DOI: 10.13137/2039-8646/33605

# *'Pedagogical Stylistics' through Corpora in the University Classroom: A Case-Study*[*]

Piergiorgio Trevisan[**]
Dipartimento di Studi Umanistici
Università di Trieste
*ptrevisan@units.it*

Claudia Montico[***]
Pordenone
*cla.montico@gmail.com*

Abstract

*How can corpus linguistics be used for pedagogical purposes? How can it contribute to the study and appreciation of specific literary aspects? The aim of this paper is to show how students can become researchers, and use authentic language data to discover tendencies and thematic trends through inductive procedures: in other words, how awareness/consciousness can be triggered in a practical way, without being overtly deductive. To start with, some activities that were carried out in the language classroom are described: these include the introduction of crucial corpus methods such as keyness, collocations and concordances, followed by their application to a specific case-study, the novel* Pride and Prejudice *by Jane Austen. These activities proved very useful to raise the students' awareness regarding the use of particular language patterns by the English author. After that, the steps of a specific 'methodological journey' carried out by a student in her BA dissertation are presented: here, Austen's novel has been compared to the rest of her literary production first, and to a reference corpus composed by a sample of English literature (the Imaginative Writing Section of the British National Corpus) later. Overall, the activities described show that corpus methods can prove successful both for testing previously-formed assumptions about specific aspects of literature, and for improving language skills more generally.*

Parole chiave

Glottodidattica / Language Education; Lingua e Letteratura Inglese / English Language and Literature; Stilistica Pedagogica / Pedagogical Stylistics; Linguistica dei Corpora[1] /

[*] *Titolo: "Stilistica pedagogica" attraverso i corpora nell'aula universitaria: un caso di studio.*
[**] Ai fini di legge i paragrafi 1, 2 e 5 sono da attribuire a Piergiorgio Trevisan.
[***] Ai fini di legge i paragrafi 3 e 4 sono da attribuire a Claudia Montico.

[1] La *linguistica dei corpora* è un approccio allo studio del linguaggio che utilizza strumenti computazionali per l'analisi di tendenze linguistiche all'interno di grandi dati testuali memorizzati sui computer. I *corpora* possono avere dimensioni diverse: i primi erano costituiti da circa un milione di parole, oggi se ne trovano facilmente alcuni che superano il miliardo

CORPUS LINGUISTICS[2]; APPRENDIMENTO / LEARNING; JANE AUSTEN.

## 1. INTRODUCTION

At least from 1967, linguists have started to use corpora to address a plethora of language phenomena[3]: starting from large collections of data, key areas of enquiry like critical discourse analysis[4], translation studies[5], linguistic variation[6], translanguaging and metrolingualism[7], to quote but a few, have greatly benefited from the potential offered by software tools analysing very big amounts of language.

Crucially, notions coming from corpus linguistics have also been applied to the study of narrative. Louw[8], for example, has shown how corpora can be used to test the researcher's intuitions about symbolism in literature; Culpeper[9] has analysed *Romeo and Juliet* by focusing on particular keywords; Semino and Short[10] have shed important light on speech and thought presentation categories; Stubbs[11] has used quantitative methods to study Conrad's *Heart of Darkness*. More recently, Mahlberg has investigated notions such as 'psycholinguistic reality'[12] and 'mind modelling'[13] in a number of Dickens's novels.

Corpus methods have also proved crucial in 'Telecinematic Stylistics'[14]: Bednarek, for instance, has studied key aspects of televisual characterisation such as 'expressivity'[15],

di parole. La diffusione di questo approccio negli ultimi quarant'anni ha rivoluzionato la linguistica in molte delle sue declinazioni.

[2] Corpus linguistics is an approach to the study of language that uses computational tools for the analysis of linguistic trends within large textual data stored on computers. Corpora can have different sizes: the first ones were made up of about 1 million words, while today it is not rare to have corpora exceeding a billion words. The spread of this approach in the last 40 years has revolutionised several subfields in linguistics.

[3] In 1967, the 'Brown Corpus' was released at Brown University, US. Composed of about one million words belonging to various text genres of American English, it is considered the first collection of language aimed at making it possible to scientifically study the frequency and distribution of word categories in everyday language use.

[4] See BAKER 2006.

[5] See KOHN 1996; KRANICH 2014.

[6] See BIBER 1988, 2012.

[7] See PENNYCOCK, OTSUJI 2015; GARCIA, WEI 2014.

[8] See LOUW 1997.

[9] See CULPEPER 2009.

[10] See SEMINO, SHORT 2004.

[11] See STUBBS 2005.

[12] See MAHLBERG 2014.

[13] See MAHLBERG, STOCKWELL, DE JOODE, SMITH, O'DONNEL 2016.

[14] The expression *'telecinematic stylistics'* refers to studies applying stylistics methods to the analysis of discourse in film and television. See, in particular, HOFFMANN, KIRNER-LUDWIG 2020.

[15] See BEDNAREK 2011a.

the 'stability of the televisual character'[16], the construction of 'nerdiness'[17]; Pavesi[18] has focused on the role of demonstratives in film dialogues; Quaglio[19] has analysed the linguistic properties of dialogue in *Friends.*

The use of corpus linguistics methods for addressing literary and televisual aspects has often been referred to as 'corpus stylistics'[20]: tools like 'keyness', 'collocation', 'concordance', 'n-grams' etc., have indeed greatly improved the possibility to access crucial aspects of style, i.e. «the way in which language is used in a given context, by a given person, for a given purpose, and so on»[21]. As remarked by Wales[22], corpus stylistics investigations have helped identifying typical ways of using language that do not hold in one text only, but can be found across a number of different texts in a corpus[23] .

When corpus methods are applied to a specific text, the text under investigation is usually compared to a large 'reference corpus', which constitutes the 'norm' against which specific features of the target text 'stand out'. The selection of a proper reference corpus is crucial for the significance of the analysis. As Mahlberg remarks[24]:

> *We cannot simply assume that large general corpora constitute what makes 'ordinary' language so that we can contrast it with 'creative' language that stands out in an individual text compared to the large corpus, and that makes a piece of language 'literary' as opposed to 'ordinary'*

Depending on the particular research questions, the target corpus may therefore be compared to different sets of reference corpora: if the main purpose is to study the peculiar language of a particular TV show, for example, a reference corpus collecting language from other TV shows will be used[25]; on the other hand, if the purpose of the study is to look for potential differences between the language spoken by characters in movies and the language spoken by real people, a reference corpus containing real

[16] See Bednarek 2011b.
[17] See Bednarek 2012.
[18] See Pavesi 2020.
[19] See Quaglio 2009.
[20] See McIntyre, Walker 2019.
[21] See Leech, Short 1981, p. 11.
[22] See Wales 2001, p. 371.
[23] See Mahlberg 2007, p. 221.
[24] See Mahlberg 2007, p. 221.
[25] The *Sydney Corpus of Television Dialogue*, for example, contains approximately 275,000 words of dialogue from 66 US TV series. It can therefore be used as a reference 'norm' for the language of TV series. For further information, see The Sydney Corpus of Television Dialogue (SydTV) (see Web sites).

language interactions will clearly be more suitable[26].

Starting from these premises, the aim of this paper is to introduce a case-study dealing with corpus linguistics use in the University classroom. Specifically, the analysis of some aspects of Jane Austen's *Pride and Prejudice* (*P&P* henceforth) are presented as they were investigated during the class work first, and in a student's BA dissertation later. Given the specific, 'hands-on' teaching objectives of the activities, and the beginners status of the students, it was decided to only focus on the way in which corpus tools can be used to identify peculiar thematic features of *P&P*, compared to Austen's production at large. The thematic areas thus identified were then investigated by means of specific tools applied to *P&P* and to a narrative reference corpus extracted from the larger *British National Corpus* (BNC henceforth). Finally, students were introduced to other analytical possibilities through the discussion of recent studies that apply corpus methods to Austen's narrative[27].

