

**THE LEVEL OF PHYSICALITY
IN JANE AUSTEN'S MAJOR NOVELS,
THEIR FILM ADAPTATIONS AND HER PRIVATE LETTERS**

KRISTĪNA KORNELIUSA

University of Latvia

***Abstract:** The current paper explores the physicality in Jane Austen's novels, their film adaptations, and her personal letters. The nouns referring to human body are examined in terms of frequency and context. Corpus linguistics is used as a method, involving the use of in-built online corpora, offline microcorpus creation, and the use of part-of-speech tagger and a concordancer. The results allow to conclude that, firstly, to Austen, referring to the human body was less appropriate in private written communication than in fiction; secondly, the film adaptations use some words in a literal sense, while the novels – in a figurative sense.*

***Keywords:** Corpus linguistics, fiction, human body, Jane Austen, physicality*

1. Introduction

In the biography *Jane Austen*, Carol Shields (2001: 166) argues that Jane Austen's spinsterhood and sexual innocence was reflected in the ways she referred to the human body in her novels. However, Shields does not draw any comparison between the frequency and context in which this or that item referring to the human body is found in Austen's works and her letters to her relatives. Besides, since the adaptations of Austen's novels are much more recent in history than the novelist's writings, it can be presumed that the adaptations are more overt and liberal in terms of references to the human body.

The aim of the current research is to explore the references to the human body in Jane Austen's novels, her personal letters and the scripts of film adaptations of her works. In order to achieve this aim, corpus linguistics method was employed: a microcorpus containing the letters of Jane Austen was created; for examining the texts of her novels, an embedded Lextutor Corpus *Jane Austen* was used; as to the film adaptations, the option of creating a virtual corpus in BYU Corpora was applied. The research question which is going to be answered with the help of these corpora is the following: what is the difference between the frequencies and the context in which the nouns referring to the human body are used by Austen in her novels and letters, and by the screenwriters in the film adaptations of her novels?

The nouns referring to the human body are chosen as the linguistic feature to be researched in the three corpora, because these linguistic items are the most overt reference to human physicality. Also, they are often used in a figurative sense, which gives me the opportunity to compare literal and figurative references.

2. Literature review

2.1. Defining the corpus and its criteria

In *What is a corpus?* from *Web as Corpus: Theory and Practice* (Gatto 2014) and *What Is a Corpus* from *Corpus Linguistics: The Basics* (McEnery et al. 2006), the authors provide a text corpus definition, and describe the corpus criteria: authenticity, representativeness, balance and sampling, and size. Quoting McEnery and Wilson (2001), Gatto (2014: 8) indicates that a corpus is a set of more than one text, while McEnery et al. (2006) specify that this set is characterized by electronic format and authenticity. The latter is also called a “naturally-occurring language” by Sinclair (1991: 171), who argues that corpus should also be representative of the “language or language variety” it is compiled in (ibid.). Gatto further describes both sampling and balance, while McEnery and Wilson (2006: 5) point out that not all corpora are balanced, e.g., the corpora of specialized texts.

Applying these characteristics to the corpora used here, it has to be said that all of them are electronic, authentic, representative of the field they are dedicated to - Jane Austen’s language and the language used in the film adaptations of her work; they are comparatively small, and, according to McEnery and Wilson’s remark, not balanced, if compared to a general corpus.

2.2. Defining meaning, context and co-text

The corpus research in literary studies “[enables] insights not possible through intuition alone” (Wahid 2011: 105) – it provides quantitative evidence for any supposition made during a close reading. Moreover, given the amount of texts which has to be analysed in the current paper, the systematisation of the texts allows us to extract both quantitative and qualitative data more effectively. In order to specify what is meant by qualitative data here, such concepts as context and meaning (Moon 2010) and context vs co-text (Thornbury 2010; Ferrin et al. 2011) have to be explained.

According to Moon (2010: 202), “corpora indicate the context in which we find the word”. Meanwhile, Thornbury (2010: 270) argues that corpora are more applicable for indicating co-text rather than context, while Ferrin et al.’s study (2011) allows us to see how much it is possible to rely on corpus data (co-text) and the historical background (context). Co-text is defined as “the set of features necessary to bind the variables introduced” (Ferrin et al. 2011: 1491), i.e., it refers more to cohesion and coherence or text linguistics, while context answers the questions “who, where, when, what, how and why” and puts the text in a physical situation (ibid.); such an approach refers to discourse rather than text linguistics.

