CS585 Natural Language Processing - HW2: Parsing CFGs

Group Members: Cristobal Sarome (A20416473), Akhil Suryadevara(A20391322)

Possible Improvements to the grammar

In order to enhance parsing capacity we added some recursive rules including conjunctions, like “s->s fconj s” and “np->np fconj np”, where fconj are the terminals and, or, etc.

This simple modifications allows the parser to identify a big amount of sentences that weren’t possible before, for example: “he drives and she steals” and “he steals a watermelon and a jetcar”.

To solve the problem of agreement we could create separate rules for different cases, for example “S->PN-3sing PV-3sing” and “S->PN-other PV-other”. The same approach could be taken regarding plural and singular determiners: in the current state the parser identifies “a watermelons” as a constituent even though it is not grammatically correct. To remediate these situations we would have to multiply the grammar rules considering every case. We didn’t actually make this modification because it is impractical to create so many rules for an exercise, it is important to notice this issue.

Conjunctions need to be added in order for the compound sentences( like ‘and’, ‘but’). Sentences which start with a VP needs to be added. More terminal words words need to be added in order for more data and accuracy.

Improvements implemented in the simple2.gr file

We added a grammar where conjunction sentences do work now for example “he drives” used to work and “she drives” used to work but “he drives and she drives” doesn’t work in simple.gr

Now conjunction sentences which have the word “and”, “or” or “but” works in simple2.gr

For example we added the following rules in the grammar file simple2.gr

s->s fconj s

np -> np fconj np

vp ->vp fconj vp

fconj -> and | or | but

This allow the algorithm to successfully parse sentences like the following

“she drives or he drives”

“she lives and drives”

“he steals a jetcar and a watermelon”

“she drives and he steals”