The rest of the paper is organised as follows: section 2 presents the materials and the methods used in the research. Sections 3 and 4 present and discuss the data collected with two different software applications, Section 5 wraps the discussion up and introduces some concluding remarks.

## 2. Materials and methods

Two corpus tools previously introduced in the classroom were used in this study: *LancsBox* – developed at the University of Lancaster by Vaclav Brezina and Tony McEnery – and *W-Matrix* – developed at the same University by Paul Rayson[28]. Although the two software offer similar analytical tools, it was decided to use both of them as *W-Matrix* also contains a 'semantic tagging' function which, at the time, was not available in *LancsBox.*

[26] See, for example, Forchini, Seracini, Poli 2021.

[27] Previous studies investigating Jane Austen's production with corpus methods include Burrows (see Burrows 1987), that focuses on the role of particular function words as a general tool for constructing the characters' idiolect; Fischer-Starcke (see Fischer-Starcke 2010), who studies the impact of various features of text on literary meaning in *Northanger Abbey*; Bianchi (see Bianchi 2020), who investigates the role of 'suspended quotations' for the construction of style in Dickens and Austen.

[28] For detailed information on the two tools, see #LancsBox: Lancaster University Corpus Toolbox and Wmatrix Corpus Analysis and Comparison Tool (see Web sites).

Additionally, *W-Matrix* also incorporates the Imaginative Writing Section of the BNC among its default corpora. This gave the students the possibility to use a reference corpus which worked as a 'norm' sample of British literature. Specifically, the study was conducted by following these steps:

1. Initially, a *P&P* text file was downloaded from the *Project Gutenberg* database and saved as a *.txt* file. The file was then uploaded onto *LancsBox* and used to introduce the most important tools of quantitative analysis: 'Concordance', 'Keyness', 'Collocation', 'N-grams'. Since *LancsBox* contains the '*Jane Austen Corpus* 2019' among its default corpora, it was chosen to use it as a reference corpus to investigate keyness in *P&P*. As will be specified in part 3, keyness analysis made it possible to identify lexical items which are significantly more present in the novel than in the rest of Jane Austen's production.
2. To further analyse these findings, the same *P&P* file was then uploaded onto *W-Matrix* in order to also run analyses by using the BNC Sampler Written Image as reference corpus. Since this BNC sub-corpus is considered representative of British fiction in the last two centuries[29], students had the possibility to realise that some keywords, identical to those previously identified in *LancsBox*, could really be considered peculiar to Jane Austen's narrative.
3. In following classes, more robust evidence was collected by resorting to the semantic tagging function of *W-Matrix*: specifically, this function allows to identify the key semantic areas of a specific corpus compared to another one. As will be argued in the next sections, the same areas of meaning were proved statistically significant both when the corpus was analysed on its own and when it was compared to the BNC Sampler Written Image.
4. Finally, it was hypothesised that a POS analysis (Part of Speech Tagging) could help shedding more light on the data previously collected. In particular, the

[29] See BNC SAMPLER: XML EDITION (see Web sites) for more detailed information.

> POS function in *W-Matrix* made it possible for students to gain awareness regarding the most frequently used parts of speech in *P&P*. By doing so, they could also understand that grammar words tend to always appear first in frequency lists, regardless of the text analysed: therefore, they mainly focused on the content words occuring after the function words.

Thanks to the activities described above, students had the possibility to realise at least two important facts: first of all, observations should not begin with too many restrictions determined by previous assumptions; secondly, corpus linguistics and literary stylistics can powerfully complement each other but the 'quantitative' can never replace the 'qualitative'[30].

The next sections of this paper describe the steps of the 'methodological journey' later carried out by a student in her BA dissertation, Claudia Montico, starting from Michael Stubb's crucial assumption[31] that

> *Even if quantification only confirms what we already know, this is not a bad thing. Indeed, in developing a new method, it is perhaps better not to find anything too new, but to confirm findings from many years of traditional study, since this gives confidence that the method can be relied on.*

## 3. IDENTIFYING THEMES THROUGH LANCSBOX

As anticipated above, *P&P* was initially compared to a reference corpus comprising all Jane Austen's narrative production. Specifically, a keyness analysis of words was conducted to find out which lexical items were more frequent in the novel.

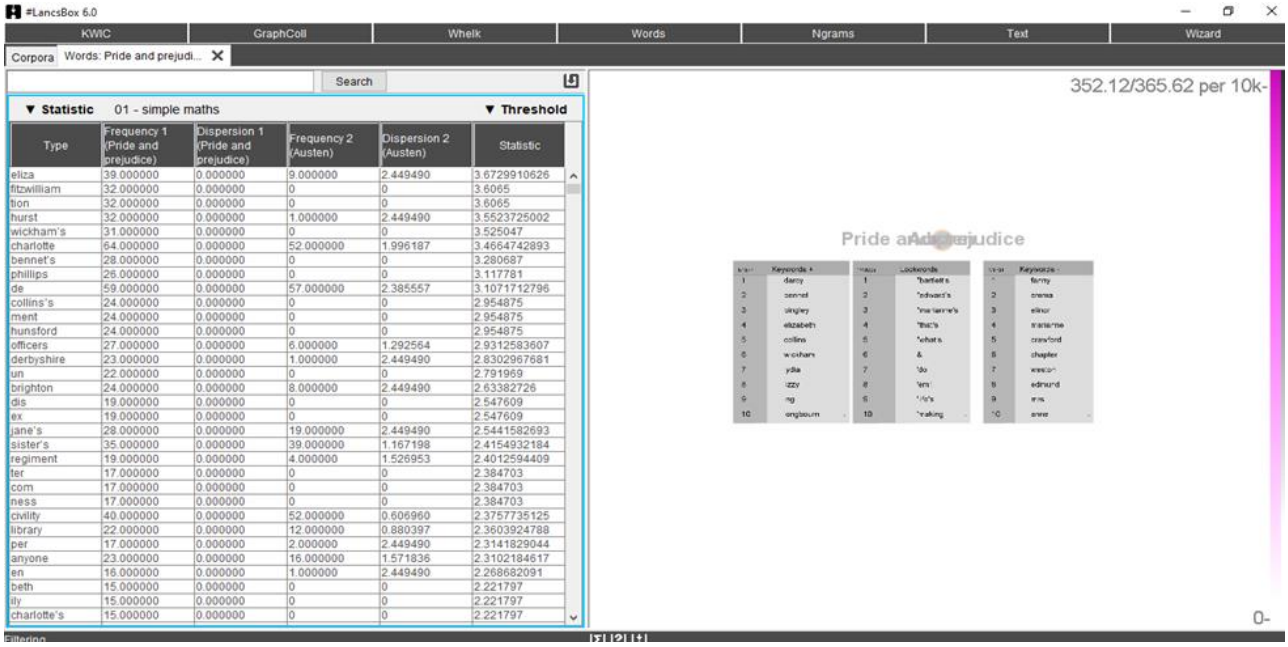
Not surprisingly, the first results in the list regarded proper names of characters like *Eliza*, *Charlotte* and *Fitzwilliam*, since these characters belong exclusively to *P&P* and are not part of any other Austen's novel.

However, this analysis also offered the possibility to focus on words that may not be easily recognised as 'key' with a traditional qualitative analysis alone. In particular, the author of the dissertation found it interesting to note that three very frequent lexical items were:

[30] See MAHLBERG 2007, p. 219.
[31] See STUBB 2005, p. 6.

