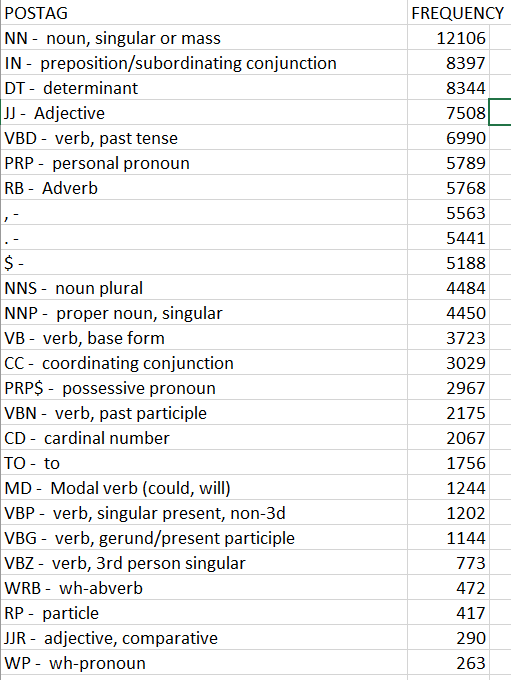
HW 4

QTM 446 W

Gabriel Wang

POSTAG and DEPREL are two important terms to Distant Reading and quantitative analysis of articles. POSTAG is pat-of-speech tag, and DEPREL is Dependency Relations. The POSTAG and DEPREL we use in PC-ACE are a pre-existing tag Penn Treebank or Brown Corpus Tag and Stanford DERPREL. (Franzosi)

The three output tables we are going to use in this report to analyze the sample of Song of Ice and Fire are the POSTAG Frequency table, the FORM by POSTAG Frequency table and the DEPREL table.



First, let’s look at the POSTAG Frequency table. The frequencies of VBD (verb, past tense), VBN (verb, past participle), dominate over those of the present verb forms. This is an expected conclusion as the whole book Song of Ice and Fire was about a story setting back to the ancient time.

Now move on the FORM by POSTAG table. We want to explore the top 5 most used nouns (analyzed in last week’s assignment), verbs, and adjective words other. After performing query on the FORM POSTAG table. The result is as follow. As mentioned in the analysis of POSTAG Frequency table, the past tense verbs worth analyzing the most. (note: I manually combined lexeme).

|  |  |  |
| --- | --- | --- |
| POSTAG | FORM | Frequency |
| VBD - verb, past tense | be | 1203 |
| VBD - verb, past tense | have | 998 |
| VBD - verb, past tense | say | 376 |
| VBD - verb, past tense | do | 181 |
| VBD - verb, past tense | make | 128 |
| VBD - verb, past tense | tell | 116 |
| VBD - verb, past tense | thought | 109 |

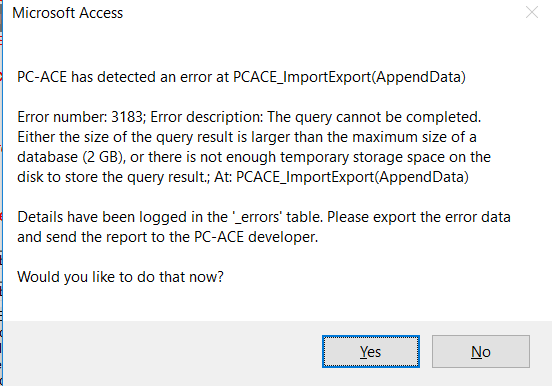
Those common ones like ‘be’, ‘have’, ‘do’, and ‘make’ don’t really tell much about the corpus. However, ‘say’, ‘tell’, ‘thought’ reveal some important information. Those three words are always used to describe interaction between two people or the one’s inner activities. Indeed, Song of Ice and Fire (referred as SIF) is well known for its complex relationship between characters. The description of interactions and inner activities helps author to organize and the reader to sort out the story line.

The table for top adjective words.

