Melissa Mosier

7/28/2022

**Corpus Statistics**

There are many interpretations of the Alice in Wonderland books by Lewis Carroll, some more popular than others. Most believe he was documenting the way children tell stories, processing the way their imagination works. There are various TV shows and movies that explore Wonderland as an adventure through a child’s psyche or imagination. A less popular interpretation is that Lewis Carroll may have been unwell, or that the stories he produced about Alice were the result of hallucinations or psychotic breaks. This interpretation led to some darker media, like the American McGee Alice and Alice: Madness Returns video games, which explored the inner mental workings of a young woman going through a psychotic break caused by traumatic life events.

Of Lewis Carroll’s texts, we will focus on Alice’s Adventures in Wonderland. Curious to explore this latter interpretation, I retrieved Delusion and Dream: An Interpretation in the Light of Psychoanalysis of Gradiva by Sigmund Freud – a description of adults experiencing delusion. Here I will seek to answer the question of whether Alice’s Adventures in Wonderland shares words or phrases expressed by people experiencing delusions. Further curiosity prompted me to retrieve The Little Prince by Antoine de Saint-Exupéry. The question I will attempt to answer here is whether Lewis Carroll’s text shares words or phrases in common with a story purely exploring a child’s imaginative adventure.

To start, it is worth nothing that these three texts have vastly different lengths.

* The Little Prince has 90305 characters.
* Alice’s Adventures in Wonderland has 144395 characters.
* Delusion and Dream has 345344 characters.

Normalization of text length for frequency will be very important in this project.

In processing these texts, I chose to tokenize, lowercase, and remove stopwords and non-alphabet symbols. The texts were mostly written in the same tense, had simple vocabularies, and were novels so I chose not to lemmatize.

I began by tokenizing the words, which changed the number of items we were working with by changing our perspective from characters to words:

* The Little Prince now had 21198 tokens.
* Alice’s Adventures in Wonderland now had 66632 tokens.
* Delusion and Dream now had 33493 tokens.

Especially Lewis Carroll’s writing tends to have random words capitalized, so I lowercased all the words to normalize them. I am making the choice that capitalization will not differentiate the words I am looking for. I also assumed that there would be more similarities than differences between the texts, so I used the same process for each text.

**Frequency**

I have included a normalized frequency of the most common words in each text. I decided to normalize the frequencies because the lengths of the three texts are so different that raw numbers are not comparable. I did this without removing stopwords and symbols for a raw comparison of each text, and then again removing stopwords and symbols.

|  |  |  |
| --- | --- | --- |
| Delusion and Dream | Alice’s Adventures in Wonderland | The Little Prince |
| (',', 0.0727)  ('the', 0.0611)  ('of', 0.0351)  ('to', 0.0239)  ('.', 0.0234)  ('and', 0.0215)  ('in', 0.0199)  ('a', 0.0189)  ('he', 0.0124)  ('that', 0.0122) | (',', 0.0721)  ('the', 0.0482)  ("'", 0.0390)  ('.', 0.0291)  ('and', 0.0241)  ('to', 0.0214)  ('a', 0.0188)  ('she', 0.0162)  ('it', 0.0160)  ('i', 0.0159) | ('.', 0.0739)  (',', 0.0488)  ('the', 0.0458)  ("''", 0.0337)  ('``', 0.0258)  ('i', 0.0256)  ('to', 0.0217)  ('a', 0.0192)  ('and', 0.0164)  ('of', 0.0158) |

Despite understanding Zipf’s Law, and knowing that these are the least significant characters or symbols in the texts, I was hoping that this would mean the texts might actually be similar.

