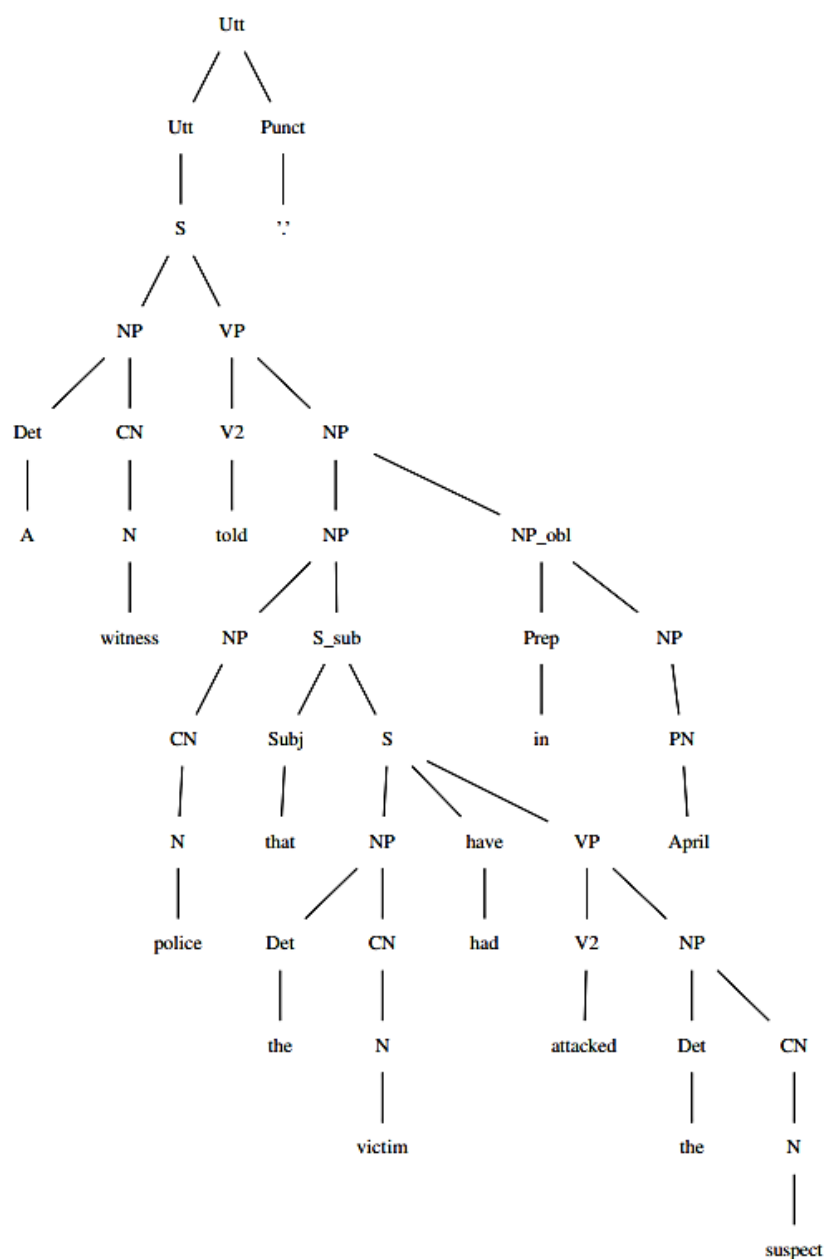


The two trees differ in the grouping placements. In more detail, the generated tree divide the sentence in 4 different groupings (there, be_V, Adv and NP), whereas the hand-made one divides the sentence in 2 (NP and VP). The punctuation in the generated tree is dependent to the final utterance (Utt), while in the hand-made tree it is a part of the sentence (S). Moreover, the handmade tree seems to have a “extra” group below NP, CN (common noun phrase), while in the hand-made tree we don’t see that.

Sentence 2: A witness told police that the victim had attacked the suspect in April.

