The Use Of Superlatives And Degree Boosters In Jane Austen's Novels:

Applying Corpus Methods To The Study Of Literature

Luz Andrea Alvariño Gabeiras

English Department

Albert-Ludwigs-Universität Freiburg

Corpus-Linguistic Methods

Table of Contents

The Use Of Superlatives And Degree Boosters In Jane Austen's Novels: Applying

Corpus Methods To The Study Of Literature

1. Introduction

The use of exaggerated characters as a tool for parody and ironic writing has been studied as one of the main defining features of Jane Austen's fiction (Harding 1968; Morini 2016; Mudrick 2018); with authors like Wijitsopon (2013) having found evidence of degree boosters being among the most frequent semantic categories in the author's writing. Following this, the present paper will attempt to bridge the gap between corpus methods and literary research, in order to answer the initial research question of whether Austen's use of hyperbolic language for characterization has resulted in a higher number of degree boosters and maximizers in her novels when compared to her contemporaries, as the previous literature seems to suggest.

In order to fulfill this aim, the structure of the paper will be as follows: Chapter 2 will serve as a theoretical background to Austen's literary style and features and the grammatical features related to degree boosting, as well as some of the research in the field of corpus stylistics. In Chapter 3, the methodology applied in the case study will be thoroughly explained, and its results subsequently presented in Chapters 4 and 5. Lastly, some concluding remarks and a summary of the findings will appear in Chapter 6.

2. Literature Review

2.1. Austen's Irony, Caricature and Exaggeration

Jane Austen (1775-1817)[1] is widely regarded as one of the most influential figures of the turn of the nineteenth century, alongside authors like "Fanny Burney, Maria Edgeworth, and Elizabeth Inchbald" (Morini 2006: 81), who also lived and wrote in the period spanning from 1750 to 1850. However, Austen possessed a series of distinctive characteristics that may have set her apart from her contemporaries, such as her wit and affinity for parodying the popular literary genres of her time.

Irony, parody and most recently, caricature, have been studied as some of the major features of Jane Austen's novels. In *Northanger Abbey* and *Sense and Sensibility*, irony is prevalent and central to the meaning of the story: where *Northanger Abbey* parodies the gothic novel, *Sense and Sensibility* subtly pokes fun at the novel of sensibility (Mudrick 2018: 28, 46). The effect is accomplished through what Morini (2016) calls 'ventriloquizing,' that is, ideas are presented from the mouth of one or several characters, who are made to embody a series of characteristics and then put up for ridicule (Morini 2016: 34–35). These exaggerated characters have also been denominated 'caricatures,' standing in opposition to normal characters, who are "full and natural portraits of imaginable people" (Harding 1968: 83). Austen herself refers to certain characters as 'caricatures' in her own work, for example: "*Mrs. John Dashwood was a strong **caricature** of himself;—more narrow-minded and selfish*" (Austen 2005: 5–6 as cited in Ferguson 2018; emphasis added). Moreover, as the effect of caricature is accomplished through "exaggeration" and "ironic mockery" (Harding 1968: 84), it could be suggested that Austen's use of hyperbolic caricatures could result in a higher rate of degree boosting words in the author's works (Flynn & Katz 2017).

---

[1] (Southam 2019; Warren 2018).

2.2. Grammatical Aspects

'Degree boosting words' can be divided into intensifiers (e.g. *completely*), and comparative or superlative adjectives and adverbs. Quirk et.al. classify intensifiers within the broader category of 'adverbials,' "a functional category" not to be mistaken with 'adverb phrases[2].' (Quirk et al. 1985: 49, 52) Within adverbials, subjuncts can be further subdivided in emphasizers and intensifiers; the major difference between them being that emphasizers "do not require that the constituent concerned should be gradable," whereas when "the constituent emphasized is indeed gradable, the adverbial takes on the force of an intensifier" (Quirk et al. 1985: 583). Quirk et.al. distinguish two further types of intensifiers, that can be observed in Figure 1.
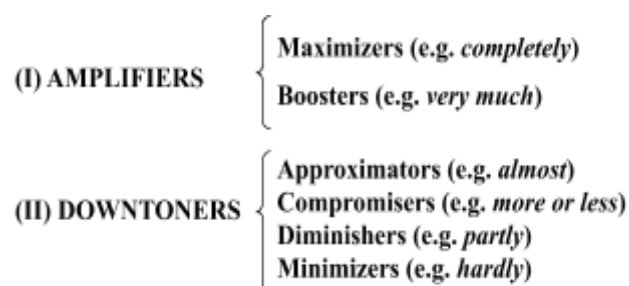

Figure 1. Subtypes of intensifiers. Reprinted from *A Comprehensive Grammar of the English Language* (589-590), by R. Quirk et.al. Copyright [1985] by Longman.

