

Authorial Attribution: The Brandon Sanderson Test

Brittany Passmore

Brigham Young University

Authorial Attribution: The Brandon Sanderson Test

Authorial attribution is a growing aspect of forensic linguistics; many linguists have dedicated immense amounts of study and research to improving the analysis of authorship so as to provide both qualitative and quantitative data for use in federal and state judicial courts. For instance, various linguists have acknowledged issues with quantifying the “‘sufficiency’ of data” (Juola 2013); quantifying “register variation, . . . dialectal variation, . . . [and] idiolectal variation” (Grant 2007); comparing the “nature and length of the text” (Turell 2010); distinguishing and counting linguistic markers; or finding appropriate tools for analyzing authorship (Guillén-Nieto et al. 2008). “It is the linguists’ responsibility to create the theoretically sound hypotheses, test these hypotheses and perform the empirical evaluation of our own methods. It is the linguists’ responsibility to recognize junk science before it gets to court” (Chaski 2001). In other words, linguists are eager to determine how we can truly use authorial attribution in an effective, scientific way that can be appropriately and reliably used in court cases.

One of the most difficult questions lingering in this field, however, is how linguists should distinguish and analyze intentional mimicry and plagiarism when identifying authorship. “Disguise in written texts is often a considerable obstacle for the identification of the writer . . . if it is recognised at all. In fact, if disguise is not recognised as such, the analyst can very easily be misguided in their search for hints towards the possible author” (Marko 2017). Virtually all writers at one time or another purposefully try to adopt a writing voice or style which is not their own, especially creative writers, although Marko insists that “few people seem to be aware of their own writing style” (2017). Some authorship attribution studies have focused on creative writers; in the case of J.K. Rowling, for instance, who wrote a detective novel under a

pseudonym, an investigator hired a linguist to confirm whether Rowling truly had written *A Cuckoo's Calling* (Juola 2013; Juola 2015). The linguist did find enough reasonable confirmation to attribute the detective novel to Rowling, who later admitted that she had indeed written the book. This study led Juola to conclude that “we’re not even conscious of many of these [language] choices” (Juola 2013) that set apart our linguistic fingerprint. Another study analyzed the disputed authorship of the 15th Oz book, written after the original author’s death. The author was able to determine who had written the 15th book, even though both authors “wrote in the same genre and about the same setting” (Binongo 2003). These studies all show that despite the sometimes intentional disguise inherent in creative writing, authors can still leave behind noticeable style markers.

Still, despite these success stories, navigating intentional mimicry and register shifts can be a difficult task. How, then, do linguists best navigate identifying authors whose career has required them to consistently adopt idiolects not of their own? In this research paper, I will address this concern with a specific analysis of a similar situation to the Oz study in order to explore what linguists can determine accurately in such complicated situations. In my study I will analyze how Brandon Sanderson purposefully mimicked Robert Jordan’s style in finishing *The Wheel of Time* series and how much of a literary fingerprint he left while doing so. I will also explore how these principles can be applied to the field of forensic linguistics and what these findings mean for cases involving disputed authorship and accusations of plagiarism.

The Case of Brandon Sanderson

Robert Jordan published his first novel in *The Wheel of Time*, an epic fantasy series, in January of 1990. He continued to write and publish the series until he died on September 16, 2007 (Robert Jordan 2019). By that point, only eleven books had been released, and the 12th was

only two-thirds complete. Fans were devastated, desperate to know the rest of the story. Harriet, Robert Jordan's editor and wife, had promised her dying husband that she would hire a writer to finish the series for him. That writer she turned to was Brandon Sanderson, an author from Lincoln, Nebraska, who had already published a few epic fantasy books of his own, including *Elantris* and the Mistborn series. Sanderson accepted Harriet's offer and took on the task of splitting Jordan's manuscript into three books, filling in the gaps, and stringing all the pieces together in a way that would please the many fans and blend in with the late author's style.

In the end, the fans were largely pleased, including myself. However, I have noticed with interest that some readers found Sanderson's Wheel of Time books to still have a distinct style from Jordan's own writing. These readers remarked how Jordan's elaborate, elongated descriptions faded as Sanderson often cut to the action and "wrapped things up." On the other hand, I once heard Sanderson himself explain at a writing conference about how working on The Wheel of Time series was a rare case where he intentionally tried to mimic every part of Robert Jordan's style as best as he could. Professional authors such as Jordan and Sanderson "mimic" other voices constantly as part of their career, writing the dialogue of diverse characters in a believable fashion. However, how well is even a trained writer able to maintain another voice? Do "wordprints" remain, despite a writer's best efforts? While there was certainly no plagiarism in the case of Sanderson and Jordan, perhaps valuable lessons can be learned in exploring the possibilities and limitations of authorial attribution.

