A Corpus-Stylistic Analysis of Mitchell's *Gone with the Wind* and Hemingway's *A Farewell to Arms*

Kristina Bujanova

A Thesis Presented to

the Department of Literature, Area Studies and European Languages

UNIVERSITY OF OSLO

May 2013

A Corpust-Stylistic Analysis of Mitchell's*Gone with the Wind* and Hemingway's *A Farewell to Arms*

Kristina Bujanova

Acknowledgements

I would like to express my gratitude to Professor Hilde Hasselgård for being a kind, patient and professional supervisor, and always finding time for me and my thesis.

Special thanks to my wonderful parents Irena and Victor for being there for me: caring, supporting, loving, encouraging and believing.

Table of Contents

Lists of Tables

1. Introduction

This master's thesis presents a corpus-stylistic analysis of Margaret Mitchell's novel *Gone with the Wind* and Ernest Hemingway's novel *A Farewell to Arms.* The novels will be analysed according to a corpus stylistic approach, outlined by Semino and Short (2004) and further developed by Michaela Mahlberg (2012) when analysing a corpus of Dickens' texts. A corpus, in general, is a collection of written and spoken texts. The current thesis focuses only on two textsin electronic form. The major difference between a corpus stylistic analysis and a stylistic or literary analysis is that the analysis is based on the recurrent word combinations found in the text by the corpus software, but not the manual analysis of a stylistician or literary scholar.Mahlberg sees corpus stylistics as "a way of bringing the study of language and literature closer together" (2007: 3). Corpus linguistic methods can contribute to the study of literature and bring to light individual qualities of the texts. Although the programme identifies the frequent combinations, the results should be interpreted and that is what constitutes the essence of this thesis.

Aims and scope

The overall hypothesis of the thesis is that a corpus stylistic approach of Michaela Mahlberg can help in interpreting the novels*Gone with the Wind* and*A Farewell to Arms*.

Research questions to be taken up in the study:

- Are the functional groups introduced by Mahlberg relevant to the discussion of the novel?
- Are longer N-grams associated with particular characters in *Gone with the Wind and A Farewell to Arms*?
- How do 8-, 7-, 6-, 5-, 4-, 3-grams reveal the characters of the novels *Gone with the Wind and A Farewell to Arms*?
- Are the N-grams particular to the novel's styles or they highlight the general tendencies of the fictional texts?
- Do the keywords found in the novels pinpoint *the fictional world* and *thematic signals?*

- How do patterns from the novels (particularly 5- and 4-grams) compare to those of the mixed-genre Brown corpus?

Material and Methodology

As mentioned above, the thesis is devoted to the analyses of Mitchell's novel *Gone with the Wind*and Hemingway's novel *A Farewell to Arms*. Both novels are written by American authors in approximately the same time period, besides, they are set during the war and reflect romance. In order to carry out the corpus-stylistic analyses, the texts of the novels in electronic, searchable form wereneeded. The files containing the novels*Gone with the Wind*and *A Farewell to Arms*were downloaded from the following websites:http://webreading.ru/prose_/prose_classic/margaret-mitchell-gone-with-the-wind.html;http://webreading.ru/prose_/prose_classic/ernest-hemingway-a-farewell-to-arms.html(last accessed 5. May 2013).The electronic version of the novel *Gone with the Wind* comprises 419, 649 words, while the text file of *A Farewell to Arms* consists of 88, 610 words.

8-, 7-, 6- grams with a minimum frequency of fourare first identified by the programme*AntConc[1]*. Then these extracted N-grams areoutlined in the descending frequency anddiscussed in reference to the character they can be ascribed to. Afterwards the 5-, 4- and 3-grams are found in the novels and grouped according to the functions they perform. The last step of the analysis is to search for the keywords with the help of software package *WordSmith[2]*. When the analysis of both novels is completed, the results got from*Gone with the Wind* will be compared to the results from *A Farewell to Arms.*

Finally, the findings from *Gone with the Wind* and *A Farewell to Arms* will be compared against 5- and 4-grams frequently encountered in the Brown Corpus. The Brown Corpus is chosen for the comparison to *Gone with the Wind* and *A Farewell to Arms* because it contains 500 samples of 2, 000 words each, covering American English as used in the year 1961 (McEnery&Hardie 2011:97). It was created by Henry Kucera and W. Nelson Francis at Brown University, Providence, Rhode Island. The Brown Corpus comprises 15 text categories: A. Press: reportage (44 texts), B. Press: editorial (27 texts), C. Press: reviews (17 texts), D. Religion (17 texts), E. Skill, trades and hobbies (36 texts), F. Popular lore (48 texts), G. Belles lettres, biography, essays (75 texts), H. Miscellaneous (government and other official documents) (30 texts), J. Learned and scientific writings (80 texts), K. General fiction (29 texts), L. Mystery and

---

[1]Developed by Laurence Anthony, Associate Professor at Waseda University in Japan.Available at: http://www.antlab.sci.waseda.ac.jp/antconc_index.html
[2]Designed byDrMike Scott.Version 5.0 © 2008 Available at the University of Oslo.

detective fiction (24 texts), M. Science fiction (6 texts), N. Adventure and western fiction (29 texts), P. Romance and love story (29 texts), R. Humor (9 texts) (Lindebjerg1997).

Margaret Mitchell's novel was first published in 1936 and Ernest Hemingway's novel – in 1929, but there is no publically available reference corpus from that time period; therefore, the Brown Corpus was considered a viable option as a basis for comparison in the discussion of the novels.

The organization of the thesis

The thesis comprises four chapters. The first chapter presents a theoretical framework for the corpus stylistic analysis: the research on clusters as pointers to local textual functions conducted by Mahlberg, findings of Michael Stubbs and Isabella Barth (2003) on recurrent phrases as text-type discriminators and the role of keywords discussed by Stubbs. The second chapter deals with the corpus stylistic analysis of the novel *Gone with the Wind* in regard to the methodology described above. The third chapter, in turn, focuses on the corpus stylistic analysis of the novel *A Farewell to Arms.* The fourth chapter presents the comparison of the N-grams from the novels and the Brown Corpus, followed by the contrastive analysis of recurrent N-grams from two novels.

2. Corpus Stylistic Analysis: Theoretical Background

The thesis presents the analyses of two novels – Margaret Mitchell's *Gone with the Wind* and Ernest Hemingway's *A Farewell to Arms.* The conducted research is based on Michaela Mahlberg's corpus stylistic analysis of Dickens' novels and Stubbs and Barth's analysis of recurrent phrases functioning as text-typediscriminators. These studies are intentionally selected to explore how corpus tools can contribute to the stylistic analysis of the novels: whether they help to interpret the novels and show peculiarities of the authors' styles or highlight general tendencies among two novels. Firstly, Mahlberg's studies will be touched upon, then Stubbs and Barth's research will be discussed, followed by a description of the keyword analysis.

2.1 Leech and Short's Style in Fiction

Before Mahlberg's research findings will be presented, we will take a look at an important publication for literature in linguistics that Mahlberg uses in her study – *Style in Fiction* (1981). Leech and Short claim that

> One major concern of stylistics is to check or validate the intuitions by detailed analysis, but stylistics is also a dialogue between literary reader and linguistic observer, in which insight, not mere objectivity, is the goal. Linguistic analysis does not replace the reader's intuition […] (1981: 5).

The linguists explain that "style" expresses how language operates within specific context, by a certain person, for a specific aim (ibid: 10). They use the terms proposed by the Swiss linguist Saussure *langue* and *parole. Langue* denotes the system employed by speakers of a particular language, whereas *parole* signifies the applications of this system in specific occasions. "Style, then, pertains to *parole*: it is selection from a total linguistic repertoire that constitutes a style." (ibid:11).

Leech and Short assert that authorial style is not necessarily easy to generalize, but to find the distinctive features of bigger domains like a genre or an epoch is an even more difficult task (ibid: 12). They, in turn, concentrate on the style of texts. As they point out,

We normally study style because we want to explain something, and in general, literary stylistics has, implicitly or explicitly, the goal of explaining the relation between language and artistic function (ibid: 13).

Leech and Short analyse the styles of prose texts, applying "linguistic and stylistic categories: lexical categories, grammatical categories, figures of speech, and cohesion and context" (ibid: 75). The corpus-stylistic analysis implemented in this thesis focuses on the recurrent N-grams found in Mitchell's and Hemingway's novels. For that reason, Mahlberg's corpus stylistic analysis of Dickens' fiction on the basis of the frequent functional groups will be described next.

2.2 Mahlberg's Corpus-Stylistic Analysis of Dickens' Fiction

In her article *Corpus stylistics: bridging the gap between linguistic and literary studies* (2007) Michaela Mahlberg clearly describes the interface between corpus linguistics and literary stylistics, thus, clarifying 'corpus stylistics'.

