Evidence of Language Decline in Terry Pratchett's Writing: Effects of Posterior Cortical

Atrophy on the Lexicon

Luz Andrea Alvariño Gabeiras

English Seminar, Albert-Ludwigs Universität Freiburg

Language Processing Across the Lifespan

Dr. Alice Blumenthal-Dramé

March 31, 2021

Table of Contents

Tables and Figures

1.   Introduction

Changes to the lexicon, often in the form of deficits in vocabulary size or increased repetition and vagueness, are a well documented feature of dementia-like diseases such as Alzheimer's Disease (Bryan & Maxim 2007a; Hirst & Feng 2012; Pekkala et al. 2013). Published writers present an unique oportunity for analyzing the development of these features, as their lexicon can be easily accessed through their works and tracked across time. Studies on lexical deterioration caused by dementia in literary authors have often focused on works by Agatha Christie (Lancashire & Hirst 2009; Le et al. 2011; Lancashire 2015) and Iris Murdoch (Garrard et al. 2005; Le et al. 2011; Lancashire 2015), using corpus-based methodology to assess changes in the rate of word types, lexical tokens and n-grams across time. The present paper will attempt to conduct a similar study, this time focusing on British author Terry Pratchett in order to assess whether the progression of his rare form of Alzheimer's Disease had a negative impact on his lexicon. The research will focus on three separate parameters: (1) vocabulary size, (2) lexical repetition, and (3) vagueness.

In order to satisfy this aim, the structure of the paper will be as follows. Chapter 2 will provide a theoretical overview of Alzheimer's Disease and its effects on the lexicon, as well as an outline of some of the previous research conducted on authors with dementia. Chapter 3 will contain the case study, beginning with an introduction to Terry Pratchett and his struggle with the disease, followed by a full account of the methodology and finally, the full results and conclusions to the study.

2.   Theoretical Background

2.1.   Alzheimer's Disease and the Lexicon

Before delving into the possible effects of Alzheimer's Disease (AD) on the lexicon, it is worth noting that there is not a clear boundary between cognitive changes derived from healthy aging and those that come from a certain pathology, like AD or dementia (Bryan & Maxim 2007b: 2), and that both Alzheimer's Disease and healthy aging may lead to problems with lexical retrieval (Pekkala et al. 2013: 2). Due to length restrictions, this term paper will not explore the effects of aging on the lexicon[1], rather focusing on the impact that AD, and more especifically Posterior Cortical Atrophy, might have on an individual's vocabulary.

Alzheimer's Disease is a form of dementia that commonly appears in old age, with a "gradual onset and continuing cognitive decline" (Bryan & Maxim 2007a: 76). Although the disease affects every patient differently, language deficits are a prevalent symptom. Some of the major deficits impact aspects of the lexicon such as naming and word-finding, with "evidence that, in the early stages of the disease, access to the semantic lexicon is the main problem … but, as the disease progresses, specific items in the lexicon are lost" (81). Alzheimer's can therefore lead to a 'vague' and 'reduced' vocabulary, with some loss of low-frequency words (Hirst & Feng 2012: 357; Bryan & Maxim 2007a: 98). The meaning of words, that is 'semantic memory', may also be impacted (Albert 2007: 107).

Posterior Cortical Atrophy (PCA) is "an atypical form of Alzheimer disease with relatively young age at onset" (Schott & Crutch 2019: 52), with the main effects impacting motor development and the vision, whereas 'episodic memory', 'personality' and 'behavior' remain "relatively preserved" (Schott & Crutch 2019: 55). Although language can be affected by PCA, causing symptoms like 'agraphia' (inability to write), 'alexia'(inability to read), and

---

[1] Some research on the subject can be found in Wulff et al. (2021) and Verhaeghen (2003) among others. Ramscar et al. (2013) criticise the notion of decline altogether.

'ataxia' (slurred speech)[2] (Schott & Crutch 2019: 55), "language impairment is typically considered a late feature" (56). Some patients, however, described struggles with 'word retrieval' with the most distinct being "anomia[3], reduced phonemic fluency and slowed speech rate" (Crutch et al. 2013: 460). An example of a PCA patient with language dysfunction is presented in Figure 1.

Figure 1

Drawing of a clock in a patient with PCA, and writing of the sentence "'Eu estou aqui no escritório do Dr. Krause' ('I am here at the office of Dr. Krause')" Taken from Souza et.al. (2018); cropped.
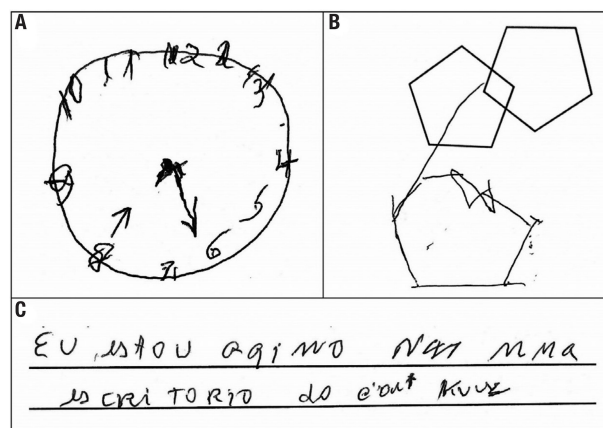
In the experiment, "the patient presented: repetitions of letters … omitted words … inconsistent spacing between words, and mixed capital and lowercase letters" (Souza et.al. 2018), writing the sentence as '*Eu estou aqui mo nas mma es cri torio do dout kuuse*'.