Methods

To better determine the reliability and validity of authorial attribution tests, I analyzed texts by both Robert Jordan and Brandon Sanderson, as well as certain portions of Sanderson's writing that purposely mimicked Jordan's. I selected a 5,000 word passage from Robert Jordan's

last published novel before his death—book eleven of The Wheel of Time series, *Knife of Dreams*, published October 11, 2005, and 315,163 words long (The Wheel of Time Series 2019). I then selected a 5,000 word passage from Brandon Sanderson’s novel, *The Way of Kings*, the first book in his epic fantasy series The Stormlight Archive, published August 31, 2010, and 383,389 words long (The Stormlight Archive 2019). I also selected two 5,000 word passages from book twelve of The Wheel of Time series, *The Gathering Storm*, which was published October 27, 2009, completed by Brandon Sanderson, and 297,502 words long (The Wheel of Time Series 2019). The passages written either solely by Robert Jordan in *Knife of Dreams* or solely by Brandon Sanderson in *The Way of Kings* acted as my baseline knowledge when comparing the passages from *The Gathering Storm*.

I chose these passages from these specific books because they were written around the same time, are about the same length, and reflect very similar genres. These factors allow me to test how accurately authorial attribution can identify an author under such similar circumstances. One of the passages in *The Gathering Storm*, selected from the chapter “Night in Hinderstap,” features a prominent protagonist named Mat and is knowingly authored by Brandon Sanderson according to his own experience recorded on his blog. The other selection from *The Gathering Storm*, a chapter focusing on the character Egwene, was likely partly written or at least heavily influenced by Robert Jordan and his notes according to Sanderson’s blog (Sanderson 2013). While knowing the likely authors of these excerpts may seem to introduce some “bias” to my study, this knowledge was only taken into consideration at the end of my research so that I could confirm the accuracy of my tests and determine how reliable authorial attribution can be under cases of intentional mimicry and disguise.

After selecting these four passages, I transcribed them into Microsoft Word and saved them as plain text files. I imported these files into AntConc and determined the 25 most frequent words in each passage. I used Microsoft Word to process each text sample and found the maximum, minimum, mean, and average sentence lengths in each passage. I copied the text of each passage into Haiyang Ai's Lexical Complexity Analyzer to obtain data on various syntactic structure patterns in the texts. Finally, I studied the passages for noticeable differences in punctuation and used the "find" function in Microsoft Word to help me quantify my findings about em dashes, ellipses, the serial comma, exclamation points, and italics into usable data. Following these linguistic tests, I compared the results of each test and looked for similarities and differences to determine how much of each author's wordprint remained in *The Gathering Storm* and attempted to identify which passage was written by Robert Jordan (if any) and which was written by Brandon Sanderson.

Results

Word Frequency

After running all four text samples through AntConc to determine the 25 most frequently used words, I compiled the data into a chart and found some interesting patterns to indicate which author most likely wrote each sample from *The Gathering Storm*. While all four texts had overall rather similar results in word choice, comprising mainly function words, the two texts from *The Gathering Storm* appeared to mirror Sanderson's text from *The Way of Kings* more than either of them did Jordan's text from *Knife of Dreams*.

The Egwene sample from *The Gathering Storm* and the sample from *The Way of Kings* showed a similar pattern of the most frequent function words (see figure 1). The first four words for each were *x* (meaning apostrophes), *the*, *to*, and *a*. Similarly, the Mat sample from *The*

Gathering Storm also listed *x* and *the* as its first two words, then *and* and *to* as the third and fourth. Additionally, each of these passages listed the main character's name as the fifth or seventh most frequent word in the list. Jordan's passage, on the other hand, listed *the*, *x*, *to*, and *and* as the four most frequent function words. Also, the main character's name appeared as the 24th most frequent word. These patterns, though slight, indicate that Sanderson was more likely the author of both text samples of *The Gathering Storm*.