In the present paper, the focus will be strictly on 'amplifiers,' as the main topic is exaggerated language. Examples of 'maximizers' and 'boosters' can be found in Table 1.

---

[2] "A formal category, whose members frequently function as adverbials" (Quirk et al. 1985: 49).

Table 1

Examples of Maximizers and Boosters

Type of Intensifier	Examples
Maximizers	*absolutely, altogether, completely, entirely, extremely, fully, perfectly, quite … thoroughly, totally, utterly; in all respects;* the intensifying use *of most*
Boosters	*very, badly, bitterly, deeply, enormously, far, greatly, heartily, highly, intensely, much, severely, so, strongly, terribly, violently, well; a great deal, a good deal, a lot, by far;* exclamatory *how;* the intensifying use of *more*

Note: Examples taken from Quirk et.al. (1985: 591).

In addition, most adjectives and some adverbs are "gradable and … denote comparative and superlative degree" (Biber et al. 1999: 521). The marking can be done 'inflectionally' by using "the suffix *-er* for the comparative and *-est* for the superlative" (Bauer 2014: 51) or 'phrasally,' with the "degree adverbs" *more* and *most* (Biber et al. 1999: 521) as seen in Table 2.

Table 2

Marking of Gradable Adjectives

	Comparative	Superlative
Inflectional	*bigger*	*biggest*
Phrasal	*more hard-working*	*most hard-working*

Whether a higher rate of either intensifiers or marked adjectives can be found in Austen's literature –relating to her use of exaggeration in irony– is hardly possible to prove without relying on the methodological approaches native to corpus linguistics.

2.3. Corpus Methods in Literary Research

Corpus linguistics can be described as "a set of procedures, or methods, for studying language," based around "any collection of more than one text" (McEnery & Hardie 2011: 1–2; McEnery & Wilson 2001: 30), most commonly in "machine-readable form" (McEnery & Wilson 2001: 30). Rather than being constrained to a single area of language, "corpora have been shown to be highly useful in a range of areas of linguistics," including "stylistics and text linguistics" (McEnery & Hardie 2013: 731; McEnery & Wilson 2001: 117).

'Stylistics' refers to the study of a set of textual characteristics belonging to a single author or genre, referred to as 'style,'

> based on the assumption that authors have a choice between different ways of putting
> things, for instance … between using long and short sentences." (McEnery & Wilson
> 2001: 117)

In the last decade, computational sciences have made mass corpora possible and easily accessible, giving an "opportunity for corpora to play a role in literary criticism[3]" (Louw 2014: 240), and giving way to the emergence and popularization of 'corpus stylistics.' (Leech & Short 2007; Mahlberg 2013; Mahlberg 2014)

Some corpus stylistics studies have been conducted on Austen's own work. Firstly, Fischer-Starcke's (2010) compared a Jane Austen corpus to a reference corpus of her contemporaries in a series of separate studies, which used corpus techniques like "keyword analysis or "phraseological research;" (Fischer-Starcke 2010: 36) coming to conclusions like the characterization of "the protagonists in [*Northanger Abbey*] … by way of their reading habits," or "the dominant role of irony in [*Northanger Abbey*] as shown through its "topic textuality" (Fischer-Starcke 2010: 199). In a similar way, Wijitsopon (2013) examined the

---

[3] For a more complete review of corpus-based literary research see, for instance the bibliography by Mahlberg & Wynne (2006), as well as Biber (2011).