Some issues arise from the language features of PCA patients regarding the present study. Agraphia, alexia, ataxia, and anomia, as well as phonemic fluency, are all most characteristic of spontaneous speech and therefore hard to measure in a literary corpus. Furthermore, significant differences between PCA patients and control groups are not found for word repetition or type/token ratio[4] (Crutch et al. 2013: 465), which will be the focus of

---

[2] Definitions from"Agraphia" (2020).
[3] The "persistent inability to find the correct word" (Baldo & Dronkers 2009).
[4] The "number of different words/total number of words" (Crutch et al. 2013: 263).

our corpus study later on, although some patients did show impairments in sentence repetition.

2.2. Previous Research

Some research has been conducted measuring the impact of Alzheimer's on authors' literary output. Garrard et al. (2005) studied the effects of AD on novelist Iris Murdoch's writing, comparing three books written at different stages of the author's life; all books were converted to word lists, and the type/token ratio was measured, finding that the author's lexicon was reduced in her last work when compared to the two earlier ones (Garrard et al. 2005). Lancashire & Hirst (2009) looked into Agatha Christie's works, measuring the number of words used, as well as the number of repeated n-grams and the occurrences of vague words (e.g. "thing"); finding that "the richness of the vocabulary … declines with her age at composition", word types falling, repeated phrases increasing and indefinite words increasing (Lancashire & Hirst 2009: 3–4). Le et al. (2011) used a similar methodology to compare Iris Murdoch and Agatha Christie to healthy author P.D. James, measuring aspects like 'vocabulary size' (e.g. type/token ratio), 'lexical repetition' (e.g. 2-10 n-grams), 'lexical specificity' (e.g. number of indefinite nouns), and word class deficit (e.g. word class proportions); ultimately finding that the authors with dementia reported "a major loss in vocabulary … increase in repetition … [and] pronounced increase in fillers (Le et al. 2011: 457). Lancashire (2015) studied the vocabulary of three authors with dementia – Iris Murdoch, Ross Macdonald, and Enid Blyton – comparing them to three healthy authors – Frank Baum, R.A. Freeman and James Hilton; using specialized software to analyze word types and tokens, as well as phrasal repetition, finding that both Macdonald and Murdoch showed "impaired vocabulary and increased phrasal repetition", although Enid Blyton did not (Lancashire 2015: 81, 95). Finally, Vogel (2021) analyzed a corpus of author Terry Pratchett's works and looked at his pronoun use, embedding words and lexical variety; the results

showing a linear decrease in the type/token ratio correlating with age, with a "significant decrease in lexical variety" (539).

In conclusion, decreasing rates of word types in relation to word tokens, increases in the number of repeated phrasal types (n-grams), and more frequent indefinite words and expressions emerge as a marked characteristic of cognitive impairment in authors with dementia. Deficits to the lexicon have also been observed in patients with Alzheimer's Disease, although they are less common in Posterior Cortical Atrophy, which impacts mainly motor and visual abilities. The following case study will attempt to provide some evidence to the hypothesis that, as the previous research suggests, changes in vocabulary size, lexical repetition and vagueness will appear in Terry Pratchett's later works, written after his PCA diagnosis.

3.   Case Study

3.1.   Terry Pratchett

Sir Terence Pratchett, commonly known as 'Terry Pratchett' was an acclaimed British author and later advocate for Alzheimer's Disease; well known for his "humorous fantasy and science fiction", particularly the colossal series *Discworld,* comprising 41 parts (Eldridge 2021; Penguin Random House 2019). After publishing his first work in 1971, the author went on to write more than seventy books during his career, including several novels, short stories, theater plays, nonfiction including diaries, and poetry (Penguin Random House 2015), up until his passing in 2015.

In 2007, at the cenit of a wildly successful career, Pratchett was diagnosed with Posterior Cortical Atrophy (PCA; see section 2.1), prompting him to become outspoken about both the disease and the right to a humane death. The author reported that he had lost the ability to drive, and to type (Nuckpang 2008a; 2008b) as well as experiencing "sight problems, and difficulty with topological tasks, such as buttoning up a shirt" (Pratchett 2015: 14). The disease particularly affected his typing, with mistakes in his writing appearing as one of the initial symptoms, which he first attempted to rectify by dictating to his assistant (Pratchett 2015: 14), and later on by using software like 'TalkingPoint' which provides speech-to-text recognition ("Talking Point" 2007). Although towards the end of his life the author even sought assistance for giving speeches (Pratchett 2015), he described himself as

> the only person suffering from posterior cortical atrophy which, for some unknown reason, still leaves me able to write – with the help of my computer and friend – bestselling novels. (Pratchett 2013)

Interestingly, it is evident that when describing his experiences with the disease, Pratchett emphasized the mechanical and visual issues it caused him while drawing importance away from difficulties with the faculty of language itself, showing an

understandable reluctance to admit any cognitive or language impairment. One's introspection, however, is often insufficient to draw accurate conclusions on language use. The following study employs corpus methods to analyze two separate corpora of Pratchett's novels, pre and post-diagnosis, in order to gather some evidence on how much of the author's language, particularly his lexicon, was ultimately affected by the disease.