Figure 1

25 Most Frequent Words

Gathering Storm: Egwene		Gathering Storm: Mat		Knife of Dreams		The Way of Kings	
<i>Word</i>	<i>Frequency</i>	<i>Word</i>	<i>Frequency</i>	<i>Word</i>	<i>Frequency</i>	<i>Word</i>	<i>Frequency</i>
x	350	x	415	the	256	x	340
the	277	the	345	x	165	the	317
to	154	and	141	to	145	to	126
a	121	to	137	and	138	a	115
of	113	Mat	132	a	115	he	110
and	109	of	107	he	108	of	99
her	92	and	100	of	77	and	95
Egwene	75	in	69	that	69	Kaladin	93
that	75	he	68	was	69	his	73
she	72	his	60	had	65	was	64
was	72	the	52	his	65	in	57
you	63	t	51	in	60	they	54
in	62	was	48	s	50	had	49
it	57	said	47	they	46	it	48
with	55	it	46	but	45	on	46
said	50	Talmanes	45	as	44	bridge	45
not	44	them	44	not	44	that	44

would	41	that	43	for	43	at	41
I	40	you	43	on	41	t	41
be	39	as	41	it	40	one	36
but	37	at	36	her	36	with	36
had	36	had	35	with	36	as	35
is	36	him	35	but	31	said	34
Ferane	34	on	35	Mat	30	you	34
for	34	s	35	would	30	s	33

Sentence Length

After processing the average sentence lengths in each passage and compiling the data, I noticed a similar pattern to the results of the data on word frequency. The baseline text sample for Sanderson's novel had an average sentence length of 9.9 words per sentence, and the baseline text sample for Jordan's novel had an average sentence length of 14.2 words per sentence, a difference of 4.3 words on average. The Egwene sample from *The Gathering Storm* had an average of 12.2 words per sentence, which lies fairly evenly between both baselines. This result makes it difficult to determine which author would have more likely written the passage about Egwene. However, the passage does appear to be .1 average words closer to Jordan, so we could assume that Jordan is the more likely author. The passage about Mat and Hinderstap from *The Gathering Storm* had an average of 10.8 words per sentence, which is much closer to Sanderson's baseline than Jordan's, indicating that Sanderson at least authored this second passage from *The Gathering Storm*. Please refer to figure 2 for a graphical summary of the data.

Figure 2

Sentence Length

	Gathering Storm: Egwene	Gathering Storm: Mat	Knife of Dreams	The Way of Kings
Number of Sentences	404	457	350	500
Average Words per Sentence	12.2	10.8	14.2	9.9

Syntactic Structure Patterns

The Ai Haiyung Lexical Analyzer dissected a variety of linguistic features about each passage, ultimately giving me data points in 25 categories under the areas of lexical density and sophistication, number of different words, type-token ratio, and word variation (see figure 3). I compared the data points from both texts of *The Gathering Storm* to determine which author's baseline they resembled most (see figure 4). Some of these linguistic features were not able to distinguish between the two authors, but overall, both passages were more similar to Jordan's baseline than Sanderson's. The passage about Mat and Hinderstap had 16 markers line up closer to Jordan's baseline, and the passage about Egwene had 12. Contrary to the two previous tests, these results indicate that the most likely author of both passages from *The Gathering Storm* was Robert Jordan, and the passage about Mat indicates this even more strongly than the passage about Egwene.

Figure 3

Linguistic Analysis

Syntactic Feature	Gathering Storm: Egwene	Gathering Storm: Mat	Knife of Dreams	The Way of Kings
Lexical Density	0.48	0.51	0.51	0.5
Lexical Sophistication 1	0.38	0.4	0.37	0.39
Lexical Sophistication 2	0.48	0.47	0.48	0.49
Verb Sophistication 1	0.18	0.15	0.17	0.18
Verb Sophistication 2	25.56	21.19	21.81	27.21
Corrected Verb Sophistication 1	3.57	3.25	3.3	3.69
Number of Different Words	1132	1001	1144	1198
Number of Different Words (First 50)	42	41	43	44
Number of Different Words (Expected Random 50)	41.8	41.4	44.1	41.6
Number of Different Words (Expected Sequence 50)	38.4	39.3	40.5	38.9
Type-Token Ratio	0.21	0.19	0.22	0.23
Mean Segmental Type-Token Ratio	0.79	0.8	0.82	0.81
Corrected Type-Token Ratio	10.97	9.66	11.26	11.65
Root Type-Token Ratio	15.51	13.66	15.93	16.48
Bi-Logarithmic Type-Token Ratio	0.82	0.8	0.82	0.83
Uber Index	20.64	19.07	21.07	21.5
Verb Variation 1	0.42	0.36	0.42	0.44
Squared Verb Variation 1	139.7	118.44	137.58	161.09
Corrected Verb Variation 1	8.36	7.7	8.29	8.97
Lexical Word Variation	0.4	0.33	0.4	0.41
Verb Variation 2	0.13	0.12	0.12	0.14
Noun Variation	0.41	0.32	0.42	0.41
Adjective Variation	0.08	0.06	0.07	0.07