(sentence 8 in Conllu file and in EnglishSentences.pdf)

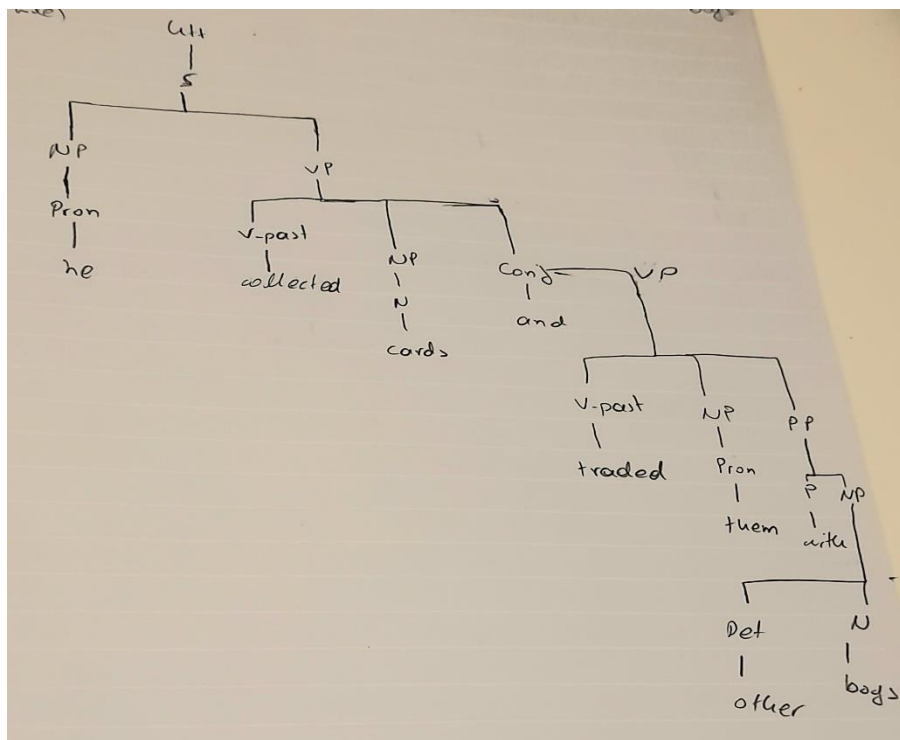
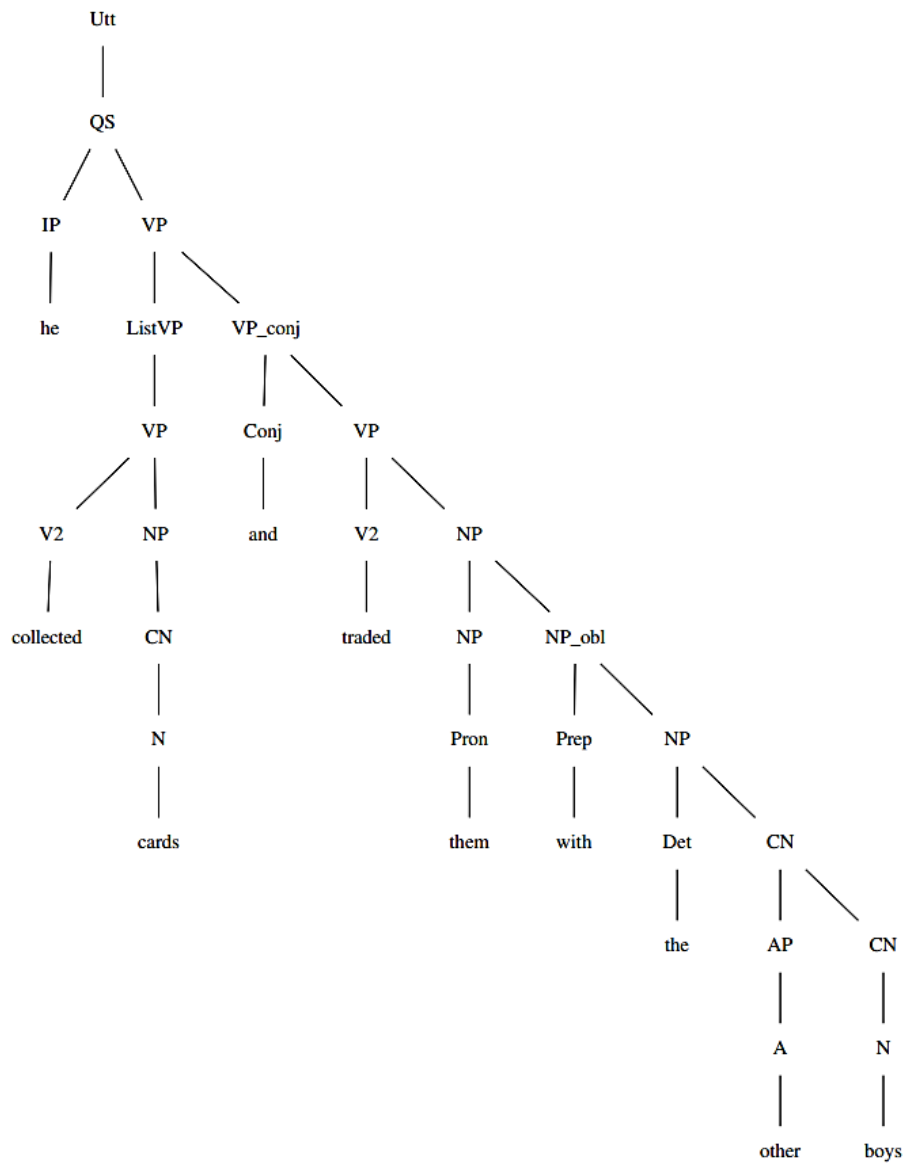




