INNERVATE
Leading student work in English studies

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

School of English

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19ᵗʰ Century fiction.

1. Introduction:

With the advent of personal computers in the 1990s, and several new corpus tools and resources, Corpus Linguistics has become further equipped for making innovative contributions to the study of literature. Headed under the banner of Corpus Stylistics, the application of this relatively new methodology provides a means of examining a literary text as a whole, drawing conclusions that can support, confirm or refute critique based on close textual analysis. Corpus linguistics studies naturally recurring language through the use of a 'corpus', a body of language pertaining to relevant criteria.

Although this approach has made it possible to examine literature on a larger scale, by using the new CLiC Tool that makes it possible to search in-quotes, this article aims to make a focus on dialogue. Bringing together Corpus work on pragmatics and Sociolinguistic work on politeness, the study will explore the form and function of polite language, taking a look at this in a reference corpus of 19ᵗʰ Century fiction and primary corpus of *Pride and Prejudice*. This article makes reference to the findings of an earlier study that uses keyword analysis to explore *civility* as a thematic component in the novel (Mahlberg and Smith 2010). The article builds from their keyword analysis by looking at recurring patterns in dialogue, classifying them for elements of politeness and then dividing them into functional sub-groups. The aim of this investigation is to find further explanation for Mahlberg and Smith's findings, which indicate that despite the social expectations associated with *civility*, it can be used "strategically" (Mahlberg and Smith 2010: 453) with sometimes only the "appearance of goodness" (ibid: 454). Given that politeness also appears in the

Jessica Mateo

keyword list, and features as part of Emsley's (2005) definition of civility, quoted in their article (Mahlberg and Smith 2010: 454), a study of politeness formulae appears to be the most natural next step.

This article begins with a discussion of the background to Corpus linguistic work of formulaic language, Sociolinguistic politeness strategies, and outlines past studies related to politeness functions. Following this, I will summarize the method and discuss the results. Finally, I will conclude the article, summarizing the key points.

2. Politeness: Defining and Identifying the Elusive Concept.

The foundations for this research stem from three areas: Corpus approaches to phraseology and links to politeness; Sociolinguistic work on politeness strategies; and work into the labelling of functions. Links *Pride and Prejudice* share with politeness have been identified by various critics such as Michaela and Smith's (2010) who refer to, Emsley's definition of civility as telling of how closely this novel and politeness are linked. Writing that "[c]ivility clearly has a great deal to do [with] maintaining social niceties even when one does not feel like being polite" (2005: 91), suggests an atmosphere of obligation in the novel. This mirrors the economic context outlined by Markley (2013: 80), that revolves around "money, inheritance law, and the values and obligations". Copeland refers to this as a "fear of debt" driving the novel (1997: 138). With politeness and civility as the "outward manifestation of real goodness" (Emsley 2005: 90, qtd. in Mahlberg and Smith 2010: 454), politeness strategies used in a society that is all about correct behaviour, may accordingly reflect equally false presentations. Putting on a politeness 'mask' may have allowed the female characters to pursue rich husbands.

2.1. Corpus Approaches to Politeness, Pragmatics and Phraseology

Although Corpus linguistic approaches towards politeness are scarce, there has already been some linkage made between politeness formulae and multi-word units, or 'clusters', as I will refer to them. Although politeness is a very "elusive" concept (Watts et al., 1992: 11 qtd. in Mills 2003: 57) that is difficult to sufficiently define, most of us will be certain that we can recognise what is 'polite' (Watts 2003: 1). Thus, most work into politeness could be considered subjective. However, there are 'functions' that people would likely agree on. Many studies refer to this as the "expressions" of politeness (McCarthy and Carter 2006: 18; Altenberg 1998: 109; and Adolphs and Carter 2013: 33). Few actually refer to 'politeness formulae' directly (Altenberg 1998: 110; and Mahlberg 2013: 80). Identification of certain word combinations such as, *if I may* that have been identified to "act" as politeness formulae (Altenberg 1998: 110), makes way to suggest some stronger connection between form and function, where the form may supersede function in personal assessment of politeness when they are used in abundance. In this way sarcasm and mock politeness can be performed, where the form and function don't traditionally match up.

Many studies have explored formulaic language, having revealed, "much of our lexical output consists of multi-word units" based on preferences of expression that language users commonly adopt (Adolphs and Carter 2013: 23). Different studies use different terminology to describe this, but ultimately agree that they can work like "building blocks" (Mahlberg 2012: 79; Biber 2002: 444) that can create fictional worlds. 'Lexical-bundles' (Biber, Conrad and Leech 2002: 444), as a variation of clusters, work with criteria making them more statistically significant, while clusters work with a top-down approach, that samples based on

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19th Century fiction.

Jessica Mateo

highest to lowest occurring frequency (see Mahlberg 2013, chapter 3) making them more appropriate for literature.

2.2. A Sociolinguistic Perspective of Politeness

Sociolinguistic research has taken a far deeper approach to politeness than Corpus linguistics, though not always in terms of multi-word expressions. Although Brown and Levinson (1978) do refer to polite formulae (1978: 43), there is a greater emphasis on politeness strategies. Detailed instruction-like descriptions are included for each strategy, using concepts such as 'stress' (1978: 104) to suggest that any number of formulaic expressions could fulfil these functions. The over-arching message built by Brown and Levinson (1978) is that politeness revolves around avoidance of misunderstandings that could cause dispute (1978: 103-116). However, although data can be found to fit this study's model, some have criticised it for neglecting to consider the role of inference in conversation as a politeness, or impoliteness strategy (Culpeper, 1996, qtd. in Mills 2003: 83).

Mills discusses Brown and Levinson's model in detail, adding that this indicates that "politeness may function as a way of […] hiding one's real intentions" (2003: 60), presenting a metaphorical 'mask'. A connection between identity and the notion of 'face', has been touched upon by some studies (Wray, 2008: 69, qtd. in Mahlberg 2013: 50; Culpeper 2011: 13; Goffman 1967, 1999/1967, via Mills 2003: 58), which suggests that politeness and impoliteness concerns the maintaining or damaging of the hearer's sense of self (Culpeper 2011: 13).

Considering this, politeness and impoliteness may be more closely linked than we would usually assume. Impoliteness has not been explored to the same extent as politeness, but it seems apparent that if "impoliteness often involves a

clash with expectations […] [in] particular contexts" (Culpeper 2011: 14), then impoliteness is ultimately defined by what is not politeness. Even polite behaviour though can be treated negatively, characterising its users as being 'standoffish', 'haunty' or 'insincere' (Watts 2003: 1), making polite and impolite talk seem interchangeable, dependant on context and personal assessment.