'officers', 'regiment', 'civility'. All three terms gravitate around the semantic areas of 'military life', 'order', 'discipline', 'respect' and are peculiar to *P&P* but not to the other novels by the same author. Indeed, a common pattern of meaning among these areas was frequent during the Georgian period: the word 'civility' was mainly linked to military life as a synonym for 'good breeding' and 'politeness', as also pointed out by Huf[32]

> *In the context of the officer corps, civility appears to have been understood as behaviour which did not offend associates or companions, and it was this which forms the basis of acceptance by other officers.*




| Type | Frequency 1 (Pride and prejudice) | Dispersion 1 (Pride and prejudice) | Frequency 2 (Austen) | Dispersion 2 (Austen) | Statistic |
|---|---|---|---|---|---|
| eliza | 39.000000 | 0.000000 | 9.000000 | 2.449490 | 3.6729910626 |
| fitzwilliam | 32.000000 | 0.000000 | 0 | 0 | 3.6065 |
| tion | 32.000000 | 0.000000 | 0 | 0 | 3.6065 |
| hurst | 32.000000 | 0.000000 | 1.000000 | 2.449490 | 3.5523725002 |
| wickham's | 31.000000 | 0.000000 | 0 | 0 | 3.525047 |
| charlotte | 64.000000 | 0.000000 | 52.000000 | 1.996187 | 3.4664742893 |
| bennet's | 28.000000 | 0.000000 | 0 | 0 | 3.280687 |
| phillips | 26.000000 | 0.000000 | 0 | 0 | 3.117781 |
| de | 59.000000 | 0.000000 | 57.000000 | 2.385557 | 3.1071712796 |
| collins's | 24.000000 | 0.000000 | 0 | 0 | 2.954875 |
| ment | 24.000000 | 0.000000 | 0 | 0 | 2.954875 |
| hunsford | 24.000000 | 0.000000 | 0 | 0 | 2.954875 |
| officers | 27.000000 | 0.000000 | 6.000000 | 1.292564 | 2.9312583607 |
| derbyshire | 23.000000 | 0.000000 | 1.000000 | 2.449490 | 2.8302967681 |
| un | 22.000000 | 0.000000 | 0 | 0 | 2.791969 |
| brighton | 24.000000 | 0.000000 | 8.000000 | 2.449490 | 2.63382726 |
| dis | 19.000000 | 0.000000 | 0 | 0 | 2.547609 |
| ex | 19.000000 | 0.000000 | 0 | 0 | 2.547609 |
| jane's | 28.000000 | 0.000000 | 19.000000 | 2.449490 | 2.5441582693 |
| sister's | 35.000000 | 0.000000 | 39.000000 | 1.167198 | 2.4154932184 |
| regiment | 19.000000 | 0.000000 | 4.000000 | 1.526953 | 2.4012594409 |
| ter | 17.000000 | 0.000000 | 0 | 0 | 2.384703 |
| com | 17.000000 | 0.000000 | 0 | 0 | 2.384703 |
| ness | 17.000000 | 0.000000 | 0 | 0 | 2.384703 |
| civility | 40.000000 | 0.000000 | 52.000000 | 0.606960 | 2.3757735125 |
| library | 22.000000 | 0.000000 | 12.000000 | 0.880397 | 2.3603924788 |
| per | 17.000000 | 0.000000 | 2.000000 | 2.449490 | 2.3141829044 |
| anyone | 23.000000 | 0.000000 | 16.000000 | 1.571836 | 2.3102184617 |
| en | 16.000000 | 0.000000 | 1.000000 | 2.449490 | 2.268682091 |
| beth | 15.000000 | 0.000000 | 0 | 0 | 2.221797 |
| ity | 15.000000 | 0.000000 | 0 | 0 | 2.221797 |
| charlotte's | 15.000000 | 0.000000 | 0 | 0 | 2.221797 |

Figure 1. Keyness analysis (Fonte: LancsBox)

Since these words were so frequent in *P&P*, Montico considered it important to observe their 'behaviour' more closely: she therefore decided to resort to the Concordance tool in LancsBox, which provides the possibility to study all the occurrences of a particular item in a text, with a span of some words on the left and some words on the right. Importantly, this analytical procedure is 'corpus-driven'[33], since the ideas regarding what to investigate qualitatively were prompted by the previous quantitative analysis.

[32] See Huf 2017, p. 56.

[33] Even though corpus-driven and corpus-based approaches can give researchers complementary insights, they are regarded as slightly different. Specifically, corpus-driven approaches are generally used for various semantic information, whereas corpus-based analyses play a very important role for obtaining more grammatically-oriented data.

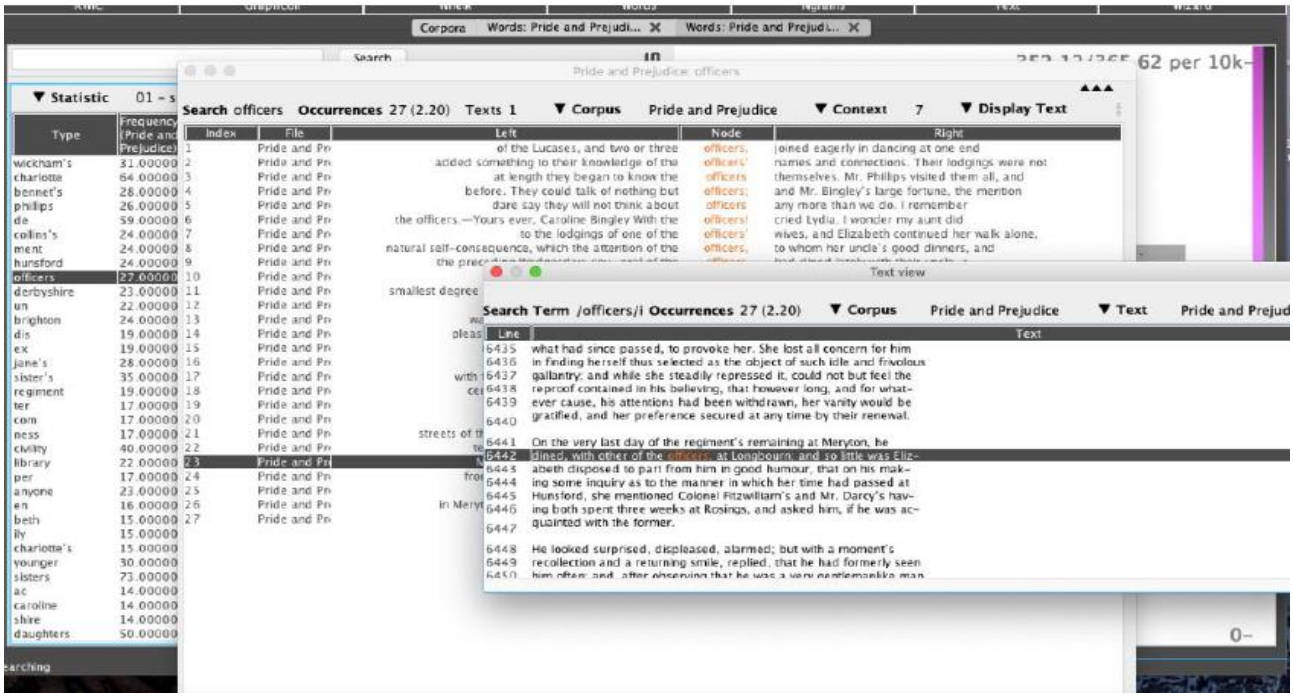



Figure 2. Concordance analysis of 'Officers' (Fonte: LancsBox)

As can be seen, not only does the software offer the possibility to study the chosen word within a particular word-span (shot in the middle), but it also makes it possible to extend the investigation to a larger portion of text surrounding that word (shot in the foreground). At this point, a more qualitative oriented analysis can be conducted. What Montico could infer thanks to the observation of these words in their actual context of use integrates her previous knowledge regarding values dominating British society at the beginning of the 19th century. In particular, among the concordance lines of the word 'civility', she found it quite indicative that the character Mr Darcy pays the 'utmost civility' to Elizabeth (one of the female protagonists) by pointing out how his aunt (Lady Catherine) couldn't have bestowed her kindness on a more grateful recipient, Mr. Collins. Notably, the student noticed, 'utmost civility' is paid by Mr. Darcy with respect to Elizabeth, a character belonging to a lower social rank. The word 'civility' is therefore used by Jane Austen in this context not so much with connotations related to 'citizenship' and 'civil order', as may perhaps be expected, but rather as a sort of 'social trademark' characterising the speech of people belonging to

higher social ranks: in other terms, as an essential politeness device creating social cohesion in specific occasions.