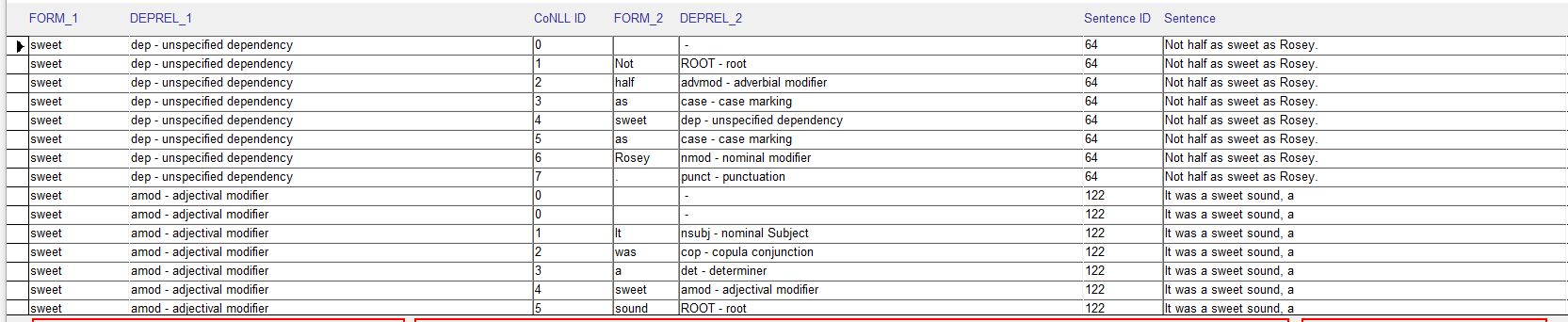
|  |  |  |
| --- | --- | --- |
| POSTAG | FORM | Frequency |
| JJ - Adjective | back | 168 |
| JJ - Adjective | like | 164 |
| JJ - Adjective | Ser | 162 |
| JJ - Adjective | old | 154 |
| JJ - Adjective | only | 122 |
| JJ - Adjective | red | 109 |
| JJ - Adjective | black | 104 |

This shows some wired results. First, the word ‘back’ is also recognized as verb, noun and adverb, so the result for back isn’t significant. ‘Ser’ is a common way to spell ‘sir’ in medieval Latin text which should be recognized as a noun instead of adjective. The same misrecognition also happened to ‘maester’ , a SFI word for ‘master’. The word ‘old’, ‘red’ and ‘black’ are closely related to three major places in sample, ‘Old Town’, ‘Black Water’ and ‘Red Keep’.

Since the sample data is too large and PC-ACE gave error saying that the memory size is too small:



I used the ‘Search Tool’ to manually analyze the dependency relationship. The key word I picked were generated from the previous analysis of POSTAG. They were ‘lord’, ‘ser’, ‘said’, ‘tell’, ‘think’, ‘ maester’, ‘dead’, ‘half’, ‘great’, ‘sweet’, ‘dark’, ‘gold’, ‘high’ etc. The result is shown below. I show one of the query I performed on the search.



The only results came from this labor work was the search of some adjective and adverb with strong negative or positive tendency, like ‘sweet’, ‘great’, ‘dark’, ‘gold’ etc.

|  |  |  |
| --- | --- | --- |
| FORM\_1 | FORM\_2 | FREQEUNCY |
| sweet | sister | 66 |
| sweet | girl | 46 |
| sweet | she | 44 |
| sweet | Lady | 30 |
| sweet | summer | 22 |
| sweet | ser | 2 |

|  |  |  |
| --- | --- | --- |
| FORM\_1 | FORM\_2 | FREQEUNCY |
| dark | his | 127 |
| dark | He | 78 |
| dark | beast | 58 |
| dark | Dothraki | 34 |
| dark | iron | 22 |
| dark | she | 4 |

|  |  |  |
| --- | --- | --- |
| FORM\_1 | FORM\_2 | FREQEUNCY |
| great | wall | 77 |
| great | Maester | 67 |
| great | stone | 48 |
| great | chamber | 38 |
| great | king | 2 |
| great | Khaleesi | 30 |

Those tables tell a lot about Martin’s (the author’s) tendency: positive word comes with female characters, negative word with male characters, and Martin hardly never deify anybody, except the great Khaleesi – Daenerys Taghryan, the rightful heir to the Iron Throne, mother of dragon. After I looked into the specific sentences, the only two great kings are two ancient Taghryan kings who conquer the Westeros.

There are more interesting combinations that definitely worth take a look at once I figured out how to solve the memory overflow bug.

Source:

Franzosi, Roberto, TIPS files on POSTAG and DEPREL.