2.3. Politeness in action: Functions

Despite the subjective nature of politeness, there are functions identified over and again. These can be summed up as *reassurance* (Altenberg 1978: 105; Brown and Levinson 1978: 103, 106, 112) *indirectness/vagueness/hedging* (Adolphs and Carter 2013: 33, 35; Mills 2003: 141, McCarthy and Carter 2006: 21; Brown and Levinson 1978: 116), *monitoring* (McCarthy and Carter 2006: 20), *agreement* (Altenberg 1998: 105) and *thanks* or *apology* (Mahlberg 2013: 80). These functions support Brown and Levinson's (1978) stance that politeness involves FTA-minimizing strategies, that *attend to the hearer* (103)*, exaggerate* (104) such things as approval, *intensify interest* (106) and so on.

3. Methodology

Given that some have commented on the similarities between fictional and real speech (see Mahlberg 2013: 75-76), the goal of this study is to apply the aforementioned background to the fictional dialogue of *Pride and Prejudice* and 19th Century Fiction. Thus, in the current study the following methods will be used:

   (a) keyword analysis to re-evaluate the role of *politeness* and *civility* when applied to a larger reference corpus.

Jessica Mateo

(b) identify politeness clusters from lists generated via 5-gram cluster

searches in both corpora.

(c) classify politeness clusters for function according to newly fashioned

labels.

3.1. Keywords and Concordance Lines

In order to first verify the findings this study is based on, I generated keywords

from *Pride and Prejudice* against a 19th Century reference corpus. The tool used

to accomplish this was CliC (http://clic.nottingham.ac.uk:8080/index.html)

searched for in full-text. While in Mahlberg and Smith's study they used 18

novels, excluding Austen from the reference corpus, this study includes some

Austen works, excluding only *Pride and Prejudice* itself.

After grouping keywords that are associated with politeness (see Table 1),

I then ran the three most relevant keywords (i.e. *civility*, *politeness* and

behaviour) through a concordance search (full-text) of *Pride and Prejudice*.

3.2. Clusters: Classifying for Politeness

Using a reference corpus (19CC) of 28 19th Century Novels located in the CLiC

Tool (comprised of 4,391,329 words, 1,553,986 in-quotes) I ran a search for 5-

gram clusters. 5-gram refers to clusters (i.e. repeated sequences of words)

spanning 5 words. The same process was applied to the *Pride and Prejudice*

(PrPrC) corpus (comprised of 121,747 words, 52,988 in-quotes) leaving a total of

261 Clusters for PrPrC and 4,999 clusters for the 19CC (from in-quotes).

Using a top-down approach, I classified the first 545 clusters from the

19CC cluster list. While using the entire list would be ideal, it is impractical. In

addition, the sample from 19CC needs only to be large enough for comparison.

The same process was applied to the entire PrPrC cluster list since 261 clusters
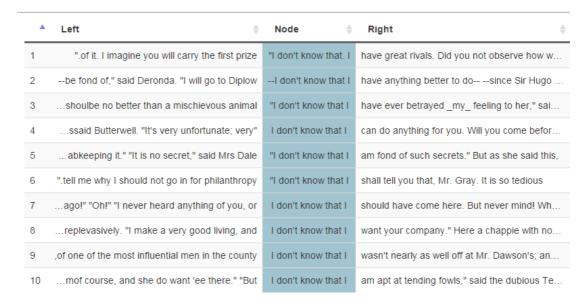
are manageable and necessary given that *Pride and Prejudice* is central to the study. Initially this classification is done on face-value, however, the more ambiguous clusters are checked using concordance lines. For example from the 19CC:

I don't know that I

Figure 1: Concordance search results in CLiC for *I don't know that I*, in the 19CC in-quotes.

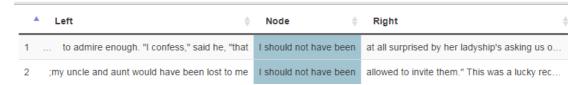|	Left	Node	Right
1	"'.of it. I imagine you will carry the first prize	"I don't know that. I	have great rivals. Did you not observe how w…
2	--be fond of," said Deronda. "I will go to Diplow	--I don't know that I	have anything better to do-- --since Sir Hugo …
3	…shoulbe no better than a mischievous animal	"I don't know that I	have ever betrayed _my_ feeling to her," sai…
4	…ssaid Butterwell. "It's very unfortunate; very"	I don't know that I	can do anything for you. Will you come befor…
5	…abkeeping it." "It is no secret," said Mrs Dale	"I don't know that I	am fond of such secrets." But as she said this,
6	"'.tell me why I should not go in for philanthropy	"I don't know that I	shall tell you that, Mr. Gray. It is so tedious
7	…ago!" "Oh!" "I never heard anything of you, or	I don't know that I	should have come here. But never mind! Wh…
8	…replevasively. "I make a very good living, and	I don't know that I	want your company." Here a chappie with no…
9	,of one of the most influential men in the county	I don't know that I	wasn't nearly as well off at Mr. Dawson's; an…
10	…mof course, and she do want 'ee there." "But	I don't know that I	am apt at tending fowls," said the dubious Te…

Units containing *I don't know* have been identified as commonly polite (Adolphs and Carter 2013: 33;, Altenberg 1998: 105, and McCarthy and Carter 2006: 20-21) and as a sign of indirectness, hedging and reassurance. Figure 1 shows this variation expresses doubt, where the speaker hedges to reduce the directness of the statement. This protects the face of the speaker and hearer.

However, not all concordances confirm politeness as shown by an example taken from the PrPrC:

I should not have been

Jessica Mateo

Figure 2: Concordance search results on CLiC for *I should not have been* in PrPrC in-quotes.

	Left	Node	Right
1	... to admire enough. "I confess," said he, "that	I should not have been	at all surprised by her ladyship's asking us o...
2	;my uncle and aunt would have been lost to me	I should not have been	allowed to invite them." This was a lucky rec...

When examining the cluster list from the 19CC, modality stood out as a common component, leading to suggest that politeness and modality may be strongly linked. However, upon comparison to PrPrC, where modality was not so striking, this gave way to double-checking modality clusters. *Should not* in Figure 2 is used in the more literal than figurative sense that is seemingly common in polite language (e.g. *I must ask you to*).

