A Mixed-Method Reading of *Leaf by Niggle*:

Documenting a token distribution analysis of “picture” and “tree”

**The approach**

Craig A. Boyd has pointed out that *Leaf by Niggle* could be interpreted an allegory of purgatory and Thomistic values, overcoming sloth in life and on the way to heaven (Boyd). We perhaps can also see it as a morality story, much like the morality plays of the medieval era, which Tolkien studied – think of *Everyman* and the quest to climb to heaven. Everyman represents an idea, that of *every man*, every Christian. Other morality play characters are named in this fashion; think of *Castle of Perseverance’s* Belial, World, and Good and Bad Angel. In this paper, therefore, I will read *Leaf by Niggle* as it is presented, and assume the story to be reflective of a journey through earth/mortal life to heaven. I am interested in the two topics of *picture* (that is to say, the *painting* by Niggle) and *tree*. They are not, on the whole, in opposition with each other, despite one being a mere representation of the other. They work orthogonally, because from the tree comes the picture, and from the picture Niggle is drawn to the tree.There is no one without the other; without heaven, one cannot dream.

I approach this short study using distant reading mixed with close reading, also called “mixed-method reading.” The computer program R will be used to identify where *picture* occurs and where *tree* occurs in *Leaf by Niggle,* and in what quantities. This will be presented in both table and graph format. I myself will then use close reading observations to draw conclusions from the quantitative analysis. The combination of the distant reading and close reading styles means that we have a more complete view of the story. Because this method is something new to Tolkien studies, I consider this to be an exploratory, documentary research project, rather than one which responds extensively to previous research. But of course I also hope this brings something new to our close reading of *Leaf by Niggle*. At the end of my analysis, I briefly explain what natural language processing can bring to the field of Tolkien studies, and how texts outside of *Leaf by Niggle* could benefit from being studied computationally and from a mixed-method view.

My source code can be viewed on GitHub at shorturl.at/pEPW9. There are extensive comments in the code explaining each step of the analysis.

**The close reading**

Presbyterian pastor Timothy Keller wrote a book called *Connecting your work to God’s work* (Keller)*.* In it, he reiterates the popular point that Tolkien’s worldview was shaped by his Catholic faith - even *The Lord of the Rings*, the well-loved trilogy he had written. It has extensive symbolism, Keller says. The other point that Keller makes (à la Tom Shippey, too) about Tolkien was that he was a perfectionist; he was never done with his work (Keller) (Shippey). This tweaking of the work is to ‘niggle’ at it. It should not be a surprise, therefore, that our character takes on the name of his personal trait. At one point in time, Tolkien was working on *The Lord of the Rings*, a massive volume: it had languages, it had genealogy, it had geography; Tolkien was not making any forward progress, so he took a break to write a short story, *Leaf by Niggle.*

*Leaf by Niggle* is as close to a biography of Tolkien as we can get. His first love, as we know, was linguistics and philology. When he created his languages, particularly Qenya, he was constantly going back to revise them, *and* he was concurrently going back to revise the mythology to support the languages. That was a lot of labour, and it might have gone on indefinitely. And that, in a nutshell, is the leaf and the tree of his life over those years. He began by illustrating the few leaves, i.e. the few beautiful languages that come to him, but he soon became overwhelmed by the want to draw the entire tree, i.e. the history of his worlds.

And any moderate fans of Tolkien will know that Tolkien spent many drafts and many years to get *The Lord of the Rings,* and any other work, just right. A Christian might look at this as a lived experience of the promises of heaven. The lesson that Tolkien is conveying to us, and that Keller has pointed out, is that as Christians, we may devote our lives to a great vision that never becomes fulfilled and that never finishes being accomplished. Many of us do not see the fruits of our labour, whether it be from a life well-lived or a life poorly-lived. We live and live and live until we die. We *only* have one leaf to show for our entire life, when we had initially reached for an entire tree. And this is where Tolkien’s message becomes clear, that our faith helps us realize our dream of a whole tree in eternal life. We cannot ever finish painting the tree – we see the tree – but it will be realized only with faith.

**The distant reading**

This section looks at the occurrences of "picture" (i.e. the painting) and "tree" in *Leaf by Niggle* in an attempt to understand how an inverse relationship of where they occur might reflect the path of the plot (earth to paradise). This section is broken down into the steps that I followed in performing the distant reading. You can follow along using the GitHub link I provided above to do your own analysis, but you do not need to follow along to understand the process.

​1. I imported the original short story into R as a text (.txt) file.

2. I cleaned the text by:

* removing all punctation
* removing spaces between words
* putting all text into lowercase
* putting all text into a single string, and
* indexing the string list of words for easy searching.

3. I performed basic statistical analyses on the text by:

* identifying how many times "picture" occurs, and at which index positions
* identifying how many times "tree" occurs, and at which index positions
* identifying the number of unique word types in short story
* plotting the frequency of words, and comparing the top 10 frequently occurring words with Zipf's law) ​ ​ ​ ​

Chart

Description automatically generated

Figure 1. Frequency graph of the top 10 words in "Leaf by Niggle."

Zipf's law says that the frequency of any word in a corpus is inversely proportional to its rank or position in the overall frequency distribution. In other words, the second most frequent word will occur about half as often as the most frequent word. ​ Unsurprisingly, "the" occurred the most and "he" occurred second most. *Leaf by Niggle* does not follow Zipf's law very well, since 265 is not half of 354, and 235 is not half of 265. These were the values given by R in a chart. Even if we just ignored the numbers given to us by R, by the printing of word frequencies in the chart, we can visually inspect the graph to see that the relationship is certainly *not* decreasing by half each time. ​

4. I accessed and compared word frequency data.

When comparing the word frequency of the top two words in the text, "the" and "he," we see that "the" occurs 1.335849 times more frequently than "he." Below is the table that R produces for relative frequencies of every single word in Leaf by Niggle (i.e. number of occurrences/total number of words): ​ ​ ​ ​ ​ ​ ​ ​ ​ ​

Table

Description automatically generated

Figure 2. A relative frequency table for some of words in "Leaf by Niggle."