For the second sentence, the main differences are again in the placement of the group parts. This time the sentence is divided in both in 2 (NP and VP) with the only difference being that the punctuation is a part of the Utterance (Utt) in the generated tree and part of the sentence (S) in the handmade one. A second sentence is a part of the VP in both. However, it belongs to a different category in the generated tree (S_sub), while in the hand-drawn it is part of the sconj part “that”. Also, the phrase “in April”, is divided differently, under differently titled groups and it’s part of the main NP in the generated tree, while it’s part of the second Sentence in the handmade one.

Sentence 3: he collected cards and traded them with the other boys

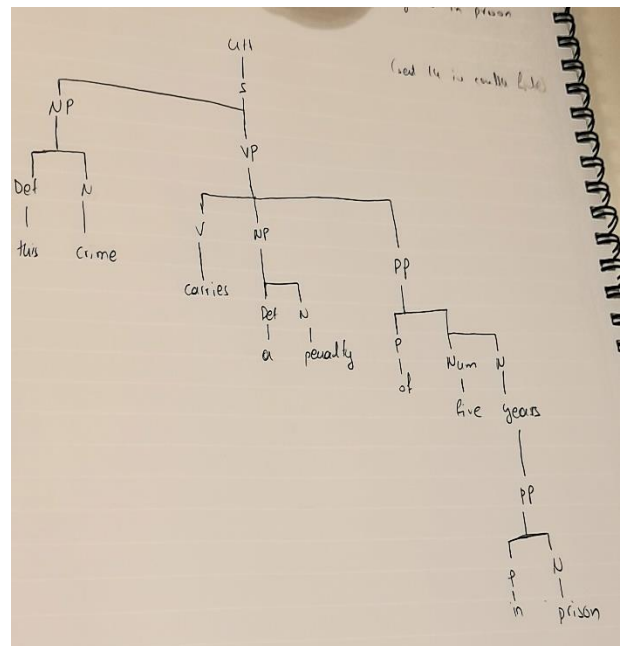
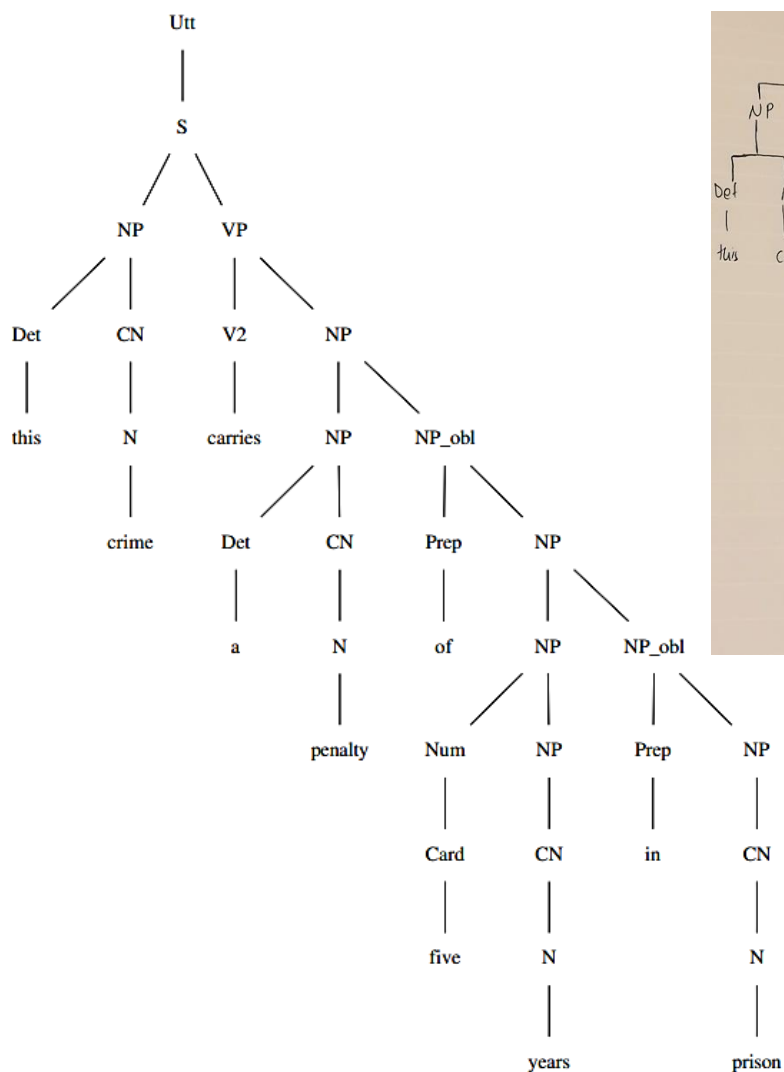
(sentence 13 in Conllu file and in EnglishSentences.pdf)



Just like the previous 2 sentences, the main difference with this one is the grouping placements. In the generated tree, we have a QS divided in IP and VP, while in the hand-drawn one, we have a S divided in NP and VP. I think there aren't any differences between the trees other than the grouping names/labels. The generated one seems to be more analytical, but I would say they follow the same structure.

Sentence 4: this crime carries a penalty of five years in prison

(sentence 14 in Conllu file and in EnglishSentences.pdf)