2.3. Co-text and collocations

Co-text is the text surrounding the query. In order to analyse it, it is necessary to look at collocations.

According to McEnery and Hardie (2012: 122-123), “the term *collocation* denotes the idea that important aspects of the meaning of a word (or another linguistic unit) are not contained within the word itself, considered in isolation, but rather subsist in the characteristic associations that the word participates in, alongside other words or structures with which it frequently co-occurs”. The

queries used for the current research are the nodes, while the words surrounding them in a certain position are collocates (idem: 124).

3. Corpora description

Table 1 summarises the basic information about the corpora used for the current analysis

	Lextutor Corpus	BYU Corpus	Microcorpus created by the author of the paper
Title	Jane Austen	Jane Austen Film Adaptations	Letters of Jane Austen
Size (no. of words)	648 000	58 999 (note: the sizes of six subcorpora added up)	77 520 (note: the sizes of three subcorpora added up)
Text types	Major novels by Jane Austen: <i>Sense and Sensibility</i> , <i>Pride and Prejudice</i> , <i>Emma</i> , <i>Mansfield Park</i> , <i>Persuasion</i> , <i>Northanger Abbey</i>	The scripts of six film adaptations: <i>Emma</i> (1996), <i>Mansfield Park</i> (2007), <i>Northanger Abbey</i> (2007), <i>Persuasion</i> (2007), <i>Pride and Prejudice and Zombies</i> (2016), <i>Sense and Sensibility</i> (1995)	Personal letters written by Jane Austen to her sister Cassandra Austen, her relative Anna Lefroy and her niece Frances Knight
Source/Way of data extraction	An embedded corpus available from https://www.lexutor.ca/conc/eng/	A virtual corpus created by the author of this paper by selecting texts from Movie Corpus on BYU Corpora (available from https://www.english-corpora.org/movies/)	An offline corpus created by the author of the paper by extracting texts from <i>Letters of Jane Austen</i> (2013) (available from https://www.gutenberg.org/files/42078/42078-h/42078-h.htm)
Purpose of use	To identify the frequency and context of words referring to the human body in Jane Austen's novels	To identify the frequency and context of words referring to the human body in film adaptations of Jane Austen's novels	To identify the frequency and context of words referring to the human body in Jane Austen's letters to her relatives

Table 1. The description of the corpora analysed

These corpora have the following additional characteristics.

The Lextutor Corpus has a column with the name of the novel the concordance line was taken from. The results can be sorted according to the name of the novel in the alphabetical order, as shown in Figure 1.

Line	Text Snippet	Novel
061.	...e flow of spirits, an't I? But I want to set your HEART at ease as to Mrs. S. - My representation, y	[emma]
062.	...tley's absence, or when it came to the point her HEART would have failed her, and she must have put	[emma]
063.	...reat. - However, I must say, that Robert Martin's HEART seemed for him, and to me, very overflowing;	[emma]
064.	...doleful disappointment of five weeks back! Such a HEART - such a Harriet! Now there would be pleasur	[emma]
065.	...e you joy in person." He thanked her with all his HEART, and continued some time to speak with serio	[emma]
066.	...entered into with doubtful feelings, with half a HEART. I thought it my duty as a friend, and older	[emma]
067.	...dear Sir Thomas, with all my faults I have a warm HEART; and, poor as I am, would rather deny myself	[mansfield]
068.	...oliage of summer, was enough to catch any man's HEART. The season, the scene, the air, were all fa	[mansfield]
069.	..."It is a pleasure to see a lady with such a good HEART for riding!" said he. "I never see one sit a	[mansfield]
070.	...ld never happen again. Fanny went to bed with her HEART as full as on the first evening of her arriv	[mansfield]
071.	...rty on each side of the road," without elation of HEART; and it was a pleasure to increase with thei	[mansfield]
072.	...onsideration of others, that knowledge of her own HEART, that principle of right, which had not form	[mansfield]
073.	...s what may tranquillise every care, and lift the HEART to rapture! When I look out on such a night	[mansfield]
074.	..., and nurse, the despondence that sunk her little HEART was severe. The grandeur of the house astoni	[mansfield]
075.	...with the advantage of knowing half the scenes by HEART already, he did now, with the greatest alacr	[mansfield]
076.	...ife's speeches, I would undertake him with all my HEART." "With all your partiality for Cottager's w	[mansfield]
077.	...I like it." Fanny's eyes followed Edmund, and her HEART beat for him as she heard this speech, and s	[mansfield]

Figure 1. The arrangement of concordance lines according to the novel title

This feature allows not only to compare the data obtained from this corpus from the ones from other corpora, but also to compare the results across the novels themselves, and it was used when calculating the results.