3.2.   Methodology

In order to carry out the case study that follows, two corpora of Pratchett's novels and short stories were compiled (Appendix I):

- The 'Pre Alzheimer's Disease Pratchett Corpus' (PAP) consists of 35 of the author's novels belonging to his series *Discworld*, spanning years 1986 to 2006, with about 3 million words.

- The (Post) 'Alzheimer's Disease Pratchett Corpus' (ADP) includes 9 novels and 4 short stories, also from *Discworld,* from the author's PCA diagnosis in 2007 to his ultimate passing from it in 2015. It falls just short of 1 million words.

Replicating methods used in previous research (Garrard et al. 2005; Lancashire & Hirst 2009; Le et al. 2011; Lancashire 2015; Vogel 2021), three main aspects were measured in the corpora:

Vocabulary Size. Using the AntConc function 'Word List' (Anthony 2020) , the word type to word token ratio (TTR) was calculated for all novels in both corpora:

Lexical types (or word types) are the different unlemmatized vocabulary items in a text (e.g. have, has, having). The total lexical tokens (or word tokens) of a text comprise its wordcount. (Lancashire 2015: 81)

The proportion between lexical types and lexical tokens[5] can be used as a measure for vocabulary size and lexical diversity (Youmans 1990). In this case, the TTR was plotted[6] against the novels' year of publication; rather than assuming the relationship between the progression of the disease and the TTR to be linear, a local regression model (LOESS) was applied, which also effectively compensated for the low number of observations (>1000) and the outliers present in the data (Wickham et al. 2017).

Lexical Repetition. Using the function 'Clusters/N-Grams', the number of n-gram types[7] was counted in all novels of both corpora. The query looked for 3 and 4-grams[8] with a minimum frequency of 2, in order to retain the 'lexical bundle' definition (three words or more) while avoiding uncommon structures, as per Biber et al. (1999: 992–993). Due to their much smaller length resulting in a very low frequency of n-grams, the four short stories present in ADP (marked in Appendix I) were excluded from the analysis. The number of n-gram types present in each novel was compared to the number of word types; then, the percentage of n-gram types (per total number of tokens in the corpus) was calculated for each novel, in order to account for differences in length.

Vagueness. Instances of 'indefinite' expressions (Lancashire & Hirst 2009: 3) were counted for both corpora, in an attempt to see if the author's vocabulary became more vague after his diagnosis. The words and phrases chosen were: *thing, things, something, anything, some, stuff, a few, a lot* ("Conversations" 2008; Lancashire & Hirst 2009).

---

[5] "The number of unique lemmatized word-types divided by the total number of word-tokens" (Le et al. 2011: 440).

[6] All graphs and statistical models used in the present paper were obtained through RStudio (RStudio Team 2020).

[7] That is, if the form "there was a" was repeated 100 times, it would be counted 1 time as a repeated chunk.

[8] "Also called multi-word expressions (or MWEs) or lexical bundles" ("N-grams" 2019).

3.3.   Results

First, the vocabulary size and diversity of both corpora was measured through the type/token ratio (TTR). Table 1 offers an overview of the counts extracted, beginning with the mean of word types and TTR per novel, as well as the total number of word types and tokens found in each corpus, pre-diagnosis (PAP) and post-diagnosis (ADP).
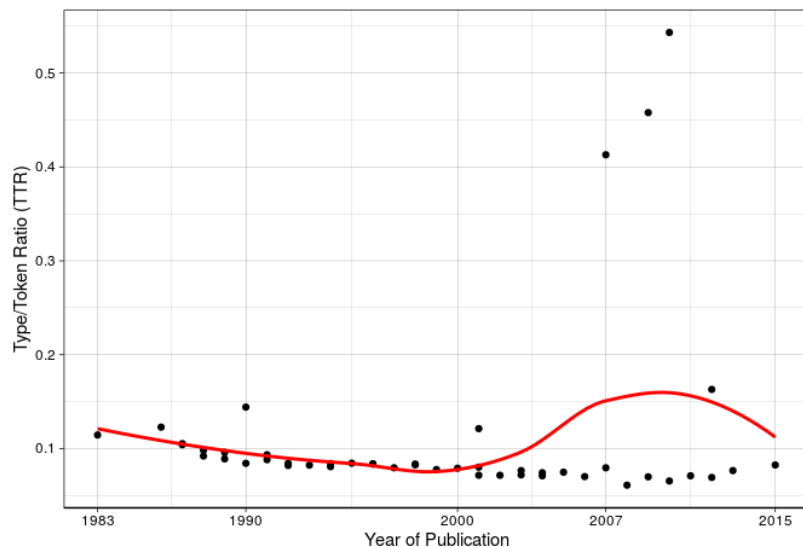
Table 1

Vocabulary Size Counts per Corpus

	Mean (per novel)		Total (corpus)	
	Word Types	TTR	Word Types	Word Tokens
PAP	7787.4	0.089	42995	3211626
ADP	7826.8	0.082	24851	977443

In order to see the progression through time, the TTR of each novel was plotted against its year of publication, as shown in Figure 2:

Figure 2

*Progression of TTR across Terry Pratchett's career.*

The three short stories in the ADP corpus – published in 2007, 2009 and 2010 respectively (Appendix I) – were much shorter than the other texts, causing abnormally high

type/token ratios. When excluding them from the data, a different trend appeared, as shown in Figure 3.