3.2.1 Cut-off points for the 19CC cluster list

Since this is a top-down approach that collects data based on highest to lowest frequency, cut-off points for the 19CC are neccessary. I used the same 28 texts, provided in .txt format from The University of Nottingham Moodle Page and ran a 5-gram cluster search in Word Smith tools. The limitations of both CliC and Word Smith Tools make it impossible for me to see how many texts a cluster occurs in while still searching in-quotes. Thus, the findings (recorded in a Table 4) are not ideal given that they are based on searches in full-text, but still enables me to judge which clusters may be too infrequent. Thus, the cut-off points placed on the 19CC politeness cluster list being a **minimum occurrence frequency of 6** and a **minimum text frequency of 3** eliminated a further 10 clusters leaving the 19CC list at 135.

3.3. Clusters: Function

Considering the work already done on politeness functions, as reviewed above, I have fashioned a list of 12 politeness function labels (see definitions in Appendix 7.1.), that I applied to the two politeness cluster lists (see Tables 3 and 4). I then used the online Log Likelihood calculator (http://ucrel.lancs.ac.uk/llwizard.html), in order to judge comparison. The p-value used by the Log Likelihood is 0.0001, with a critical value of 15.13, meaning that any result below this figure will be considered insignificant.

Using the calculator takes into account the total number of words in both corpora in full-text with the total frequency of clusters according to their sub-divided function groups. Since four of the categories could not be applied to both corpora, 13 clusters from the 19CC have been omitted from the final comparison. This leaves the study with a total of 51 PrPrC clusters and 122 19CC clusters.

4. Results and Analysis

4.1. The Thematic Components: *behaviour*, *civility* and *politeness*

As remarked by Mahlberg, a common starting point for the analysis of any novel is to begin by generating key words (2010: 295). Although most of this work has been done for me by Mahlberg and Smith (2010), their analysis paid closer attention to *civility* than *politeness*. Also, by using a larger reference corpus (28, as opposed to 18), I must verify the keyness of these words. Table 1 shows a drop from 38 to 44 for *civility* since the inclusion of additional books. However, since they remain key, the drop is not detrimental to the current study.

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19th Century fiction.

Jessica Mateo

Table 1: Keywords list grouped for association with *politeness* in *Pride and Prejudice*, against the 19th C corpus (full-text).

	Keyword	Frequency	19th C Ref. Corpus Frequency	Keyness	p-value
40	Behaviour	53	107	186.08	0.0000000000
44	Civility	42	63	165.98	0.0000000000
79	Amiable	36	175	77.2	0.0000000000
94	Manners	43	314	65.59	0.0000000000
101	Agreeable	45	359	62.81	0.0000000000
201	Politeness	17	94	33.07	0.0000000060

Table 1 reiterates Mahlberg and Smith's findings that indicate that concepts surrounding *civility* and *politeness* are major themes in the novel (2010: 455), when compared to the 19CC. Upon viewing the concordance searches (see Figure 3), these confirmed that the most remarkable occurrences were negatively charged collocates. Just as Mahlberg and Smith (2010) identified that *civility* could be; *sneering*, *cold* or *distant*, a concordance list of *politeness* revealed a similarly negative semantic prosody. Politeness could be *cold and ceremonious* (Fig. 3: line 14), *extraordinary* (line 17)*,* or an *exertion* (line 12). While collocates for *politeness* do not entirely lean towards negativity, the context in concordance lines suggests underlining feelings of obligation or duty and hidden negative inferences. For example, Elizabeth's quote: "Mr Darcy is all politeness" (line 1), carries clear undertones implying that Darcy is not saying what he really thinks. Considering that the cluster analysis (below) reveals that politeness is used to disguise true, often negative, feelings, this quote highlights that although characters use politeness to perform social expectations, hiding true feelings, Elizabeth at least appears to resent this, using the same polite guise to make an implied FTA. This is especially pronounced when aimed at Darcy who has been noted to "put truth-telling above ordinary politeness" (Todd 2013: 154). Taking into account the economic context critics have commented on, the female characters in particular are likely to feel the strain of obligation concerning the

securing of suitable matches, which obstructs direct FTAs. Instead more covert

impoliteness would be a necessary alternative, clearly evident in this example.

Figure 3: Concordance lines (1-17) for politeness in *Pride and Prejudice* (full-text)

	Left	Node	Right
1	to oblige us for one half-hour." "Mr. Darcy is all	politeness,"	said Elizabeth, smiling. "He is, indeed; but, con...
2	...her's manners there was something better than	politeness;	there was good humour and kindness. Mr. Dar...
3	...unctual to his time, and was received with great	politeness	by the whole family. Mr. Bennet indeed said littl...
4	...tion of him. She received him with her very best	politeness,	which he returned with as much more, apologi...
5	...Bennet by admiring Mrs. Phillips's manners and	politeness.	He protested that, except Lady Catherine and ...
6	...e of their entertainment, and the hospitality and	politeness	which had marked their behaviour to their gues...
7	...oved for the night; and Mrs. Bennet, with great	politeness	and cordiality, said how happy they should be t...
8	...; for Mrs. Bennet, with more perseverance than	politeness,	protested he must be entirely mistaken; and Ly...
9	lost none of his recent civility; and, to imitate his	politeness,	she began, as they met, to admire the beauty of
10	...lined, and they parted on each side with utmost	politeness.	Mr. Darcy handed the ladies into the carriage; a...
11	fault to find. They could not be untouched by his	politeness;	and had they drawn his character from their ow...
12	...gh it could not be equalled, by some exertion of	politeness	on their side; and, consequently, that it would ...
13	... is the matter?" cried he, with more feeling than	politeness;	then recollecting himself, "I will not detain you ...
14	...hen contrasted with the cold and ceremonious	politeness	of her curtsey and address to his friend. Elizab...
15	...chemes, which the good humour and common	politeness	of Bingley, in half an hour's visit, had revived. On
16	...h high importance, received her with the utmost	politeness.	After sitting for a moment in silence, she said v...
17	...cience told me that I deserved no extraordinary	politeness,	and I confess that I did not expect to receive

Showing 1 to 17 of 17 entries

It is true, as Mahlberg and Smith note (2010: 453), that *civility* can be seen as

positive (e.g. *utmost* or *perfect civility*), just as politeness can be *utmost* (line 10,

16) or *great* (line 3, 7)*.* This does present some expectation of genuine politeness

in the novel, though it does not entirely repress signs that masked impoliteness

may be present. One explanation for the fluctuating nature of these themes

showing both negative and positive occurrences could be found in the keyword

"*I should like to know…*": A corpus analysis of politeness phrases in
Pride and Prejudice and in a reference corpus of 19th Century fiction.