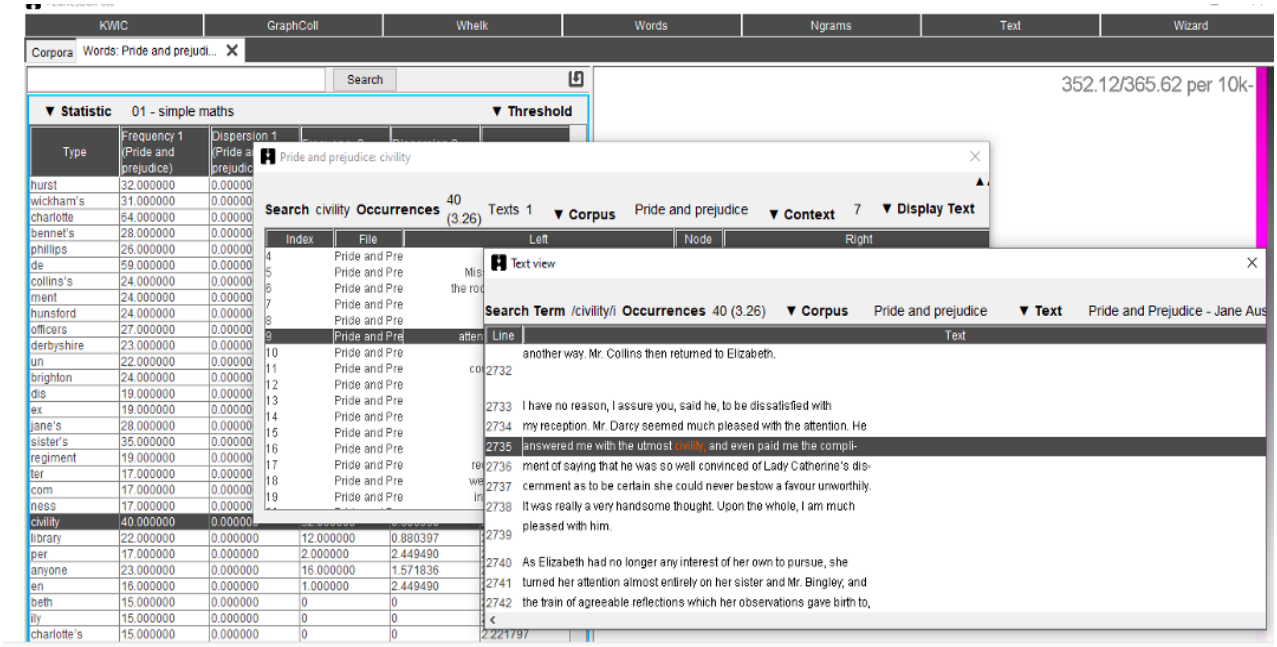



Figure 3. Concordance analysis of 'Civility' (Fonte: LancsBox).

As far as the word 'officers' is concerned, the main semantic area it is related to in the novel is that of 'social prestige': as remarked by Michael Glover[34], it is still unclear whether an officer could be addressed as a gentleman due to his role in the army or if he was genuinely a gentleman beyond being a soldier. In Huf's words[35]:

> *Acceptance and entry into an institution underpinned by an entrenched sense of gentlemanliness could be expected to encourage the view that one was either a gentleman already, or became a gentleman by becoming an officer. [...] By specifically outlining gentlemanliness as a criterion for holding a commission, a recommendation taking this form appears to assume that the applicant was already, in some sense, a gentleman.*

Besides being soldiers, 'officers' therefore acquired the right to be addressed with the honorific 'gentleman': not rarely, then, the social prestige they enjoyed made them become suitable parties in 'combined marriages'. Lydia's marriage with Wickham in *P&P* is a clear example of this: thanks to Darcy, who pays a commission for Wickam's career advancement in the army, a combined marriage between the two is organised.

[34] See Glover 1980, p. 233.
[35] See Huf 2017, pp. 45-46.

This is also crucially important for 'wiping out' potential family reputation issues, thus making it possible for Lydia's sisters to also get married.

Thanks to the observation of the Concordance results, it was therefore possible for Montico to realise that the use of the words 'officers', 'regiment' and 'civility'in *P&P* seems to slightly deviate from its primary thematic area of 'authority', 'law' and, to some extent, 'order': rather, this use of lexis seems more related to notions of social cohesion or to the embodiment of 'suitable options' for young women of marriageable age. Therefore, the characters Austen produces may be considered thematically emblematic of the rigorous order of the Georgian period.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 35 | Concordance | sisters | 107 | 0.09 | 20 | 0.01 + | 136.02 | 3.37 |
| 36 | Concordance | dear | 155 | 0.13 | 56 | 0.03 + | 135.67 | 2.42 |
| 37 | Concordance | may | 189 | 0.16 | 89 | 0.04 + | 131.88 | 2.03 |
| 38 | Concordance | any | 262 | 0.23 | 165 | 0.07 + | 131.13 | 1.61 |
| 39 | Concordance | Mr._Wickham | 60 | 0.05 | 0 | 0.00 + | 128.93 | 7.85 |
| 40 | Concordance | Miss_Bennet | 59 | 0.05 | 0 | 0.00 + | 126.78 | 7.83 |
| 41 | Concordance | every | 163 | 0.14 | 69 | 0.03 + | 125.49 | 2.19 |
| 42 | Concordance | been | 514 | 0.45 | 487 | 0.22 + | 124.45 | 1.03 |
| 43 | Concordance | manner | 91 | 0.08 | 15 | 0.01 + | 121.64 | 3.55 |
| 44 | Concordance | Meryton | 54 | 0.05 | 0 | 0.00 + | 116.03 | 7.70 |
| 45 | Concordance | herself | 201 | 0.17 | 113 | 0.05 + | 116.02 | 1.78 |
| 46 | Concordance | feelings | 81 | 0.07 | 13 | 0.01 + | 109.36 | 3.59 |
| 47 | Concordance | as | 822 | 0.71 | 963 | 0.43 + | 107.58 | 0.72 |
| 48 | Concordance | should | 246 | 0.21 | 172 | 0.08 + | 106.02 | 1.46 |
| 49 | Concordance | by | 570 | 0.49 | 599 | 0.27 + | 105.46 | 0.88 |
| 50 | Concordance | opinion | 61 | 0.05 | 4 | 0.00 + | 104.36 | 4.88 |
| 51 | Concordance | most | 184 | 0.16 | 105 | 0.05 + | 104.34 | 1.76 |
| 52 | Concordance | soon | 161 | 0.14 | 82 | 0.04 + | 103.76 | 1.92 |
| 53 | Concordance | behaviour | 48 | 0.04 | 0 | 0.00 + | 103.14 | 7.53 |
| 54 | Concordance | Pemberley | 46 | 0.04 | 0 | 0.00 + | 98.84 | 7.47 |
| 55 | Concordance | Mrs._Gardiner | 46 | 0.04 | 0 | 0.00 + | 98.84 | 7.47 |
| 56 | Concordance | ladies | 74 | 0.06 | 13 | 0.01 + | 96.49 | 3.46 |
| 57 | Concordance | pleasure | 86 | 0.07 | 21 | 0.01 + | 96.37 | 2.98 |
| 58 | Concordance | daughters | 56 | 0.05 | 4 | 0.00 + | 94.28 | 4.75 |
| 59 | Concordance | added | 71 | 0.06 | 12 | 0.01 + | 94.00 | 3.51 |
| 60 | Concordance | subject | 75 | 0.06 | 15 | 0.01 + | 92.59 | 3.27 |
| 61 | Concordance | Rosings | 42 | 0.04 | 0 | 0.00 + | 90.25 | 7.34 |
| 62 | Concordance | much | 240 | 0.21 | 183 | 0.08 + | 89.92 | 1.34 |
| 63 | Concordance | than | 273 | 0.24 | 228 | 0.10 + | 86.64 | 1.21 |
| 64 | Concordance | civility | 40 | 0.03 | 0 | 0.00 + | 85.95 | 7.27 |
| 65 | Concordance | Charlotte | 57 | 0.05 | 7 | 0.00 + | 84.14 | 3.97 |
| 66 | Concordance | acquaintance | 46 | 0.04 | 2 | 0.00 + | 83.89 | 5.47 |
| 67 | Concordance | affection | 56 | 0.05 | 7 | 0.00 + | 82.23 | 3.95 |
| 68 | Concordance | their | 423 | 0.37 | 440 | 0.20 + | 80.56 | 0.89 |
| 69 | Concordance | hope | 121 | 0.10 | 60 | 0.03 + | 80.18 | 1.96 |
| 70 | Concordance | Lydias | 37 | 0.03 | 0 | 0.00 + | 79.50 | 7.16 |