Figure 3

Progression of TTR across Terry Pratchett's career; short stories excluded.
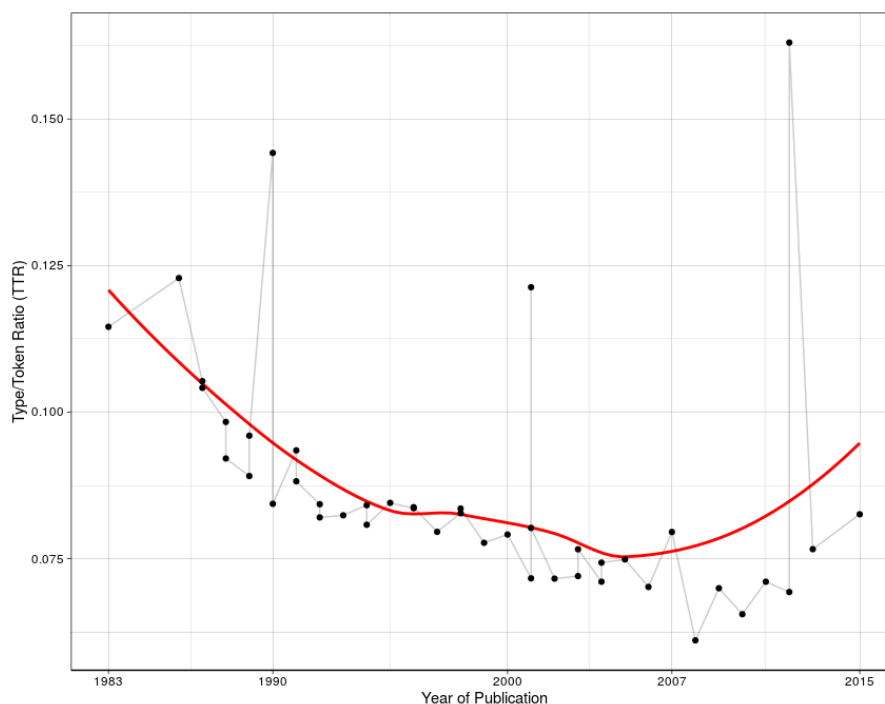
When processed using a linear regression model, the year of publication as a predictor for TTR (excluding the short stories) reported a p value of 0.01093, being moderately significant.

Secondly, lexical repetition was measured by observing the amount of 3 and 4-grams in each of the novels for either corpus.

Table 2

Lexical Repetition Counts per Corpus

	Mean (per novel)		Total (corpus)		
	N-Gram Types	Word Types	N-Gram Types	Word Types	Word Tokens
PAP	8149.914	7787.4	460448	42995	3211626
ADP	9917.556	7826.8	127389	24851	977443

In the linear regression model used, year of publication as a predictor for number of repeated lexical types (n-grams) reported a p value of 0.00247. The progression of n-gram and word types across Pratchett's career can be visualized in Figure 4.
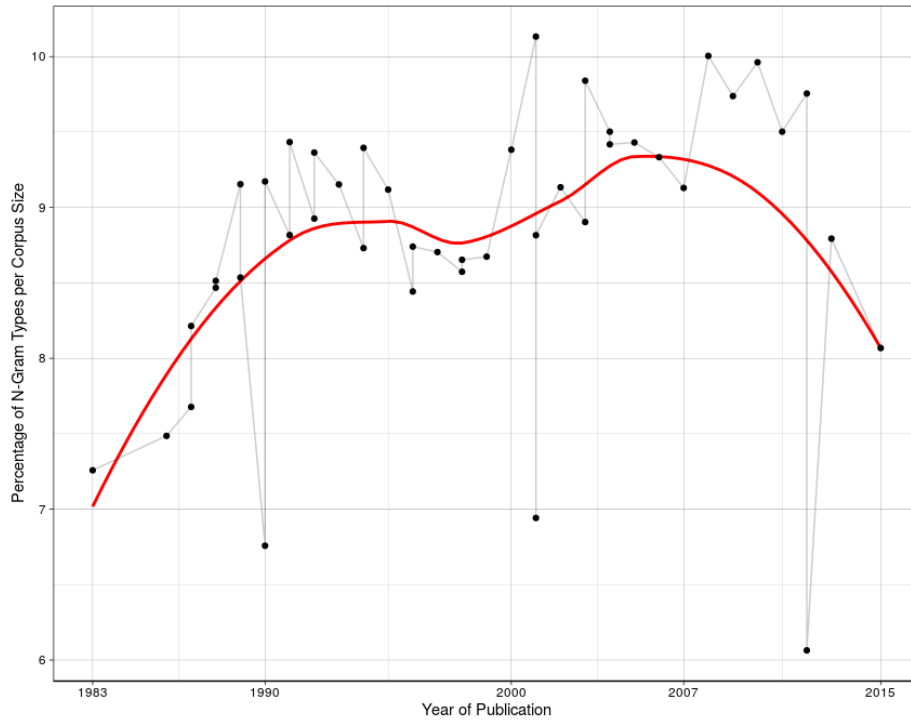
Figure 4