'lexical patterns' of Austen's novels, using "'keyness', 'collocation' and "cluster'" analysis (Wijitsopon 2013: 44) on a Jane Austen corpus, and compared it to two reference corpora: the BNC, and a contemporary corpus; finding that degree boosters ranked as the most frequent semantic field in Austen's work, and that degree adverbs were the most common grammatical categories, pointing at the author's usage of "exaggerated discourse … [which] is likely to encourage readers to feel that the part of the text they are reading cannot be interpreted at face value" (Wijitsopon 2013: 52). Wijitsopon (2018) and Ostade (2014) have also elaborated additional corpus studies of both Austen's novels and her letters respectively.

3. Methods and Data

The methods applied in the present study drew on the techniques of inter-textual corpus stylistics. A corpus of Jane Austen's novels was compared to a reference corpus of her contemporaries, and subsequently analyzed using the software tools TagAnt and AntConc, as well as RStudio (Anthony 2015; Anthony 2019; RStudio Team 2016). The dependent variable studied was the authors' use of degree boosters and maximizers. To equalize some of the independent variables, only female contemporary authors and fiction novels were used as a reference corpus, the remaining independent variable being the differences in authorship and style.

The *Jane Austen* corpus selected was compiled by Sinclair and Rockwell[4] (2020) and contains the following of her novels:

- *Love And Friendship* (1790).

- *Lady Susan* (1805).

- *Sense and Sensibility* (1811).

- *Pride and Prejudice* (1813).

- *Mansfield Park* (1814).

- *Emma* (1815).

- *Northanger Abbey* (1818).

- *Persuasion* (1818).

For the reference corpus, henceforth called *Contemporaries* corpus, fiction novels were selected from three of Austen's contemporaries: Frances Burney, Maria Edgeworth and Elizabeth Inchbald (Morini 2006: 81). From Burney:

- *Evelina* (1778).

---

[4] Accessible online at https://voyant-tools.org/?corpus=austen.

- *Cecilia* (1782).

- *Camilla* (1796).

- *The Wanderer* (1814).

From Edgeworth:

- *Belinda* (1801).

- *Leonora* (1806).

And finally, from Inchbald:

- *A Simple Story* (1791).

- *Nature and Art* (1796).

All the texts in the *Contemporaries* corpus were obtained through The Online Books Page, edited by Ockerbloom (2006), and supported by the University of Pennsylvania. In table 3, a comparison of corpora's token size corpus can be observed.

Table 3

Number of Tokens per Corpora

Corpus	Tokens
Jane Austen	864782 tokens
Contemporaries	15766689 tokens

All the texts were downloaded in TXT format, and tagged using Laurence Anthony's 'TagAnt' software (2015), which uses TreeTagger to annotate text with Part-Of-Speech (POS) information. The tagged corpora were then processed using the corpus analysis tool AntConc in version 3.5.8, (Anthony 2020) two of its tools in particular:

a. **Keyness.** Keyword lists can show which words are statistically overrepresented in the target corpus when compared to a reference corpus.. First, the *Jane Austen* corpus was loaded onto AntConc, then, the *Contemporaries* corpus was loaded as a reference corpus using the tool preferences (Figure 2) and the statistical constraints were selected, in this case, using the default settings.
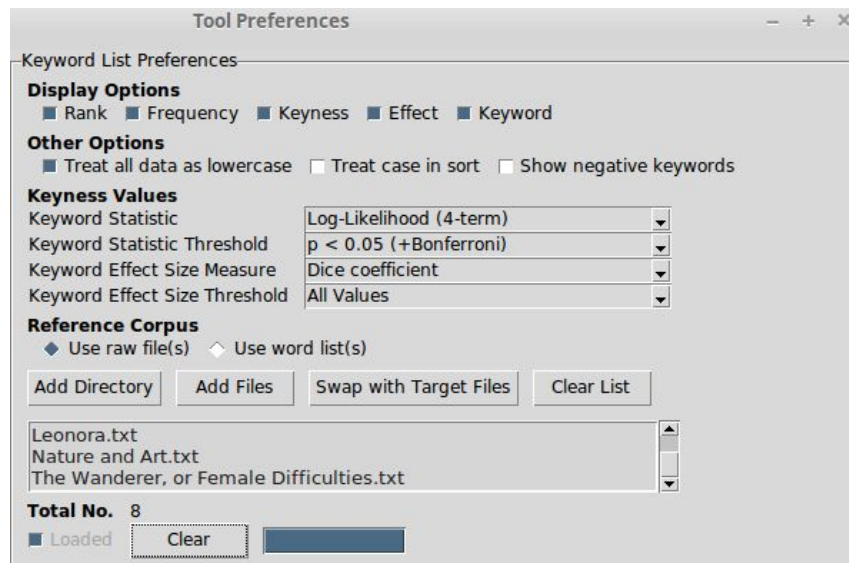