The BYU virtual corpus consists of six subcorpora (see Figure 2).

LIST NAME ↓	# TEXTS ↓	# WORDS ↓	FIND KEYWORDS ● SPECIFIC ○ FREQ	CREATED ↓
EMMA	1	13,954	NOUN VERB ADJ ADV N+N ADJ+N	54 d
MANSFIELD PARK	1	8,863	NOUN VERB ADJ ADV N+N ADJ+N	54 d
NORTHANGER ABBEY	1	9,615	NOUN VERB ADJ ADV N+N ADJ+N	54 d
PERSUASION	1	8,544	NOUN VERB ADJ ADV N+N ADJ+N	54 d
PRIDE AND PREJUDICE	1	8,632	NOUN VERB ADJ ADV N+N ADJ+N	52 d
SENSE AND SENSIBILITY	1	9,391	NOUN VERB ADJ ADV N+N ADJ+N	54 d

Figure 2. The structure of the compiled BYU virtual corpus with Jane Austen film adaptations

Each of them contains the script of one film adaptation of a different novel. The adaptations were selected according to their year of release - the most recent available in the Movie Corpus were preferred. Thus, the time gap between the film adaptations and the original novels was increased.

The microcorpus *Letters of Jane Austen* has been divided into three subcorpora, according to the addressee of the letter. The quantitative information about the subcorpora is given in Table 2.

Subcorpus (file) name	No. of letters	No. of tokens
jane-cassandra	62	66742
jane-anna	9	4631
jane-fanny	5	6147

Table 2. The structure of the microcorpus Letters of Jane Austen

The tokens were calculated using the Lancaster University Statistics tool (available at <http://corpora.lancs.ac.uk/stats/toolbox.php>). All the texts are taken from *The Letters of Jane Austen*, edited by Sarah Chauncey Woolsey (2013).

It should also be added that this microcorpus is annotated, as some of the research nodes are subject to conversion (e.g., *hand – to hand*). The primary interest if this research are nouns - they need to be separated from the verbs whose forms can coincide with those of the nouns. Therefore, the corpus files were inserted into the Simple Part of Speech Tagger. A small annotated fragment is added below for illustration:

MY_NNP DEAR_NNP ANNA_NNP ,_PPC [/LRB 31_CD]/RRB —I_NNP
 am_VBP very_RB much_JJ obliged_JJ to_TO you_PRP for_IN sending_VBG
 your_PRPS MS. NNP It_PRP has_VBZ entertained_VBN me_PRP extremely_RB
 ;_PPS indeed_RB all_DET of_IN us_PRP . PP I_PRP read_VBD it_PRP aloud_RB
 to_TO your_PRPS grandmamma_NN and_CC Aunt_NNP Cass_NNP ,_PPC
 and_CC we_PRP were_VBD all_DET very_JJ much_JJ pleased_JJ . PP The_DET
 spirit_NN does_VBZ not_RB droop_JJ at_IN all_DET . PP Sir_NNP Thos_NNP
 . PP ,_PPC Lady_NNP Helen_NNP ,_PPC and_CC St. NNP Julian_NNP are_VBP
 very_RB well_RB done_VBN ,_PPC and_CC Cecilia_NNP continues_VBZ to_TO
 be_VB interesting_JJ in_IN spite_NN of_IN her_PRPS being_NN so_IN amiable_JJ
 . PP It_PRP was_VBD very_RB fit_VB you_PRP should_MD advance_VB
 her_PRPS age_NN . PP I_PRP like_IN the_DET beginning_NN of_IN
 Devereux_NNP Forester_NNP very_RB much_JJ ,_PPC a_DET great_JJ deal_NN
 better_RB than_IN if_IN he_PRP had_VBD been_VBN very_RB good_JJ or_CC
 very_RB bad_JJ . PP A_DET few_JJ verbal_JJ corrections_NNS are_VBP all_DET
 that_IN I_PRP felt_VBD tempted_VBN to_TO make_VB ; PPS the_DET
 principal_NN of_IN them_PRP is_VBZ a_DET speech_NN of_IN St. NNP
 Julian_NNP to_TO Lady_NNP Helen_NNP ,_PPC which_WDT you_PRP see_VBP
 I_PRP have_VBP presumed_VBN to_TO alter_VB . PP As_IN Lady_NNP H. NNP
 is_VBZ Cecilia_NNP 's_POS superior_JJ ,_PPC it_PRP would_MD not_RB be_VB
 correct_JJ to_TO talk_VB of_IN her_PRPS being_NN introduced_VBD . PP
 It_PRP is_VBZ Cecilia_NNP who_WP must_MD be_VB introduced_VBN . PP
 And_CC I_PRP do_VBP not_RB like_IN a_DET lover_NN speaking_NN in_IN
 the_DET 3rd_CD person_NN ; PPS it_PRP is_VBZ too_RB much_JJ like_IN
 the_DET part_NN of_IN [/LRB 263_CD]/RRB Lord_NNP Overtley_NNP ,_PPC
 and_CC I_PRP think_VBP it_PRP not_RB natural_JJ . PP If_IN you_PRP
 think_VBP differently_RB ,_PPC however_RB ,_PPC you_PRP need_VBP not_RB
 mind_VB me. NN I_PRP am_VBP impatient_JJ for_IN more_JJR ,_PPC and_CC
 only_RB wait_VB for_IN a_DET safe_JJ conveyance_NN to_TO return_VB
 this_DET . PP
 Yours_NNP affectionately_RB ,_PPC
 J. NNP A_NNP . PP