(without stopwords, only alpha-numerical characters)

|  |  |  |
| --- | --- | --- |
| Delusion and Dream | Alice’s Adventures in Wonderland | The Little Prince |
| ('one', 0.0083)  ('gradiva', 0.0069)  ('dream', 0.0063)  ('hanold', 0.0058)  ('delusion', 0.0048)  ('however', 0.0043)  ('yet', 0.0043)  ('norbert', 0.0041)  ('us', 0.0038)  ('pompeii', 0.0037) | ('said', 0.0367)  ('alice', 0.0315)  ('little', 0.0101)  ('one', 0.0078)  ('know', 0.0070)  ('like', 0.0067)  ('went', 0.0066)  ('queen', 0.0059)  ('thought', 0.0058)  ('time', 0.0054) | ('little', 0.0364)  ('said', 0.0275)  ('prince', 0.0261)  ('one', 0.0187)  ('planet', 0.0097)  ('like', 0.0081)  ('flower', 0.0077)  ('good', 0.0069)  ('time', 0.0063)  ('sheep', 0.0060) |

However, upon removing stopwords and non-alphabetical symbols, it became very clear that Alice’s Adventures in Wonderland and The Little Prince share many of their most common words. I have no doubt that expanding these lists to more than 10 would produce similar results. I could see drawing a similarity between Freud’s “dream” and “delusion” to Carroll’s “thought”, but “dream” and “delusion” are literally in the title of that work. I do not think it is fair to make that connection, and thus it seems that Freud’s work documenting delusions and dreams may not actually be as similar to Lewis Carrol’s adventure as one might be led to believe by common interpretations of Alice’s Adventures in Wonderland.

While compelling, this is not conclusive. We examine phrases now rather than just words.

**Bigrams**

The bigrams scored by frequency shows a rank of the most frequent words, normalized by the number of words in the document. The bigrams scored by PMI how the probability of two words appearing together. As in, the words themselves may not actually be common, but when the first word appears the second word has a high likelihood of also appearing.

**Bigram Frequency**

In processing these bigrams, I used the word lists that still had stopwords and alphanumerical symbols for each text. As before, there was little difference between the texts – mostly punctuation, function words, and words that referred back to the title or the main characters. After producing a list of bigram frequencies, I removed non-alphabetical tokens and stopwords. Shown below is a rank of the most frequent words, normalized for comparison between texts.

|  |  |  |
| --- | --- | --- |
| Delusion and Dream | Alice’s Adventures in Wonderland | The Little Prince |
| (('norbert', 'hanold'), 0.0011)  (('project', 'gutenberg-tm'), 0.0008)  (('project', 'gutenberg'), 0.0004)  (('\_g.\_', 'p'), 0.0003)  (('young', 'archæologist'), 0.0003)  (('thousand', 'years'), 0.0003)  (('two', 'thousand'), 0.0002)  (('gutenberg-tm', 'electronic'), 0.0002)  (('young', 'lady'), 0.0002)  (('zoë', 'bertgang'), 0.0002)  (('let', 'us'), 0.0002)  (('years', 'ago'), 0.0002)  (('electronic', 'works'), 0.0002)  (('first', 'time'), 0.0002)  (('noon', 'hour'), 0.0002)  (('strada', 'di'), 0.0002)  (('\_g.\_', 'p.'), 0.0001)  (('archive', 'foundation'), 0.0001)  (('casa', 'di'), 0.0001)  (('gutenberg', 'literary'), 0.0001) | (('said', 'alice'), 0.0034)  (('mock', 'turtle'), 0.0016)  (('march', 'hare'), 0.0009)  (('thought', 'alice'), 0.0007)  (('white', 'rabbit'), 0.0006)  (('alice', 'thought'), 0.0003)  (("'of", 'course'), 0.0003)  (('alice', 'said'), 0.0003)  (('poor', 'alice'), 0.0003)  (('alice', 'replied'), 0.0002)  (('alice', 'looked'), 0.0002)  (('king', 'said'), 0.0002)  (('little', 'thing'), 0.0002)  (('poor', 'little'), 0.0002)  (('alice', 'began'), 0.0002)  (('cried', 'alice'), 0.0002)  (('good', 'deal'), 0.0002)  (('oh', 'dear'), 0.0002)  (('beautiful', 'soup'), 0.0001)  (('golden', 'key'), 0.0001) | (('little', 'prince'), 0.0086)  (('good', 'morning'), 0.0009)  (('conceited', 'man'), 0.0004)  (('long', 'time'), 0.0004)  (('boa', 'constrictor'), 0.0003)  (('one', 'day'), 0.0003)  (('prince', 'said'), 0.0003)  (('drawing', 'number'), 0.0003)  (('glass', 'globe'), 0.0003)  (('little', 'man'), 0.0003)  (('said', 'nothing'), 0.0003)  (('far', 'away'), 0.0002)  (('fresh', 'water'), 0.0002)  (('one', 'never'), 0.0002)  (('prince', 'went'), 0.0002)  (('went', 'away'), 0.0002)  (('boa', 'constrictors'), 0.0002)  (('little', 'bushes'), 0.0002)  (('never', 'knows'), 0.0002)  (('shall', 'look'), 0.0002) |