In order to understand what corpus linguistics and stylistics have in common, Mahlberg gives an account of both disciplines. She quotes Wales (2001: 371)[3] who states that style is always "distinctive" as it reflects linguistic peculiarities which in turn are displayed through register, gender or period. For that reason, stylistics employs linguistic categories. As Short (1996: 1) points out, stylistics is at the interface of two disciplines – linguistics and literary criticism. Besides, some linguists argue that stylistics involves too much interpretation, whereas literary critics, on the contrary, believe that stylistics lack interpretation due to the linguistic framework it draws on.

Corpus linguistics explores quantitative data, words in their contexts to identify patterns, and annotation. Thus, the corpus linguistic methods can be used in different fields for various purposes. Mahlberg turns to the book of Semino and Short (2004:8) who propose that corpus approaches can be applied in the study of style. Further Mahlberg describes common ground between corpus linguistics and literary stylistics. She highlights that both disciplines focus on the correlation between meaning and form. As she explains, stylistics is preoccupied with how people say things, moreover, corpus linguistics also asserts that what is said depends on form. Nevertheless, the two disciplines tend to concentrate on different areas. Stylistics seeks

---

[3] References to previous research quoted by Mahlberg have been given as they appear in her articles; the originals have not been checked

distinctive features in a text or a group of texts and it also examines linguistic variations that express creative ways of language use. Corpus linguistics, in turn, searches for repeated uses of language across the texts. Mahlberg points out that corpus linguistics and literary stylistics are related as "creative" language is always notable when it deviates from linguistic norms (ibid:221).

Turning to the corpus stylistic approach Mahlberg employs in her research, she follows the description of style provided by Leech and Short (1981: 74) who claim that all texts are marked by individual properties and these properties can vary from text to text even though the author is the same. She further expands on their claim, saying that corpus stylistics can be applied to distinguish "tendencies, intertextual relationships, or reflections of social and cultural contexts (ibid.). What is more, corpus stylistics can contribute with descriptive tools that pinpoint characteristic features of a text. Leech and Short (1981) introduce categories that can be useful in a stylistic analysis like "the high number of adjectives in a text, or the predominance of simple sentences, the fact that nouns tend to be postmodified by prepositional phrases, and so on" (Mahlberg 2007: 222).

Mahlberg emphasizes two important factors about frequency and function words. Firstly, she explains that frequency helps to draw comparison of words in one text. Mike Scott's software *WordSmith* (Scott 2004) has a function called KeyWords that makes a list of words "which are unusually frequent (or infrequent) in a text as compared to the reference corpus" (ibid: 223). These key words are basically proper nouns, content words signifying what the text is about and function words. "Scott (2004-2006: 116) points out that the function words thrown up by KeyWords may indicate stylistic features" (ibid.). Secondly, Mahlberg underlines that despite being frequent, function words tend to occur in collocations; therefore, they affect the collocational patterns of content words. She adds also that it is necessary to look at the co-occurrence of words in a sequence. In her corpus project Mahlberg calls these sequences "clusters", which appear to be the most frequent three, four and five-word clusters found in the Dickens' corpus. As Mahlberg highlights, all of them contain function words and have a stylistic value.

Furthermore, Mahlberg explores local textual functions of clusters in the Dickens corpus. As Mahlberg clarifies, local textual functions describe lexical items according to the functions they perform in texts (ibid: 224). Mahlberg chose 23 texts consisting of 4.5 million words that she obtained from Project Gutenberg. In addition, she uses a corpus of 29 novels by 18 authors from the nineteenth century (19C) also from Project Gutenberg which comprises 4.5 million

words. Mahlberg notes that most of the texts are novels, but not all of them belong to this genre. "From a corpus stylistic perspective, clusters can be taken as pointers to meanings and textual functions" (ibid: 225). Other terms for "clusters" are "n-grams" or "lexical bundles" (e.g. Biber et al. 1999) (ibid.).

Mahlberg makes an important observation: the longer the clusters are, the lesser is their frequency. For example, two most frequent three-word clusters in the Dickens corpus are *out of the* (1,210 occurrences) and *as if he* (1,158 occurrences), whereas two most frequent five-word clusters are *as if he had been* (90) and *his hands in his pockets* (90). In addition, Mahlberg concludes that longer clusters can be linked to specific texts. She mentions the example of eight-word cluster *not to put too fine a point upon* encountered 14 times and which refers to the character from *Bleak House* – MrSnagby. Mahlberg cites Stubbs (2005: 6) who explains that "quantification" only proves the fact that was already discussed by stylisticians, still it also proves that the method employed by corpus stylistics works (ibid: 228).

When analysing*Bleak House*, Mahlberg selects five functional groups that characterise Dickens' style, namely:

1. Labels

2. Speech clusters

3. Body part clusters

4. *As if* clusters

5. Time and place clusters

In order to find the functional groups, Mahlberg identifies 66 positive key clusters with the help of *WordSmith* at first. Mahlberg compares these functional groups to 19C. The first four functional groups represent most of key clusters, still time and place clusters are found both in *Bleak House* and 19C.

Above all, Mahlberg presents distribution of cluster types in *Bleak House*. The novel consists of 350, 000 words and *WordSmith* identifies 97 cluster types with a minimum frequency of five. Labels dominate in the novel due to the fact that these are not just names, but also speech and body part clusters that indicate a particular character. Mahlberg states the following: "[…] it is important to note that the clusters are only pointers to more detailed questions of analysis;

clusters alone provide an incomplete picture. The patterns of which they are part can be very flexible and blend in with their contexts in various ways" (ibid: 236).

To sum up, Mahlberg stresses again the importance of meaning and form for both corpus linguistics and literary stylistics. She believes that corpus stylistics can succeed in characterizing texts as local textual functions distinguish textual features from different points of view.

In her recent book Mahlberg (2012) argues that "clusters can be interpreted as textual building blocks for fictional worlds. These building blocks have particular functions in the creation of characters" (ibid: 26). "They contribute building blocks of fictional words that characterise wider areas of meanings [...] (ibid: 67). Mahlberg provides a detailed account of cluster types.

2.2.1 Speech Clusters

Speech clusters are defined on the basis of surface features: they contain a first- or second-person pronoun, i.e. the clusters are put into one group because they are moresimilar to one another than the other clusters (2012: 75).

Speech clusters that fall under definition carry interpersonal functions and represent mostly interaction between characters. Carter and McCarthy (2006:835) also stress that clusters reveal how interpersonal meanings are implemented by speakers and listeners (ibid.). Mahlberg emphasises that speech clusters found in fiction distinguish from real-life speech as they are more "polished" despite the fact that some of speech clusters include pauses, repetitions, interruptions, etc. (ibid.). Nevertheless, speech described in a novel tends to be like real speech because it contributes to the perception of the characters that readers anticipate to sound real. Mahlberg refers to Oostdijk (1990) who undertakes a corpus study and finds out that "a number of features in fictional dialogue [...] resemble features of spoken language, e.g. frequent use of imperatives, interrogatives, and exclamatory phrases' (ibid.).

Further Mahlberg discusses four main functions of speech clusters: negotiating information, turn-taking, politeness formulae, and first-person narration (2012: 76). She finds 108 speech clusters which are used 5 or more times in at least one text in the Dickens' Corpus (DCorp). All in all, they are present in more than one text, otherwise they would be put into the category Labels.

The examples of speech clusters which imply negotiating information are as follows: *you don 't mean to say* and *I don 't know what*. As Mahlberg explains, the former is the speaker's surprise and disbelief, while the latter signifies that Kate struggles to express her thoughts. Carter and McCarthy (2006: 835) underline that clusters containing the verbs *know*, *mean* and *think* are highly frequent: *you know, I don't know what, I mean, do you think*. Thus, the interpersonal meanings they carry point to shared or new knowledge and degrees of certainty about the world (Carter & McCarthy 2006: 835) (ibid: 77).

As for turn-taking, "in spoken language we find clusters that function to elicit responses including forms of backchanneling from the listener or to hand the turn over from one speaker to another" (cf. Carthy& McCarthy 2006: 836).Mahlbergmentions the example *you don't mean to say*, pointing out that it serves both as negotiating information and also prompting a response from another character. She also highlights that such clusters as *I was going to say* and *I should like to ask* function as "prefaces" to turns. Carter &McCarthy(2006: 837) define them as "drawing attention to what the speaker is about to say" (ibid: 79). Mahlberg clarifies that form and function do not necessarily correspond; therefore, each cluster should be seen in context. For instance, *Iwas going to say* marks a preface, but it can bring out other implications.

Besides, politeness formulae occur among speech clusters in greetings – *am delighted to see you*, *how do you do*; expression of gratitude – *am much obliged to you*, requests – *will you have the goodness*; apologies – *I beg your pardon*.

