Yitong Wang
CSCI-GA-2590—Natural Language Processing
Assignment 4
Mar 11th 2015

1. Try the noun group / verb group patterns, available as chunkPatterns.txt, on four sentences from a newspaper. How many groups did it get right? How many did it miss? How many wrong ones did it identify? (These patterns aim to tag tensed verb groups and base form verbs (infinitives). They are not intended to mark present or past participles appearing by themselves.)

Sentence #1: The company has appointed Fred Fernandez as the vice president.

Annotating The company as constit cat=np
Annotating has as constit cat=vgroup
Annotating Fred Fernandez as constit cat=np
Annotating the vice as constit cat=np
Annotating president as constit cat=np

Wrongly parsed: The "has appointed" was missed, and instead parsed "has" as a verb in the verb group.

Sentence #2 : Polls show a majority of Americans in favor of legalization of medical marijuana.

Annotating Polls as constit cat=np
Annotating show as constit cat=vgroup
Annotating a majority as constit cat=np
Annotating Americans as constit cat=np
Annotating favor as constit cat=np
Annotating legalization as constit cat=np
Annotating medical marijuana as constit cat=np

Right: 4, Wrong: 3
The part "A majority of Americans" should be parsed as one noun-group, instead of two pieces, and "in favor of" should be parsed as verb group.

Sentence #3: Our workers and our people deserve not to have to see this color, which reminds them of the evil actions of Islamic State.

Annotating Our workers as constit cat=np
Annotating our people as constit cat=np
Annotating deserve as constit cat=vgroup
Annotating have as constit cat=vgroup-inf
Annotating see as constit cat=vgroup-inf
Annotating this color as constit cat=np

Annotating reminds as constit cat=vgroup
Annotating them as constit cat=np
Annotating the evil actions as constit cat=np
Annotating Islamic State as constit cat=np

This is a correctly parsed sentence.

SENTENCE #4: There were floral prints and soaring dragons, and as a finale there were Ben Stiller and Owen Wilson doing a "Zoolander" strut.

Annotating There as constit cat=np
Annotating were as constit cat=vgroup
Annotating floral prints as constit cat=np
Annotating dragons as constit cat=np
Annotating a finale as constit cat=np
Annotating were as constit cat=vgroup
Annotating Ben Stiller as constit cat=np
Annotating Owen Wilson as constit cat=np
Annotating Zoolander as constit cat=np
Annotating strut as constit cat=np

Missed: 1
The "doing" in the near end should be parsed, instead of being left out.

#2. Extend the patterns to allow perfect tenses (of the form "have" + past participle) to the verb groups. Note that this should include progressive perfect ("have been eating") and passive perfect ("has been eaten"); the overall pattern of English tenses is shown below. In addition, allow quantifiers in the noun group ("five assignments") and pre-nominal nouns ("my afternoon tea"). In extending the noun group, think about which modifiers can occur together in a single noun group and the order in which they can occur, and allow for these combinations in your rules.. Construct some test sentences to check these patterns.

Sentence#1: I have been eating afternoon tea all day.

Annotating I as ngroup
Annotating have been eating as constit cat=vgroup
Annotating afternoon tea as ngroup
Annotating all day as ngroup

Sentence#2: I have eaten breakfast.

Annotating I as ngroup
Annotating have eaten as constit cat=vgroup
Annotating breakfast as ngroup

Sentence#3: I have eaten five bagels.
Annotating I as ngroup
Annotating have eaten as constit cat=vgroup
Annotating five bagels as ngroup

#3. Retry the four sentences from the newspaper.

Sentence #1: The company has appointed Fred Fernandez as the vice president.

Annotating The company as ngroup
Annotating has appointed as constit cat=vgroup
Annotating Fred Fernandez as ngroup
Annotating the vice president as ngroup

Sentence #2: Polls show a majority of Americans in favor of legalization of medical marijuana.

Annotating Polls as ngroup
Annotating show as constit cat=vgroup
Annotating a majority as ngroup
Annotating Americans as ngroup
Annotating favor as ngroup
Annotating legalization as ngroup
Annotating medical marijuana as ngroup

Sentence #3: **Our workers and our people deserve not to have to see this color, which reminds them of the evil actions of Islamic State.**

Annotating Our workers as ngroup
Annotating our people as ngroup
Annotating deserve as constit cat=vgroup
Annotating have as constit cat=vgroup-inf
Annotating see as constit cat=vgroup-inf
Annotating this color as ngroup
Annotating reminds as constit cat=vgroup
Annotating them as ngroup
Annotating the evil actions as ngroup
Annotating Islamic State as ngroup

Sentence #4: **There were floral prints and soaring dragons, and as a finale there were Ben Stiller and Owen Wilson doing a "Zoolander" strut.**

Annotating There as ngroup
Annotating were as constit cat=vgroup
Annotating floral prints as ngroup

Annotating dragons as ngroup
Annotating a finale as ngroup
Annotating were as constit cat=vgroup
Annotating Ben Stiller as ngroup
Annotating Owen Wilson as ngroup
Annotating Zoolander as ngroup
Annotating strut as ngroup

The major improvement in performance is the catch in parsing different verb tense. For the first sentence, "has appointed" was missed because the verb group does not take the present perfect tense into consideration. After modification, this tense was caught and parsed perfectly.

4. One complication with verb groups is that they may include embedded adverbials, such as "has *rarely* been criticized. Select a text corpus and find two examples of such embedded adverbials. A well-designed grep search should help; so can the [Google ngram viewer](). You do *not* have to extend your patterns to handle these examples.

Sentence#1: I rarely speak about my sexuality preference.

Annotating I as ngroup
Annotating speak as constit cat=vgroup-inf
Annotating my sexuality preference as ngroup

Sentence#2: I almost have the chance to get into the program.

Annotating I as ngroup
Annotating have as constit cat=vgroup-inf
Annotating the chance as ngroup
Annotating get as constit cat=vgroup-inf
Annotating the program as ngroup