Figure 2. *Keyword List Preferences in AntConc.*

The resulting keyword list was then manually processed to find overrepresented degree boosting words.

b. **Word List.** The Word List tool was used to find comparative and superlative adjectives. The same process was applied to both corpora; the tagged corpus was loaded onto AntConc, and the token definition altered to append the tag symbol '_' using the global settings, then three different search queries were conducted. First, comparative adjectives and adverbs were searched using the lines '*er_JJR,' which returned the list shown in Figure 3, and '*er_RBR'[5]. This process was replicated for superlative adjectives and adverbs using the lines '*est_JJS' and '*est_RBS.'

---

[5] Full list of TreeTagger tags available online at
https://www.laurenceanthony.net/software/tagant/resources/treetagger_tagset.pdf.

Secondly, the boosters and maximizers suggested by Quirk et.al. (see Table 1 in Chapter 2) were individually looked up; for additional maximizers and boosters, the line'*ly_RB' was utilized. For all queries obtained with the Word List tool, the results were sorted by word end and found using the 'Search Only tool. The list contained both the word and its frequency in the corpus.
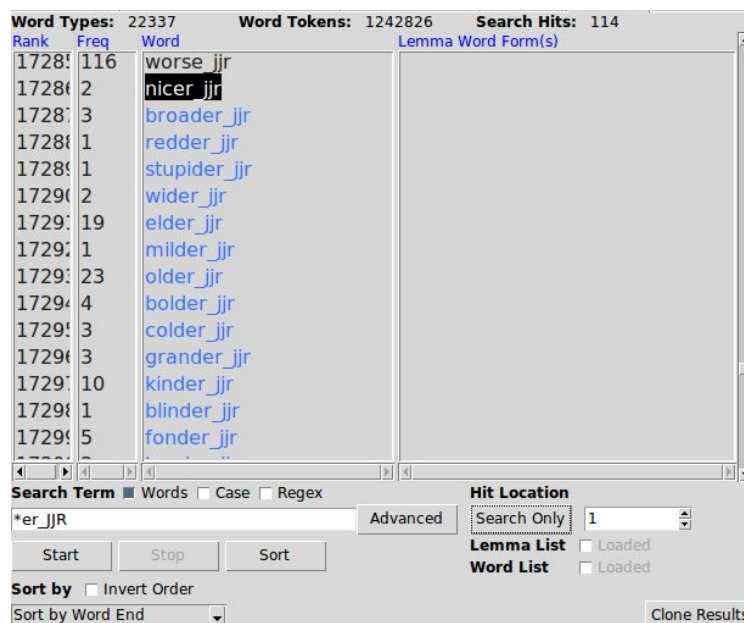

Figure 3. *Example of Word List obtained using AntConc.*

The results obtained through the Word List tool were recorded in a data collection table with seven columns, as shown in Figure 4.

corpus	ntoken	word	POS	degree	process	rawfreq
Jane Austen	864782	Brighter	_adj	Comparative	Marking	6
Contemporaries	15766689	Vastly	_adv	Booster	Intensifier	86

Figure 4. *Example Of Unprocessed Data Collection Table Items.*

Once all words were recorded in the table, it was exported as a comma separated CSV file, and processed with RStudio. The columns 'word' and 'POS' were combined, and an additional 'relfreq' column was added, using the formula:

$$rawfreq \; x \; 10,000 \; / \; ntoken \; = \; relfreq$$

Which calculates the relative frequency of each of the vectors in the table by ten thousand

words. The resulting table can be seen below:

	corpus	ntoken	word	degree	process	rawfreq	relfreq
150	Jane Austen	864782	Brighter_adj	Comparative	Marking	6	0.06938165

Figure 5. *Example Of Processed Data Collection Table Item.*

Using the relative frequency of each of the words, the frequency of degree boosting

words in each of the corpora was compared, as well as the prominence of each of the studied

processes.