First-person narration clusters reflect the definition of speech clusters mentioned above, i. e. they comprise a first-person pronoun: *I could not help observing*, *Ifelt that I was* (ibid: 82). Mahlberg elucidates that first-person clusters cannot be considered just as direct speech as they may stand for indirect thought representation or bring to light the point of view expressed by a first-person narrator (ibid.). However, speech clusters mainly externalise fictional characters.

2.2.2 Body Part Clusters

Body part clusters evidently include body part nouns, performing different functions. Mahlberg's investigation builds on Korte's (1997) framework on body language in literature (2012: 100). "To reflect the reader's general knowledge of body language, Korte suggests a system based on 'modal' and 'functional' classes. By modal classes Kortemeans "kinesics", actual

movements of face, eyes and body and automatic reactions, "haptics", a tactile behaviour, and "proxemics", spatial distances between people (cf. Korte 1997: 38) (ibid).

Functional classes, in turn, are represented by "externalisers, […] forms of body language that provide information on characters relating to "relatively stable mental conditions (such as psychopathological states, attitudes, opinions, values, personality traits)" (ibid.). Besides, there are also "emotional displays" which express temporary emotions and psychological conditions. Apart from externalisers and emotional displays, there are also "illustrators, regulators, and emblems". "Illustrators" are found in speech to "emphasize and structure, complete, and support a verbal utterance" (Korte 1997: 45) (ibid: 102). "Regulators" mark turn-taking, but they do not have connotation in contrast to illustrators. "Emblems" denote certain cultural phenomena, for example, nodding signifies a positive answer. According to Korte, body language used in literature can be analysis with the help of grids. These grids show "a cross-classification of modal and functional categories" (ibid.). Korte looks at 80 British novels from the 16 century to the present time and finds out that modal-functional classes are encountered evenly in relation to the novels from the 19th century, whereas in the 20th century novels there are less gestural externalisers and automatic reactions revealing emotions.

Korte (1997: 185) relates these tendencies to the use of conspicuous gestures for characterisation and character identification and conventional repertoires for emotional displays that were тоге strongly observable in the 19th century, a point that is also reflected in John's (2001) account of popular melodrama (2012: 103).

Additionally, Mahlberg explains that body language in literature is closely connected to cultural and historical context. Vretto (1999/2000: 413) analyses psychology and habit in the 19[th]century and the cluster *putting one's hands into one's pockets*is common there, which is also the case in Dickens' novels (ibid: 109). 19th century drama had also impact on the novels; therefore, certain stereotypes were employed to depict melodramatic moments (ibid.).

To analyse body language clusters, Mahlberg speaks of the body language cline from contextualising to highlighting functions. "Contextualising functions show how character information is less strikingly presented as part of a larger textual picture, while highlighting functions illustrate how prominence is given to character information" (ibid: 110). Mahlberg underlines that highlighting functions correspond to labels as they emphasize a character's feature which is repeated throughout the texts, thus, making himor her conspicuous and important due to interpretation and the narrator's comments. Contextualising functions, in turn, add information about a character, for example, used in the form of prepositional phrases or -

ingclauses to express the body language (ibid.). A cluster performing the contextualising function is *with his back to*. Mahlberg clarifies that it is the only cluster that appears in both DCorp and 19C, furthermore, it describes the character's position, but it does not relate to a particular character, Mostly the cluster is immediately followed the noun *fire*, but there are also found variations like *the bedroom fire* or *the smoked chimney piece* (ibid: 111).

Regarding body language clusters functioning as labels and thematic clues, Mahlberg cites examples of the following clusters – *her apron over her head*, *(with) his hand to his(forehead)*, and *(with) his head against the (wall)* (ibid: 119). AfferyFlintwinch, Mrs Clennam's maid in *Little Dorrit*. She becomes aware of mysterious things going on in the house, but her husband Jeremiah persuades her that she is only dreaming (ibid: 122).

2.2.3 *As if* Clusters

Apart from speech clusters and body language clusters, there are also found clusters containing collocations of *as if*.Mahlberg asserts that

> The analysis of collocations approaches *as if* from a lexico-grammatical angle and contributes further detail to the areas of meanings of the text-driven fictional world. In this fictional world, patterns *as if* are associated with the voice of the narrator who draws the reader's attention to features of the fictional characters (2012: 128).

She finds out that most of the clusters go together with function words, for instance, *as if he were going*. Lexical verbs appearing with *as if* are*see*and *feel – as if I had seen, felt as if I had, I felt as if I had.* Besides, Mahlberg remarks on the frequent use of *as if* clusters in fiction in general.

In order to obtain information about the characters and understand the narrator's comments, the left-collocates should be looked at. Mahlberg studies the left-collocates with at least 20 occurrences and sorts the lexical verbs in them according to the meanings they carry (ibid: 132):

1) Action verbs *made, turned, went, stood, shook, stopped, came*

2) Body part nouns *head, hand, hands, eyes, face, back, mouth, arms, arm, lips*

3) Settings *door, fire, room, side, chair, wall*

4) Manner *manner, way, air, seemed*

5) LOOK *looked, looking, look, looks*

6) SPEAK *said, speaking, spoke, voice*

7) FEEL *felt, feel*

8) Other *time, moment, man, round, down, again, little, almost, very, great, now, well, quite, still*

Mahlberg's classification shows that collocates are linked to the characters' body language. They add lexically driven categories to Korte's (1997) framework which is described in Section 2.2.2. Mahlberg explains that the cline she introduces when dealing with body part clusters can be appropriate to the analysis of *as if* clusters. The cline demonstrates how clusters contribute to characterisation. "At one end, there are examples that state what is compared – the text reports an action before it is commented on. At the other end of the cline, the narrator only implicitly presents the features or actions that are compared by means of an *as if* comparison (ibid: 133).

2.2.4 Labels

Labels are the clusters which contain characters' names or pin down attributes of particular characters (Mahlberg 2012: 152). For that reason, it is expected that labels would comprise characters' names. Nevertheless, Mahlberg finds out that only 157 out of 406 include proper names, while the majority does not, still they fulfill specific functions that fall into six groups: Reporting Speech Labels, Speech Labels, Body Language Labels, Attributes, Long Clusters, and Name of People and Places (ibid: 153).

Reporting Speech Labels and Speech Labels reveal the significance of character speech. The former comprises a reporting verb like **returned**mrssparsitwith a or mrsdombey**said**mrdombey, *kate my dear* **said**mrs.Mahlberg explains that it is difficult for Mr Dombey to express his feelings even to the member of his family, therefore, he formally addresses his wife. In the case *kate my dear said mrs.* the further context shows that it is used by Mrs Nickleby (ibid: 154).Mahlberg exemplifies a Speech Label:*I expect the judgement shortly* is encountered five times in *Bleak House* and it is uttered by Miss Flite (ibid: 155).

As concerns body language labels, there are two types of them: the first one including a character's name (*mrmeagles shaking his head*) and the second one reflecting the body language illustrating a particular character (*and his nose came down* – Rigaud in *Little Dorrit*) (ibid.).

Attributes, in turn, demonstrate peculiarities of fictional characters – *young lady with the black the very queer small boy, the man with the camp-stool* (ibid.). Long clusters are the ones which prove to be variations of the longer cluster. "For instance, the waiter who ought to is related to three other five-word clusters, as all are part of the eight-word cluster *the waiter who ought to wait upon us*" (ibid: 156). The sixth group named Names of people and place stands for clusters that do belong to the previous five groups, these are – *mrjarndyce of bleak house, mrpickwick and his friends, mr chick and miss tox.*

Here again Mahlberg speaks of contextualising and highlighting functions labels perform. She concludes that labels that include character attributes serve highlighting functions, whereas labels contributing to the situations' descriptions carry contextualising functions. Moreover, the cline of functions labels have signifies how labels and other types of clusters are interconnected. "Speech Labels are related to Speech clusters. Similarly, Body Language Labels are related to body language clusters" (ibid: 158).

To sum up, Mahlberg observes that among 23 texts in the Dickens' Corpus, there is only one text that lacks clusters, 11 of them include all five functional groups, but five shorter texts, consisting of fewer than 34, 000 words, have labels as their only cluster type. Labels also turn out to be the most frequent cluster category in Dickens' texts. Mahlberg explains that apparently longer texts contain more clusters because shorter ones with a high number of clusters may "appear too obtrusive" (ibid: 70).

2.3 Mahlberg's Keyword Analysis

Turning to one of Mahlberg's earlier publications, the article *Corpus Linguistics and the Study of Nineteenth-Century Fiction*(2010) emphasises that the analysis of a novel starts from compiling keywords. Mahlberg cites an example of *Pride and Prejudice,* when she and Catherine Smith examine keywords in this novel with the help of *WordSmith Tools*and compare them against other keywords found in a corpus of novels from eighteen novels from the nineteenth century. Thus, they identify the words which are particular of *Pride and Prejudice*and then they can analyse how these keywords are used in the novel. Besides, they search for the same keywords in the reference corpus in order to find out their meaning in other novels. Mahlberg and Smith also suggest looking at 'suspensions' that reveal body language. For that reason, they use the annotated electronic version of *Pride and Prejudice*that recognizes quotation marks and other punctuation indicators. Mahlberg emphasises again the significance of clusters

as they reveal a certain pattern and carry functional relevance. Above all, clusters bring to light the externalization of character that was described by Juliet John (2001). "She discusses the relevance of the ostension of the private and the depiction of transparent character. John argues that Dickens depicts emotions in exaggerated ways and in his narrative prose gestures and actions add to externalize character, whereas the mind only takes a marginal place" (Mahlberg 2010).