Jessica Mateo

list. Table 1 shows *behaviour* (keyword number 40) as the highest keyword in this category, suggesting that *behaviour*, good or bad, is the main focus of this novel, rather than *politeness* or *civility* that from the outset appear more positively than neutrally charged. After all, politeness and civility are behavioural acts that compliment this broader theme. A look at this in concordance lines (see Figure 4 in Appendix) confirms behaviour as a usually neutral concept (i.e. *general* (lines 1, 23)), often belonging to a character (i.e. *his* (lines 4,5,9,10,18, 19, 21, 26-31) and *her* (lines 7, 16, 17)).

Thus, a brief keyword analysis creates expectations that, despite the grouping of the keywords: manners (94), amiable (79), agreeable (101), behaviour (40), politeness (201) and civility (44), that produce "notions such as 'good manners', 'social etiquette', 'social graces' and minding your Ps and Qs'" (Culpeper 2011: 36), are strikingly more negative, yet indirect than would be clear at first.

4.2. Politeness Functions

From applying the newly fashioned politeness function labels (see Tables 3 and 4) some comparisons can be drawn by looking at the initial frequency results (see Table 2). What is most apparent are that the categories for **Give Assurance** (17) and **Clarify** (9) are the most frequent in *Pride and Prejudice* politeness, while in the 19th-Century reference corpus **Give Assurance** (35) and **Hesitation filler** (26) are the most commonly used.

After using the Log likelihood calculator (LLC), which provides a statistical measurement from which I can gauge which functions have been used more, **Give Assurance** (LLC 59.02) and **Clarify** (LLC 28.88) are confirmed to be statistically significant functions used in higher proportional frequency than is seemingly common in 19th Century fictional politeness. This would lead to the

suggestion that PrPr politeness forms revolve a lot more around monitoring meaning in conversation to avoid misunderstandings and to regularly reassure one another in subtle or by more direct means. Given that to **Clarify** may seek to elicit some positive reinforcement or reassurance from the listener, proves that these two functions are often provoked by one another in conversation. This may explain why there are significantly fewer occurrences of **Hesitation filler** and **Thanks, compliment** functions, since monitoring a conversation should ideally reduce the need for additional softening statements or clear signs of agreement to avoid FTAs.

Table 2: Cluster frequencies in-quotes for PrPrC and 19CC politeness function groups and Log Likelihood calculation (PrPr to 19CC).

Function	PrPrC (121,747 words)	19CC (4,391,329 words)	Log Likelihood
Cluster total:	51	122	
Give Assurance (GA)	17	35	59.02
Hesitation Filler (HF)	5	26	10.16
Apology, Commitment (AC)	3	3	13.52
Thanks, Compliment (TC)	6	16	18.45
Clarify (C)	9	22	28.88
Address (A)	3	0	21.68
Hesitation filler/Apology (HF/Ap)	7	16	23.19
Apology, Commitment/Hesitation filler (AC/HF)	1	3	2.89

On the other hand, the results for the 19CC in Table 2 suggest that **Hesitation Filler** (LLC 10.16) and **Apology, Commitment** (LLC 13.52) are two more common features of 19th Century fictional politeness (considering their lower LLC). Thus, it could be assumed that fictional characters from this period tend more often to use FTA-minimizing strategies that feature phrases that latch onto a statement in order to appear less direct and thus aggressive or face-threatening. Interestingly, the weaker presence of these other FTA-minimizing functions could suggest that the main protagonists in *Pride and Prejudice* (i.e. those who feature the most in dialogue) are more direct, using a high proportion

Jessica Mateo

of **Give assurance** to counteract this. Similarly to the keyword analysis, this further grounds a negative attitude towards politeness and obligation, which justifies the unusual way in which this novel approaches politeness. Contextually, the following quotation by Elizabeth in reply to Darcy's proposal presents this negativity:

> "I such cases as this, it is, I believe, the established mode to express a sense of obligation for the sentiments avowed, however unequally they may be returned. It is natural that obligation should be felt, and if I could FEEL gratitude, I would now thank you. But I cannot – […]"

> (*Pride and Prejudice*, Chapter 34, p. 186)

In her reply, Elizabeth touches on what, I argue, is the fundamental cause of both the unusually negative semantic prosody to the keywords *civility* and *politeness*, and for the striking differences in functional politeness strategies in the two corpora. It is obligation. Instead of using clusters that are non-committal (i.e. hedging) or make statements more subtle (i.e. **Hesitation filler**), the characters often **Give assurance** and **Clarify** in abundance, which could be interpreted as mock politeness. Thus, impoliteness in the guise of socially acceptable politeness formulae aims predominately to hide any clear threats to face when in social situations. Obligation appears to be a driving force that makes masking the character's true feelings essential to interaction in this novel, especially given that the economic context and pursuit of wealth through marriage is such a high priority. Darcy's offer, and truth-telling direct FTA on Elizabeth, prompts her to reveal her true feelings. She acknowledges that society would expect her to express polite gratitude, but in reality she refuses to perform this obligation.

The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

School of English

5. Conclusion

The current study has shown that clusters are a useful method of exploring politeness in fictional discourse. Just as "speech clusters show some similarities with features of real speech" (Mahlberg 2013: 75), so too can speech clusters exhibit similar features of politeness. New labels have been necessary to apply to an older context, but the foundations for these labels were founded in Sociolinguistic and Corpus related studies of today, showing that even though 19th century politeness may appear archaic, it is more similar to today's standards than is initially clear. However, one finding of this research suggested that the relationship between form and function may not always correlate, transforming seemingly polite statements into sarcasm or mock politeness. This relates predominately to Culpeper's theory that "inference […] is the level at which a great deal of linguistic politeness and impoliteness occurs" (1996, summarized in Mills 2003: 83).

