

Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study

Ian Lancashire* and Graeme Hirst†

University of Toronto, *Department of English and †Department of Computer Science
Ian.Lancashire, Graeme.Hirst@utoronto.ca

Introduction

Alzheimer's disease leads to changes in language production at all levels — lexical, syntactic, and discourse — that are different to or markedly greater than those observed in normal aging (Maxim & Bryan 1994). For example, whereas “the lexicon continues to expand indefinitely until death or illness intervenes”, the “semantic and then phonological output lexicon” becomes progressively “inaccessible” in Alzheimer's disease (Maxim & Bryan 1994: 3, 24). And while in healthy aging, semantic retrieval speed deteriorates and hence “the number of ‘indefinite’ words may increase” (Maxim & Bryan 1994: 46), as may the number of repeated phrases, Nicholas et al. (1985) demonstrated that both indefinite words and repetitions occur significantly more often in the language of Alzheimer's patients than in that of healthy people of similar age and level of education.

These facts suggest that an assessment of dementia might be based in part on an analysis of a diachronic corpus of writing by the patient. Garrard et al. (2005) compared three works by the British novelist Iris Murdoch, whose diagnosis of Alzheimer's disease was confirmed post mortem. Her final novel, which was written during her decline, had a much smaller vocabulary than novels from her early and middle years. In addition, a small sample suggested that her sentences were syntactically simpler.

Here, we analyze the vocabulary of the British mystery writer Agatha Christie, who, although never diagnosed, was also believed to have suffered from dementia in her final years even as she continued to write. Our analysis, on a much larger and more-representative corpus than that which Garrard et al. used, concentrates on vocabulary-richness measures and opens a project that will also look at syntactic and discourse-level aspects of her texts.

Background

Agatha Christie (1890–1976), in a 53-year writing career, crafted about 85 novels and plays. Bodley Head published her first novel, *The Mysterious Affair at Styles*, in 1920, and Collins her last, *Postern of Fate*, in 1973. The public had bought 400 million copies of her works by 1975 and 2 billion by 1990. Her typical detective novel gives readers a

Presented at the 19th Annual Rotman Research Institute Conference, *Cognitive Aging: Research and Practice*, 8–10 March 2009, Toronto,

problem to be solved. Conversational dialogue and spare narration carry the reader along. She provides clues and diversions and maintains suspense until her sleuth unveils a usually elusive solution that ends the book. Her novels each contain between 55,000 and 75,000 words. Her custom, until the last books, was to work out the plot meticulously beforehand in a notebook, and to write the last chapter—where her detective laid out the solution—first. Her notes on *Lord Edgeware Dies* and *Evil under the Sun* “are almost identical to the finished article” (Thompson 2007: 369).

Janet Morgan, a biographer trusted by the Christie family, says that after *Elephants can Remember*, written when Christie was 81, “her powers really declined” (Morgan 1984: 370). When subsequently writing *Postern of Fate*, she reportedly found it “harder than ever to concentrate”: this last book “nearly killed her” (Morgan 1984: 371). Her preoccupation with old people and their memories in both *Elephants can Remember* and *Postern of Fate* reflects more on her personal circumstances than on crime, murderer, and clues. Readers have complained about inconsistencies in character and plotting in both these late works. Much of *Postern* digresses into Christie’s past memories and current problems, and the murderer is an afterthought. Her agent directed her to editorial help, and her husband Max and her secretary, Mrs Daphne Honeybone, “tidied it up”: Christie’s daughter Rosalind then asked Collins “to press for no more books”. Morgan concluded, “Physical and mental decline is sad” (pp. 371–72). By the time of *Elephants can Remember*, Christie had aged considerably, having fallen and broken a hip (Thompson 2007: 464, 473–74). Four years later, friends reported her thin and “frail”; she had angry fits (in one she cut off all her hair) and did not always make sense in conversation (Thompson 2007: 483). Although she was never assessed for dementia, her last novels reveal an inability to create a crime solvable by clue-detection according to the rules of the genre that she helped to create.

Material and Methods

Fourteen Christie novels written between ages 34 and 82 were digitized, and digitized copies of her first two mysteries, *The Mysterious Affair at Styles* (age 28) and *The Secret Adversary* (age 32), were taken from Project Gutenberg. After all punctuation, apostrophes, and hyphens were deleted, each text was divided into 10,000-word segments. The segments were then analyzed with the software tools *Concordance* and the *Text Analysis Computing Tools (TACT)*.

We performed three analyses of the first 50,000 words of each novel:

1. Like Garrard et al. (2005), as a simple measure of vocabulary size and richness, we counted the number of different words used.
2. As a second measure of vocabulary richness, we counted the number of different maximal phrase-types (i.e., word n -grams) that were repeated. These are defined by word-length and frequency. For example, if in a given text we saw 5 occurrences of “all sorts of things” and 7 occurrences of both “all sorts of” and “all

sorts”, we would count this as two repeated phrase-types, not three, because all occurrences of “all sorts” are contained in the longer phrases.