In another article *A Case for corpus stylistics. Ian Fleming's Casino Royale* (2011) written together with Dan McIntyre, Mahlberg stresses the benefit of combining qualitative and quantitative approaches in the analysis of novels. Mahlberg and McIntyre introduce the findings of Bill Louw (1993) who claims that computer corpora could be used to identify semantic prosodies which in turn reveal irony. Mahlberg and McIntyre carry out a corpus-stylistic analysis of Ian Fleming's novel *Casino Royale*. They assert that corpus linguistic techniques contribute a lot in the analysis of prose texts as they tend to be long; however, the manual analysis cannot be substituted by the corpus-based analysis. In addition, a corpus-stylistic analysis may reflect the peculiarities of the author's style as whole texts are taken into account.

Mahlberg and McIntyre employ WMatrix software tools which generate the automatic semantic analysis of texts. Those keywords that are not found in the reference corpus are of special interest for the analysis. Firstly, Mahlberg and McIntyre search for keywords in*Casino Royale*and compare them against keywords from the fictional prose found in the BNC. They "suggest that keywords might be viewed as signals for the building of fictional worlds as well as triggers for thematic concerns of the novel" (2011: 207). On the basis of keywords they grouped semantic domains. All in all, they got three broad categories like the names of characters, the words that signify "aboutness" of the text and the words that reflect the style of the novel. Further Mahlberg and McIntyre explore "thematic signal" keywords and "fictional world" keywords. They describe the difference between the words "table" and "gambler". The former has a concrete meaning therefore it belongs to "fictional world", while the latter conveys additional information as "gambler" also means a person who likes taking risks, thus, it reveals one of the novel's themes – taking risks. "Fictional world" keywords tend to be text-centred, while "thematic signal" keywords are reader-centred as they require interpretation. Moreover, Mahlberg and McIntyre suggest subgroups such as *characters* and *settings and props.* The linguists point out that keywords may be ascribed to different categories due to their polysemy.

Speaking of key semantic domains, WMatrix provides the tags, clarifying the semantic domains the words belong to. Mahlberg and McIntyre got 20 semantic domains that they

compare to the semantic domains encountered in the BNC. Computer-based key semantic domains have both similarities and differences with the groups proposed by Mahlberg and McIntyre. The domain "Anatomy and physiology" covers the keyword group "body part nouns". In addition, the domain "Geographical names" overlaps with their group "Places", still it also contains words meaning location like *room.* What is more, some domains have different classification than the keyword groups established by Mahlberg and McIntyre. For example, "Paper documents and writing" contained words *cards* and *notes,* though they clearly show the casino context. Besides, some domains include keywords that seem to belong to text-centred categories, but they turn out to be reader-centred.Mahlberg and McIntyre conclude that

> The analysis of keywords is based on a more limited set of words thanthose covered by the semantic domain analysis, but the analysis is more detailed,taking into account more information concerning themes and relationships in the text in question. Key semantic domains have the advantage of grouping together a greater number of words that may not even show in a keyword analysis (even if we had worked with the full keyword list) (ibid: 223).

Functional groups of clusters, Labels, Speech clusters, Body part clusters, *As if* clusters, Time and place clusters, and *Other*clusters, proposed by Michaela Mahlberg will be used in the analysis of novels *Gone with the Wind* and *A Farewell to Arms.* They will be applied to the discussion of 5-grams, 4-grams and 3-grams. As noted above, Mahlberg uses the term *cluster*, while in the present study the term *N-gram* is used instead. In addition, the keyword analysis will be implemented and if relevant the keyword groups established by Mahlberg and McIntyre (2011) will be applied.

2.4 Stubbs and Barth's Study on Recurrent Phrases as Text-Type Discriminators

Another study that is pertinent to the present thesis is carried out by Michael Stubbs and Isabel Barth (2003). They analyse recurrent phrases in different types of texts namely, *Fiction, Belles and Learned* using data from the Brown, LOB, Frown, and FLOB corpora. That is, they look at one-million-word corpora of American and British English from 1961 and 1991. Stubbs and Barth's goal is to identify word chains that function as text-type discriminators. They regardreduced forms *n't ,'m*, etc. as a part of the word they are written together, e.g. *didn't,*

I'm.Besides, they underline that fiction has a number of informal reduced forms (ibid: 65).The personal pronoun *I* was found in their top 50 words in fiction and the chain *I don't* also frequently occurs in the other works of fiction and it is included in the top 50 chains. The verbs that carry their lexical meanings also turn out to be frequently used in the novel: *said, know, want*. The word *know* also occurs in the top 50 word-forms when Stubbs and Barth use a stop list of 100 word-forms (ibid: 64).According to Stubbs and Barth's results, the three word chain *you want to*occurs 61 times in fiction works. Moreover, Stubbs and Barth introduced a pattern for *of*-phrase – *DET N of* (ibid: 72). One of the top 30 examples, that are frequent in fiction looked by the linguists, was the phrase *a lot of* (ibid: 70). In addition, Stubbs and Barth take up an important issue: they conclude that in such phrases like *a number of* "the delexicalization" takes place as the word loses its full lexical meaning (ibid: 71). Stubbs and Barth find out that positional phrases with the patterns *the N of and PREP the N of* like in the phrases *the edge of; in the middle of; in front of the*, etc. are found in the top 10 (ibid: 72).They claim that "FICTION is characterized by a verbal style, by past tense verb forms, and by frequent vocabulary from the lexical fields of saying, looking, thinking and wanting" (ibid: 79). What is more, when describing three-word chains, the linguists saw a tendency that time and place expressions made over a quarter of all three-word chains.

Stubbs and Barth operate with the term "chains" which basically means the same as "N-grams" and "clusters." They consider a reduced form to be a part of the word it appears together, while this form is counted as a single word in the current thesis. Additionally, Stubbs and Barth investigate chains of different lengths separately and this method will be employed, while the author of this thesis explores how N-grams can discern characteristic features of particular novels.

2.5 Keywords: Their Role in Corpus Linguistics

As Marina Bondi points out "Keywords are not necessarily a key to culture, however: they may facilitate understanding of the main point of a text, constituting chains of repetition in text" (2010: 3). Keywords carry significant implications of a text or culture, thus, in quantative analysis they appear to be more frequent or infrequent in a particular text or corpus than in a reference corpus (Scott 1997; Baker 2004; Scott&Tribble 2006 cited by Bondi 2010:3).

According to Michael Stubbs (2010: 21), keywords have three senses: cultural, statistical and lexico-grammatical. The second sense is of greater importance for this master's thesis as it is

displayed in corpus linguistics. Stubbs explains that "keyness is a textual matter" because certain words are linked to specific texts or text-types and "intellectual areas (such as medicine and natural science) (ibid: 25). A function of Mike Scott's software package *WordSmith Tools* (1998) is to generate word-lists or lists of n-grams from a text, and further compare them to other text samples. "Content words directly indicate the propositional content of texts" (ibid.). Stubbs also notices that these n-grams are a part of "global textual cohesion, but not textual structure" (ibid: 26). To demonstrate this, Stubbs presents two small case studies. In the first case study he analyses the transcripts of the Hutton Inquiry and compares it to the BNC as a reference corpus. The keywords he gets are mostly proper names and words that characterize the main themes of the transcripts. Nevertheless, Stubbs underlines that the search of longer recurrent combinations provides a better understanding of the content and the formal nature of the discourse. The latter is clearly revealed through phraseology. Stubbs cites an example of the most frequent 5-gram can I take you to that "is part of the polite, formal, cautious, public usage of Lord Hutton himself" (ibid: 27).

In the second case study Stubbs looks at a 30-million-word corpus of newspaper articles from 1996 and 2006. These articles were chosen by Friedrich (2007) as they included the following words-forms – islam, muslim/s, middle east. The BNC corpus was taken as a reference corpus. Through this study, Stubbs makes important observations: firstly, the same topic is described in different ways – the business periodical concentrates on the financial side, whereas the tabloid presents the topic from the personal perspective of the soldiers who are involved. Secondly, "textual collocates" (Mason&Platt 2006) become visible (ibid.). Stubbs clarifies that the software manages to identify words that fit into same semantic fields, that in turn, signifies "how homogeneous the vocabulary is across a text" (ibid.). Moreover, the meanings of a word is obtained from its interrelation to other words, but not from the denotation in the world.