In addition, a comparison between the fictional politeness of *Pride and Prejudice* against a 19th Century reference corpus has revealed truths about the novel that correspond to critical writing. Having exhibited an approach to politeness that differs from the 19CC, *Pride and Prejudice* can be seen to project thematic components onto the way its characters use politeness. Unlike many of the texts used in the 19CC, *Pride and Prejudice* has a noticeable emphasis on the role of society, which I argue the major themes (i.e. *behaviour*, *civility* and *politeness*, identified via the Keyword analysis) revolve around. Markley's critique on the economic context of the novel, which comments on the roles of "money, inheritance law, values and obligations", epitomizes the overarching roots of this plot that this study of politeness leads to confirm.

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19th Century fiction.

Jessica Mateo

A key word analysis implies that *Behaviour* may be most major theme to the novel, supported by expectations of correct behaviour as the foundation of social rules. Social gatherings, after all, facilitate the pursuit and negotiation of marriage settlements, an important aspect linked to the economic context. Furthermore, in consideration of Emsley's definition of *civility* that says the "maintaining [of] social niceties [is prioritised] even when one does not feel like being polite […]" (2005: 91), this brings to the fore, as stated above, that form and meaning are not always correlated. Mills theory that politeness can be used to mask intentions or true feelings (2003: 60) is reinforced by the findings of this study, politeness is often used as a front, made apparent by the unusual abundance of **Give Assurance** and **Clarify** functions in the novel. This implies that the protagonists are noticeably more direct than usual, though no less outwardly polite, making it possible to hide true feelings. Thus, the emphasis on social expectations, which assert feelings of obligation and duty on the characters is, as touched on by critics, projected into the politeness functions used as strategies to achieve social and economic goals.

Word Count: 3,809

6. References

- Adolphs, S. and R. Carter (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal*. New York: Routledge.
- Altenberg, B. (1998). "On the Phraseology of Spoken English: The Evidence of Recurrent word-combinations. In A. P. Cowie (ed.). *Phraseology, Theory, Analysis, and Applications*. Oxford: Oxford University Press, 101-122.

- Austen, Jane. (2008). *Pride and Prejudice*. London: Penguin Group.

- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press Ltd.

- Biber D. and S. Conrad and G. Leech (2002). *Student Grammar of Spoken and Written English*. Essex: Pearson Education Limited.

- Biber D. and S. Conrad and R. Reppen (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

- Brown, P. and S. C. Levinson (1978). *Politeness: Some Universals in Language Use.* Cambridge: Cambridge University Press.

- Copeland, E. (1997). "Money". In E. Copeland and J. McMaster (eds.). *The Cambridge Companion to Jane Austen.* Cambridge: Cambridge University Press, 127-143.

- Culpeper, J. (2011). *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.

- Hunston, S. (2011). *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. New York: Routledge.

- Mahlberg, M. (2012). "The Corpus stylistic analysis of fiction – or the fiction of Corpus stylistics". In J. Mukherjee and H. Huber (eds.). *Corpus Linguistics and Variation in English: Theory and Description.* Amsterdam: Rodopi, 77-95.

- Mahlberg, M. (2013). *Corpus Stylistics and Dickens Fiction*. New York: Routledge.

- Mahlberg, M. (2010). "Corpus Linguistics and the Study of Nineteenth-Century Fiction', *Journal of Victorian Culture,* 15: 292-298

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19ᵗʰ Century fiction.

Jessica Mateo

- Mahlberg, M. and C. Smith (2010). "Corpus approaches to Prose Fiction: Civility and Body Language in *Pride and Prejudice*". In B. Busse and D McIntyre (eds.). *Language and Style*. Basingstoke: Palgrave, 449-467.

- Markley, R. (2013). "The Economic Context". In J. Todd (ed.). *The Cambridge Companion to 'Pride and Prejudice'*. Cambridge: Cambridge University Press, 79-96.

- McCarthy, M. and R. Carter (2006). "This, that and the other: Multi-word clusters in spoken English as visible patterns of interaction". In M. McCarthy (ed.). *Explorations in Corpus Linguistics*. Cambridge: Cambridge University Press, 7-26.

- Mills, S. (2003). *Gender and Politeness*. Cambridge: Cambridge University Press.

- Scott, M. and C. Tribble (2006). *Textual Patterns: Key words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins Publishing Co.

- Todd, J. (2013). "The Romantic Hero". In J. Todd (ed.). *The Cambridge Companion to 'Pride and Prejudice'.* Cambridge: Cambridge University Press, 150-161.

- Watts, R. J. (2003). *Politeness*. Cambridge: Cambridge University Press.

- Emsley, S. (2005). *Jane Austen's Philosophy of the Virtues*. New York: Palgrave Macmillan.

7. Appendix

Figure 4: Concordance results 1-53 on CLiC (full-text) for keyword Behaviour

"*I should like to know…*": A corpus analysis of politeness phrases in
Pride and Prejudice and in a reference corpus of 19ᵗʰ Century fiction.