\*\*I have highlighted in black anything I do not wish to compare.

Although it is difficult to compare because it seems that our Delusion and Dream text retrieved directly from the Gutenberg website is riddled with references to the Gutenberg project, there is a clear difference in vocabulary and story telling. Delusion and Dream is an account of delusions, which led me to think it would be similar enough to Alice’s Adventures in Wonderland because they are both in third person perspective. It does appear that Delusion and Dream is a more clinical text, however.

If we focus on imagery and descriptive words, we still see a stark difference between Delusion and Dream, and the other two works. From this, I believe that in addition to Delusion and Dream being a different genre of text, the difference also comes from an age gap in the characters and the audience. Those experiencing delusions as recorded by Freud describe adult things – ‘electronic works’, ‘young [people]’, and time. The fantastic characters and colorfully described items that people often use to describe Lewis Carroll’s work as being descriptive of psychotic breaks – ‘white rabbit’, ‘march hare’, ‘mock turtle’, ‘golden key’, etc. - actually seem more similar to characters described in the other children’s book, The Little Prince – with its ‘boa constrictors’ and ‘glass globes’. The frequent occurrence of the word “little” in both also remind me that these are both stories about small children.

**IF I HAVE TIME BEFORE SUBMITTING THIS, RUN AGAIN WITH THE SECOND FILTER!!**

**Bigram PMI Score**

In processing these bigrams, I used the word lists that still had stopwords and alphanumerical symbols for each text. As before, there was little difference between the texts – mostly punctuation, function words, and words that referred back to the title or the main characters. After producing a list of bigram frequencies, I removed non-alphabetical tokens and stopwords. Shown below is a rank of the most frequent words, normalized for comparison between texts.