To sum up, Stubbs agrees that Scott's method based on frequency and distribution of keywords is easily applicable as semantically related keywords reveal propositional content, but it lacks information about text structure, therefore, leaving room for other techniques.

As stated above, the keyword analysis will be carried out in the thesis. The novels *Gone with the Wind* and *A Farewell to Arms* will be compared to the Brown Corpus and the obtained keywords will be examined.

3. Corpus Stylistic Analysis of Margaret Mitchell's Novel *Gone with the Wind*

The thesis presents a corpus stylistic analysis of Margaret Mitchell's novel *Gone with the Wind*, which was first published in 1936. The opening of the novel is set in 1861, a short time before American Civil War begins. The main characters in the story are Scarlett O'Hara, Rhett Butler, Ashley and Melanie Wilkes. In the beginning of the novel Scarlett O'Hara is a young lady who lives with her parents and sisters. She loves Ashley Wilkes, but she gets to know that Ashley will be engaged to his cousin Melanie. Scarlett tells him about her love to make him change his mind, but Ashley explains that he is supposed to marry Melanie and he cannot make Scarlett happy. Accidentally, Rhett Butler overhears their conversation. To spite Ashley, Scarlett marries Charles Hamilton, Melanie's brother. Soon after he dies at war. Then Scarlett gives birth to Wade and after being depressed staying at Tara, her mother Ellen sends her to Atlanta to stay with Melanie and Aunt Pitty. When Atlanta falls, Scarlett together with Melanie, her son Wade, her servant Prissy and Melanie's son Beau drives home to Tara. Rhett helps them to escape from town, but then leaves them in the forest and joins the army. During the war they stay at Tara, working hard and hardly surviving. When the war ends, Scarlett needs to pay high taxes for the plantation, but she does not have money; therefore, she goes to Atlanta to ask Rhett to give her a credit or become his mistress. As Rhett is in jail, he cannot help her, but on her way to Aunt Pitty's house she meets Frank Kennedy, her sister's fiancé. By trick, she manages to marry Frank and clear debts. Once when driving from her mill, Scarlett gets attacked. Ku Klux Klan decides to revenge her. Frank and Ashley are members of Ku Klux Klan and when fighting, Frank gets

killed. Rhett proposes to her and she accepts. They get married and after a while she gives birth to their first daughter – Bonnie. At the age of four, she dies when trying to jump over a cross bar on her pony. When staying in Marietta, Scarlett gets a message from Rhett about Melanie being ill. She returns to Atlanta and finds Melanie dying. When talking to Ashley it dawns upon her that she never loved Ashley and Ashley always loved Melanie. Having understood this, Scarlett becomes aware that she actually loves Rhett Butler and she runs home to tell him about it, but Rhett confesses that he does not love her anymore, he loved her from their first meeting at Twelve Oaks, but her blind devotion to Ashley made his love pass. In the end Rhett leaves Scarlett, but she still hopes that she can return to him and decides to go to Tara.

3.1 8-Grams Present in the Novel *Gone with the Wind*

The corpus stylistic analysis of Mitchell's novel *Gone with the Wind*will start with a discussion of 8-grams. The minimum frequency is set to three because 8-grams occurring five times have not been found. Michaela Mahlberg explored a corpus of 4.5 million words; therefore, she got the 8-grams repeated five times. The programme*AntConc* returns only two 8-grams which are outlined in Table 3.1.

Table 3.1 8-grams in *Gone with the Wind*

Rank	Frequency	8-gram
1	7	For the first time in her life she
2	4	I won't think of it now she

The most frequent 8-gram *For the first time in her life she*is an adverbial of time with the third-person pronoun *she* which is associated with Scarlett. The 8-gram implies what dawns upon Scarlett:

1. And now, *for the first time in her life*, she was facing a complex nature.

2. *For the first time in her life*, she hated Tara, hated the long red road that led down the hill to the river, hated the red fields with springing green cotton.

3. She looked into Gerald's putty-colored face and, *for the first time in her life*, she saw him unshaven, his once florid face covered with silvery bristles.

4. She fell to trembling and, *for the first time in her life*, she saw people and events as something apart from herself, saw clearly that Scarlett O'Hara, frightened and helpless, was not all that mattered.

5. *For the first time in her life she* was regretting things she had done, regretting them with a sweeping superstitious fear that made her cast sidelong glances at the bed upon which she had lain with Frank.

6. *For the first time in her life she* had met someone, something stronger than she, someone she could neither bully nor break, someone who was bullying and breaking her.

7. *For the first time in her life she* had felt alive, felt passion as sweeping and primitive as the fear she had known the night she fled Atlanta, as dizzy sweet as the cold hate when she had shot the Yankee.

Examples 1-2 show Scarlett's reaction when she gets to know that Ashley is going to marry Melanie. Scarlett was convinced that Ashley would propose to her and when she first learns the news about his proposal, she feels despair and even hates Tara. The 7-gram is encountered further when Scarlett notices how her father Gerald has changed after her mother's death (Example 3). Example 4 signifies that Scarlett can think of others, not just herself. She is afraid of war, but so are other women and men. As seen from Examples 2, 5, 6, 7, the adverbial opens the sentences, thus, focus is on the adverbial. In the fifth example Scarlett regrets that she cheated Frank in order to get married to him and save Tara. In the six and seventh examples Scarlett realizes that Rhett is stronger than her and she is willing to surrender for the passion she feels for him.

The presence of the third-person pronoun *she* signifies that it is Scarlett O'Hara who utters a colloquial phrase *I won't think of it now*. The context of the 8-gram reveals the narrator's comment how Scarlett reacts when she is determined to forget about the situation:

8. *"I won't think of it now,"* she told herself and hurried her steps.

9. *"I won't think of it now,"* she said desperately, burying her face in the pillow.

10. *"I won't think of it now,"* she thought grimly, summoning up her old charm.

11. *"I won't think of it now," she* said again, aloud, trying to push her misery to the back of her mind, trying to find some bulwark against the rising tide of pain.

It is striking that the utterance is used in key moments of the novel. Scarlett first says the phrase when she realizes that she can be regarded like Belle, the prostitute, as she intends to ask money from Rhett in exchange for being with him (Example 8). The next situation when Scarlett resigns to think happens after India and Mrs. Elsing see Ashley and Scarlett hugging. She is afraid that India tells Melanie about this incident (Example 9). The last two utterances describe how Scarlett feels when Rhett leaves her alone. She wants to stop him at first, but then she realizes that it is better to wait till tomorrow when she gets to her childhood home Tara.

The 8-grams are clearly associated with Scarlett, thus, Mahlberg's finding that longer clusters are text-specific and reveal the features of certain characters holds true in *Gone with the Wind*. The next step in the analysis is to examine what information the 7-grams carry.

3.2 7-Grams Present in the Novel *Gone with the Wind*

7-grams are going to be looked at in this section. There are seven 7-grams with a minimum frequency of four that are present in *Gone with the Wind*, which are presented in Table 3.2.

Table 3.2 7-Grams in*Gone with the Wind*

Rank	Frequency	N-gram
1	10	for the first time in her life
2	7	I won't think of it now
3	5	she was going to have a baby
4	5	Do you think I'm going to
4	5	It did not occur to her that
5	4	the first time in her life

6	4	like a duck on a June bug

Most of the 7-grams prove to be associated withScarlett. If one looks closer at the contexts where these 7-grams are used, then it becomes obvious that they describe Scarlett's nature and thoughts. The most frequent 7-gram *for the first time in her life* is connected to Scarlett's revelations about other people, her feelings or actions. Seven uses of the 7-gram are seen as parts of the 8-grams in Section 3.1, but other three occur as the 7-grams.

12. At the sound of his voice, Melanie turned and *for the first time in her life* Scarlett thanked God for the existence of her sister-in-law.

13. Melanie flew at Ashley like a small determined dove and pecked him *for the first time in her life*.

14. And in the heavy silence that fell between them, she felt that she really understood him *for the first time in her life*.

Scarlett is glad that Melanie is with her when Rhett approaches them (Example 12). Example 14, in turn, points that Scarlett understands how Ashley feels for the first time in her life. When Melanie dies, she realizes that Ashley has always loved Melanie and only because of her, he could be strong. The only use the adverbial not referring to Scarlett is seen in Example 13. Ashley does not want to run Scarlett's mills, but Melanie forces him. The example brings to light Melanie's blind love for her dear friend Scarlett. She feels indebted to Scarlett; therefore, she does not want her husband to hurt Scarlett. In addition, the 7-gram*the first time in her life* is used as a part of the 8-grams as presented in Examples 5-7 in Section 3.1. However, it is used in a different context in one occasion:

15. It was *the first time in her life she* had been sorry for anyone without feeling contemptuous as well, because it was the first time she had ever approached understanding any other human being.