Jessica Mateo

	Left	Node	Right
1	…m was Mrs. Bennet, whose dislike of his general	behaviour	was sharpened into particular resentment by his…
2	… listened in silence, but was not convinced; their	behaviour	at the assembly had not been calculated to plea…
3	…tantly to Jane, leaving her own and her relations	behaviour	to the remarks of the two ladies and Mr. Darcy;
4	…e occurs before we discuss the discretion of his	behaviour	thereupon. But in general and ordinary cases be…
5	…ible that if such an idea had been suggested, his	behaviour	during the last day must have material weight in…
6	…ted that "he had never in his life witnessed such	behaviour	in a person of rank-- --such affability and conde…
7	…ng cousin no ill-will, and should never resent her	behaviour	as any affront, seated himself at another table w…
8	be in the wrong, she could no more explain such	behaviour	than her sister. Mr. Collins on his return highly g…
9	… the soul by a thousand tender recollections. His	behaviour	to myself has been scandalous; but I verily belie…
10	But we are none of us consistent, and in his	behaviour	to me there were stronger impulses even than p…
11	…e kingdom-- --provided that a proper humility of	behaviour	is at the same time maintained. You must theref…
12	…ospitality and politeness which had marked their	behaviour	to their guests. Darcy said nothing at all. Mr. Be…
13	…voice that marked his displeasure, "to resent the	behaviour	of your daughter. Resignation to inevitable evils …
14	…oads, were all to her taste, and Lady Catherine's	behaviour	was most friendly and obliging. It was Mr. Collin…
15	…bstinate if I still assert that, considering what her	behaviour	was, my confidence was as natural as your sus…
16	if she feels it, it will easily account for her	behaviour	to me; and so deservedly dear as he is to
17	…which she honours us during your stay here. Her	behaviour	to my dear Charlotte is charming. We dine at Ro…
18	…less Miss Bingley's designs on him were, by his	behaviour	to his cousin, for whom he was evidently destin…
19	any symptom of love; and from the whole of his	behaviour	to Miss de Bourgh she derived this comfort for …
20	…cided. From that moment I observed my friend's	behaviour	attentively; and I could then perceive that his pa…
21	…hes, but his eagerness to grasp at anything. His	behaviour	to herself could now have had no tolerable moti…
22	inclination ever to see him again. In her own past	behaviour,	there was a constant source of vexation and re…
23	…ed to him all the improprieties of Lydia's general	behaviour,	the little advantage she could derive from the fri…
24	…a sameness to disgust and weary. In his present	behaviour	to herself, moreover, she had a fresh source of …
25	…ever been blind to the impropriety of her father's	behaviour	as a husband. She had always seen it with pain;
26	… as they walked, "is not quite consistent with his	behaviour	to our poor friend." "Perhaps we might be decei…
27	…in over the perverseness of the meeting. And his	behaviour,	so strikingly altered-- --what could it mean? Th…
28	of it." "I was never more surprised than by his	behaviour	to us. It was more than civil; it was really
29	…lf called on to say something in vindication of his	behaviour	to Wickham; and therefore gave them to unders…
30	be imaginary, she could not be deceived as to his	behaviour	to Miss Darcy, who had been set up as a
31	there was scarcely an eye which did not watch his	behaviour	when he first came into the room. In no counten…
32	…for the welfare of his friend. Elizabeth's collected	behaviour,	however, soon quieted his emotion; and as Mis…

33	...g her feelings in criticisms on Elizabeth's person	behaviour,	and dress. But Georgiana would not join her. He...
34	...d particularly interested them both. The look and	behaviour	of everybody they had seen were discussed, ex...
35	...forward; and he might imagine, from my father's	behaviour,	from his indolence and the little attention he has...
36	...ouring. "I told you, the other day, of his infamous	behaviour	to Mr. Darcy; and you yourself, when last at Lon...
37	and that she cannot be too much guarded in her	behaviour	towards the undeserving of the other sex." Eliza...
38	him any alarm. I am so grieved for him! His	behaviour	was attentive and kind to the utmost. He _was_ ...
39	... Charlotte informs me, that this licentiousness of	behaviour	in your daughter has proceeded from a faulty d...
40	...thoughts to the obligations which Mr. Gardiner's	behaviour	laid them all under. "For we must attribute this h...
41	not tell you how little I was satisfied with her	behaviour	while she staid with us, if I had not perceived,
42	enough to say before) how much I like him. His	behaviour	to us has, in every respect, been as pleasing as
43	...at she had known on first witnessing his altered	behaviour	in Derbyshire. The colour which had been drive...
44	...them with tolerable ease, and with a propriety of	behaviour	equally free from any symptom of resentment or...
45	...bjects that must deaden them more. Mr. Darcy's	behaviour	astonished and vexed her. "Why, if he came onl...
46	...y, with an expression of half-laughing alarm. His	behaviour	to her sister was such, during dinner time, as sh...
47	...ce, she yet received pleasure from observing his	behaviour.	It gave her all the animation that her spirits could
48	...y great spirits; she had seen enough of Bingley's	behaviour	to Jane, to be convinced that she would get him
49	you are not entitled to know mine; nor will such	behaviour	as this, ever induce me to be explicit." "Let me
50	...re ill-founded, formed on mistaken premises, my	behaviour	to you at the time had merited the severest repr...
51	...row." This was a sad omen of what her mother's	behaviour	to the gentleman himself might be; and Elizabet...
52	...d early withstood, and as for my manners-- --my	behaviour	to _you_ was at least always bordering on the u...
53	...in love." "Was there no good in your affectionate	behaviour	to Jane while she was ill at Netherfield?" "Deare...

Table 3: The classification for function in the politeness clusters (in-quotes) from *Pride and Prejudice* (PrPrC).

	Politeness Cluster	Freq.	%	Classification
1	you would not wish to	3	0.006	C
2	but i hope you will	3	0.006	C
3	you may depend upon my	3	0.006	GA
4	assure you that i have	3	0.006	GA
5	i assure you it is	3	0.006	GA
6	i am afraid he has	3	0.006	HF/Ap
7	i am sure i shall	3	0.006	GA
8	i do assure you that	3	0.006	GA

Jessica Mateo

9	i am sorry to say	2	0.004	HF/Ap
10	that i am sure i	2	0.004	GA
	i am sure and i	2	0.004	GA
	i thought i should have	2	0.004	HF/Ap
	i am sure i do	2	0.004	GA
	i hope we shall be	2	0.004	C
15	i am very glad you	2	0.004	TC
	you must give me leave	2	0.004	AC
	yourself but i hope you	2	0.004	C
	should be sorry to think	2	0.004	HF/Ap
	any possible way that would	2	0.004	GA
20	would not wish to be	2	0.004	C
	i assure you that i	2	0.004	GA
	am very glad to hear	2	0.004	TC
	last i was sure you	2	0.004	C
	i am much mistaken if	2	0.004	C
25	i wish with all my	2	0.004	TC
	and i am sure i	2	0.004	GA
	i am glad you are	2	0.004	TC
	not take the trouble of	2	0.004	HF/Ap
	i am very glad to	2	0.004	GA
30	i dare say i shall	2	0.004	HF
	i must confess that i	2	0.004	AC/HF
	i should be sorry to	2	0.004	HF/Ap
	must give me leave to	2	0.004	AC
	and i wish with all	2	0.004	TC
35	sure i do not know	2	0.004	C
	have the honour of seeing	2	0.004	A
	i am afraid you will	2	0.004	HF/Ap
	i have the honour of	2	0.004	A
	my dear miss elizabeth that	2	0.004	A
40	i shall be glad to	2	0.004	GA
	i am sure i know	2	0.004	GA
	i am inclined to think	2	0.004	HF
	may depend upon my not	2	0.004	GA
	i am sure you must	2	0.004	GA
45	i hope you will be	2	0.004	C
	i thank you again and	2	0.004	TC
	you must allow me to	2	0.004	AC
	man i am afraid he	2	0.004	HF
	i dare say she is	2	0.004	HF
50	am inclined to think that	2	0.004	HF
	in any possible way that	2	0.004	GA

School of English

Table 4: The classification for function in the politeness clusters (in-quotes) from the 19CC, with a column indicating the number of texts each cluster appears in (full-text in Word Smith Tools).