4. Analysis and Results

Beginning with keyness, at least twelve adverbial intensifiers were found among the overrepresented words in the *Jane Austen* corpus extracted with the Keyword List tool (Table 5).

Table 5

Overrepresented Intensifiers in the Austen Corpus

Rank	Word	Keyness (log-likelihood)
3	*very*	3940
53	*quite*	904
105	*particularly*	191
110	*well*	1597
111	*altogether*	99
119	*exceedingly*	103
120	*certainly*	440
148	*perfectly*	293
216	*thoroughly*	73
512	*highly*	115
560	*especially*	117
593	*undoubtedly*	52

Of the whole corpus, *Very* was the third most disproportionately frequent word, with a keyness of 3940. *Quite* was also among the top hundred overrepresented words, with 904.

Moving onto the data recorded using the Word List function; first, the relation between the mean raw frequency of the total degree words for each corpus and the corpus size rendered a p-value < 0.00001[6], calculated with the chi-square test in RStudio. Secondly, in Table 4, we can see a summary of the raw counts for each process and the total of degree boosting words in each corpora, as well as the mean relative frequency of the words, to contextualize the findings in terms of the difference in size between corpora.

---

[6] Significant at $p < .05$.

Table 4

Degree Boosting Words Found in Each Corpora

	Jane Austen Corpus		Contemporaries Corpus	
	Raw. Count	Mean Rel. Freq.	Raw. Count	Mean Rel. Freq.
Suffix Marking	*240*	*0.17*	*311*	*0.008*
Adverbial Intensifier	*48*	*3.42*	*56*	*0.220*
Total Nº of Words	*288*	*0.71*	*367*	*0.041*

Although the raw counts were higher for the *Contemporaries* corpus, the relative frequency was much higher in the *Jane Austen* corpus (Figure 6).
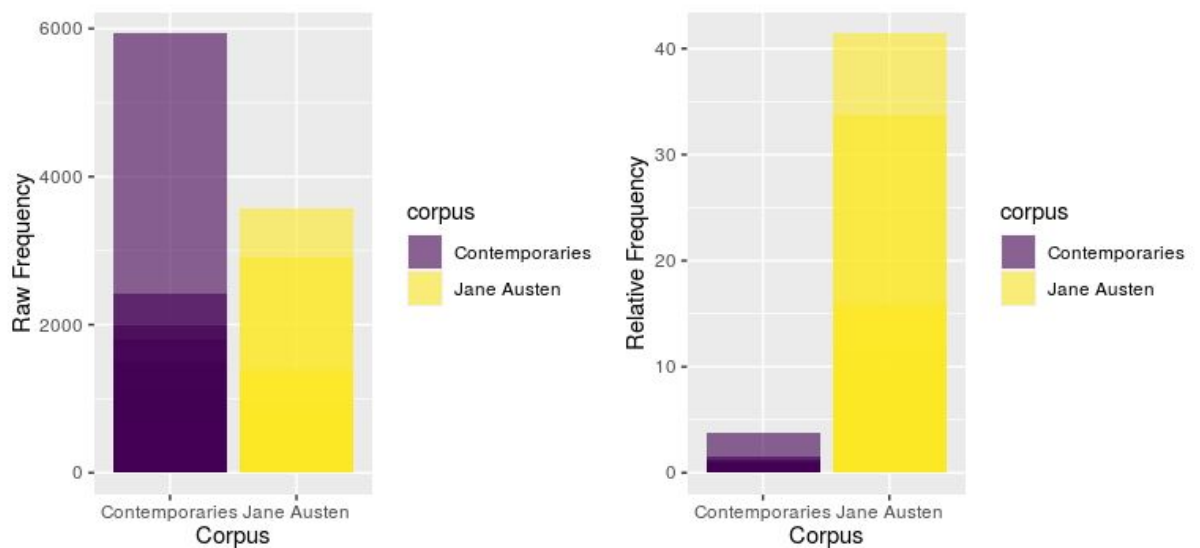
Figure 6. *Barplot of the Proportion Of Degree Boosting Words Across Corpora.*