Example 15 alludes to the last scene in the novel when Rhett confesses about his love for their daughter Bonnie. As he saw Scarlett in Bonnie, he spoilt her. It was very important for Rhett that Bonnie loved him back, whereas Scarlett never did and always preferred Ashley instead.

As describe above, the 7-gram *I won't think of it now* has occurred together with the pronoun *she* in four instances. In three other cases the 7-gram is followed by the sentences revealing Scarlett's thoughts like *I can't stand it now;I'll think of it later when I can stand it. I*

can't stand it if I do; .The 7-grams and their contexts highlight Scarlett's way of living:every time there is an unpleasant situation Scarlett cannot stand, she tries to forget it.

Besides, the 7-gram *she was going to have a baby* is also significant in the novel as Melanie announces that she is going to have Ashley's baby. In addition, it is linked to the men's perception of pregnant women –

16. But what Alex was thinking as he peered at her in the twilight was that her face had changed so completely he wondered how he had ever recognized her. Perhaps it was because *she was going to have a baby*. Women did look like the devil at such times.

This observation points out that women looked like devils for men, besides, women always needed to conceal their pregnancy, wear wide dresses and stay at home when they were highly pregnant. Women were embarrassed to show their condition to others.

Regarding the interrogatory phrase*Do you think I'm going to,* it is striking that it mainly characterizes Rhett.

17. *Do you think I'm going to* have my Bonnie ashamed of her father*?*
18. *Do you think I'm going to* have her humiliated like Wade for things that aren't her fault but yours and mine?
19. *Do you think I'm going to* let my daughter grow up outside of everything decent in Atlanta?
20. Good God, *do you think I'm going to* let her grow up and associate with the riffraff that fills this house?

As seen from the examples above, Rhett is worried about Bonnie's future. He wants her to be accepted by the old society of Southerners as he considers the Yankees to be *the riffraff*. He is aware of his and Scarlett's faults, but he does not want Bonnie to suffer for the mistakes of her parents.

The 7-gram *It did not occur to her that* proves to reveal Scarlett's unawareness of different things. For example,

21. *It did not occur to her that* if she married Ashley she would automatically be relegated to arbors and front parlors with staid matrons in dull silks, as staid and dull as they and not a part of the fun and frolicking.
22. *It did not occur to her that* the idea was Rhett Butler's.

23. *It did not occur to her that* Ellen could not have foreseen the collapse of the civilization in which she raised her daughters, could not have anticipated the disappearing of the places in society for which she trained them so well.

24. *It did not occur to her that* Ellen had looked down a vista of placid future years, all like the uneventful years of her own life, when she had taught her to be gentle and gracious, honorable and kind, modest and truthful.

25. Melanie looked up like a good little soldier awaiting a command and so tense was the situation *it did not occur to her that* for the first time Rhett was calling her familiarly by the name which only family and old friends used.

Example 21 shows that Scarlett considers marriage to be dull as then she would be deprived of any entertainment and just sit with other matrons instead of flirting with beaus. Example 22 refers to the situation, when Scarlett decides to come out of mourning because she is tired of sitting at home. As it is remarked by the narrator, the idea has been expressed by Rhett, but Scarlett supported his view and took courage to implement it. Examples 23 and 24 reflect Ellen's vision of a wife's life. She is expected to be *gentle and gracious, honorable and kind, modest and truthful.*Scarlett does see any value of what mother has taught her as these virtues appear to be useless for her. She admits that she would better learn how "to plow or chop cotton like a darky." The only use of the 7-gram associated with Scarlett, but not Melanie is demonstrated in Example 25. Rhett addresses Melanie as *Miss Melly* when he pretends that he brought drunken Ashley home from Belle Watling's house (the brothel). By doing this, Rhett saves Ashley from jail because he is suspected to be a member of the Ku Klux Klan.

The last 7-gram found in *Gone with the Wind*is*like a duck on a June bug*. It is a phraseological unit, a similein particular, meaning "an instant pursuit", moreover, it is "American colloquialism, attributed to the South" (*The Free Dictionary*). The following examples provide the contexts where the 7-gram occurs:

26. Pa is coming tomorrow and he's going to land on me *like a duck on a June bug*,answered Scarlett dolorously.

27. But if I tried to draw a draft on it, the Yankees would be on me *like a duck on a June bug* and then neither of us would get it.

28. And if you so much as swear at them, much less hit them a few licks for the good of their souls, the Freedmen's Bureau is down on you *like a duck on a June bug*."

29. It'll bring the Yankees down on us *like a duck on a June bug*.

As seen in Example 26, Scarlett uses the simile when she speaks of her father Gerald. She knows that when he is in the town, he is going to visit her at once and criticize her for coming out of mourning. Three other instances, in turn, signify that the Southerners regard the actions of the Yankees and the Freedmen's Bureau to be immediate.

To sum up, 7-grams pinpoint not just the main heroine Scarlett, but also Melanie and Rhett. In addition, there is a phraseological unit that refers to the Yankees. Compared to the 8-grams, 7-grams are more numerous and their frequency is higher.

3.3 6-Grams Present in the Novel *Gone with the Wind*

As there are not that many 8- and 7-grams in the novel, it is worth searching for 6-grams as well. There are forty-one recurrent 6-grams in *Gone with the Wind*. Nineteen of them are irrelevant to the discussion of the 6-grams as they belong to 7-grams and were already discussed in the previous section. These are *the first time in her life, first time in her life she, I won't think of it*, etc. The relevant 6-grams are presented in Table 3.

Table 3.3 6-Gramsin*Gone with the Wind*

Rank	Frequency	N-gram
1	11	if it hadn't been for
2	8	I don't want to hear
3	7	It did not occur to her
4	5	at the top of her voice
5	5	Do you mean to tell me
6	5	The Girl of the Period Saloon
7	5	For a moment she could not
8	4	could think of nothing to say
9	4	for the first time in his

10	4	for to tote the weary load
11	4	him for the first time in
12	4	I am coming home to you
13	4	I don't see why
14	4	in the back of the wagon
15	4	into the back of her mind
16	4	it would be to know that
17	4	One more victory and the war
18	4	she could not bring herself to
19	4	the first time she had ever
20	4	walk every step of the way
21	4	When This Cruel War Is Over

Taking into consideration the found 6-grams, they do not seem to reveal much about the characters. The most frequent 6-gram is a part of the conditional *if it hadn't been for* that is used by different characters. The most frequent right collocate of the 6-gram is *Ashley*. The contexts of two instances point to Ashley's kind nature: he helped Alex and Tony Fontaine to get back home on their first leave and he helped Tony to escape after the murder of Scarlett's father's former overseer Jonas Wilkerson. The third instance, in turn, refers to Scarlett as she considers him to be "the obstacle" in her love life. She thinks that if Ashley were not around her, then she would notice how much Rhett loved and cared for her. Ashley himself acknowledges that he acted foolishly when he let Scarlett go to Atlanta and marry Frank in order to save Tara. Besides, he emphasizes that without Scarlett's help, he would "have gone into oblivion." Other contexts are related to Scarlett: Melanie is thankful to Scarlett for staying with her when she gave birth to Beau, whereas MrsMerriwether disapproves of Scarlett's driving alone and, thus, exposing herself to almost being raped. For that reason, her daughter's husband Rene needed to protect her honour with Ku Klux Klan. In addition, Scarlett blames the Slatterys for killing her mother as she got infected with typhoid after curing them.

A closer look at the contexts of the 6-grams shows that the 6-gram *I don t want to hear* is used by MrsMerriwether when she disapproves of Rhett Butler, Dr. Meade says this phrase when he protests against Melanie's transportation to Macon or Scarlett's leaving Melanie all alone. Scarlett uses this phrase when she does not want to her about Rhett's pains. Thus, this 6-gram cannot be ascribed to a particular character.

The 6-gram*It did not occur to her* has appeared as the 7-gram with the conjunction *that*. However, there are two occasions when it is used together with the infinitive clauses.

30. *It did not occur to her* to plan just what they would do if Ashley should declare his love for her in unmistakable words.
31. *It did not occur to her* to question whether Tara was worth marrying Frank.

Example 30 signifies that Scarlett did not think what they would do if Ashley confessed that he loved her. Example 31 reveals the enormous signifance of Tara in Scarlett's life.

The 6-gram *at the top of her voice* can be ascribed to three female characters – Mrs. Meriwether, Scarlett and Prissy. MrsMerriwether always expresses her views at the top of her voice. Scarlett cries at the top of her voice rememberinglost youth and dreams. Prissy, in turn, screams at the top of her voice as she seems to be easily frightened and does not hide her emotions. However, *at the top of his voice* also occurs in the novel, though only twice: Gerald O'Hara, Scarlett's father, "bawled orders *at the top of his voice*" and then he was also "singing *at the top of his voice*" when riding to Tara. The 6-grams points to Gerald's Irish character, otherwise it were women who talked or cried at the top of their voices.