	Politeness Cluster	Freq.	Num. of texts (full-text):	%	Classification
1	i should like to know	48	18	0.003	HF
2	i should like to see	29	16	0.002	HF
3	it seems to me that	29	10	0.002	He
4	i should like to have	24	14	0.001	HF
5	i am sorry to say	21	17	0.001	HF/Ap
6	i am sure you will	21	12	0.001	GA
7	very much obliged to you	18	9	0.001	TC
8	you don't mean to say	18	8	0.001	C
9	to tell you the truth	17	11	0.001	HF
10	i don't know what you	17	10	0.001	C
	i do not know that	17	8	0.001	HF/Ap
	i should like to be	17	11	0.001	HF
	i don't know what i	17	10	0.001	C
	much obliged to you for	16	10	0.001	TC
15	i have no wish to	16	7	0.001	GA
	i did not mean to	15	8	0.001	GA
	i am sure you are	14	7	0.001	GA
	what do you mean by	14	11	0.001	C
	i do not wish to	14	11	0.001	GA
20	be so good as to	13	9	0.001	AG
	i should be glad to	13	10	0.001	GA
	i should like to hear	13	10	0.001	HF
	i am much obliged to	12	8	0	TC
	am much obliged to you	12	8	0	TC
25	i don't think i could	12	6	0	HF/Ap
	be so kind as to	11	9	0	AG
	i shall be glad to	11	9	0	GA
	i dare say i shall	11	7	0	HF/Ap
	and that sort of thing	11	3	0	He
30	i am sure you would	11	8	0	GA
	i should like to go	11	8	0	HF

"I should like to know…": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19th Century fiction.

Jessica Mateo

	i should be sorry to	11	7	0	HF/Ap
	if you will let me	11	7	0	C
	i think it would be	11	9	0	HF
35	and all that sort of	11	6	0	He
	i am very glad to	11	8	0	GA
	i am glad to hear	11	5	0	GA
	i should like to do	11	9	0	HF
	i am so glad you	11	7	0	GA
40	i don't know that i	10	10	0	HF/Ap
	i do not know whether	10	8	0	HF/Ap
	very glad to see you	10	8	0	TC
	seems to me to be	10	7	0	He
	i beg your pardon sir	10	9	0	AG/Ap
45	i am very much obliged	10	6	0	TC
	i have no doubt you	10	9	0	GA
	i am sure she will	10	5	0	GA
	you may depend upon it	10	7	0	GA
	i give you my word	10	8	0	GA
50	i don't know whether i	10	7	0	HF/Ap
	it is a sort of	10	7	0	He
	i think we had better	9	6	0	HF
	i am sure she would	9	6	0	GA
	i do not know how	9	9	0	C
55	i am obliged to you	9	7	0	TC
	i beg your pardon i	9	5	0	AG/Ap
	i don't know about that	9	5	0	HF/Ap
	don't know what you mean	9	8	0	C
	if you will allow me	9	7	0	HF
60	you will allow me to	9	7	0	HF
	am very much obliged to	9	6	0	TC
	i am sure it is	9	5	0	GA
	do you mean to say	9	8	0	C
	and i should like to	9	7	0	HF
65	very good of you to	9	6	0	TC
	i do not mean to	9	6	0	HF/Ap
	i'm sure i don't know	9	6	0	AC/HF
	i thought it would be	9	7	0	C
	but i should like to	8	10	0	HF
70	tell you the truth i	8	7	0	HF
	i hope you will be	8	6	0	C
	i do not pretend to	8	4	0	GA

	i should like you to	8	7	0	IR
	should like to see the	8	6	0	HF
75	i should so like to	8	4	0	HF
	obliged to you for your	8	6	0	TC
	i had the pleasure of	8	7	0	TC
	i hope you will not	8	7	0	C
	i am glad of it	8	6	0	TC
80	i am afraid we must	8	6	0	AC/HF
	to tell the truth i	8	6	0	HF
	i am sorry to hear	7	5	0	HF/Ap
	if i can be of	7	6	0	GA
	i beg your pardon for	7	4	0	AG/Ap
85	i may be allowed to	7	6	0	C
	i shall be very glad	7	6	0	GA
	should like to know what	7	6	0	HF
	and i am sure i	7	4	0	GA
	do you mean to do	7	7	0	C
90	but it seems to me	7	5	0	He
	i quite agree with you	7	5	0	GA
	i ought to have done	7	5	0	HF/Ap
	don't know what to say	7	5	0	HF/Ap
	i beg your pardon miss	7	4	0	AG/Ap
95	i am afraid i must	7	5	0	AC/HF
	i am afraid it is	7	5	0	HF/Ap
	will you have the goodness	7	5	0	C
	don't mean to say that	7	5	0	C
	don't mean to say you	7	4	0	C
100	i don't pretend to know	6	5	0	GA
	i should have thought you	6	6	0	C
	i don't think i shall	6	3	0	HF
	i am sure i should	6	4	0	GA
	been very kind to me	6	6	0	TC
105	not that i mean to	6	4	0	GA
	i think i ought to	6	6	0	AC
	i dare say we shall	6	3	0	HF
	i dare say it will	6	5	0	HF
	good of you to come	6	5	0	TC
110	i am sure there is	6	4	0	GA

Jessica Mateo

	i am sure i never	6	4	0	GA
	am not sure that i	6	3	0	HF
	shall be most happy to	6	4	0	TC
	i should not like to	6	4	0	HF/Ap
115	and i am sure she	6	5	0	GA
	i am very fond of	6	6	0	GA
	have no doubt of it	6	4	0	GA
	i can assure you that	6	4	0	GA
	i thought you would be	6	5	0	C
120	and i hope you will	6	6	0	C
	i must ask you to	6	4	0	AC
	you mean to tell me	6	4	0	C
	i am glad you have	6	5	0	TC
	i am afraid you have	6	5	0	HF/Ap
125	should not have thought it	6	4	0	HF
	i must go back to	6	5	0	AC
	i am sure he will	6	6	0	GA
	i hope he will be	6	4	0	C
	i shall be happy to	6	4	0	GA
130	am sure she will be	6	3	0	GA
	and to tell the truth	6	5	0	HF
	i don't know how it	6	6	0	HF
	i should be very glad	6	6	0	GA
	i wish you would not	6	5	0	IR
135	if you have no objection	6	4	0	C

7.1. Function Labels: Definitions

- **Give Assurance (GA)** – Clusters that aim to reassure the listener.
  Common lexical components may be: *sure*, *shall*, *assure*, *may depend*,
  glad, *if*, as well as intensifiers such as *so*, *very*.