| | Item | O1 | %1 | O2 | %2 | LL | LogRatio | |
|---|---|---|---|---|---|---|---|---|
| 1 | Concordance | Elizabeth | 554 | 0.48 | 7 | 0.00 + | 1120.97 | 7.25 |
| 2 | Concordance | her | 2136 | 1.85 | 1708 | 0.77 + | 735.80 | 1.27 |
| 3 | Concordance | Mr._Darcy | 242 | 0.21 | 0 | 0.00 + | 520.00 | 9.87 |
| 4 | Concordance | Jane | 256 | 0.22 | 14 | 0.01 + | 451.66 | 5.14 |
| 5 | Concordance | of | 3340 | 2.89 | 3885 | 1.75 + | 448.37 | 0.73 |
| 6 | Concordance | not | 1505 | 1.30 | 1318 | 0.59 + | 434.10 | 1.14 |
| 7 | Concordance | to | 3821 | 3.31 | 4802 | 2.16 + | 380.83 | 0.62 |
| 8 | Concordance | be | 1226 | 1.06 | 1088 | 0.49 + | 343.87 | 1.12 |
| 9 | Concordance | Mrs._Bennet | 136 | 0.12 | 0 | 0.00 + | 292.23 | 9.03 |
| 10 | Concordance | which | 535 | 0.46 | 319 | 0.14 + | 287.48 | 1.69 |
| 11 | Concordance | Mr._Collins | 129 | 0.11 | 0 | 0.00 + | 277.19 | 8.96 |
| 12 | Concordance | sister | 176 | 0.15 | 20 | 0.01 + | 265.71 | 4.08 |
| 13 | Concordance | Lydia | 122 | 0.11 | 0 | 0.00 + | 262.15 | 8.88 |
| 14 | Concordance | she | 1704 | 1.48 | 1971 | 0.89 + | 233.25 | 0.74 |
| 15 | Concordance | Lizzy | 94 | 0.08 | 0 | 0.00 + | 201.98 | 8.50 |
| 16 | Concordance | Darcy | 92 | 0.08 | 0 | 0.00 + | 197.69 | 8.47 |
| 17 | Concordance | have | 823 | 0.71 | 789 | 0.35 + | 193.73 | 1.01 |
| 18 | Concordance | bingley | 87 | 0.08 | 0 | 0.00 + | 186.94 | 8.39 |
| 19 | Concordance | Mr._Bingley | 87 | 0.08 | 0 | 0.00 + | 186.94 | 8.39 |
| 20 | Concordance | such_a | 143 | 0.12 | 25 | 0.01 + | 186.83 | 3.46 |
| 21 | Concordance | Lady_Catherine | 81 | 0.07 | 0 | 0.00 + | 174.05 | 8.29 |
| 22 | Concordance | Wickham | 80 | 0.07 | 0 | 0.00 + | 171.90 | 8.27 |
| 23 | Concordance | though | 222 | 0.19 | 96 | 0.04 + | 167.72 | 2.16 |
| 24 | Concordance | am | 317 | 0.27 | 192 | 0.09 + | 166.98 | 1.67 |
| 25 | Concordance | Longbourn | 77 | 0.07 | 0 | 0.00 + | 165.45 | 8.21 |
| 26 | Concordance | Mr._Bennet | 76 | 0.07 | 0 | 0.00 + | 163.31 | 8.20 |
| 27 | Concordance | had | 1162 | 1.01 | 1341 | 0.60 + | 160.33 | 0.74 |
| 28 | Concordance | however | 128 | 0.11 | 26 | 0.01 + | 156.93 | 3.25 |
| 29 | Concordance | such | 214 | 0.19 | 96 | 0.04 + | 156.37 | 2.10 |
| 30 | Concordance | him | 762 | 0.66 | 772 | 0.35 + | 155.93 | 0.93 |
| 31 | Concordance | Miss_Bingley | 69 | 0.06 | 0 | 0.00 + | 148.26 | 8.06 |
| 32 | Concordance | Kitty | 69 | 0.06 | 0 | 0.00 + | 148.26 | 8.06 |
| 33 | Concordance | Netherfield | 67 | 0.06 | 0 | 0.00 + | 143.97 | 8.01 |
| 34 | Concordance | very | 464 | 0.40 | 402 | 0.18 + | 136.85 | 1.15 |

Figure 4. Keyness analysis (Fonte: WMATRIX).

## 4. COLLECTING MORE EVIDENCE THROUGH W-MATRIX: WORDS, MEANINGS, PARTS OF SPEECH

Montico's investigation continued by comparing *P&P* to the Imaginative Writing Section (IWS henceforth) of the BNC by means of the software tool W-matrix. The primary aim of this analysis was to investigate whether the three thematic areas previously identified held valid on a wider comparison with British narrative more in general. First of all, a keyness analysis was therefore conducted with these two corpora. Not surprisingly, the first results included a list of grammar words ("of", "not", "to", etc.) and of characters' proper names (v. Figure 4).

Quite interestingly, however, the first words appearing after the characters' names refer to 'behaviour' and to 'feelings'. 'Civility' is among them, with a very high log-likelihood (85.95)[36]: this confirmed the crucial role of this semantic area in the novel. Additionally, the student realised that the term '*Acquaintances*' also features a very high log-likelihood score (83.89) and is often combined with the term 'Civility' in the novel: indeed, the process of making new acquaintances, e.g. the opportunities to socialise, is a pivoltal driving device for the plot in the whole novel. In order to socially relate with new acquaintances, the characters need to display their 'good breeding' manners, even when it is clear that they may despise them. The Concordance tool is very useful for retrieving examples dealing with this:

> *She answered him with cold* ***civility****. He sat down for a few moments, and then getting up, walked about the room. Elizabeth was surprised, but said not a word. After a silence of several minutes, he came towards her in an agitated manner, and thus began: In vain I have struggled. It will not do. My feelings will not be repressed. You must allow me to tell you how ardently I admire and love you. Elizabeth's astonishment was beyond expression.*

Here, despite being disgusted by Darcy's behaviour, Elizabeth still treats him with 'civility', thus sticking to the social 'etiquette' that is a 'politeness trademark' in the novel. At this point, Montico found it useful to also investigate *P&P*'s 'semantic profiling' by means of another tool embedded in W-matrix: Semantic tagging (see Section 2 for

[36] Statistically speaking the log-likelihood is a term that suggests the measures of goodness of fit of a statistical model to a sample data for given values of the unknown parameters. The concept has been applied to Corpus Linguistics studies in order to give scientific value to research.

details). Specifically, this was aimed at identifying possible meanings revolving around the notion of 'civility'.