As observed in Table 4, adverbial intensifiers in the *Jane Austen* corpus had the highest relative frequency. When further subdividing that the respective sub-processes we found that relative frequencies were higher in the *Jane Austen* corpus overall (observe differences in the Y axis) and that maximizers were more frequent in both corpora, (Figure 7) although the difference is hard to see due to a few outliers with uncommonly high relative frequencies (i.e. *very, most,* and *more*).
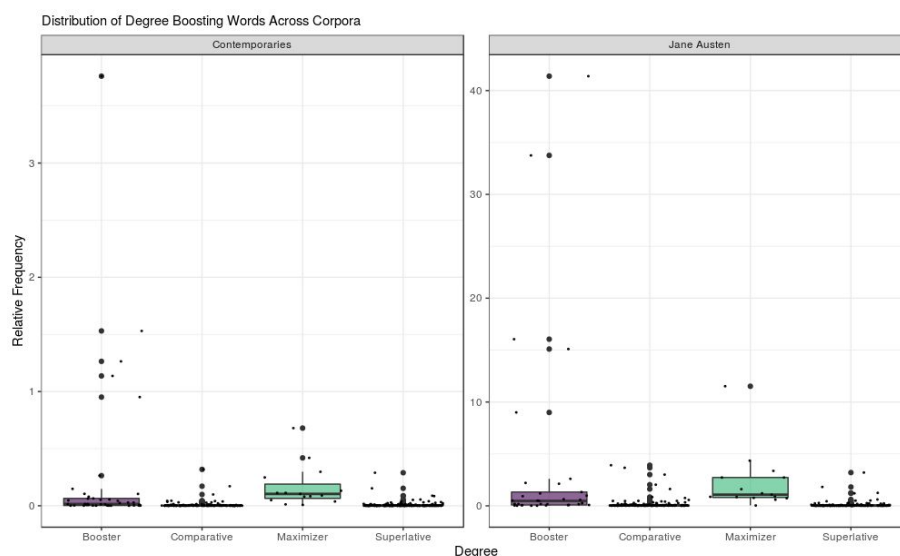
Figure 7. *Relative Frequency Proportion Boxplot for Each Degree Boosting Process in the Contemporaries and Jane Austen Corpus.*
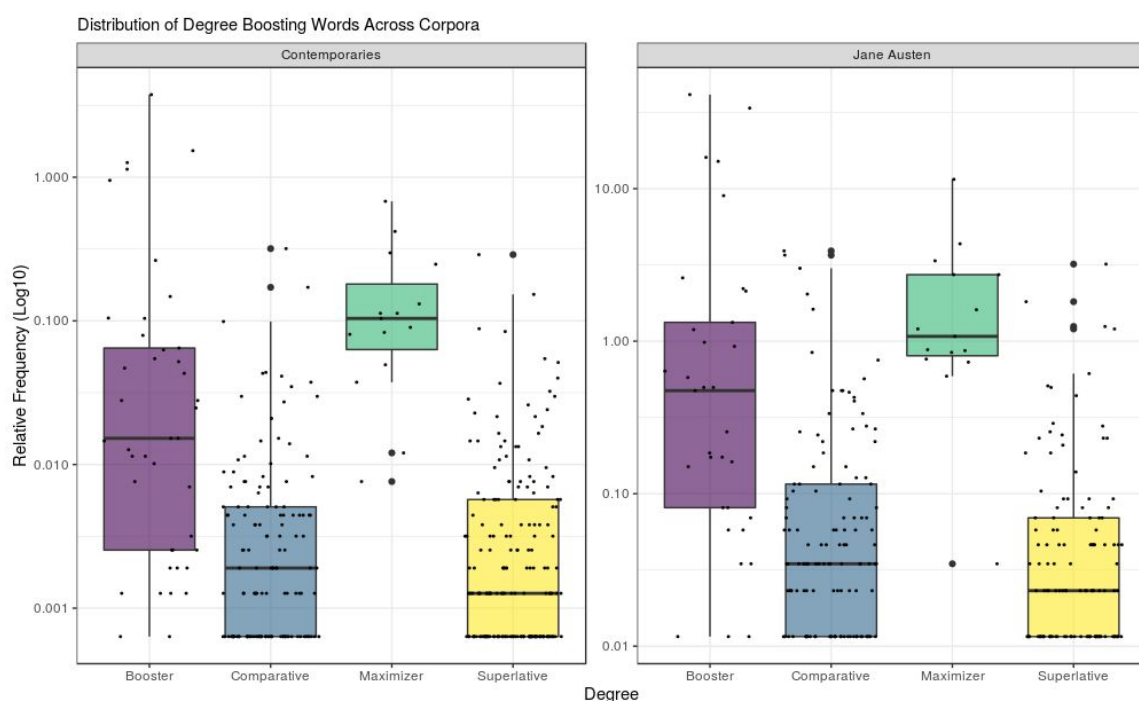