4. Procedure

4.1. Setting up a list of n nodes

Due to such limitations of research as time and volume, the list of nodes to be checked in the corpora had to be narrowed down. It was decided to focus on the words provided by Carol Shields in her biography of Jane Austen. Shields (2001: 166-167) listed the nouns related to human body which she could (or, on the contrary, could not) find in Austen's novels: *brain, head, toe, finger, hips, thighs,*

shin, buttocks, kidney, intestine, womb, navel, tooth, chin, ankle, nose, ear, leg, wrist, eyebrow, eyelash, breast, eye, heart, hand, face, skin. These words were arranged in the alphabetical order and their absolute frequency was checked in all the three corpora. Those which could not be found in any of the corpora were excluded from further analysis. The final list of nodes can be seen below:

ankle(s)	Eyebrows	heart(s)	wrist
brain(s)	Eyelashes	leg(s)	
Breast	face(s)	nose	
Chin	finger(s)	skin	
ear(s)	hand(s)	tooth (teeth)	
eye(s)	head(s)	toe	

It has to be noted that, for most of the nodes, both the singular and the plural form were checked. The frequencies of each were summed up together. For further research, the differences between the frequencies and the context of the singular and the plural forms could be explored, because some preliminary qualitative observations were made, which will be touched upon in the Discussion of the results section, but cannot be elaborated upon in more detail in this paper, due to space limitations.

4.2. Frequency normalisation

Due to the different corpora sizes, all the absolute frequencies obtained for the chosen nodes were normalised, i.e., the frequency was adjusted per a fixed number of words. The base of 10 000 was chosen - the corpora sizes being comparatively small. The absolute frequencies in the subcorpora were also normalised. The wordcounts for Austen's novels were taken from Eleanor Turner's (2016) blog article.

4.3. Extracting collocates

The definition of collocation, node, and collocate was given in section 2.3. In the empirical part, a list of collocates for the nodes *heart(s)* and *head(s)* was drawn up. These nodes were selected because *heart(s)* was the most frequent word found in Lextutor and BYU Corpus, while *head(s)* was the most frequent word found in the *Letters of Jane Austen* microcorpus. The collocates in the position -1 (a pre-noun position) - noun modifiers or determiners - were adjectives, possessive adjectives, articles. Only the absolute frequencies for the collocates of nodes *head(s)* and *heart(s)* were extracted, because it is the ranking of the collocates that mattered. The collocates for other nodes are a subject for further research.

5. Discussion of the results

I have found out that Jane Austen's novels contain more nouns related to the human body than the film adaptations or her private letters – both in terms of absolute and normalised frequency (see Table 3).

Novels		Adaptations		Letters	
Word	Freq.	Word	Freq.	Word	Freq.
heart(s)	496	heart(s)	55	head(s)	26
eye(s)	446	hand(s)	18	eye(s)	21
hand(s)	349	eye(s)	17	hand(s)	18
head(s)	237	head(s)	16	heart(s)	15
face(s)	161	brain(s)	14	face(s)	13

Table 3. The frequency data for the selected nodes
(five most frequent items for each corpus)

This can be explained by the fact that both film scripts and letters can be compared to direct speech, while the novels also contain the narration. The difference in normalised frequency between the novels and the adaptations is comparatively minor - 27.9 and 23.9 per 10 000 words respectively. In the letters, however, the normalised frequency of nouns referring to the human body is only 15.6. Apparently, the written communication, even if among family members, was highly formal, and referring to the human body too often was not considered appropriate. The fictional genre, on the other hand, provided more freedom.