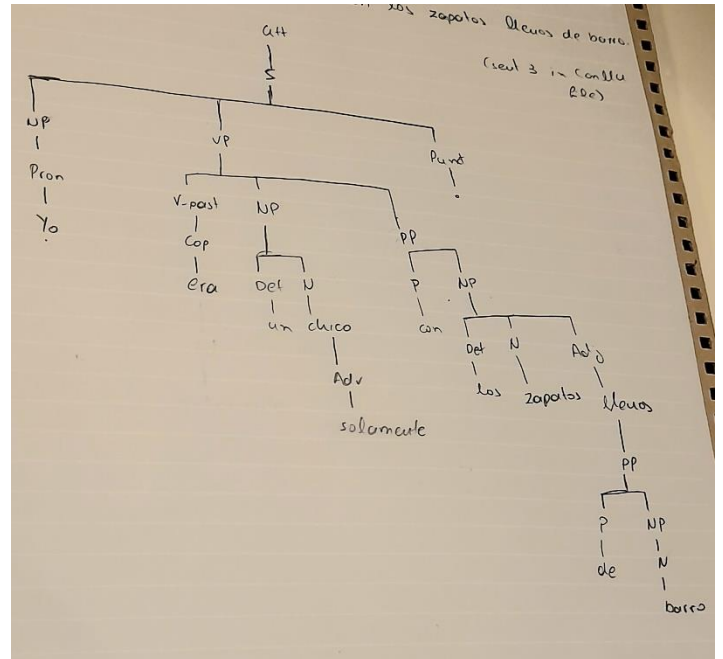
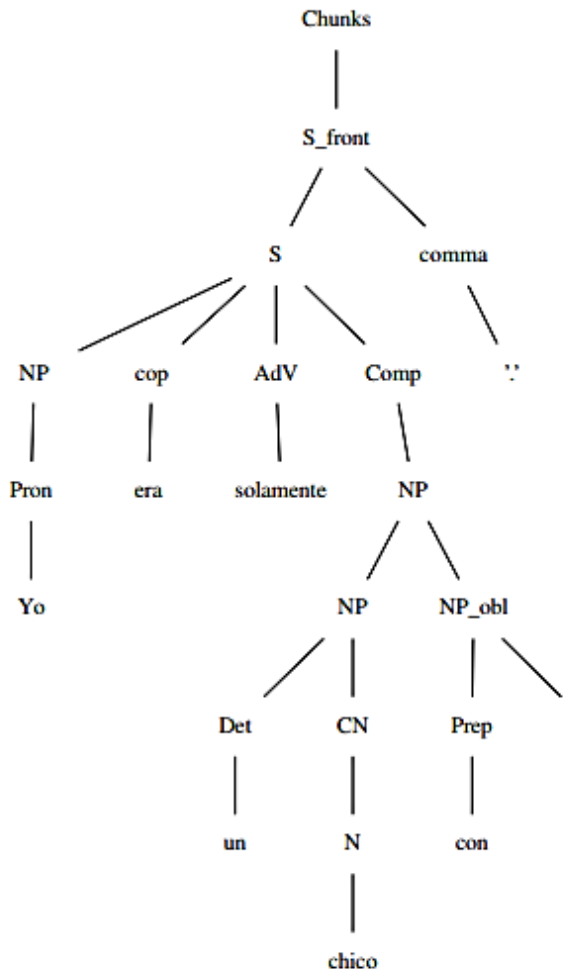


For the last sentence, I would say that the only main difference is the grouping in the last part ("in prison"). In the generated tree, the phrase "in prison" is part of the NP_obl, while in the hand-made one it's part of the NP and the Noun ("five"). Other than that I would say that there are some minor differences in the labeling, since the generated tree seems to be more analytical.

Spanish:

Sentence 1: Yo era solamente un chico con los zapatos llenos de barro.

(sentence 3 in Conllu file and in SpanishSentences.pdf)

