Melissa Mosier

8/18/2022

**Polarity Analysis**

For this homework, we will carry on our analysis of the books Alice in Wonderland by Lewis Carroll and The Little Prince by Antoine de Saint Exupéry. We discovered in our last homework assignment that these two books were more similar than an analysis of delusion – Delusion and Dream: An Interpretation in the Light of Psychoanalysis of Gradiva by Sigmund Freud – either because Freud’s was a clinical text or because, despite what many may argue in an analysis of Alice in Wonderland, it is not actually about mental illness. We will assume moving forward that both of these books – Alice in Wonderland and The Little Prince – are about innocent adventures through a child’s imagination and perform a parts of speech analysis on these texts.

**Cleaning and Analysis**

These texts did not require much cleaning prior to tokenization. I did find that, when I ran code the first time, adjectives like “little” and “Little” were counted separately. Considering how often this specific word appears in both texts, and how random capitalization can throw off part of speech counts, I decided to lowercase all words.

I found that apostrophes appeared in my Alice in Wonderland lists. This was not a concern for The Little Prince. I was not bothered by the appearance of “n’t” as an adverb, since “not” will be important in later sentiment analysis, but the classification of “ ‘s”, “ ‘m”, “ ‘ve”, and “ ‘re” as adverbs did not make much sense to me at first. I also found that text like “ ‘exactly so” and “ ‘all right” had apostrophes in it. Admittedly I did try to use regEx to remove them, but when it did not work research showed me that they were auxiliary verbs or that their phrase counts were unaffected (phrases were not counted twice), I did abandon that attempt. Thus, apostrophes and other punctuation were not removed from my text. This did not affect my parts of speech analysis.

**Results and Interpretation**

Given the snafu I ran into above with the apostrophes, I do wish I could have performed an evaluation of the POS tagging. My concern was that, given the fantastic nature of these books, words are often made up or real words are used incorrectly (or just oddly). I also did lowercase the words to group words together that should not have been separated, but it may have made nouns more difficult to recognize.

Alice text:

* “alice soon”, “alice cautiously”, and “here alice” were considered adverb phrases even though alice is a noun.
* “alice” appears as both an adjective 14 times, an adverb 49 times, and a verb 24 times. It is correctly identified as a noun 308 times.

Prince text:

* “i” was identified as part of adverb phrases.

My remaining concern is with how small the text is, especially The Little Prince. Many of the adjective and adverb phrases only appear once. These plots display the frequencies of phrases on the y-axes – the most frequent phrases appear once or twice. In Alice in Wonderland, phrases never repeated more than 15 times; in The Little Prince, never more than six.

|  |  |
| --- | --- |
| Alice in Wonderland, **Adjective** Phrase Frequency | Alice in Wonderland, **Adverb** Phrase Frequency |
| The Little Prince, **Adjective** Phrase Frequency | The Little Prince, **Adverb** Phrase Frequency |

* As an example done in class with Crime and Punishment, we were able to produce 1636 adjective phrases and 1189 adverb phrases.
* Alice in Wonderland produced 236 adjective phrases and 253 adverb phrases.
* The Little Prince produced 183 adjective phrases and 110 adverb phrases.

I am concerned about the effect our small sample size will have on a sentiment analysis. I will address that in more detail later.

Here I will list some exploratory analysis for both texts: top 50 adjective phrases (by frequency), top 50 adverb phrases (by frequency), top 50 adjectives, top 50 adverbs, top 50 nouns, and top 50 verbs. Highlighted text is not necessarily problematic, but were discussed above.