Progression of n-gram and word types across Terry Pratchett's career; short stories excluded.
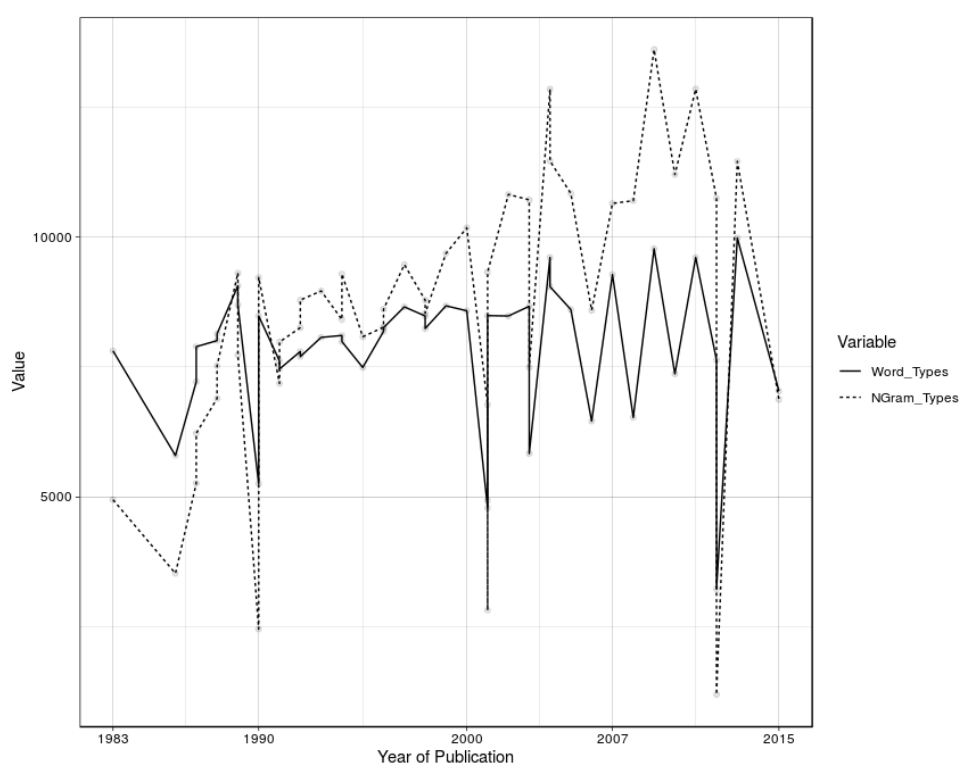
The percentage of n-gram types per corpus size (total number of tokens) can be seen in Figure 5. A trend line calculated using a local regression model (LOESS) was added for ease of understanding.

Figure 5

Progression of the percentage of n-gram types per word tokens (corpus size) across Terry Pratchett's lifespan; short stories excluded.

Finally, vagueness was analyzed through a simple frequency count of a few indefinite words and expressions for both corpora. The results can be observed in Figure 6.

Figure 6

Frequency of vague/indefinite words and expressions in each corpus.
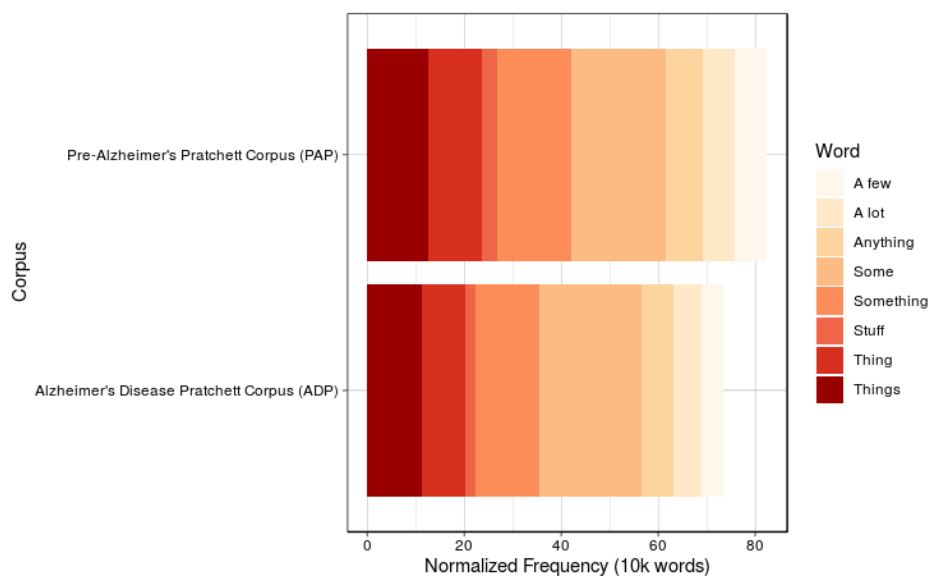
3.4. Discussion

As introduced in Chapter 2, types of dementia like Alzheimer's Disease impact the lexicon, leading to a reduced and more vague vocabulary, as well as increased lexical repetition. Previous research has found that authors with dementia show decreasing type/token ratios (TTR), increasing rates of repeated phrasal types (n-grams) and higher frequencies of vague or indefinite words across time, in line with the progression of the disease. The present study has found evidence that follows this overall trend, but also some contradicting results.