|  |  |  |
| --- | --- | --- |
| Delusion and Dream | Alice’s Adventures in Wonderland | The Little Prince |
| (('hundred', 'miles'), 13.4389)  (('honeymoon', 'travellers'), 12.4389)  (('sigmund', 'freud'), 12.4389)  (('public', 'domain'), 12.3800)  (('united', 'states'), 11.9795)  (('del', 'fauno'), 11.7760)  (('literary', 'archive'), 11.7300)  (('wilhelm', 'jensen'), 11.6504)  (('albergo', 'del'), 11.3974)  (('metal', 'brooch'), 11.3800)  (('almost', 'perpendicularly'), 11.2690)  (('manifest', 'dream-content'), 11.1170)  (('di', 'meleagro'), 11.0697)  (('technical', 'terms'), 11.0604)  (('archive', 'foundation'), 11.0101)  (('di', 'mercurio'), 10.9442)  (('thousand', 'years'), 10.8540)  (('\_g.\_', 'p'), 10.7555)  (('gutenberg', 'literary'), 10.7300)  (('strada', 'di'), 10.7071)  (('casa', 'di'), 10.6827)  (('childhood', 'friendship'), 10.6239)  (('del', 'sole'), 10.5743)  (('native', 'city'), 10.5644)  (('\_g.\_', 'p.'), 10.4274)  (('set', 'forth'), 10.4129)  (('years', 'ago'), 10.3565)  (('open', 'window'), 10.1239)  (('noon', 'hour'), 9.9795)  (('electronic', 'works'), 9.9114) | (('play', 'croquet'), 11.7685)  (('golden', 'key'), 11.6392)  (('kid', 'gloves'), 11.5721)  (("'of", 'course'), 10.2621)  (('white', 'kid'), 10.1246)  (('beautiful', 'soup'), 9.9871)  (('three', 'gardeners'), 9.9726)  (('march', 'hare'), 9.9441)  (('good', 'deal'), 9.6690)  (('cheshire', 'cat'), 9.5986)  (('trembling', 'voice'), 9.1835)  (('mock', 'turtle'), 9.1475)  (('next', 'witness'), 9.1246)  (('feet', 'high'), 9.1055)  (('white', 'rabbit'), 9.0295)  (('great', 'hurry'), 8.8717)  (('right', 'size'), 8.7461)  (('offended', 'tone'), 8.7096)  (('oh', 'dear'), 8.5721)  (('another', 'moment'), 7.9398)  (('little', 'golden'), 7.5461)  (('came', 'upon'), 7.3311)  (('poor', 'little'), 6.3311)  (('little', 'door'), 5.7096)  (('little', 'thing'), 5.4168)  (('poor', 'alice'), 5.1612)  (('thought', 'alice'), 4.8932)  (('cried', 'alice'), 4.8876)  (('alice', 'replied'), 4.7141)  (('said', 'alice'), 4.3959) | (('glass', 'globe'), 11.5642)  (('boa', 'constrictor'), 10.6712)  (('boa', 'constrictors'), 10.6712)  (('thousand', 'miles'), 10.3012)  (('fresh', 'water'), 10.2017)  (('six', 'years'), 9.6491)  (('far', 'away'), 9.1111)  (('drawing', 'number'), 9.0191)  (('conceited', 'man'), 8.8638)  (('went', 'away'), 8.4330)  (('good', 'morning'), 8.1950)  (('never', 'knows'), 8.1660)  (('long', 'time'), 8.0497)  (('shall', 'look'), 6.7063)  (('little', 'bushes'), 6.3604)  (('little', 'prince'), 6.3525)  (('prince', 'went'), 5.4252)  (('one', 'day'), 5.3621)  (('said', 'nothing'), 4.5716)  (('one', 'never'), 4.5437)  (('little', 'man'), 4.0384)  (('prince', 'said'), 2.2329) |

**In Alice’s Adventures in Wonderland, does Lewis Carroll use words or phrases expressed by people experiencing delusions?**

It does not appear that Lewis Carroll’s work shares phrases with Freud’s analysis of Dreams and Delusions. This could be because Alice in Wonderland is a novel, or because the audience was younger. I posit, however, that those choosing to interpret Alice’s Adventures in Wonderland as an allegory for mental illness are wrong. According to what I found in this analysis, Alice’s Adventures in Wonderland do not share the same topics – time, real life events and places – that are found in adult delusions.

**The Little Prince is written from the point of view of an adult recounting a vivid dream they had as a child, to explore their childhood imagination. Does Alice’s Adventures in Wonderland have more in common with this text or Delusion and Dream?**

The findings here do tell me that Alice’s Adventures in Wonderland share more in common with the childhood tale of The Little Prince, drifting to sleep and exploring his imagination through dreams. I grew up believing that Alice’s Adventure in Wonderland was about a young girl’s dreams as she slept on an outing in the park (from the presence of animals), and given those are also present in The Little Prince, I can believe that to be accurate. The presence and description of strange objects – ‘beautiful soup’, ‘golden key’, ‘glass globe’, etc. – are also shared between the two works. I find this to be a confirmation that Alice’s Adventures in Wonderland is more an exploration of childhood imagination than of psychosis.