Grammar Adaptations for the Spanish Language

To generate the structure trees using the command line for the English phrases, I used the English.dbnf file – a file that includes the grammatical phrase structure for the English Language. However, in order for me to be able to generate phrase trees for the Spanish Language using the command line, I had to make some adjustments to the file and, while having the English one as the main base, create a new one for my chosen language, titled Spanish.dbnf.

In more detail, I, first, translated the tokens in the content and structure lexicon from English to Spanish (Nouns, Adjectives, Verbs, Adverbs, etc.). For instance, Nouns like *#token N house president moon Earth girl girls wine beer* changed to *#token N casa presidente luna Tierra niña niñas vino cerveza.* And verbs like *#token V sleeps sleep slept walk walks walked sing sings sang sung move moves moved happen happens happened leave leaves left run runs ran* changed to *#token V duerme dormir dormido caminar camina caminado cantar canta cantado mover mueve movido pasar pasa pasado dejar deja dejado correr corre corrido.* In the structure lexicon – besides translating the tokens (English negation: #token neg not n't vs Spanish negation*: #token neg no*) – I added more punctuation symbols under #token punct (English: #token Punct . ! ? vs Spanish: #token Punct . ! ? ¡ ¿ , ; : -), in order to capture the Spanish language more effectively (e.g. *"Who are they?"* vs "¿Quiénes son?"). Moreover, I changed the structure for the conjunction phrases, since Spanish tends to have more variety on how sentences can be joined:

English

#token Conj_pre both either neither

If you notice there in no specific structure for the English Language. Spanish on the other hand have a "stricter" structure for conjunction phrases. Thus, we need specific phrase rules like these ones →

Spanish

Conj_pre ::= y Conj Utt comma punct Conj_pre # cc head punct conj

Conj_pre ::= Conj Utt comma punct Conj_pre # cc head punct conj

Conj_pre ::= o Conj Utt comma punct Conj_pre # cc head punct conj

Conj_pre ::= ni Conj Utt comma punct Conj_pre # cc head punct conj

Conj_pre ::= ya sea Utt Conj_pre # head conj

Conj_pre ::= ya Utt Conj_pre # head conj

Conj_pre ::= ya bien sea Utt Conj_pre # head conj

Conj_pre ::= bien Utt Conj_pre # head conj

Conj_pre ::= ni Utt Conj_pre # head conj

Among other changes, I also altered the structure of the possessive phrases in order to match Spanish sentences from:

NP_poss ::= NP 's # head case

NP_poss ::= Pron_poss

To:

NP_poss ::= NP Pron_poss # head det

NP_poss ::= NP_de NP # case head

And the negation from:

VP_to ::= neg? to VP # advmod mark head

To:

VP_to ::= neg? Infinitive VP # advmod mark head

Other changes include – aside from translation – changes in positions of the sentence tags (*nsubj head* to *head nsubj*) and parts (*SC ::= VP_to → SC ::= VP_infinitivo)*.

Overall, I have to admit that the main phrase structure for English matches with Spanish. However, I think its worth noticing that wherever I altered the Sentence (S) structure to match the Spanish grammar the evaluation percentage was lower than the one with the English grammar. For example in Spanish, it's common to skip the pronoun in a sentence like (comi pasta.) but in English the pronoun should be always present (I ate pasta.). When a change like that was made in order to include that phenomenon in Spanish the evaluation accuracy dropped to 59%, whereas without it, it remained to 62%.