3. We counted the number of occurrences of the vague, indefinite words “thing”, “anything”, and “something”.

Results

Table 1 displays total counts for vocabulary size and repeating phrases, and the percentage of words that are indefinite nouns in the first 50,000 words of each novel.

Novel	Age at composition	Word-types	Repeated phrase-types	Indefinite words (%)
<i>Styles</i>	28	5027	7623	0.27
<i>Adversary</i>	32	5576	7320	0.39
<i>Ackroyd</i>	34	4833	7905	0.48
<i>Orient</i>	43	4692	8184	0.30
<i>Appointment</i>	47	4985	8071	0.39
<i>Curtain</i>	50	5131	7846	0.43
<i>Towards Zero</i>	51	4941	7680	0.50
<i>Announced</i>	59	5181	7596	0.45
<i>Destination</i>	63	5442	7426	0.36
<i>Ordeal</i>	67	4440	8221	0.58
<i>Clocks</i>	72	4827	8014	0.61
<i>Endless</i>	76	4159	8559	0.78
<i>Frankfurt*</i>	79	5583	7418	0.71
<i>Nemesis</i>	80	4631	8103	0.66
<i>Elephants</i>	81	3762	8821	1.02
<i>Postern</i>	82	4275	8249	1.23

*A thriller (not a mystery) that was written with the help of book research.

Table 1. Counts for vocabulary (word-types) and repeating phrases, and percentages of indefinite nouns in the first 50,000 words of 16 Christie novels.

Vocabulary size. The richness of the vocabulary of Christie’s novels declines with her age at composition. The three novels that she wrote in her 80s, *Nemesis*, *Elephants*, and *Postern*, have a smaller vocabulary than any of the analyzed works written by her between ages 28 to 63. Word-types in the first 50,000 words of her novels fall by one-fifth between ages 28–32 and 81–82. *Elephants Can Remember*, written when she was 81, exhibits a staggering drop in vocabulary, almost 31%, compared with *Destination Unknown*, written 18 years earlier. Some 15,000 words shorter than *Nemesis* and *Postern of Fate*, which preceded and followed it, *Elephants* appears to register the onset of a profound writing block. Possibly Christie’s broken hip, the year before, was a factor. A linear regression on the decline in vocabulary with age approaches significance [$F(1,14) =$

3.95, $p = .066$], and is highly significant when the outlier, *Frankfurt*, is removed (see discussion below) [$F(1,13) = 9.80, p < .01$].

Repeated phrases. The number of different repeating phrase-types in the first 50,000 words in Christie's novels increases with age, again implying a decline in the lexical richness of her writing. The increase with age approaches significance [$F(1,14) = 4.06, p = .064$], and again is highly significant when the outlier, *Frankfurt*, is removed [$F(1,13) = 8.47, p < .015$].

Indefinite words. Christie's use of vague, indefinite "thing" words increases significantly with age from 0.27% of her word-count in *Styles* (1920) to 1.23% in *Postern* (1973) [$F(1,14) = 22.6, p < .0005$]. *Frankfurt* is not an outlier in these data, and excluding it makes very little difference to the analysis.

Discussion

Her family's testimony about Christie's otherwise undiagnosed physical and mental decline offers an explanation for these data: encroaching dementia, as in the case of the English novelist Iris Murdoch that Garrard et al. (2005) studied. Our analysis suggests, in addition, that repeating phrases and, in particular, indefinite-term usage (not used by Garrard et al.) are significant markers.

Outlier. *Passenger to Frankfurt* has the largest vocabulary of all the works we analyzed. Unlike Christie's other works, it is a thriller, not a detective mystery, conceived, written, and researched in her early to mid 70s. Subtitled "An extravaganza", it draws on books by political thinkers that she requested of her publishers. On receiving her manuscript, they were doubtful about bringing it out because it differed so much from her detective fiction. Much of the vocabulary in *Passenger to Frankfurt* comes from her reliance on these sources. We therefore exclude it as an outlier from our tests for vocabulary richness. Nonetheless, we observe that it was not an outlier with regard to indefinite words.

Ongoing work. We will next analyze Christie's texts for known syntactic and discourse-level characteristics of Alzheimer's language. For comparison, we will also carry out parallel analyses of the works of several writers who are not suspected of dementia in old age.

Conclusion

While few present-day patients have a large online diachronic corpus available for analysis, this will begin to change as more individuals begin to keep, if only by inertia, a lifetime archive of e-mail, blogs, professional documents, and the like. While the diversity of topics and genres in such an archive brings methodological problems to the analysis (as observed with literary genre in *Passenger to Frankfurt*), we can nonetheless foresee the possibility of automated textual analysis as a part of the early diagnosis of Alzheimer's disease and similar dementias.

References

Garrard, P., Maloney, L.M., Hodges, J. R., & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, 128.2, 250–60.

Maxim, J. & Bryan, K. (1994). *Language of the Elderly: A Clinical Perspective*. London: Whurr.

Nicholas, M., Opler, L. K., Albert, M. L., Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech and Hearing Research*, 28: 405–10.

Morgan, J. (1984). *Agatha Christie: A Biography*. London: Collins.

Thompson, L. (2007). *Agatha Christie: An English Mystery*. London: Headline Review.