When comparing the results across the subcorpora of the Letters of Jane Austen corpus, it was found out that, in terms of frequency per 10 000 words, Austen's letters to her niece Fanny contains the most nouns referring to the human body, while the letters to her sister Cassandra, contrary to the expectation that Austen would discuss physical matters more overtly with her, contain the least (see Table 4).

Cassandra	Anna	Fanny
15	15.1	22.8

Table 4. The summed-up relative frequency data of the selected nodes
across the letter addressees

Nevertheless, context has to be taken into account as well. Despite the fact that the frequency of nouns referring to the human body in the adaptations is smaller, the supposition that physical references in modern cinema are more common than in Austen novels proved to be correct. For example, the word *brain* (and also *brains*), which is used in Austen's novels as a word standing for mind and thought, in *Pride and Prejudice* and *Zombies* refers to the actual physical human brain. Compare:

- (1) [...] though Kitty might in time regain her natural degree of sense, since the disturbers of her brain were removed, her other sister, from whose disposition greater evil might be apprehended, was likely to be hardened in all her folly and assurance by a situation of such double danger as a watering-place and a camp. (*Pride and Prejudice*: the novel)

(2) Daughters do not dance well with masticated brains. (*Pride and Prejudice and Zombies*: the film adaptation)

Similarly, while the word *breast* (used only in the singular) in the novels refers to emotional experiences, the 2007 adaptation of *Northanger Abbey* quotes a passage from the novel *Monk* by Matthew Gregory Lewis, where female breasts are mentioned:

(3) Poor Mr. Woodhouse little suspected what was plotting against him in the breast of that man whom he was so cordially welcoming (*Emma*: the novel)

(4) At this moment a tame Linnet flew towards her, nestled its head between her breasts, and nibbled them in wanton play. (*Northanger Abbey*: the film adaptation; a quote from *Monk* by M. G. Lewis)

It should also be added that, according to normalised frequency, the novel *Pride and Prejudice* contains the smallest number of the researched nouns referring to the human body (see Table 5), while its 2016 adaptation is the most *physical* of all the films selected (see Table 6). This can also be explained by the addition of the zombie element in this adaptation.

Emma	Mansfield Park	Northanger Abbey	Persuasion	Pride and Prejudice	Sense and Sensibility
21,3	25.8	34.2	25.8	18.5	29.2

Table 5. The summed-up relative frequencies of the nodes analysed across the novels

Emma	Mansfield Park	Northanger Abbey	Persuasion	Pride and Prejudice	Sense and Sensibility
15.8	26	25	14	41.7	25.6

Table 6. The summed-up relative frequencies of the nodes analysed across the novels

The most frequent of all the researched nodes in the novels and the adaptations turned out to be the node *heart(s)*, while in the letters – *head(s)*. It can be presumed that Austen's novels deal more with feelings, while the letters are more reserved and deal with mind and thought. For these two nodes, which turned out the most frequent in different corpora, a list of collocates in the position -1 was extracted (see Table 7).

Heart(s)						Head(s)					
Novels		Adaptations		Letters		Novels		Adaptations		Letters	
Word	Fr.	Word	Fr.	Word	Fr.	Word	Fr.	Word	Fr.	Word	Fr.
Her	118	my	18	my	5	her	68	your	3	my	5

Table 7. The most frequent collocates in -1 position for heart(s) and head(s) in each corpus

It was found that the most common collocate for both nodes in Jane Austen's novels is *her* (i.e., the collocation *her heart*, *her head*). Since Austen's protagonists and focalizers of the story are women, the prevalence of this possessive adjective is understandable. In the film adaptations and the letters, it is the possessive adjective *my*, and also *your* – in collocation with *head* – that are the most common, because in these texts, the speakers and the writer respectively refer to themselves or their partner in conversation in the first and second person, and third parties are referred to much more seldom.