|  |  |  |  |
| --- | --- | --- | --- |
| Alice in Wonderland | | The Little Prince | |
| top 50 adj. phrases | top 50 adverb phrases | top 50 adj. phrases | top 50 adverb phrases |
| so much 8  very curious 6  very glad 5  very much 4  very little 4  very likely 3  very good 3  so many 3  only wish 3  too much 3  once more 3  very tired 2  quite natural 2  very deep 2  n't much 2  very few 2  as much 2  so grave 2  always ready 2  very uncomfortable 2  almost wish 2  very short 2  very difficult 2  quite as much 2  quite silent 2  certainly too much 2  not much 2  so large 2  n't very civil 2  very interesting 2  very sleepy 1  so very remarkable 1  so very much 1  too dark 1  rather glad 1  no longer 1  too large 1  too small 1  not much larger 1  really impossible 1  almost certain 1  very nice 1  now only ten 1  too slippery 1  very fond 1  very small 1  quite surprised 1  quite dull 1  now more 1  so desperate 1 | as well 15  very soon 7  as soon 6  very politely 5  so very 4  down here 4  very much 4  just as well 4  very well 3  so far 3  too far 3  just then 3  back again 3  well enough 3  down again 3  very carefully 3  so often 3  not quite 3  very nearly 3  n't quite 3  as long 3  just now 3  very slowly 2  alice soon 2  very earnestly 2  not even 2  so alice 2  as hard 2  rather not 2  very gravely 2  very humbly 2  very angrily 2  so easily 2  quite as 2  certainly too 2  rather timidly 2  'all right 2  n't very 2  'exactly so 2  asleep again 2  never even 2  alice cautiously 2  n't even 2  never before 1  never once 1  suddenly down 1  here alice 1  not much 1  not here before 1  too long 1 | so i 6  so much 6  so small 5  very much 5  not able 4  very small 4  not important 3  very useful 2  once more 2  too much 2  so important 2  very first 2  too small 2  so little 2  so sad 2  very busy 2  very simple 2  very beautiful 2  much more difficult 2  very strange 2  very different 2  so weak 2  not worth 2  too heavy 2  greatly pleased 1  scarcely enough 1  more isolated 1  most extraordinary 1  too overpowering 1  very dangerous 1  very cumbersome 1  too old 1  very surprised 1  not so small 1  so many 1  so good 1  much greater 1  no longer interested 1  too tall 1  too short 1  now good 1  now bad 1  generally fair-to-middling 1  not little 1  as big 1  so big 1  strictly correct 1  too many 1  very tedious 1  very easy 1 | so then 5  far away 4  not even 4  not so 3  very much 3  very well 3  just so 3  once again 3  very slowly 2  much too 2  no longer 2  not yet 2  certainly very 2  not very 2  very rarely 2  too often 2  just alike 2  too long 2  n't much 1  certainly very much 1  now i 1  not disobey 1  little crossly 1  so often 1  already very sickly 1  very far away 1  very far 1  not really 1  only once 1  over again 1  inoffensively upward 1  as soon 1  too late 1  so closely 1  not much 1  so hard 1  there no 1  too far away 1  extremely ugly 1  really very 1  i soon 1  peacefully away 1  very closely 1  i not 1  easily enough 1  very quickly 1  very close 1  just as 1  now completely 1  not only 1 |

Alice in Wonderland

|  |  |  |  |
| --- | --- | --- | --- |
| top 50 adjectives | top 50 adverbs | top 50 nouns | top 50 verbs |
| little 127  other 40  great 39  much 34  large 33  more 33  last 33  first 32  white 30  such 27  poor 26  thought 24  good 24  long 24  same 23  next 20  curious 19  sure 19  old 19  right 18  low 14  high 14  alice 14  mad 14  whole 13  many 12  best 11  beautiful 11  own 10  glad 10  small 10  few 9  different 9  dear 9  least 9  afraid 8  ready 8  oh 7  golden 7  larger 7  wish 7  enough 7  deep 6  nice 6  dry 6  bright 6  melancholy 6  offended 6  full 6  sharp 6 | n't 204  very 139  not 135  so 110  again 83  then 82  quite 52  now 50  alice 49  just 49  as 48  here 48  never 45  only 42  down 38  once 32  back 32  well 30  too 26  up 26  rather 25  soon 25  away 23  yet 22  ever 21  even 19  however 19  first 18  much 17  indeed 16  more 16  there 15  perhaps 14  anxiously 14  hastily 14  suddenly 13  still 13  certainly 13  far 13  always 13  hardly 12  about 12  else 11  enough 11  really 10  nearly 10  angrily 9  together 9  timidly 9  close 8 | alice 308  queen 71  time 68  king 61  hatter 56  mock 56  turtle 55  gryphon 54  way 53  head 50  thing 49  voice 48  rabbit 47  duchess 42  mouse 41  tone 40  dormouse 38  'it 37  cat 35  march 34  moment 31  hare 31  nothing 30  things 30  door 30  eyes 28  'you 27  day 25  course 25  caterpillar 25  round 23  words 21  minute 21  sort 20  feet 19  anything 19  hand 19  dear 18  house 18  table 18  something 18  soup 18  court 18  use 17  question 17  side 17  jury 17  bill 17  garden 16  end 15 | said 462  was 366  had 184  be 148  do 124  is 113  's 109  know 88  were 86  have 84  went 83  did 74  see 67  began 58  'm 56  think 51  say 51  are 47  go 47  thought 46  looked 45  got 45  get 44  came 40  've 39  herself 38  been 38  're 36  made 30  found 30  put 30  looking 30  come 29  tell 28  replied 28  seemed 27  going 27  make 27  took 24  heard 24  alice 24  felt 23  added 23  getting 22  find 21  does 21  take 20  like 20  tried 19  being 19 |