Figure 8. *Relative Frequency Proportion Boxplot for Each Degree Boosting Process in the Contemporaries and Jane Austen Corpus.* Log10 Frequency.

As presented in Figure 8, a Log10 scale was applied to the proportion to solve the distortion, which led to a graph where the higher frequency of adverbial maximizers is easier to observe.

5. Discussion

As was introduced in Chapter 2, irony, parody and caricature are salient characteristics of Austen's literary work, and it has been suggested that exaggeration and hyperbole play a substantial role in achieving such techniques. The initial research question was whether this literary use of exaggeration could lead to an increased number of degree boosting and maximizing words, as shown through a comparison of Austen and some of her contemporary authors. Substantial evidence was found to at least partially answer this research question:

Not only at least twelve adverbial intensifiers were among the most overrepresented words in the *Jane Austen* corpus when compared to the *Contemporaries* corpus using keyness, but when analyzing raw and relative frequencies it was found that all degree boosting adjectives and adverbs were overall more common in Austen than in her contemporaries. This correlates with Wijitsopon (2013), who found evidence of degree boosters being the most common semantic field in Austen's novels and degree adverbs being some of the most common words in the texts.

The findings could be explained by Austen's stylistic use of exaggeration in irony, which would support the hypothesis that literary features can indeed be observed through grammatical analysis.

6. Conclusions

The initial aim of this study was to find evidence that degree boosting and maximizing words are more common in Jane Austen's novels than her contemporaries. Through the use of corpus tools such as AntConc and TagAnt, as well as statistical processing in RStudio, we found that:

- Several adverbial intensifiers are disproportionately frequent in Austen's novels when compared to her contemporaries.

- All degree boosting and maximizing words, including comparative and superlative adjectives and adverbs, as well as adverbial intensifiers are more common in Austen's novels than in her contemporaries' when their relative frequency is calculated in relation to the size of both corpora.

These findings partially support the idea that the literary use of exaggeration could lead to the increased frequency of certain grammatical characteristics in narrative texts, as Austen's use of irony could explain her increased rate of degree boosting words. However, ambiguous words such as *unspeakably* or *violently*, which can occur as degree booster adverbials but also take on other meanings, had to be excluded from the study; as well as downtoners (i.e. diminishers and minimizers) like *least*, which might also be used for hyperbolic purposes. Implementing a more inclusive and qualitative approach could improve the reliability of the findings presented in this study, as looking at the concordance of degree adjectives and adverbs could confirm whether they are applied to characters presented as 'caricatures,' as we saw on Chapter 2, or rather, they're just contributing to an overall hyperbolic tone of Austen's literary style.

Still, studies such as the present one continue to show a promising future for the use of corpus methods in literary analysis and stylistics, supporting claims that would otherwise rely on the manual close reading of texts, leading to less reliable results.

References

Anthony, Laurence. 2015. *TagAnt*. [Software] Linux. Tokyo: Waseda University. https://www.laurenceanthony.net/software.

Anthony, Laurence. 2019. *AntConc*. [Software] Linux. Tokyo: Waseda University. https://www.laurenceanthony.net/software.

Austen, Jane. 2005. *The Cambridge Edition of the Works of Jane Austen*. (Ed.) Peter Sabor, Barbara M. Benedict, Deirdre LeFaye, Edward Copeland, Pat Rogers, John Wiltshire, Richard Cronin, et al. Cambridge: Cambridge University Press.