Adverb Variation	0.04	0.04	0.05	0.04
Modifier Variation	0.12	0.1	0.11	0.11

Figure 4

Most Likely Author for Linguistic Analysis Tests

Syntactic Feature	Gathering Storm: Egwene	Gathering Storm: Mat	Difference between Jordan and Sanderson
Lexical Density	Sanderson	Jordan	0.01
Lexical Sophistication 1	Both	Sanderson	0.02
Lexical Sophistication 2	Jordan	Jordan	0.01
Verb Sophistication 1	Sanderson	Jordan	0.01
Verb Sophistication 2	Sanderson	Jordan	5.4
Corrected Verb Sophistication 1	Sanderson	Jordan	0.39
Number of Different Words	Jordan	Jordan	54
Number of Different Words (First 50)	Jordan	Jordan	1
Number of Different Words (Expected Random 50)	Sanderson	Sanderson	2.5
Number of Different Words (Expected Sequence 50)	Sanderson	Sanderson	1.6
Type-Token Ratio	Jordan	Jordan	0.01
Mean Segmental Type-Token Ratio	Sanderson	Sanderson	0.01
Corrected Type-Token Ratio	Jordan	Jordan	0.39
Root Type-Token Ratio	Jordan	Jordan	0.55
Bi-Logarithmic Type-Token Ratio	Jordan	Jordan	0.01
Uber Index	Jordan	Jordan	0.43
Verb Variation 1	Jordan	Jordan	0.02
Squared Verb Variation 1	Jordan	Jordan	23.51
Corrected Verb Variation 1	Jordan	Jordan	0.068

Lexical Word Variation	Jordan	Jordan	0.01
Verb Variation 2	Both	Sanderson	0.02
Noun Variation	Sanderson	Sanderson	0.01
Adjective Variation	Both	Both	0
Adverb Variation	Sanderson	Sanderson	0.01
Modifier Variation	Both	Both	0
Total Sanderson	9	7	
Total Jordan	12	16	
Total Both	4	2	

Punctuation

Since punctuation is well known in the linguistic field to be a fairly reliable marker of individual writers, I analyzed a number of various punctuation marks in each passage to better determine who the author of each passage in *The Gathering Storm* might be. My tests concerning em dashes, ellipses, the serial comma, exclamation points, and italics led to mixed results.

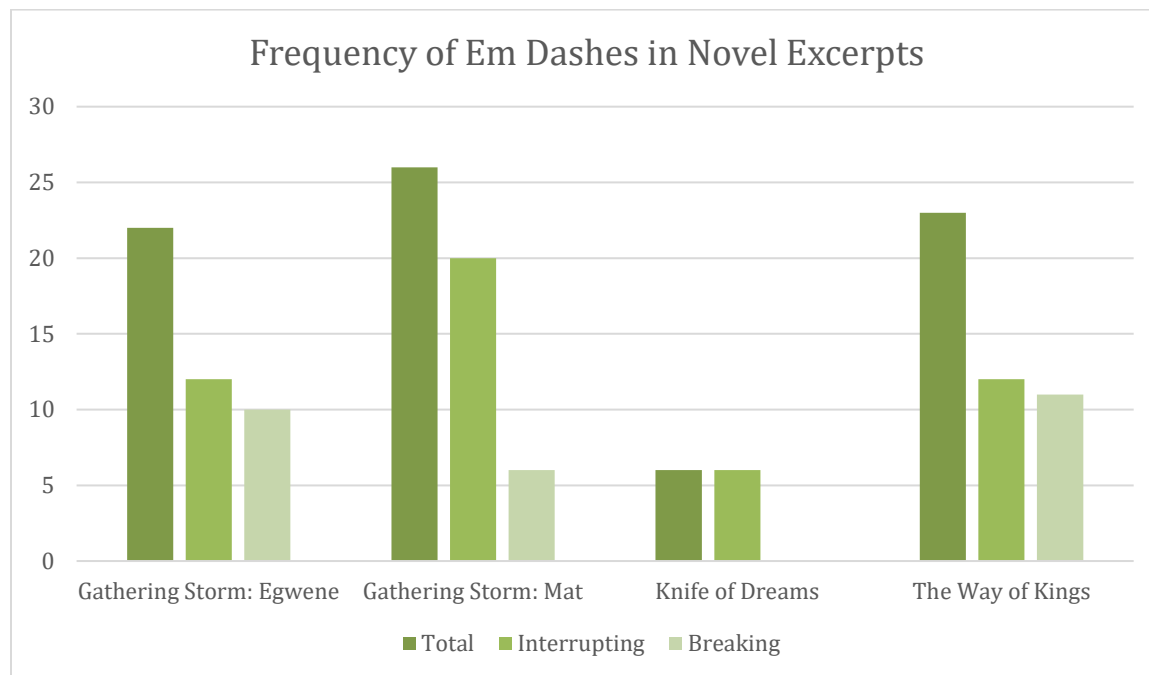
Em dashes. Writers can use em dashes in a number of different ways, and I identified primarily two different functions of the dash in Jordan and Sanderson's passages. These functions included interruptions—or detours—in the sentence as I have just demonstrated here, as well as breaks in a sentence—as I have just demonstrated here. Essentially, interruptions always consist of two em dashes, and breaks only comprise one em dash before the sentence ends.