Concerning vocabulary size, no drastic differences were observed when looking at the overall counts of word types and tokens in the corpora, as can be seen in Table 1. The pre-diagnosis corpus (PAP) was significantly larger, both in terms of time-frame and novels written during that period, containing around triple the amount of word tokens than the post-diagnosis (ADP) corpus; therefore, it is not surprising or particularly significant that PAP had double the word types than ADP. However, the mean word types per novel and the mean type/token ratio (TTR) were also quite similar between the two corpora, going against the initial hypothesis that they would be significantly lower in the post-diagnosis period.

The progression of the TTR across time shows more interesting results. When excluding the outliers from the data, as shown in Figure 3, the progression shows a downward curve from 1983 to 2007, with a steady decline in vocabulary size; then, after the diagnosis in 2007, the ratio begins to increase again. This trend could be interpreted in different ways; for instance, the initial decrease in vocabulary size might imply that the author had been suffering with the disease for years prior to his diagnosis, and upon discovering it in 2007 he began puting more effort into using varied vocabulary, or ran his novels through more thorough editing, leading to the second increase. This would fall in line with the author's tendency to over-emphasize what he refers to as an almost miraculous ability to still write 'bestsellers'

while suffering a form of dementia (see section 3.1), as he would not have wanted any signs of the disease to show through his literary work.

The results gathered on lexical repetition show a different tendency. When looking at figure 4, one can observe that the count of n-gram types is lower in earlier novels, sometimes even falling below the amount of word-types, and then increases steadily towards later years, decreasing again in Pratchett's last two works. As shown in Table 2, the mean word types per novel are almost identical for each corpus, but the mean n-gram types are significantly higher in the post-diagnosis period, contradicting the results on vocabulary size. Taking this to mean that there was an increased rate of lexical repetition in the years after Pratchett's diagnosis, a possible interpretation would be that repetition is simply harder to catch, showing through in spite of editing. However, when calculating the percentage of n-gram types per word tokens in order to account for the different size of the corpora (Figure 5), a similar trend to the one found for vocabulary size (Figure 3) can be seen, with repetition rising from 1983 to 2006, and then falling from 2007 to 2015. A comparison can be seen in Figure 7.

Figure 7

Progression in vocabulary size (left) and lexical repetition (right) across Terry Pratchett's career; short stories excluded.
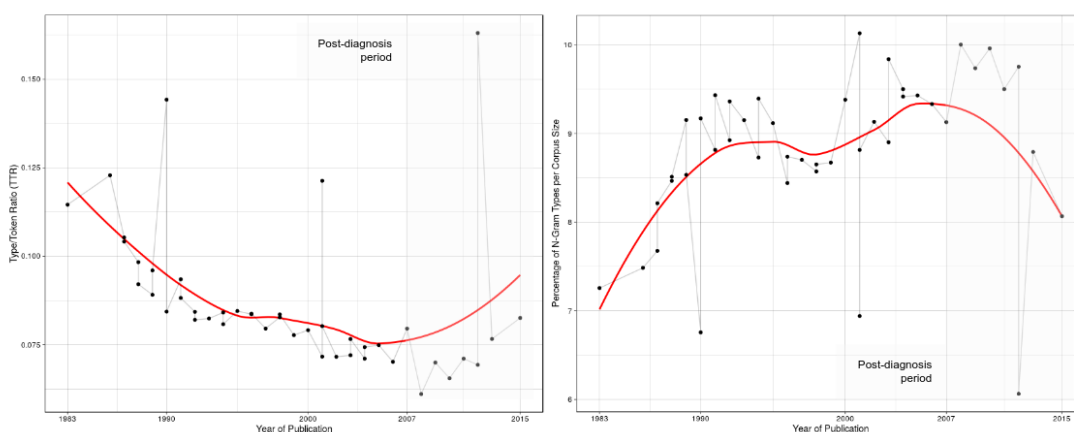
Lastly, the frequency counts gathered in order to study the 'vagueness' of each corpus found that vague expressions were overall more common in the pre-diagnosis corpus, with *thing(s)* being the most frequent in both, followed by *anything* and *some* (Figure 5). This would align with the results obtained for vocabulary size, and could be interpreted similarly, by assuming that the author put more effort into avoiding vagueness in the years after his Alzheimer's diagnosis. However, it is worth noting that the difference is not very prominent, and the frequency of indefinite expressions used is quite similar in the corpora. Further testing could search for indefinite nouns and high-frequency verbs, instead of usign a pre-made list of a few terms and expressions.

It is the author's opinion that all the results outlined above could have been better substantiated by comparing them to a reference corpus of a healthily aging author. Initially, the author had thought about using works by Douglas Adams, due to the similarity in the writers' literary style and the temporal proximity in their active writing period. The difficulty arose in finding an author who could match Terry Pratchett's extraordinarily prolific career; before his diagnosis in 2007, the author sometimes wrote up to two novels in a year, and he even published a total of nine works in the eight years he lived with the disease. Finding a single comparable author, such as Douglas Adams, proved impossible; and the author suggests that a better alternative for future studies would be to compile a corpus of several contemporary authors to be used as a reference.

3.5.   Conclusions

Using methods drawn from corpus linguistics, this paper intended to analyze the extent of lexical deficits due to Alzheimer's Disease (AD) on British author Terry Pratchett, who was diagnosed with the ailment towards the tail end of his career. After conducting the study, it was found that:

- Vocabulary size – as shown through the type/token ratio – decreased across time, up until the author's diagnosis with AD in 2007 when it began increasing. This finding contradicts previous research, which overwhelmingly found that authors with AD show a lower type/token ratio in later years.