Links to previous studies: Brown and Levinson's (1978: 103-116) strategies 1 (Notice, Attend to H), 3 (Intensify interest to H) and 6 (Avoid disagreement). Similarly, this label also matches Altenberger's reassuring expressions (1998: 105), though not exclusively in response to apologies.

- **Hesitation filler (HF)** – Clusters that aim to downplay or soften direct/blunt statements of their dialogue. Common lexical components

may include: modal verbs such as should like, or recurring phrases such
as *I dare say* or *I am afraid*.

Links to previous studies: McCarthy and Carter's reference to hedging (2006: 21)
and indirectness (2006: 33) are not directly linked, but do have some relevance
to this label. The idea of indirectness is also touched upon by Altenberg (1998:
109) as well as in Brown and Levinson's Off-record category (see Mills 2003: 59).

- **Apology, Commitment (AC)** – Cluster that express an obligation that
  may be inconvenient, disruptive or possibly rude to the listener. These
  clusters often include the modal verb *must*.

- **Thanks, Compliment (TC)** – Clusters that express thanks, and/or have a
  complimentary element towards the listener.

- **Clarify (C)** – Clusters that aim to clarify or monitor meaning in
  conversation. This is often used to avoid misunderstandings. Direct checks
  can come in the form of clusters such as *you don't mean to say,* but can
  also be more subtle such as *I hope we shall be,* which may be searching
  for some positive reinforcement or reassurance.

Links to previous studies: McCarthy and Carter's monitoring function (2006: 20)
as well as Brown and Levinson's strategy 5: Seek Agreement (1978: 112).

- **Address (A)** – Clusters that aim to address the listener or other
  characters. This label is most closely aligned to characterisation and so
  has less bearing on the study of politeness. However, this does not reject
  the fact that they do at times show polite language and thus have been
  included as a single category.

Jessica Mateo

- **Attention Getting (AG)** – Clusters that aim to alert listeners to their presence or involvement in a conversation. This is often used in small exchanges between strangers or new acquaintances.

- **Hedge (He)** – Clusters that use hedging to avoid commitment to a statement. They make statements vague. There is some overlap with **Hesitation filler**, but the first category encompasses a much broader sense of politeness than hedge, which focuses purely on vagueness and non-commitment.

- **Indirect Request (IR)** – Clusters that reduce the level of FTA when making requests. Expressions such as *wish to* are common to appear less aggressive.

The Following categories were created on the basis that some clusters could not fit into any particular label, but instead spread across at least two. This was predominately in regard to those that feature apology. Thus Apology was split into a further three categories:

- **Hesitation filler/Apology (HF/Ap)** – Clusters that express apology to the listener, but mainly soften a statement that may be face-threatening.

- **Attention Getting/Apology (AG/Ap)** – Clusters that express apology while trying to gain their attention. This is commonly seen in clusters with the expression, *beg your pardon*.

- **Apology, Commitment/Hesitation filler (AC/HF)** – Clusters that combine the indirectness of HC with the obligatory commitment of AC, with the expression of apology (e.g. *I must confess that I*).

7.2. Reference Corpus

The reference corpus includes 28 19th Century fictional texts (4,513,076 words; 1,606,974 quotes, 2,906,102 non-quotes):

- Agnes Grey (68, 197 words; 21,986 quotes; 46, 211 non-quotes)

- Antonia, or the Fall of Rome (166,517 words; 36,508 quotes, 130,009 non-quotes)

- Armadale (297,373 words; 124,828 quotes, 172,545 non-quotes)

- Cranford (70,935 words; 17,928 quotes, 53,007 non-quotes)

- Daniel Deronda (310,984 words; 103,607 quotes, 207,377 non-quotes)

- Dracula (160,538 words, 52,420 quotes, 108,118 non-quotes)

- Emma (160,312 words; 77,633 quotes, 82,679 non-quotes)

- Frankenstein (74,984 words; 13,961 quotes, 61,023 non-quotes)

- Jane Eyre (185,486 words; 78,034 quotes, 107,452 non-quotes)

- Jude the Obscure (145,169 words; 56,791 quotes, 88,378 non-quotes)

- Lady Audley's Secret (150,083 words; 62,190 quotes, 87,893 non-quotes)

- Mary Barton (160,455 words; 58,886 quotes, 101,569 non-quotes)

- North and South (182,746 words; 75,128 quotes, 107,618 non-quotes)

- Persuasion (83,225 words; 28,653 quotes, 54,572 non-quotes)

- Sybil, or the two nations (157,936 qords; 64,828 quotes, 93,108 non-quotes)

- Tess of the D'Urbervilles (149,657 words; 39,24 quotes, 110,413 non-quotes)

- The Hound of the Baskervilles (59,167 words; 32,850 quotes, 26,317 non-quotes)

- The Last Days of Pompeii (151,309 words; 57,705 quotes, 93,604 non-quotes)

- The Mill on the Floss (207,130 words; 69,695 quotes, 137,435 non-quotes)

"*I should like to know…*": A corpus analysis of politeness phrases in *Pride and Prejudice* and in a reference corpus of 19ᵗʰ Century fiction.

Jessica Mateo

- The Picture of Dorian Gray (78,702 words; 38,571 quotes, 40,131 non-quotes)

- The Professor (87,628 words; 22,061 quotes, 65,567 non-quotes)

- The Return of the Native (141,749 words; 55,469 quotes, 86,280 non-quotes)

- The Small House at Allington (257,184 words; 90,006 quotes, 167,178 non-quotes)

- The Strange Case of Dr Jekyll and Mr Hyde (25,596 words; 7,322 quotes, 18,274 non-quotes)

- The Woman in White (246,671 words; 72,515 quotes, 174,156 non-quotes)

- Vanity Fair (303,137 words; 54,573 quotes, 248,564 non-quotes)

- Vivian Grey (191,939 words; 90,777 quotes, 101,162 non-quotes)

- Wuthering Heights (116,520 words; 49,817 quotes, 66,703 non-quotes)