Jordan's excerpt from *Knife of Dreams* yielded a total of six em dashes, or three pairs of interrupting dashes. Sanderson's excerpt from *The Way of Kings* yielded a total of 23 em dashes, with six pairs of interrupting dashes and 11 breaking dashes. The Egwene excerpt from *The Gathering Storm* contained 22 em dashes (with six pairs of interrupting dashes and 10 breaking

dashes), and the Mat excerpt from *The Gathering Storm* contained 26 em dashes (with 10 pairs of interrupting dashes and six breaking dashes) (see figure 5). Because Sanderson's excerpt and the two *Gathering Storm* excerpts each yielded above 20 em dashes and Jordan's excerpt only yielded six, this test implies that Sanderson authored of both passages in *The Gathering Storm*.

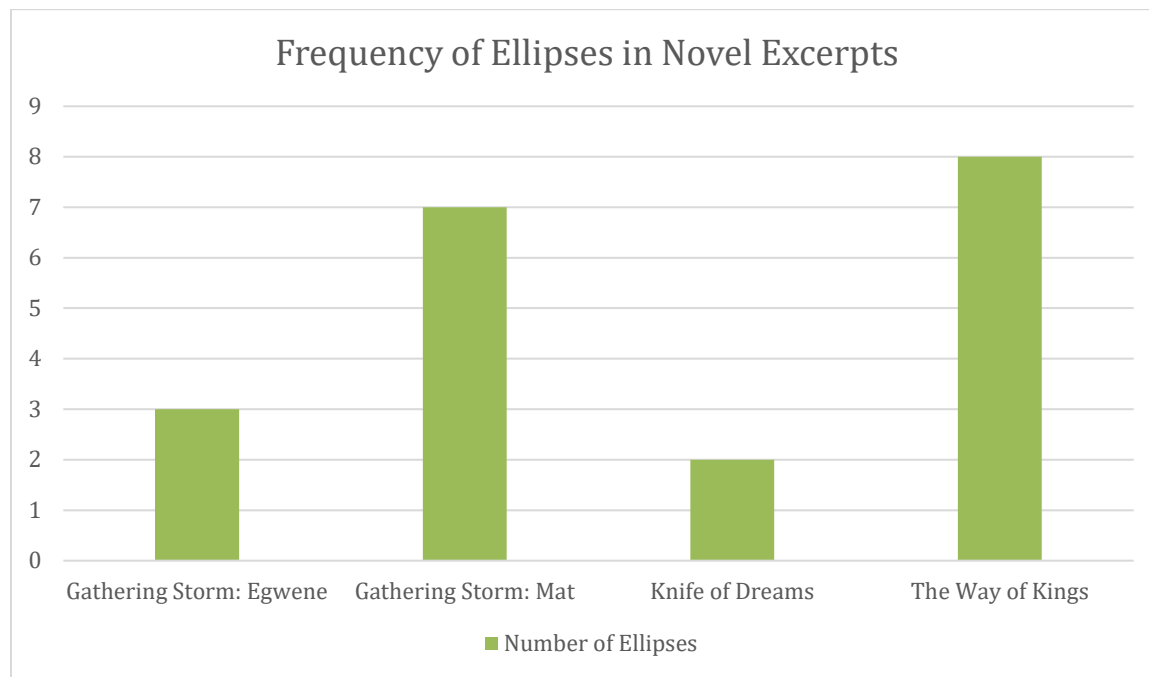
Figure 5

Frequency of Em Dashes in Novel Excerpts



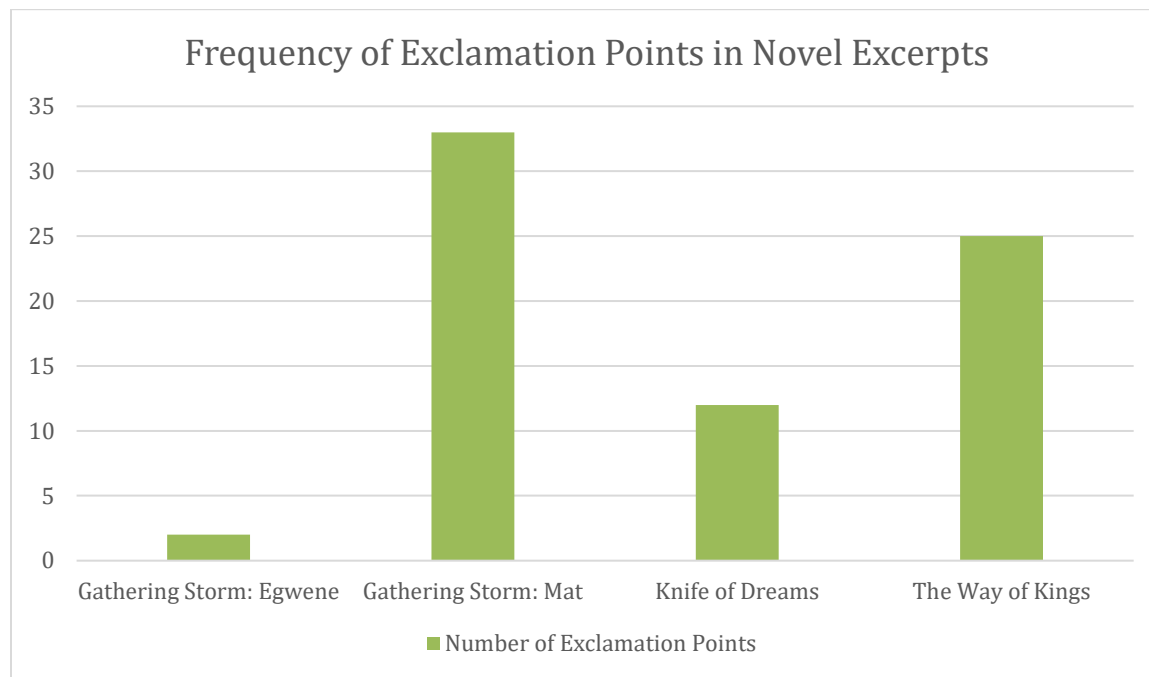
Ellipses. Ellipses, a series of three periods, are generally used to show a trailing off of one's thoughts in novels, especially in dialogue. Based on the two excerpts from Jordan and Sanderson, Sanderson used ellipses four times more often than Jordan (see figure 6). The Egwene excerpt from *The Gathering Storm* seemed to follow Jordan's pattern of having few ellipses while the Mat excerpt followed Sanderson's pattern of having more ellipses. This part of the punctuation test suggests that Sanderson authored the passage from *The Gathering Storm* featuring Mat and that Jordan authored the passage featuring Egwene.