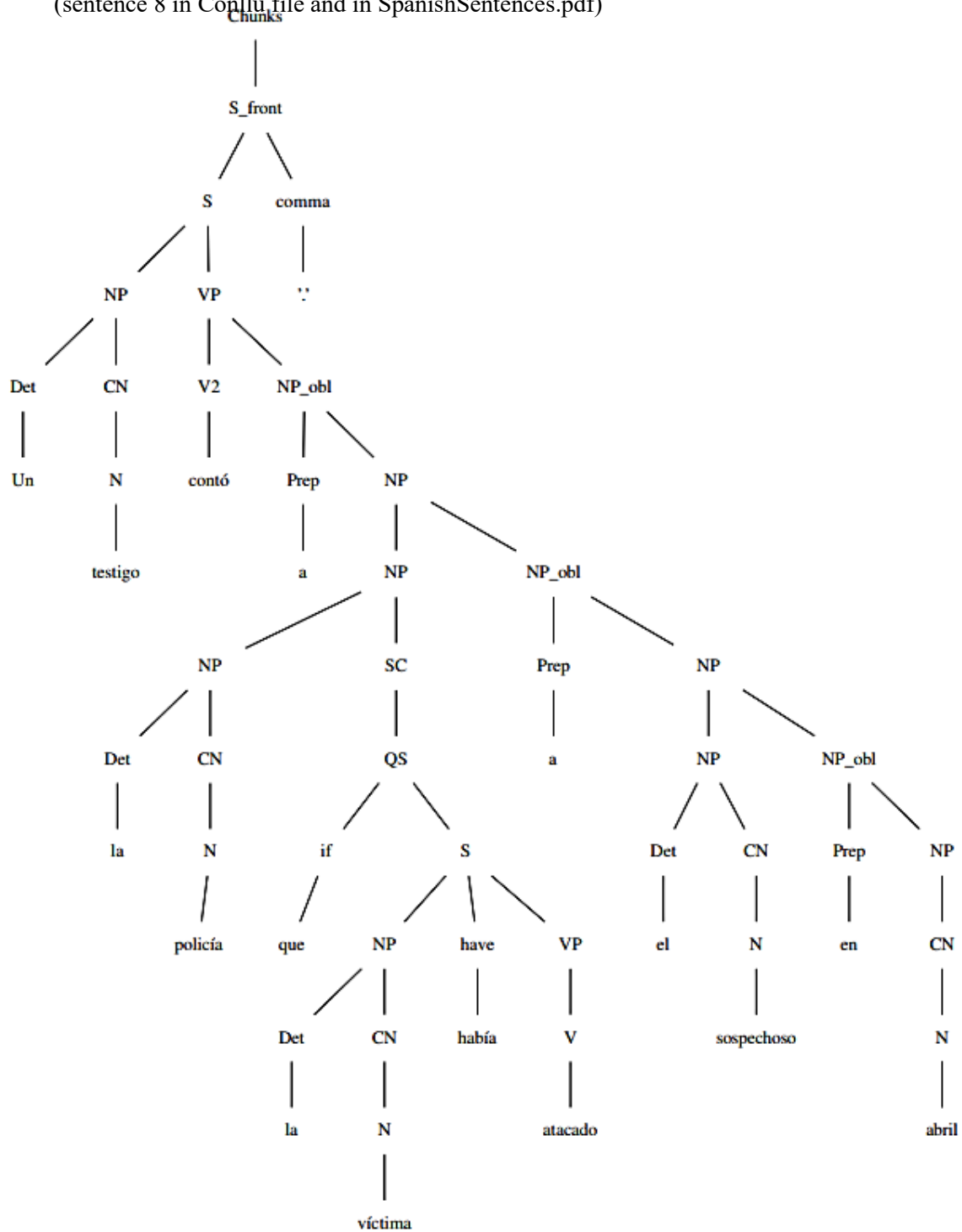


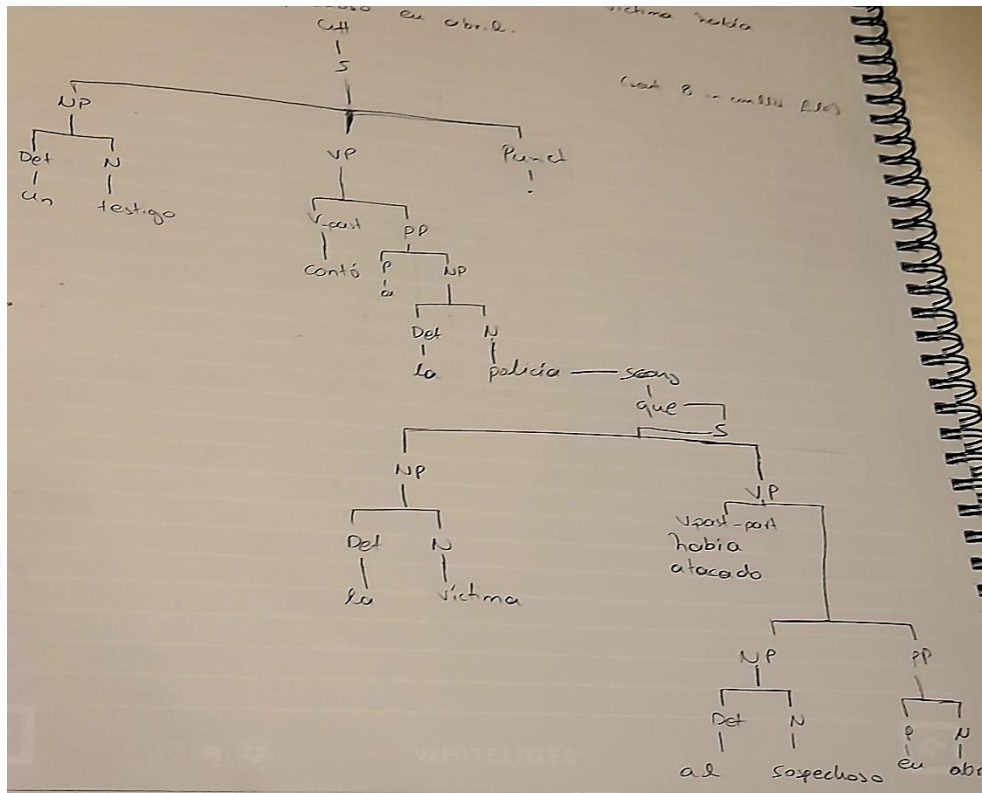
Just like in the English trees, the main differences in the Spanish ones are the placements and labels. The sentence is divided in 4 parts in the generated one, whereas it's divided in 2 parts in the hand-made one. Also, the punctuation isn't part of the sentence