As to 6-gram *Do you mean to tell me*, it is used as a way of expressing surprise, an introductory phrase with the further information to be approved by the interlocutor. Scarlett inquires whether Frank Kennedy has not asked for Suellen's hand before; Dr. Meade wonders if women discuss sexual life between them; Scarlett asks Rhett if he gave a credit to Yankees. Besides, Scarlett expresses her indignation through this phrase, when she cannot believe that Rhett does not care for her and so easily agrees to stay away from her bed.

Turning to the fourth 6-gram *The Girl of the Period Saloon*, it is definitely ascribed to Belle Watlings and the women prostitutes. Belle Watling is Rhett's mistress and she stands in contrast to Scarlett, they dislike each other, Belle sees that Scarlett is not a decent lady. Scarlett knows that Rhett spends his time with Belle and gives money to her. Towards the end of the novel readers get to know that Belle always loved Rhett and treated him well, while Scarlett hurt him and even hated him at first. Furthermore, Rhett uses this place as the alibi for the

participants of Ku Klux Klan in order to save them from prison. Apart from being a gambling house, it was also a place to get the latest gossip that was spreading around the town.

Another 6-gram is an adverbial of time occurring with the narrators words – *For a moment she could not.* The uses of the 6-gram show that it describes Scarlett and it stands in clause-initial position, thus, being a sentence opener:

32. *For a moment she could not* understand what it all meant and then, remembering that the commissary warehouses were down by the railroad tracks, she realized that the army had thrown them open to the people to salvage what they could before the Yankees came.

33. *For a moment she could not* speak or move, and though her mind shouted: He isn't dead!

34. *For a moment she could not* take in his words.

35. She looked up into Rhett's immobile face and *for a moment she could not* speak.

36. *For a moment she could not* speak and all the pretty gestures and smiles she had thought to use upon him were forgotten.

Judging from Examples 32-36, the verb the 6-gram goes together most is *speak.* Scarlett is speechless when she gets to know that Ashley is probably dead (Example 33), then she gets stunned when she suspects that Frank is dead (Example 35). In addition, she cannot speak when she realizes that the night they spent with Rhett is something usual for him (Example 36). As to Example 33, Scarlett gets frightened when Big Sam hails her on a lonely road.

The predicate *could think of nothing to say* appears as the recurrent 6-gram is well. The extended contexts of the 6-gram reveal the use of pronoun she that clearly depicts the main character – Scarlett.

37. Scarlett *could think of nothing to say*, and she felt her face growing red with annoyance.

38. She *could think of nothing to say* and so merely looked at him, wondering why men were such fools as to think women interested in such matters.

39. Scarlett was silent because she knew he was telling the truth and she *could think of nothing to say*.

40. She looked as if she expected some comment from Scarlett but the words had made little sense to her and she *could think of nothing to say*.

The above examples signify that Scarlett lacks words when her father reads her thoughts about Ashley's marriage (Example 37), later when Charles, hir first husband, talks of war, she is silent as she believes that women do not like talking about war matters (Example 38). Besides, Scarlett has nothing to say when she is confronted with the truth about Frank's intention to marry her sister Suellen, but not Scarlett (Example 39). The last occurrence of the 6-gram shows that Scarlett is narrow-minded.Grandma Robilliard tells her how the wind of change makes people adaptto the new way of life. "We're not wheat, we're buckwheat! When a storm comes along it flattens ripe wheat because it's dry and can't bend with the wind. But ripe buckwheat's got sap in it and it bends.And when the wind has passed, it springs up almost as straight and strong as before." It is an allusion to the novel's title: those Southerners who could not bend, have gone with the wind. As seen from Example 40, Scarlett does not pay much attention to Grandma's words. Though she is a strong personality who has successfully adapted herself, she is not able to think of herself as "ripe buckwheat" as she does not understand such abstruse things.

A parallel construction to the 7-gram *for the first time in her life* is found among the 6-grams – *for the first time in his.* It refers to the male characters who were Scarlett's husbands – Charles, Frank and Rhett.

41. I love you! You are the most-the most-and he found his tongue *for the first time in his* life.

42. She made him feel, *for the first time in his* old-maidish life, that he was a strong upstanding man fashioned by God in a nobler mold than other men, fashioned to protect silly helpless women.

43. At her words, his grip tightened and he began speaking rapidly, hoarsely, babbling as though to a grave which would never give up its secrets, babbling the truth *for the first time in his* life, baring himself mercilessly to Melanie who was at first, utterly uncomprehending, utterly maternal.

Firstly, it points to Charles, Scarlett's first husband, who dared to confess his love of Scarlett for the first time (Example 41). Then the phrase relates to Frank, Scarlett's second husband, who realized his strength thanks to Scarlett (Example 42). Further the 4-gram *for the first time in his life* introduces a turning point for Rhett, who reveals the truth about his deeds and relationship with Scarlett to Melanie. Example 43 shows that Rhett deeply trusts Melanie, therefore he speaks frankly to Melanie, whereas he cannot do this with Scarlett as he is afraid that she will hurt him.

The phraseological unit *for to tote the weary load* turns out to be frequent as well. It is a line from the song "My Old Kentucky Home":

44. Just a few more days *for to tote the weary load*!

Scarlett repeats the phrase in her mind to soothe herself: "

45. Just a few more steps, hummed her brain, over and over, just a few more steps *for to tote the weary load*."

Example 45 describes the situation when Scarlett together with pregnant Melanie, Wade and Prissy try to get back to Tara. Their horse is exausted and it goes very slowly, but Scarlett believes that they are going to make it.

The next 6-gram is *him for the first time in.* It has been already discussed as a part of the 7-gram *for the first time in her life* seen in Examples 13-14. Two other occurrences that have not been mentioned before relate to Ashley as Scarlett sees *him for the first time in more than two years* since he joined the army, in addition to Rhett, whom Scarlett sees *him for the first time in weeks.* This scene happens towards the end of the novel when Scarlett notices the drastic change in Rhett: "This man was not going to laugh, nor was he going to comfort her."

The 6-gram representing the future time clause *I am coming home to you* turns out to be associated with Ashley. He writes this phrase to Melanie:

46. With fingers that shook, she opened it and read:

Beloved, *I am coming home to you*-

Tears began to stream down her face so that she could not read and her heart swelled up until she felt she could not bear the joy of it.

Example 46 reveals the moment when Scarlett reads Ashley's letter to Melanie without asking her. She imagines that Ashley addresses her instead of Melanie and, without doubt, she feels extremely happy that he is coming back.

I don't see why is a negative clause that is used by Scarlett when she wonders why Rhett treats Melanie better than her. In addition, Mrs. Elsing doubts whether Rhett had any reason to enlist in the army. Afterwards the 6-gram *in the back of the wagon* follows. It is an adverbial of place. Three uses of it are found depicting the scene when Rhett helps Scarlett, Melanie, Wade and Prissy to escape from Atlanta, but then he leaves them in order to join the army. Melanie, Wade and Prissy are transported in the back of the wagon. The narrator remarks that Scarlett hates Rhett when he makes fun of the soldiers who retreat though she also considered them to be fools:

47. She knew her safety and that of the others *in the back of the wagon* depended on him and him alone, but she hated him for his sneering at those ragged ranks.

Another 6-gram, *into the back of her mind,* is ascribed to Scarlett as she pushes unpleasant memories or emotions into the back of her mind. These memories are, for instance, her mother Ellen teaching Scarlett about the moral education of servants (Example 48), or trying not to think what she is going to sacrifice if she wants to get money from Rhett (Example 49), hiding her anger at Rhett (Example 50), or getting rid of unpleasant thoughts after suffering a miscarriage.

48. But now, Scarlett pushed that admonition *into the back of her mind*.
49. I'll think of it later, and she pushed the unwelcome idea *into the back of her mind* lest it shake her resolution.
50. When he came back from New Orleans, cool and bland, she swallowed her anger as best she could, pushing it *into the back of her mind* to be thought of at some later date.
51. As she had once fled Atlanta before an invading army, so she was fleeing it again, pressing her worries *into the back of her mind* with her old defense against the world: I won't think of it now.

The part of conditional expressing wish *it would be to know that* is also repeated in the novel. The extended context reveals that the 6-gram is used with the adverbs to emphasize extent – *how awful/nice/dreadful.* The uses of the 6-gram together with the adverbs *how awful/nice* point to Melanie who speaks of Confederate Soldiers graves:

52. Their graves are somewhere up in the Yankeescountry, just like the Yankee graves are here, and oh, how awful *it would be to know that* some Yankee woman said to dig them up and.
53. But how nice *it would be to know that* some good Yankee woman-And there must be SOME good Yankee women.
54. How nice *it would be to know that* they pulled weeds off our men's graves and brought flowers to them, even if they were enemies.