Figure 6

Frequency of Ellipses in Novel Excerpts

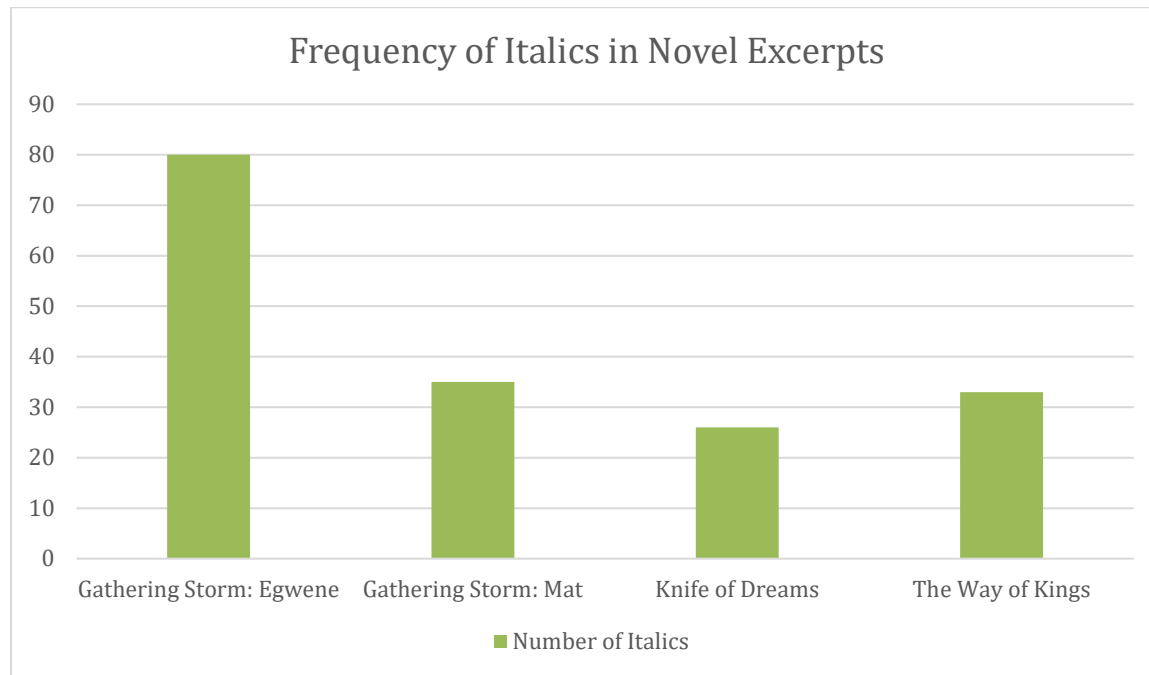
Exclamation points. Professional authors are generally aware that overusing exclamation points can make writing seem melodramatic, but the right amount can convey the tone of a novel's dialogue appropriately. Sanderson used about twice as many exclamation points as Jordan when comparing their two excerpts (see figure 7). Similar to the pattern of ellipses above, the excerpt from *The Gathering Storm* about Mat reflected Sanderson's exclamation point usage, and the excerpt about Egwene reflected Jordan's exclamation point usage. These patterns again suggest that Sanderson authored the passage about Mat and that Jordan authored the passage about Egwene.

Figure 7

Frequency of Exclamation Points in Novel Excerpts

Italics. In novels, authors tend to use italics to emphasize important words (especially in dialogue) or to set off thoughts inside the characters' heads. Sanderson and Jordan had a fairly similar pattern of how often they used italics in their excerpts, and the passage from *The Gathering Storm* about Mat followed this same trend (see figure 8). Interestingly, the passage about Egwene is an outlier and uses anywhere from two to four times as many italicized words as any of the other passages. This test seems to indicate that Sanderson may have been the more likely author of these passages since he used slightly more italics than Jordan, but the results are not so clear as the other tests.

Figure 8

Frequency of Italics in Novel Excerpts**Discussion**

Approaching these four excerpts with a variety of tests proved to be beneficial, as there was quite a range of results and possible conclusions from each individual test. Overall, the word frequency test and the punctuation tests concerning em dashes and italics seemed to point to Sanderson as the author of both passages from *The Gathering Storm*. The sentence length test and the punctuation tests concerning exclamation points and italics all seem to suggest that Sanderson authored the passage about Mat and that Jordan authored the passage about Egwene. The tests concerning syntactic structure patterns, however, ultimately point to Jordan being the author of both passages. Since a majority of the tests at least point to Sanderson being the author of the excerpt featuring Mat, we will conclude this result to be true. As the tests surrounding the passage featuring Egwene were split more evenly between the two possible authors, we may not

be able to conclude who the true author of this passage is without further testing. Considering that we know that Sanderson did in reality write the passage about Mat entirely on his own and that Jordan at least had substantial notes and drafts completed about Egwene, the outcome of these tests were, for the most part, accurate.

There may be a variety of elements influencing the outcome of these tests, however. For instance, the passage from Jordan's original excerpt had a little less dialogue than the other passages. Since the syntax of dialogue naturally differs from the descriptive syntax of narration, it is possible that this selection may have skewed the results. Additionally, most of the passages featured different characters in different situations, and the nature of these characters and situations may have affected the overall tone and construction of these passages. For instance, the passage in Hinderstap about Mat in *The Gathering Storm* is an intense scene where a "bubble of evil" has taken over a town and caused all the residents to attack each other like feral animals, leaving Mat and his party to fight for their lives to escape the madness. The other passages featured significantly less violent action, thus potentially requiring a different syntax.