The Little Prince

|  |  |  |  |
| --- | --- | --- | --- |
| top 50 adjectives | top 50 adverbs | top 50 nouns | top 50 verbs |
| little 255  good 46  more 25  other 22  old 19  great 19  small 17  much 17  first 16  last 15  true 14  able 13  beautiful 12  important 12  same 12  long 10  such 9  sure 9  sad 9  bad 9  single 9  happy 9  conceited 9  only 8  magnificent 7  difficult 7  enough 7  possible 7  extinct 7  many 6  unique 6  fresh 6  different 6  responsible 6  proud 5  second 5  new 5  big 5  invisible 5  own 5  golden 5  present 5  general 5  afraid 5  whole 4  extraordinary 4  funny 4  asteroid 4  serious 4  turkish 4 | not 166  very 81  so 78  then 74  too 43  never 41  again 36  just 33  now 31  n't 29  only 26  here 24  away 22  once 21  yes 19  always 17  also 17  back 17  much 16  more 16  even 16  still 13  there 12  ever 11  already 10  down 10  far 10  well 10  alone 9  most 9  certainly 9  perhaps 9  as 9  yet 8  first 7  really 6  ago 6  slowly 6  exactly 6  long 6  else 5  about 5  altogether 5  instead 4  scarcely 4  carefully 4  however 4  somewhere 4  no 4  longer 4 | prince 185  planet 69  flower 55  time 45  stars 41  sheep 39  fox 36  man 35  nothing 32  day 31  morning 31  king 31  geographer 23  men 22  anything 21  life 21  grown-ups 20  earth 20  drawing 18  water 18  ah 17  night 16  consequence 16  baobabs 16  desert 15  friend 15  one 15  lamp 15  people 14  question 14  thing 14  flowers 14  order 13  orders 13  snake 13  years 12  boa 12  place 12  heart 12  thorns 12  volcanoes 12  lamplighter 12  number 11  something 11  world 11  matters 11  everything 11  way 11  sunset 11  use 11 | is 298  said 195  was 193  have 143  are 136  be 106  do 83  had 79  am 56  has 38  were 37  made 37  know 36  did 35  been 32  come 32  see 30  make 29  asked 27  does 27  say 23  go 23  look 23  understand 22  answered 20  came 20  replied 19  want 18  put 18  added 17  looked 16  went 16  saw 15  took 14  live 14  think 14  rose 14  seemed 12  looking 12  set 11  draw 11  take 11  tell 11  love 11  laughed 11  let 11  seen 10  find 10  demanded 10  get 10 |

**Sentiment Analysis**

I understand that, for sentiment analysis, I will be able to use a list that is already classified as “positive” or “negative”. I am interested to see how context is handled here since there is so much surrealism in these texts. A comfort is that the texts are written as such a low reader level that the words are used simply, so they should not be so difficult to classify.

I am concerned about the size of the texts. I am considering that maybe between the whimsical nature of the texts and the short length, the overall text may be misread as positive or negative based on a small group of words or phrases being misread. Thankfully there is not a lot of repetition in either text, so the way any one phrase is interpreted should not affect the polarity of the overall text.

I am considering that for fun I might analyze both separately and then once together, just to get an idea of the polarity of a child’s imagination…if I can make that reach.

**Python Processing**

A lot of my process was done going back and forth between the texts, so the POS tagging and analysis are basically the same for both Alice in Wonderland and The Little Prince.

Import and text preparation

1. Read in the text.
2. Tokenize the text into sentences.
3. Lowercase.
4. \*\* Note here, I did attempt to apply a non-alpha filter. I explained earlier, however ,that it did not work and I came to appreciate the presence of apostrophes. I left it in the code just to be clear about what could have happened, but the filter was not effective as written in removing punctuation.
5. Apply function to tokenize words.

POS Tagging

1. Use NLTK to apply POS tags to text.
2. Identify and count adjective phrases.
   1. Graph frequencies of phrases.
3. Identify and count adverb phrases.
   1. Graph frequencies of phrases.
4. Display top 50:
   1. Adjective tokens
   2. Adverbs tokens
   3. Nouns tokens
   4. Verbs tokens
5. Print lengths of lists from steps 7-9.