- Lexical repetition – measured through the rate of repeated phrasal types per word tokens – increased from 1986 to 2007, and then began decreasing after diagnosis. This also contradicts previous research showing that lexical repetition increased as the disease progressed in authors with AD. When looking at the n-gram counts by themselves, however, it was found that they were considerably higher in the novels belonging to the post-diagnosis corpus, in spite of word types staying constant. These contradicting results might reflect a problem with the methodology used, or might occur due to the differences in corpus size.

- No significant differences were detected in the frequency of vague expressions used between both corpora. The expressions were slightly more frequent in the pre-diagnosis period, falling in line with the results obtained for vocabulary size. Further studies should be conducted analyzing a broader range of expressions, for instance indefinite nouns and pronouns.

All in all, the present paper managed to provide some interesting data to the research question of whether Terry Pratchett's struggles with PCA impacted his lexicon. Rather than showing a deficit in language abilities after diagnosis, the results may suggest that the author

paid more attention to the language used with his novels to avoid being perceived differently, and that a consistent decline can be seen in the novels written in years before his diagnosis instead. This theory, nonetheless, remains purely speculative and would need further research to be substantiated, perhaps by comparing Pratchett to a healthy contemporary author.

Sources Cited

Agraphia: Symptoms, Causes, Treatment & More. 2020. *Healthline*. https://www.healthline.com/health/agraphia (26 February, 2021).

Albert, Marilyn S. 2007. The Neuropsychology of the Development of Alzheimer's Disease. In Fergus Craik & Timothy Salthouse (eds.), *The Handbook of Aging and Cognition*, 97–133. New York: Taylor & Francis. https://doi.org/10.4324/9780203837665 (28 February, 2021).

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Bryan, Karen & Jane Maxim (eds.). 2007a. Language, communication and cognition in the dementias. In *Communication Disability in the Dementias*, 73–125. 1st edition. Wiley.

Bryan, Karen & Jane Maxim (eds.). 2007b. Health, ageing and the context of care. In *Communication Disability in the Dementias*, 1–17. 1st edition. Hoboken: Wiley.

Conversations: Being Vague. *BBC Learning English*. 2008. https://www.bbc.co.uk/worldservice/learningenglish/radio/specials/1210_how_to_con verse/page17.shtml (28 February, 2021).

Crutch, Sebastian J., Manja Lehmann, Jason D. Warren & Jonathan D. Rohrer. 2013. The language profile of Posterior Cortical Atrophy. *Journal of neurology, neurosurgery, and psychiatry* 84(4). 460–466. https://doi.org/10.1136/jnnp-2012-303309 (28 February, 2021).

Eldridge, Alison. 2021. Terry Pratchett. *Encyclopædia Britannica*. https://www.britannica.com/biography/Terry-Pratchett (22 February, 2021).

Garrard, Peter, Lisa M. Maloney, John R. Hodges & Karalyn Patterson. 2005. The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain* 128(2). 250–260. https://doi.org/10.1093/brain/awh341 (5 March, 2021).

Hirst, Graeme & V. Feng. 2012. Changes in Style in Authors with Alzheimer's Disease. https://doi.org/10.1080/0013838X.2012.668789 (28 February, 2021).

Lancashire, Ian. 2015. Vocabulary and dementia in six novelists. In Annette Gerstenberg & Anja Voeste (eds.), *Language Development: The lifespan perspective*, 77–108. Amsterdam: John Benjamins Publishing Company.

Lancashire, Ian & Graeme Hirst. 2009. Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study. https://www.researchgate.net/publication/242107161_Vocabulary_Changes_in_Agath a_Christie's_Mysteries_as_an_Indication_of_Dementia_A_Case_Study (7 March, 2021).

Le, Xuan, Ian Lancashire, Graeme Hirst & Regina Jokel. 2011. Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists. *Literary and Linguistic Computing* 26(4). 435–461. https://doi.org/10.1093/llc/fqr013 (7 March, 2021).

N-grams, multiword expressions, lexical bundles. 2019. *Sketch Engine*. https://www.sketchengine.eu/guide/n-grams-multiword-expressions/ (28 February, 2021).

Nuckpang. 2008a. *Terry Pratchett on Alzheimer's Disease (Part 1 of 2)*. https://www.youtube.com/watch?v=Z44ip5kYQ8w&t=0s&ab_channel=Nuckpang (12 February, 2021).

Nuckpang. 2008b. *Terry Prachett on Alzheimer's Disease (Part 2 of 2)*. https://www.youtube.com/watch?v=NTPdubbQISQ&ab_channel=Nuckpang (12 February, 2021).

Pekkala, Seija, Debra Wiener, Jayandra J.J. Himali, Alexa S. Beiser, Loraine K. Obler, Yulin Liu, Ann McKee, et al. 2013. Lexical retrieval in discourse: An early indicator of Alzheimer's dementia. *Clinical linguistics & phonetics* 27(12). 905–921. https://doi.org/10.3109/02699206.2013.815278 (22 February, 2021).

Penguin Random House. 2015. A Warm Welcome: Biography of Sir Terry Pratchett. *Sir Terry Pratchett*. https://www.terrypratchettbooks.com/about-sir-terry/ (14 February, 2021).

