Davis Busteed – LING 360 – Homework #2

Text string I was searching thru:

After 2112, Rush went to the United Kingdom to record A Farewell to Kings (1977) and Hemispheres (1978) at Rockfield Studios in Wales. These albums saw the band members expanding the progressive elements in their music. "As our tastes got more obscure," lead singer Geddy Lee said in an interview, "we discovered more progressive rock-based bands like Yes, Van der Graaf Generator and King Crimson, and we were very inspired by those bands and their employees. They made us want to make our music more interesting and more complex and we tried to blend that with our own personalities to see what we could come up with that was indisputably us." Increased synthesizer use, lengthy songs, and highly dynamic playing featuring complex time signature changes became a staple of Rush's compositions. To achieve a broader, more progressive sound, Lifeson began to use classical and twelve-string guitars, and Lee added bass-pedal synthesizers and Minimoog. Likewise, Peart's percussion became diversified in the form of triangles, glockenspiel, wood blocks, cowbells, timpani, gong, and chimes. Beyond instrument additions, the band kept in stride with the progressive rock trends by continuing to compose long, conceptual songs with science fiction and fantasy overtones. As the new decade approached, Rush gradually began to dispose of its older styles of music in favour of shorter and sometimes softer arrangements. The lyrics up to this point were heavily influenced by classical poetry, fantasy literature, science fiction, and the writings of writer Ayn Rand, as exhibited most prominently by their 1975 song "Anthem" from Fly By Night and a specifically acknowledged derivation in 2112 (1976). Permanent Waves (1980) shifted Rush's style of music with the introduction of reggae and new wave elements. Although a hard rock style was still evident, more synthesizers were introduced. Moreover, because of the limited airplay Rush's previous extended-length songs received, Permanent Waves included shorter, more radio-friendly songs such as "The Spirit of Radio" and "Freewill", two songs that helped Permanent Waves become Rush's first US Top 5 album. Meanwhile, drummer Neil Peart's lyrics shifted toward an expository tone with subject matter that dwelled less on fantastical or allegorical story-telling and more heavily on topics that explored humanistic, social, and emotional elements.

I recognized and highlighted 19 nominalizations. I’m no linguistics expert, so there may be some false positives and false negatives.

# **Initial Results**

After reading the guide for recognizing nominalizations from DailyWritingTips.com, I attempted to mirror those rules and wrote six regular expressions. The output of my script was as follows (highlighting added after to show which results were correct):

Found 33 nominalizations.

Results: ['After', 'singer', 'der', 'Generator', 'synthesizer', 'broader', 'older', 'shorter', 'softer', 'writer', 'Moreover', 'shorter', 'drummer', 'matter', 'percussion', 'fiction', 'fiction', 'derivation', 'introduction', 'instrument', 'Lee', 'see', 'Lee', 'use', 'use', 'expanding', 'King', 'interesting', 'playing', 'featuring', 'string', 'continuing', 'telling']

From the above results, 11 were actual nominalizations found in the text, while the remaining items were false positives. Because the script returned 33 results, the precision for this set of regular expressions is

To calculate the recall for this search, I divided the 11 true positives by the number of true nominalizations that I hand-counted within the text.

# **Adjustments**

By looking closely at both the false positives and false negatives from the search, I made the following adjustments to increase the precision and recall of the regular expressions used in the script:

1. Added s? to the end of all regular expressions to allow for plural nominalizations
2. Added a negative lookahead for the word “to” before “use” and “change” to filter out these words used as a verb
3. Added a requirement of two or more letters before the -ing and -ee suffix to filter out small nouns that will never be nominalizations
4. Added a negative lookahead to filter out the word “element” (a special -ment word that is not a nominalization)

# **Final Results**

After making the aforementioned adjustments in the regular expressions, my script gave the following output:

Found 37 nominalizations.

Results: ['After', 'members', 'singer', 'der', 'Generator', 'synthesizer', 'broader', 'synthesizers', 'older', 'shorter', 'softer', 'writer', 'synthesizers', 'Moreover', 'shorter', 'drummer', 'matter', 'compositions', 'percussion', 'additions', 'fiction', 'fiction', 'derivation', 'introduction', 'instrument', 'arrangements', 'employees', 'use', 'changes', 'expanding', 'interesting', 'playing', 'featuring', 'string', 'continuing', 'writings', 'telling']

From the above results, 19 were actual nominalizations found in the text. Similar to the initial test, I calculated scores for precision and recall.

Both scores for precision and recall increased after the regular expressions were adjusted. As I thought about the precision score, I realized that the highest number of false positives are words that end in -er. I ran the script without the regular expression for identifying nominalizations that end in -er, and although the precision increased to 60%, the recall rate dropped to 63%. I think that if I combined regular expressions with part-of-speech tagging, I could improve the performance of this script, but until then, it seems to work pretty well.