| | | | Item | O1 | %1 | O2 | %2 | | LL | LogRatio | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | List1 | Concordance | A13.3 | 1683 | 1.46 | 1256 | 0.56 | + | 653.82 | 1.37 | Degree: Boosters |
| 2 | List1 | Concordance | E4.2+ | 385 | 0.33 | 121 | 0.05 | + | 371.70 | 2.62 | Content |
| 3 | List1 | Concordance | Z99 | 2989 | 2.59 | 3788 | 1.70 | + | 287.43 | 0.61 | Unmatched |
| 4 | List1 | Concordance | X2.6+ | 293 | 0.25 | 101 | 0.05 | + | 265.45 | 2.48 | Expected |
| 5 | List1 | Concordance | S7.2+ | 239 | 0.21 | 61 | 0.03 | + | 261.53 | 2.92 | Respected |
| 6 | List1 | Concordance | A13 | 114 | 0.10 | 0 | 0.00 | + | 244.96 | 8.78 | Degree |
| 7 | List1 | Concordance | A7+ | 2198 | 1.90 | 2703 | 1.21 | + | 239.52 | 0.65 | Likely |
| 8 | List1 | Concordance | S1.1.1 | 405 | 0.35 | 237 | 0.11 | + | 222.76 | 1.72 | Social Actions, States And Processes |
| 9 | List1 | Concordance | A2.2 | 645 | 0.56 | 526 | 0.24 | + | 214.24 | 1.24 | Cause&Effect/Connection |
| 10 | List1 | Concordance | T1.1 | 114 | 0.10 | 4 | 0.00 | + | 213.36 | 5.78 | Time: General |
| 11 | List1 | Concordance | S4 | 1603 | 1.39 | 1879 | 0.84 | + | 209.39 | 0.72 | Kin |
| 12 | List1 | Concordance | S1.2.4+ | 217 | 0.19 | 77 | 0.03 | + | 192.50 | 2.44 | Polite |
| 13 | List1 | Concordance | X2.6- | 146 | 0.13 | 28 | 0.01 | + | 183.58 | 3.33 | Unexpected |
| 14 | List1 | Concordance | N5.1+ | 1067 | 0.92 | 1178 | 0.53 | + | 170.33 | 0.80 | Entire; maximum |
| 15 | List1 | Concordance | A13.2 | 347 | 0.30 | 225 | 0.10 | + | 166.89 | 1.57 | Degree: Maximizers |
| 16 | List1 | Concordance | Z8 | 18950 | 16.42 | 32536 | 14.62 | + | 159.49 | 0.17 | Pronouns |
| 17 | List1 | Concordance | A13.7 | 208 | 0.18 | 109 | 0.05 | + | 130.01 | 1.88 | Degree: Minimizers |
| 18 | List1 | Concordance | G2.2+ | 158 | 0.14 | 64 | 0.03 | + | 126.31 | 2.25 | Ethical |
| 19 | List1 | Concordance | N5+ | 570 | 0.49 | 578 | 0.26 | + | 116.36 | 0.93 | Quantities: many/much |
| 20 | List1 | Concordance | S1.2 | 82 | 0.07 | 15 | 0.01 | + | 105.18 | 3.40 | Personality traits |
| 21 | List1 | Concordance | S6+ | 813 | 0.70 | 970 | 0.44 | + | 99.56 | 0.69 | Strong obligation or necessity |
| 22 | List1 | Concordance | S1.2.6+ | 91 | 0.08 | 28 | 0.01 | + | 89.08 | 2.65 | Sensible |
| 23 | List1 | Concordance | X2.1 | 1181 | 1.02 | 1579 | 0.71 | + | 88.53 | 0.53 | Thought, belief |
| 24 | List1 | Concordance | E2+ | 498 | 0.43 | 559 | 0.25 | + | 75.40 | 0.78 | Like |
| 25 | List1 | Concordance | S3.1 | 380 | 0.33 | 396 | 0.18 | + | 71.99 | 0.89 | Personal relationship: General |
| 26 | List1 | Concordance | I1.1+++ | 37 | 0.03 | 1 | 0.00 | + | 71.09 | 6.16 | Money: Affluence |
| 27 | List1 | Concordance | S7.2- | 51 | 0.04 | 12 | 0.01 | + | 58.26 | 3.03 | No respect |
| 28 | List1 | Concordance | X5.1+ | 112 | 0.10 | 70 | 0.03 | + | 56.63 | 1.63 | Attentive |
| 29 | List1 | Concordance | Q2.2 | 1453 | 1.26 | 2171 | 0.98 | + | 55.52 | 0.37 | Speech acts |
| 30 | List1 | Concordance | Z5 | 32974 | 28.57 | 60412 | 27.15 | + | 55.41 | 0.07 | Grammatical bin |
| 31 | List1 | Concordance | A5.1+ | 617 | 0.53 | 810 | 0.36 | + | 50.57 | 0.55 | Evaluation: Good |
| 32 | List1 | Concordance | I1.1 | 146 | 0.13 | 118 | 0.05 | + | 49.31 | 1.25 | Money and pay |
| 33 | List1 | Concordance | Z6 | 2187 | 1.89 | 3485 | 1.57 | + | 48.06 | 0.28 | Negative |
| 34 | List1 | Concordance | N5.1+++ | 21 | 0.02 | 0 | 0.00 | + | 45.12 | 6.34 | Entire; maximum |

Figure 5. Semantic Tagging in *W-matrix* (Fonte: WMATRIX).

When compared to the IWS of the BNC by means of semantic keyness analysis, some areas showed a strikingly high Log Likelihood: crucially, 'Social Actions' and 'Polite' are among them. The first area, in particular, is significantly related to the words 'behaviour' and 'manners', which had already proved central in the analysis of words discussed above. Semantic areas, in other words, seemed to go 'hand in hand' with word frequency, thus corroborating the previous observations regarding the centrality of social behaviour in the novel.

Since the semantic tagging function in W-Matrix also allows the observation of Concordances, it was extremely useful to study the actual usage of words related to social behaviour in context. As Figure 6 shows, the word 'civility' is present in the examples. Specifically, among the 217 occurrences of the 'politeness' semantic area, 'civility' appears 34 times: this supports the view that, in Austen's novel, to be polite seems to largely be possible by exhibiting 'civility' traits, even if other social conducts could of course be used to address the same purpose. This trait is clearly in line with Jane Austen being one of the most prominent members of the so-called 'novel of manners'.

| 217 occurrences. | | | Extend context |
|---|---|---|---|
| time . I did not expect such a | compliment | . Did not you ? I did for you . | 1 More \| Full |
| e great difference between us . | Compliments | always take you by surprise , an | 2 More \| Full |
| ious , and his manners , though | well-bred | , were not inviting . In that re | 3 More \| Full |
| tion at St. Jamess had made him | courteous | . Lady Lucas was a very good kin | 4 More \| Full |
| the dark . There is so much of | gratitude | or vanity in almost every attach | 5 More \| Full |
| was glad to purchase praise and | gratitude | by Scotch and Irish airs , at th | 6 More \| Full |
| not think it would be a proper | compliment | to the place ? It is a complimen | 7 More \| Full |
| mpliment to the place ? It is a | compliment | which I never pay to any place i | 8 More \| Full |
| artner . Mr. Darcy , with grave | propriety | , requested to be allowed the ho | 9 More \| Full |
| ne half-hour . Mr. Darcy is all | politeness | , said Elizabeth , smiling . He | 10 More \| Full |
| e was received , however , very | politely | by them ; and in their brothers | 11 More \| Full |
| there was something better than | politeness | ; there was good humour and kind | 12 More \| Full |
| t little besides expressions of | gratitude | for the extraordinary kindness s | 13 More \| Full |
| eing her quite well . Elizabeth | thanked | him from her heart , and then wa | 14 More \| Full |
| , said Miss Bingley , with cold | civility | , that Miss Bennet will receive | 15 More \| Full |
| I wish I might take this for a | compliment | ; but to be so easily seen throu | 16 More \| Full |
| Mrs. Bennet began repeating her | thanks | to Mr. Bingley for his kindness | 17 More \| Full |
| . I mend pens remarkably well . | Thank | youbut I always mend my own . Ho | 18 More \| Full |
| te ill . That will not do for a | compliment | to Darcy , Caroline , cried her | 19 More \| Full |
| to be a sort of panegyric , of | compliment | to yourselfand yet what is there | 20 More \| Full |
| ting what my friend says into a | compliment | on the sweetness of my temper . | 21 More \| Full |
| g one argument in favour of its | propriety | . To yield readilyeasilyto the p | 22 More \| Full |
| yield without conviction is no | compliment | to the understand-ing of either | 23 More \| Full |
| t of the room , I shall be very | thankful | ; and then you may say whatever | 24 More \| Full |
| the pianoforte ; and , after a | polite | request that Elizabeth would lea | 25 More \| Full |
| lead the way which the other as | politely | and more earnestly negatived , s | 26 More \| Full |
| himself to Miss Bennet , with a | polite | congratulation ; Mr. Hurst also | 27 More \| Full |
| less in the real object of her | civility | ; Mr. Darcy looked up . He was a | 28 More \| Full |
| ll , took place . Miss Bingleys | civility | to Elizabeth increased at last v | 29 More \| Full |
| in , dear sir , with respectful | compliments | to your lady and daughters , you | 30 More \| Full |
| to be a most conscientious and | polite | young man , upon my word , and I | 31 More \| Full |
| e , and was received with great | politeness | by the whole family . Mr. Bennet | 32 More \| Full |