Penguin Random House. 2019. Books. *Sir Terry Pratchett*. https://www.terrypratchettbooks.com/books/ (13 February, 2021).

Pratchett, Terry. 2013. Dementia Blog, what's the point of it all? *Alzheimer's Research UK*. https://www.alzheimersresearchuk.org/blog/terry-pratchett-on-dementia/ (12 February, 2021).

Pratchett, Terry. 2015. *Shaking Hands with Death*. London: Corgi.

Ramscar, Michael, Peter Hendrix, Bradley Love & Harald Baayen. 2013. Learning is not decline: The mental lexicon as a window into cognition across the lifespan. *The Mental Lexicon* 8. https://doi.org/10.1075/ml.8.3.08ram (22 February, 2021).

RStudio Team. 2020. *RStudio: Integrated Development for R. RStudio*. Boston: RStudio Inc. http://www.rstudio.com/.

Schott, Jonathan M. & Sebastian J. Crutch. 2019. Posterior Cortical Atrophy. *Continuum : Lifelong Learning in Neurology* 25(1). 52–75. https://doi.org/10.1212/CON.0000000000000696 (5 March, 2021).

Souza, Ricardo Krause Martinez de, Lea Tenenholz Grinberg, Nalini Drieli Josviak, Daniel Benzecry de Almeida, Ricardo Ramina, Pedro André Kowacs, Paulo Caramelli, et al. 2018. A patient with posterior cortical atrophy due to Alzheimer's disease. *Dementia & Neuropsychologia*. Associação de Neurologia Cognitiva e do Comportamento 12(3). 326–328. https://doi.org/10.1590/1980-57642018dn12-030015 (17 February, 2021).

Talking Point. 2007. Milton Keynes (UK): GHG Software Developments Ltd. https://www.talkingpoint.uk.com/ (14 February, 2021).

Verhaeghen, Paul. 2003. Aging and vocabulary score: A meta-analysis. *Psychology and Aging*. US: American Psychological Association 18(2). 332–339. https://doi.org/10.1037/0882-7974.18.2.332 (7 March, 2021).

Vogel, Carl. 2021. Linguistic Evidence of Ageing in the Pratchett Canon. In Anna Esposito, Marcos Faundez-Zanuy, Francesco Carlo Morabito & Eros Pasero (eds.), *Progresses in Artificial Intelligence and Neural Systems* (Smart Innovation, Systems and Technologies), 531–541. New York: Springer. https://doi.org/10.1007/978-981-15-5093-5_45 (26 February, 2021).

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas L. Pendersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani & Dewey Dunnington. 2017. Smoothed conditional means. *ggplot2*. https://ggplot2.tidyverse.org/reference/geom_smooth.html (23 February, 2021).

Wulff, Dirk U., Simon De Deyne, Samuel Aeschbach & Rui Mata. 2021. Understanding the aging lexicon by linking individuals' experience, semantic networks, and cognitive performance. *PsyArXiv.* https://doi.org/10.31234/osf.io/z3ebt. https://psyarxiv.com/z3ebt/ (28 February, 2021).

Youmans, Gilbert. 1990. Measuring Lexical Style and Competence: The Type-Token Vocabulary Curve. *Style*. Penn State University Press 24(4). 584–599. https://www.jstor.org/stable/42946163 (6 March, 2021).

Appendix I
Terry Pratchett Corpora

	Corpora	
	Pre Alzheimer's Disease Pratchett Corpus (PAP)	**Alzheimer's Disease Pratchett Corpus (ADP)**
Works (chronological)	*The colour of Magic* (1983) *The Light Fantastic* (1986) *Equal Rites* (1987) *Mort* (1987) *Sourcery* (1988) *Wyrd Sisters* (1988) *Guards! Guards!* (1989) *Pyramids* (1989) *Eric* (1990) *Moving Pictures* (1990) *Reaper Man* (1991) *Witches Abroad* (1991) *Lords and Ladies* (1992) *Small Gods* (1992) *Men at Arms* (1993) *Interesting Times* (1994) *Soul Music* (1994) *Maskerade* (1995) *Feet of Clay* (1996) *Hogfather* (1996) *Jingo* (1997) *Carpe Jugulum* (1998) *The Last Continent* (1998) *The Fifth Elephant* (1999) *The Truth* (2000) *The Amazing Maurice and His Educated Rodent* (2001) *The Last Hero* (2001) *Thief of Time* (2001) *Night Watch* (2002) *Monstrous Regiment* (2003) *The Wee Free* (2003) *A Hat Full of Sky* (2004) *Going Postal* (2004) *Thud!* (2005) *Wintersmith* (2006)	*Making Money* (2007) *Minutes of the Metting to Form the Proposed Ankh-Morpock* (2007)* *Nation* (2008) *The Ankh Morpock Football Association Hall of Fame* (2009)* *Unseen Academicals* (2009) *I Shall Wear Midnight* (2010) *Sir Joshua Easement: A Biographical Note* (2010)* *Snuff* (2011) *Dodger* (2012) *The World of Poo* (2012) *Raising Steam* (2013) *Mrs Bradshaw's Handbook* (2014)* *The Sheperd's Crown* (2015)
Time frame	1983-2006	2007-2015
Author ages	36-58	59-67
Size	~3M words	~1M words

* Short stories.

The books employed for the corpora above come from the author's personal collection, as a fan.