The extraction of collocates has also confirmed that *heart* in Austen's novels and letters refers to emotions and human qualities, collocating with such words as *amiable*, *affectionate* and *joyful*, while *head* – for opinions and thinking abilities (*capacious*, *stronger*). In the adaptations, most of such adjectives were omitted from the dialogues.

However, the word *head* often did not refer either to the human body, or to mind. Often found in the collocation *the head of*, it was used in phrases like *the head of the house*, *the head of the table* etc. Also, Austen often used in her writing such phrases as *on this/that head*, which do not refer to the physical organ, but impacts the frequency of the noun considerably.

6. Conclusion

The comparison of the three corpora has allowed me to draw two main conclusions: (1) there are more references to the human body in Austen's novels than in her private letters, which implies that, for Austen, referring to the human body was less appropriate in private written communication than in fiction; (2) the context of the researched words shows that film adaptations use some words in a more literal, physical sense, while Austen does it in a figurative sense or in set expressions which have no connection in meaning to the human body; at the same time, the frequency of the researched words in the screenplays is lower, which shows that, in corpus research, frequency is not always enough to draw conclusions on language features. The second conclusion is related not only to Austen's language, but also to using corpora in research.

It has to be mentioned here that the limitations of my research, the decision to focus on nouns only and to focus just on the nouns in Carol Shields's list do not allow me to provide a deeper insight into physicality in Austen's novels, letters, and film adaptations. Also, the collocations I have analysed were extracted only for two nodes, occurring only in one position. Further research needs to be carried out, with an enlarged scope for more precise conclusions to be drawn.

References

- Ferrin, Giovanni, Lauro Snidaro, Gian Luca Foresti. 2011. *Contexts, Co-texts and Situations in Fusion Domain*. [Online]. Available: https://www.researchgate.net/publication/252027427_Contexts_co-texts_and_situations_in_fusion_domain [Accessed 2022, April 10].
- Gatto, Maristela. 2014. *Web as Corpus: Theory and Practice*. London: Bloomsbury.
- McEnery, Tony, Andrew Wilson. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

- McEnery, Tony, Richard Xiao, Yukio Tono. 2006. *Corpus-based Language Studies: An Advanced Resource Book*. [Online]. Available: <https://www.jbe-platform.com/content/journals/10.1075/ijcl.11.4.09won> [Accessed 2022, April 10].
- McEnery, Tony, Andrew Hardie., 2012. *Corpus Linguistics: Method, Theory and Practice*. New York: Cambridge University Press
- Moon, Rosamund. 2010. "What Can a Corpus Tell Us about Lexis?" in Anne O'Keefe, Michael McCarthy (eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 197-211.
- Shields, Carol. 2001. *Jane Austen*. London: Phoenix
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press
- Thornbury, Scott. 2010. "What Can a Corpus Tell Us about Discourse?" in Anne O'Keefe, Michael McCarthy (eds.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge, pp. 270-287.
- Turner, Eleanor. 2016. *Word Counts for Your Favourite Regency Romance Novels*. [Online]. Available: [https://www.janetogeorgette.com/word-counts-for-your-favourite-regency-romancenovels/#:~:text=This%20is%20a%20pretty%20similar,Park%20\(159%2C%20526\)](https://www.janetogeorgette.com/word-counts-for-your-favourite-regency-romancenovels/#:~:text=This%20is%20a%20pretty%20similar,Park%20(159%2C%20526)) [Accessed 2021, January 8].
- Wahid, Ridwan. 2011. "The Use of Corpus-based Techniques in Literary Analysis: Exploring Learners' Perceptions" in *Asiatic* 5 (1), pp. 104-128. [Online]. Available: <https://journals.iium.edu.my/asiatic/index.php/AJELL/article/download/404/380> [Accessed 2022, April 10].

The source(s) for the microcorpus Letters of Jane Austen

- Austen, Jane. 2013. *The Letters of Jane Austen*. Sarah Chauncey Woolsey (ed.). [Online]. Available: <https://www.gutenberg.org/files/42078/42078-h/42078-h.htm> [Accessed 2022, April 10]

Links to online corpora and research tools

- AntConc*. Available: <https://www.laurenceanthony.net/software/antconc/> [Accessed 2022, April 10].
- BYU Corpus*. Available: <https://www.english-corpora.org/> [Accessed 2022, April 10].
- Lextutor Web Concordance*. Available: <https://www.lexutor.ca/conc/eng/> [Accessed 2022, April 10].
- Simple POS Tagger*. Available: http://martinweisser.org/ling_soft.html#tagger [Accessed 2022, April 10].