Figure 6. Semantic tagging concordance tool of the 'Politeness' semantic area in *W-matrix* (Fonte: WMATRIX).

Finally, starting from the assumption that the relation between language and politeness may emerge at different language levels, the student was guided to also investigate the role of 'honorifics'. To do so, a Part of Speech (POS) analysis was conducted to find out potential overuses of this pattern. Since it only focuses on speech parts, this type of analysis moves away from more traditional word/semantic frequency explorations:

| | | Item | O1 | %1 | O2 | %2 | LL | LogRatio |
|---|---|---|---|---|---|---|---|---|
| 1 | List1 \| Concordance | NNB | 1631 | 1.41 | 762 | 0.34 + | 1146.85 | 2.05 |
| 2 | List1 \| Concordance | IO | 3379 | 2.93 | 3873 | 1.74 + | 477.24 | 0.75 |
| 3 | List1 \| Concordance | PPHO1 | 1490 | 1.29 | 1327 | 0.60 + | 414.75 | 1.11 |
| 4 | List1 \| Concordance | VBI | 1233 | 1.07 | 1031 | 0.46 + | 390.41 | 1.21 |
| 5 | List1 \| Concordance | VVN | 3132 | 2.71 | 3719 | 1.67 + | 390.39 | 0.70 |
| 6 | List1 \| Concordance | RG | 1263 | 1.09 | 1178 | 0.53 + | 317.24 | 1.05 |
| 7 | List1 \| Concordance | DA | 474 | 0.41 | 235 | 0.11 + | 314.16 | 1.96 |
| 8 | List1 \| Concordance | TO | 2399 | 2.08 | 2834 | 1.27 + | 304.71 | 0.71 |
| 9 | List1 \| Concordance | APPGE | 4420 | 3.83 | 6025 | 2.71 + | 299.82 | 0.50 |
| 10 | List1 \| Concordance | CST | 1226 | 1.06 | 1214 | 0.55 + | 266.31 | 0.96 |
| 11 | List1 \| Concordance | NP2 | 181 | 0.16 | 42 | 0.02 + | 208.24 | 3.05 |
| 12 | List1 \| Concordance | VM | 2838 | 2.46 | 3919 | 1.76 + | 179.42 | 0.48 |
| 13 | List1 \| Concordance | RGT | 171 | 0.15 | 62 | 0.03 + | 149.27 | 2.41 |
| 14 | List1 \| Concordance | PPX1 | 473 | 0.41 | 427 | 0.19 + | 127.85 | 1.09 |
| 15 | List1 \| Concordance | RR | 4511 | 3.91 | 7041 | 3.16 + | 120.72 | 0.30 |
| 16 | List1 \| Concordance | VBN | 516 | 0.45 | 500 | 0.22 + | 118.34 | 0.99 |
| 17 | List1 \| Concordance | IW | 1170 | 1.01 | 1483 | 0.67 + | 112.43 | 0.61 |
| 18 | List1 \| Concordance | CSA | 639 | 0.55 | 694 | 0.31 + | 107.31 | 0.83 |
| 19 | List1 \| Concordance | PPHS1 | 3038 | 2.63 | 4625 | 2.08 + | 100.51 | 0.34 |
| 20 | List1 \| Concordance | JJT | 216 | 0.19 | 145 | 0.07 + | 98.90 | 1.52 |
| 21 | List1 \| Concordance | CS | 1428 | 1.24 | 1959 | 0.88 + | 93.60 | 0.49 |
| 22 | List1 \| Concordance | VHD | 1142 | 0.99 | 1501 | 0.67 + | 93.07 | 0.55 |
| 23 | List1 \| Concordance | CSN | 272 | 0.24 | 223 | 0.10 + | 89.45 | 1.23 |
| 24 | List1 \| Concordance | DA1 | 186 | 0.16 | 123 | 0.06 + | 87.02 | 1.54 |
| 25 | List1 \| Concordance | VHI | 458 | 0.40 | 481 | 0.22 + | 84.89 | 0.88 |
| 26 | List1 \| Concordance | DDQ | 1006 | 0.87 | 1315 | 0.59 + | 84.15 | 0.56 |
| 27 | List1 \| Concordance | RGR | 165 | 0.14 | 124 | 0.06 + | 63.36 | 1.36 |
| 28 | List1 \| Concordance | PNQO | 84 | 0.07 | 39 | 0.02 + | 59.42 | 2.05 |
| 29 | List1 \| Concordance | VBDZ | 1843 | 1.60 | 2855 | 1.28 + | 52.73 | 0.32 |
| 30 | List1 \| Concordance | CCB | 950 | 0.82 | 1355 | 0.61 + | 49.69 | 0.43 |
| 31 | List1 \| Concordance | DDQGE | 59 | 0.05 | 28 | 0.01 + | 40.86 | 2.02 |
| 32 | List1 \| Concordance | DD | 478 | 0.41 | 624 | 0.28 + | 40.23 | 0.56 |
| 33 | List1 \| Concordance | RRQV | 33 | 0.03 | 9 | 0.00 + | 34.78 | 2.82 |
| 34 | List1 \| Concordance | CS31 | 60 | 0.05 | 35 | 0.02 + | 33.13 | 1.72 |

Figure 7. Part of Speech (POS) analysis (Fonte: WMATRIX).

Quite strikingly, the NNB category (e.g. titles preceding nouns) presented a very high log-likelihood score of 1146.65, which was even more surprising in comparison with the second most frequent tag, that only reached 477.

The POS findings therefore shed further light on the features observed during the ealier stages of the research and contributed to foregrounding them even more: regardless of characters, events, plot twists, etc. what seems central to Austen's fictional world revolves around the 'social behaviour' core nucleus. All the quantitative investigations carried out by the student indeed show that this area is semantically and lexically pervasive in the novel: though very useful, a qualitative analysis alone may not have revealed the contribution of these aspects to the general atmosphere of *P&P.*

## 5. Concluding remarks

This paper has discussed a case-study related to what has recently become known as 'pedagogical stylistics'[37], e.g. the use of stylistics to raise language awareness and skills in the language teaching context. In particular, a series of corpus-informed analyses have been carried out, thereby taking a corpus stylistic approach to learning tasks.

It is now widely recognised that corpus linguistics in general has had an important impact on English language teaching[38] in general, and that a growing number of stylistics sillabi use data-driven learning (DDL) to teach students how corpus tools like concordance lines or keyness analyses may be crucial for new discoveries or for answering questions about language.