​

​Plotting the relative frequency of the top 10 words gives us: ​ ​ ​ ​ ​ ​ ​ ​ ​

Chart, line chart

Description automatically generated

Figure 3. Relative frequency of the top 10 words of "Leaf by Niggle."

​ ​ ​

But as our real interest lies in the word "picture" and "tree;" we see that "picture" occurs 1.2 times more frequently than "tree." ​ Now we have to see *where* those two terms are occurring to support my argument about the journey from earth to heaven.

​ 5. I used token distribution and regular expressions ​to visualize where the words occur:

Creating a dispersion plot, or a 'token distribution plot,' will help us visually determine where in the story the words "picture" and "tree" tend to occur. When we say "where" in the story, we really mean "when;" that is, the further along in a text a word is, the more time has passed. Therefore, I will call the x-axis of such plots "novelistic time" and the y-axis will be the YES/NO occurrences of such words. YES indicates a presence of the word; NO indicates that the word is not there. There are no numerical values for the x-axis of such a graph, since a black line indicates the YES, and a blank/white line indicates a NO. ​ ​ ​ ​ ​ ​ ​

Chart, histogram

Description automatically generated

Figure 4. Token distribution plot of "picture."

Chart

Description automatically generated

Figure 5. Token distribution plot of "tree".

​ As we can see, "picture" occurs in a larger clump in the first half of Leaf by Niggle while "tree" occurs in a larger clump in the second half of Leaf by Niggle.

6. I performed some literary analysis.

Initially, after some close reading, I wished to compare the terms *painting* and *tree*, because it appeared that those were the main topics in the story, constantly moving towards each other. But the word *picture* is more often used to reference the painting, so I adjusted my search to *picture* and *tree* instead. As one can see, *tree* (and its variants, like *trees,* singular) occurs more frequently in the second half. *Picture* occurs more frequently in the first half.

A tentative argument could be made, subject to close reading analysis, that Tolkien shifts from emphasizing the more artificial word *picture* to the realized vision, *tree,* chronologically as our Niggle travels from his imperfect home to a vision of perfection.

**Combining close and distant in mixed-method reading**

We can clearly see an inverse relationship: as the number of *picture* goes down, the number of *tree* goes up. An argument could be made that when picture becomes less important to the plot, *tree*’s importance rises, and so it is mentioned more and more as Niggle encounters the realized vision.

**What I have observing about using mixed-method reading for Tolkien**

I have said in the introduction that this project was documentary and exploratory, rather than a rigorous reading. But we too have to be careful about using distant reading as a tool.

Some critics have vociferously voiced their criticism for distant reading: and maybe rightly so. Distant reading on its own, missing the flavour of context or close reading, can make erroneous conclusions. Take a fictitious example. The words “eat poison” may appear 89 times in *A fictional text* by Bard the Great*,* while “eat well” appears only 20 times. A scholar who looks at this statistic could summarize, erroneously, that “eat poison” is thus weighted more than “eat well”; and that Bard the Great wished the reader to think of dying by poisoning rather than living healthily*.* What the researcher forgot to search for, however, was the word “not”: “*not* eat poison,” “*not* eat well,” which completed reverses the sentiment of the phrases searched for.

In this research on *Leaf by Niggle*, I have tried to mix context well with distant reading. Notice that I counted using the word “picture” instead of the word “painting.” In my close reading, I have noticed that though Tolkien uses these two words to reference the same object (the painting), he uses the word “picture” on a greater frequency. After careful consideration, I find that there is no context lost if we choose to search for “picture” instead of “painting,” so no statistical errors based of context will occur in the final comparison study either.

As I mentioned in the introduction, other Tolkien texts could benefit from mixed-method reading. Long stories, like those in the *Lord of the Rings*, are an obvious first choice, because of the sheer quantity of text to be processed and analysed. For example, it certainly would not do to sit and highlight all the Germanic, Latin, and Anglo-Saxon words in *Lord of the Rings* with a highlighter nor to count them. However, a program in R, using packages like <quanteda> or <wordcount>, could do very well. The less obvious choice for mixed-method reading, however, like *Leaf by Niggle*, are short stories. Trends are easier to spot when close reading, erroneous distant reading conclusions can be quickly verified by a second or third close reading.

It comes altogether as a surprise to me that little work has been done in bringing distant reading, natural language processing, or mixed-method reading to Tolkien. Such sagas as he writes are ripe for study under a different lens, to bring out all nuances that we have previously not seen.

# Works Cited

Boyd, Craig A. *The Thomistic Virtue of Hope in Tolkien’s Leaf by Niggle*. 15 January 2019. 30 November 2021. <https://christianscholars.com/the-thomistic-virtue-of-hope-in-tolkiens-leaf-by-niggle/>.

Shippey, T. A. “J.R.R. Tolkien : Author of the Century .” *Tolkien : Author of the Century*, HarperCollins Publishers, 2000.

Keller, Timothy, and Katherine Leary Alsdorf. “Every Good Endeavor: Connecting Your Work to God’s Work.” *Every Good Endeavor*, First Riverhead trade paperback edition., Riverhead Books, 2014.