Bauer, Laurie. 2014. *Watching English Change: An Introduction to the Study of Linguistic Change in Standard Englishes in the 20th Century*. Amsterdam: Routledge.

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

Fischer-Starcke, Bettina. 2010. *Corpus Linguistics in Literary Analysis: Jane Austen and her Contemporaries*. London: Bloomsbury Publishing.

Flynn, Kathleen A. & Josh Katz. 2017. The Word Choices That Explain Why Jane Austen Endures. *The New York Times*, sec. The Upshot. https://www.nytimes.com/2017/07/06/upshot/the-word-choices-that-explain-why-jane-austen-endures.html (18 April, 2020).

Harding, Denys, W. 1968. Character and Caricature in Jane Austen. In *Critical essays on Jane Austen*, 83–105. London: Routledge & K. Paul.

Leech, Geoffrey N. & Mick Short. 2007. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London: Pearson Education.

Louw, Bill. 2014. The Role of Corpora in Critical Literary Appreciation. In Anne Wichmann & Steven Fligelstone (eds.), *Teaching and Language Corpora*, 240–251. Amsterdam: Routledge.

Mahlberg, Michaela. 2013. *Corpus Stylistics and Dickens's Fiction*. Amsterdam: Routledge.

Mahlberg, Michaela. 2014. Corpus Stylistics. In Michael Burke (ed.), *The Routledge Handbook of Stylistics*, 378–392. Amsterdam: Routledge Handbooks.

McEnery, Tony & Andrew Hardie. 2011. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

McEnery, Tony & Andrew Hardie. 2013. The History of Corpus Linguistics. In Keith Allan (ed.), *The Oxford Handbook of the History of Linguistics*, 727–747. Oxford: Oxford University Press.

McEnery, Tony & Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Morini, Massimiliano. 2016. *Jane Austen's Narrative Techniques: A Stylistic and Pragmatic Analysis*. Amsterdam: Routledge.

Mudrick, Marvin. 2018. *Jane Austen: Irony as Defense and Discovery*. Great Barrington: Berkshire Publishing Group.

Ostade, Ingrid Tieken-Boon van. 2014. *In Search of Jane Austen: The Language of the Letters*. New York: Oxford University Press USA.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive grammar of the English language*. Harlow: Longman.

RStudio Team. 2016. *RStudio: Integrated Development for R. RStudio*. [Software] Boston: RStudio Inc. http://www.rstudio.com/.

Southam, Brian C. 2019. Jane Austen. *Encyclopædia Britannica*. https://www.britannica.com/biography/Jane-Austen (8 April, 2020).

Warren, Renee. 2018. Jane Austen Biography. *Jane Austen*. https://www.janeausten.org/jane-austen-biography.asp (21 April, 2020).

Wijitsopon, Raksangob. 2013. A Corpus-Based Study of the Style in Jane Austen's Novels. *Manusya: Journal of Humanities* 16(1). 41–64.

Wijitsopon, Raksangob. 2018. A Corpus Stylistic Study of Singular and Plural Keywords in Jane Austen's Persuasion. *Kemanusiaan the Asian Journal of Humanities* 25. 1–22.

Corpus References

Ockerbloom, John Mark. 2006a. *Burney, Fanny, 1752-1840*. The Online Books Page. http://onlinebooks.library.upenn.edu/webbin/gutbook/author?name=Burney%2C%20Fanny%2C%201752-1840 (23 April, 2020).

Ockerbloom, John Mark. 2006b. *Edgeworth, Maria, 1767-1849*. The Online Books Page. https://onlinebooks.library.upenn.edu/webbin/book/lookupname?key=Edgeworth%2c%20Maria%2c%201767%2d1849&c=x (23 April, 2020).

Ockerbloom, John Mark. 2006c. *Mrs. Inchbald (Inchbald, Mrs., 1753-1821)*. The Online Books Page. https://onlinebooks.library.upenn.edu/webbin/book/lookupname?key=Inchbald%2c%20Mrs%2e%2c%201753%2d1821&c=x (23 April, 2020).

Sinclair, Stéfan & Geoffrey Rockwell. 2020. *Austen Corpus*. Voyant Tools. https://voyant-tools.org/?corpus=austen (21 April, 2020).