The tests, of course, were naturally a little more difficult to determine results from because Sanderson was purposely trying to imitate Jordan's style. This difficulty was reflected in the indeterminate results surrounding the passage about Egwene. Because so many of the tests on this text were split between determining Jordan or Sanderson as the author, it is difficult to conclude whether these results came because Sanderson added on to text already written by Jordan or whether Sanderson was successfully able to imitate Jordan's writing style in various linguistic aspects. Finally, the limited nature of this study leaves additional questions and room for error. Perhaps a few more detailed tests, an analysis of a few more passages, or even an

analysis of the full novels may have led to more accurate conclusions about which author authored which passages within *The Gathering Storm*.

Conclusion

Despite some limitations and some inconclusive results, this study has proven useful in determining some of the limitations and capabilities of authorial attribution within forensic linguistics. Because various cases in the field of forensic linguistics are bound to deal with purposeful disguise and mimicry, linguists should understand the inherent difficulty of detecting authorship in such complicated cases. When dealing with mimicry or disguise, linguists should be careful to use multiple tests and to analyze the results of these tests extensively, perhaps relying a little more on their human intuition than the numbers that computer programs will spit out.

References

- Ai, Haiyang & Lu, Xiaofei (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In Ana Díaz-Negrillo, Nicolas Ballier, and Paul Thompson (eds.), *Automatic Treatment and Analysis of Learner Corpus Data*, pp. 249–264. Amsterdam/Philadelphia: John Benjamins.
- Binongo, Jose Nilo G. (2003). Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. *Chance*, 16(2), 9–17. <http://dh.obdurodon.org/Binongo-Chance.pdf>
- Chaski, Carole E. (2001). Empirical evaluations of language-based author identification techniques. *Forensic Linguistics*, 8(1), pp. 1–65. University of Birmingham Press.
- Grant, Tim. (2007). Quantifying evidence in forensic authorship analysis. *The International Journal of Speech, Language and the Law*, 14(1), 1–25. Equinox Publishing.
doi:10.1558/ijssl.v14i1.1
- Guillén-Nieto, V., Vargas-Sierra, C., Pardiño-Juan, M., Martínez-Barco, P., & Suárez-Cueto, A. (2008). Exploring state-of-the-art software for forensic authorship identification. *International Journal of English Studies*, 8(1), 1–28.
doi://dx.doi.org.erl.lib.byu.edu/10.6018/ijes.8.1.49071
- Juola, Patrick. (August 20, 2013). How a computer program helped show J.K. Rowling write *A Cuckoo's Calling*. *Scientific American*. <https://www.scientificamerican.com/article/how-a-computer-program-helped-show-jk-rowling-write-a-cuckoos-calling/>
- Juola, Patrick. (2015). The Rowling case: A proposed standard analytic protocol for authorship questions. *Digital Scholarship in the Humanities*, 30, i100-i113. Oxford University Press.
doi:10.1093/llc/fqv040

- Lu, Xiaofei (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4):474–496.
- Lu, Xiaofei (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers's language development. *TESOL Quarterly*, 45(1):36–62.
- Lu, Xiaofei & Ai, Haiyang. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27.
- Marko, K. (2017). Strategies for disguise in written threatening communications and ransom demands: An analysis of English and German texts. *International Journal of Speech, Language and the Law*, 24(2), 243–247. doi:10.1558/ijssl.35084
- Robert Jordan. (2019, January 07). Retrieved March 2, 2019, from https://en.wikipedia.org/wiki/Robert_Jordan
- Sanderson, B. (2013, October 15). The wheel of time retrospective: The gathering storm: Writing process. [blog post]. Retrieved from <https://brandonsanderson.com/the-wheel-of-time-retrospective-the-gathering-storm-writing-process/>
- The Stormlight Archive. (2019, March 03). Retrieved March 4, 2019, from https://en.wikipedia.org/wiki/The_Stormlight_Archive
- The Wheel of Time. (2019, February 08). Retrieved March 2, 2019, from https://en.wikipedia.org/wiki/The_Wheel_of_Time
- Turell, M. Teresa. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison. *The International Journal of Speech, Language and the Law*, 17(2), 211–250. Equinox Publishing. doi:10.1558/ijssl.v17i2.211