By using DDL, then, students become researchers, and use authentic language data to discover tendencies and thematic trends through inductive procedures: in other words, the corpus work raises awareness/consciousness in a practival way, without being overtly deductive. Indeed, consciousness-raising by means of inductive practices is now an established approach in language teaching.

It goes without saying that incorporating corpus activities into the stylistics classroom is not without limitations: first of all, some corpora are not freely available so it may take quite a lot of time for the teacher to collect the materials to use in the classroom. Secondly, while corpus interfaces provide invaluable tools like concordances, collocations

[37] See Burke 2010.
[38] See Aarts, Smith-Dennis 2018.

and frequency lists, not all corpora are encoded grammatically, therefore important information emerging from the analysis of grammatical patterns may go lost.

Thirdly, the meta-language associated to the use of corpora may be off-putting for some learners – and for some teachers[39] Despite these drawbacks, as this contribution has hopefully at least in part shown, approaching the study of literature through corpus tools can be very beneficial for easily identifying recurrent words, semantic areas, parts of speech. Clearly enough, the present study is limited to concordance and to keyness analysis, therefore the findings are to be interpreted as merely an initial and partial overview of what could be done with students both in the classroom and in more independent research contexts. Crucial insights may of course be obtained by using other corpus tools: with the study of collocations, for example, students may realise that some lexical patterns in a specific narrative or poem do not reflect conventional words combination in the English language, therefore they may be empowered regarding the mechanisms of foregrounding, besides generally improving their language skills.

Future avenues of research in this field may include – but not be limited to – the creation of training materials for students of language and of literature: this could foster hands-on inductive activities in a highly motivating environment, while favoring fruitful interdisciplinary dialogue between subjects that have traditionally been considered as distinct and apart.

## References

Aarts B., Smith-Dennis E.
2018, *Using corpora for English language teaching and learning* in D. McIntyre, H. Price (eds), «Applying Linguistics: Language and the Impact Agenda», Abingdon, Routledge, pp. 163-176.

Baker P.
2006, *Using Corpora in Discourse Analysis*, London, Continuum.

Bednarek M.
2011a, «Expressivity and televisual characterisation», *Language and Literature*, 20, n. 1, pp. 3-21.

[39] See McIntyre 2019, p. 214.

2011b, *The stability of the televisual character: A corpus stylistic case study*, in R. PIAZZA, M. BEDNAREK, F. ROSSI (EDS), «Telecinematic discourse: Approaches to language of films and television series», Amsterdam, John Benjamins Publishing, pp. 185-204.
2012, «Construing 'nerdiness': characterisation in *The Big Bang Theory*», *Multilingua*, 31, pp. 199-229.

BIANCHI C.
2020, «Suspended Quotations: A Corpus Analysis of Functions», *Le Simplegadi*, Vol. XVIII (20), pp. 68-80.

BIBER D.
1988, *Variation across Speech and Writing*, Cambridge, CUP.
2012, «Register as a predictor of linguistic variation», *Corpus Linguistics and Linguistic Theory*, 8, n. 1, pp. 9-37.

BURKE M.
2010, «Why care about pedagogical stylistics? », *Language and Literature*, 19(1), pp. 7-11.

BURROWS J. F.
1987, *Computation into Criticism. A study of Jane Austen's Novels and an Experiment in Method*, Oxford, Clarendon.

CULPEPER J.
2009, «Keyness: Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet», *International Journal of Corpus Linguistics*, 14(1), pp. 29-59.

DOUTHWAITE J.
2000, *Towards a Linguistic Theory of Foregrounding*, Torino, Edizioni Dell'Orso.

FISCHER-STARCKE B.
2010, *Corpus Linguistics in Literary Analysis: Jane Austen and Her Contemporaries*, London, Bloomsbury Publishing.

FORCHINI P., POLI F., SERACINI F.
2021, *The American Movie Corpus: A Tool for the Development of Spoken Lexico-Grammatical Competence*, Milano, EDUCatt Università Cattolica.

GARCIA O., WEI L.
2014, *Translanguaging: Language, Bilingualism and Education*, New York, Palgrave MacMillan.

GLOVER M.
1980, «The Purchase of Commissions: A Reappraisal», *Journal of Society for Army Historical Research*, 58, p. 233.

HOFFMANN C., KIRNER-LUDWIG M.
2020, *Telecinematic Stylistics*, London, Bloomsbury Academic.

HUF L. D.
2017, «The Junior British Army Officer: Experience and Identity, 1793-1815», *School of Humanities – University of Tanzania*, pp. 45-56.

KOHN J.
1996, *What can (corpus) linguistics do for translation?,* in K. KLAUDY, J. LAMBERT, A. SOHÁR (eds), «Translation Studies in Hungary», Budapest, Scholastica, pp. 39-52.

KRANICH S.
2014, *Translations as a Locus of Language Contact*, in J. HOUSE (ed), «Translation: a multidisciplinary approach», Basingstoke, Palgrave MacMillan, pp. 96-115.

LEECH G., SHORT M.
1981, *Style in Fiction*, London, Longman.

LOUW B.
1997, *The role of corpora in critical literary appreciation*, in A. WICHMAN, S. FLIGELSTONE, T. MCENERY, G. KNOWLES (eds), «Teaching and Language Corpora», Harlow, Addison Wesley Longman, pp. 240-251.

MAHLBERG M.
2007, *Corpus stylistics: bridging the gap between linguistic and literary studies*, in M. HOEY, M. MAHLBERG, M. STUBBS, W. TEUBERT (eds), «Text, Discourse and Corpora: Theory and Analysis», London, Continuum.

MAHLBERG M., CONKLIN K., BISSON M. J.
2014, «Reading Dickens's characters: employing psycholinguistic methods to investigate the cognitive reality of patterns in texts», *Language and Literature*, 23, n. 4, pp. 369-388.

MAHLBERG M., STOCKWELL P., DE JOODE J., SMITH C., O'DONNELL M.
2016, «CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics», *Corpora*, 11, n. 3, pp. 433-463.

MCINTYRE D., WALKER B.
2019, *Corpus Stylistics: Theory and Practice*, Edinburgh, Edinburgh University Press.

PAVESI M.
2020, *"I shouldn't have let this happen": demonstratives in film dialogue and film representation*, in C. HOFFMANN, M. KIRNER-LUDWIG (eds), «Telecinematic Stylistics», London, Bloomsbury Academic.

PENNYCOCK A., OTSUJI E.
2015, *Metrolingualism: Language in the City*, London, Routledge.

QUAGLIO P.
2009, *Television Dialogue. The sitcom Friends vs. Natural Conversation*, Amsterdam, John Benjamins Publishing.

SEMINO E., SHORT M.
2004, *Corpus Stylistics: Speech, Writing and Thought. Presentation in a Corpus of English Writing,* London: Routledge.

STOCKWELL P., MAHLBERG M.
2015, «Mind-modelling with corpus stylistics in David Copperfield», *Language and Literature*, 24, pp. 129-147.

STUBBS M.
2005, «Conrad in the computer: examples of quantitative stylistics methods», *Language and Literature*, 14, n. 1, pp. 5-24.

WALES K.
2001, *A dictionary of Stylistics*, Harlow, Pearson Education.

## WEB SITES

BNC SAMPLER: XML EDITION (JULY 31, 2008)
<http://www.natcorp.ox.ac.uk/corpus/sampler/sampler.pdf>, sito consultato il 6.4.2022.

#LANCSBOX: LANCASTER UNIVERSITY CORPUS TOOLBOX
<http://corpora.lancs.ac.uk/lancsbox/>, sito consultato il 6.4.2022.

THE SYDNEY CORPUS OF TELEVISION DIALOGUE (SYDTV)
<https://www.syd-tv.com/>, sito consultato il 6.4.2022.

WMATRIX CORPUS ANALYSIS AND COMPARISON TOOL
<http://ucrel.lancs.ac.uk/wmatrix/>, sito consultato il 6.4.2022.