in the generated tree.

Sentence 2: Un testigo contó a la policía que la víctima había atacado al sospechoso en abril.

(sentence 8 in Conllu file and in SpanishSentences.pdf)



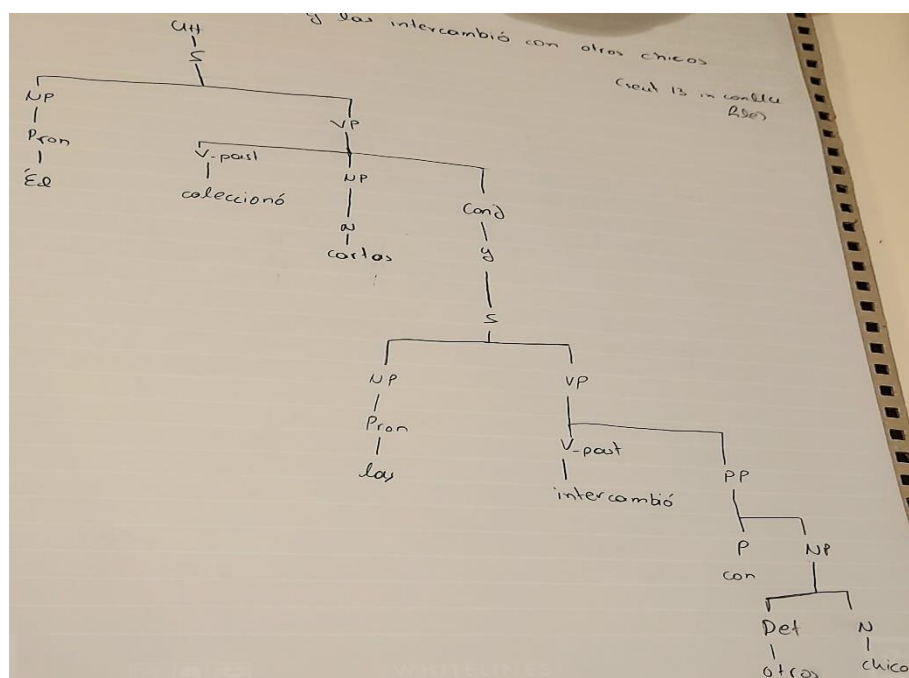


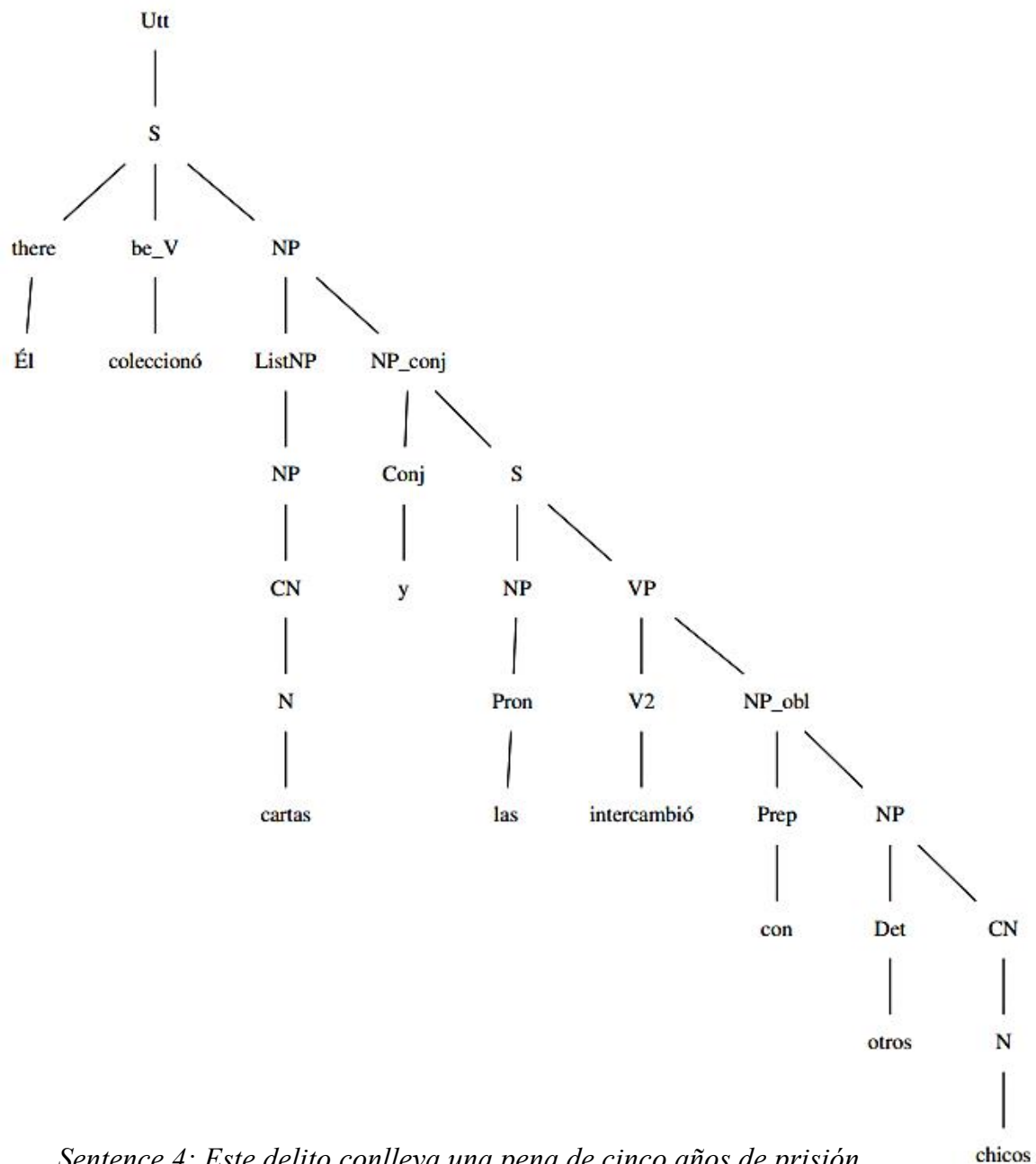
The second sentence in Spanish seems to have the same analysis and structure as the 2nd sentence in English, therefore the differences are the same as the one mentioned in the English version.

Sentence 3: *Él coleccionó cartas y las intercambió con otros chicos*

(sentence 13 in Conllu file and in SpanishSentences.pdf)

The third sentence seems to also have the same structure as the English pair. The generated one seems to be more analytical (e.g. CN label doesn't exist in the handmade tree), but I would say they follow the same structure.





Sentence 4: Este delito conlleva una pena de cinco años de prisión

(sentence 14 in Conllu file and in SpanishSentences.pdf)

Finally, for the last sentence, just like the English one, the main difference is the placement of the phrase “de cinco años”, which is part of the NP in the generated tree, while it’s a part of the VP in the handmade one, as well as the labeling of the phrases and their elements.