Melanie is merciful and she is willing to take care of the the Yankees' graves. Moreover, she believes that good people are found everywhere and also among the Yankees. Another occurrence of the 6-gram *it would be to know that* is seen when Scarlett thanks God that nor Ashley nor Frank are involved in the Klan.

55. How dreadful *it would be to know that* the Yankees might swoop down and arrest them at any minute!

The contexts 6-gram*One more victory and the war* relates indicates the optimistic mood of the Southerners when they won the battles. They tend to say: *One more victory and the war*was over! One more 6-gram that carries the similar meaning and is also recurrent is *When This Cruel War Is Over.* It was the song the Southerners sang when they gathered.

Moreover, the 6-gram *she could not bring herself to* is again ascribed to Scarlett. The clause is followed by the phrasal verb *go back*, then it describes the situation when Scarlett cannot immediately return to the room where pregnant Melanie lies. Other verbs the 6-gram occurs with are *speak*and *utter* refer to the situation when the Yankees come to Tara. In addition, the verbs *go* and *ask* appear in the 6-gram's contexts. The former verb relates to Scarlett's unwillignes to visit the Tarletons as her youth friends the Tarleton boys are dead, whereas the latter is alluding to Scarlett's mood after the miscarriage. She wants Rhett to be with her, to calm her, as this was the first child she wanted herself, but she is too proud to beg him to come.

The noun phrase *the first time* is again frequent, but in a different 6-gram - *the first time she had ever.* The 6-gram has been already encountered in Example 15. Here are other examples of the 6-gram:

56. It was *the first time she had ever* seen any woman who she knew for certain had done something to her hair and she watched her, fascinated.
57. Whenever Scarlett raised her voice in reproof, he went weak with fright as his vague childish memory brought up the horrors of *the first time she had ever* done it.
58. It was *the first time* she had ever known what Ashley was thinking when his eyes went past her, crystal clear, absent.

Example 56 signifies that Scarlett is amazed at the colour of Belle Watling's hair. Example 57 describes how Wade fears his mother Scarlett as he knows that she can be very angry with him. Example 58, in turn, implies that Scarlett becomes aware that Ashley is resigned as the world he was living in before the war has drastically changed and he could not find his place there.

Walk every step of the way describes Scarlett's will to get everything she wants, her stubborness, as she is ready to walk every step of the way to Virginia in order to be with Ashley (Example 48) or go home in order to be with her mother (Example 49), or drive with servants to supervise the mill workers (Example 61).

59. Oh, Ashley, I love you so much I'd *walk every step of the way* to Virginia just to be near you!

60. If it wasn't for Melanie, she'd start home, right this minute, if she had to *walk every step of the way*.

61. I'll *walk every step of the way* and pick up a crew of darkies somewhere.

The analysis of the 8-, 7- and 6-grams has shown that the image of Scarlett is created with the help of the time adverbials *for the first time in her life she, for the first time in her life, the first time in her life, the first time she had ever, it did not occur to her that* and *it did occur to her.* Thus, the stress is put on the things that happen to Scarlett, ideas that dawn upon her and the feelings that arise in her. Furthermore, the recurrent N-grams reveal Scarlett's attitude to the events she cannot change. This is done by the repetition of such clausesas*I won't think of it now she, I won't think of it now,into the back of her mind.* Apart from the N-grams that have been mentioned and represent negative clauses, there are other N-grams that frequently occur and describe Scarlett like *For a moment she could not*and *she could not bring herself to.* In addition to the depiction of Scarlett, the narrator uses the parallel time adverbial construction to highlight the features of male characters - *for the first time in his.* All in all, the 8-, 7- and 6-grams prove to be exclusive for the novel*Gone with the Wind* as they are associated with Scarlett, Rhett, Melanie and Ashley, though some of the N-grams turn out to be speech signals and the phraseological units that contribute to the narrative. As the discussion of the longer N-grams is completed, the next stage is to search for the 5-grams and explore whether Mahlberg's functional groups are relevant in their classification.

3.4 5-Grams Present in the Novel *Gone with the Wind*

As mentioned in Section 2.2, Michaela Mahlberg (2007) introduced six groups that characterized Dickens' style: labels, speech clusters, body part clusters, *as if* clusters, time and place clusters, and *other*. To the extent that these groups are applicable to 5-grams from Mitchell's novel *Gone with the Wind*, they will be also used here. In Table 3.4, the 5-grams in *Gone with the Wind* are grouped according to the functional group they belong to. There are sixty-two of them with a minimum frequency of five that have not occurred before as parts of longer N-grams with the same context. Fifty-eight of them are included in Table 3.4, while the

rest are found in Appendix 1. In addition, the distribution of 5-gram types is provided in Table 3.5.

Table 3.4 5-Grams in *Gone with the Wind*

Labels (L)	Speech (S)	Time and place (TP)
She did not want to (15)	I don't want to (21)	for the first time in (18)
I won't think of (14)	I don't know what (14)	at the end of the (9)
going to have a baby (10)	I'm going to have (12)	at the head of the (6)
she was going to have (10)	it hadn't been for (11)	at the top of the (6)
Don't be a fool (7)	I'm not going to (10)	in the back of the (6)
her heart in her throat (7)	I don't care what (9)	at the bottom of the (5)
I'll think of it (7)	and I don't want (9)	at the foot of the (5)
hundred and fifty thousand dollars (6)	I don't know how (9)	for the first time since (5)
the end of the world (6)	I don't see why (9)	in front of the house (5)
but she did not care (5)	I won't have you (8)	Just a few more days (5)
Don't be a goose (5)	Do you think I'm (8)	out of the house and (5)
How do you run on (5)	you don't want to (7)	
I beg your pardon Scarlett (5)	and I don't see (6)	
I can't stand it (5)	Do you think I'd (6)	
I mustn't think about (5)	I'm going to be (6)	
She looked up at him (5)	but I'm going to (6)	
the baby in her arms (5)	't know what to do (5)	
there were so many things (5)	as well as I do (5)	

thing in the world that (5)	Do you think I'd (5)	
	he didn't want to (5)	
	I didn't mean to (5)	
	I don't believe I (5)	
	I don't believe it (5)	
	I don't care if (5)	
	I don't know I (5)	

Table 3.5 Distribution of 5-gram typesin*Gone with the Wind*

Labels (L)	Speech (S)	Time and Place (TP)
19	30	11

3.4.1 Labels

The analysis of these 5-grams shows that speech 5-grams prevail among the functional groups. However, some of the 5-grams like *going to have a baby, every step of the way, the back of her mind,the first time in her, if it hadn't been, It did not occur to, Girl of the Period Saloon,a moment she could not, the first time in his, you think I'm going to, etc.* that function as labels have not been included in the table because they have been discussed above. Thus, labels are more prominent among the novel's 5-grams than speech clusters. *As if* clusters turned out to be irrelevant as there were no 5-grams beginning with this conjunction.In its context, one 5-gram can be assigned to the body part clusters, namely*at the top of*. It is used together with the pronoun *her* and the nouns *voice* or *lungs*. Its occurrences as a 6-gram have been already discussed in Section 3.3. What is more, Mahlberg's functional group *other* has not been applied as almost all 5-grams fit under labels, speech clusters, and time and place clusters. The only 5-gram that can be classified as *other* is *the baby in her arms.* It depicts several female characters who hold a baby – Prissy, Scarlett and Melanie. Due to the space limitation, it is placed under labels. Time and place clusters appear to be in the minority.

Most of the labels are associated with Scarlett, revealing her character. The 5-gram *she did not want to* refers to Scarlett as she was unwilling to be fair, display her condition, hear sympathy, think of others, sell her mills to Melanie and Ashley, marry or bother with anything unpleasant. All these verbs drawn from the contexts, where the 5-gram appears, basically show Scarlett's true face. *Going to have a baby* is ascribed to both Melanie and Scarlett. Melanie is happy to have children, but Scarlett does not want to have babies as they ruin her figure. While pregnant with Frank's baby, she even exclaims:

62. Oh, God, if only I wasn't *going to have a baby* at this of all times!

Then she could freely run her lumber mills and enjoy life as usually. Though the 5-gram *She was going to have* appears with the noun *baby,* it also occurs together with the noun *money.* Scarlett is very materialistic and she strives for wealth; therefore, she is ready to do everything to get a lot of money. The 5-gram signals also preoccupation with the future.

63. And, no matter what price she had to pay, *she was going to have* money again, more than just enough to pay the taxes on Tara.
64. Somehow, some day *she was going to have* plenty of money if she had to commit murder to get it.

Turning to the phrase *Don't be a fool*, it is often used by Scarlett and in reference to Scarlett. Scarlett uses it when she is irritated and could say it to Prissy or she also used this phrase to reassure somebody that everything was well, for example, when she speaks to Pork. By contrast, Rhett used this phrase when he wanted to make fun of Scarlett. Another parallel phrase *Don't be a goose* is used by Scarlett and addressing her. Scarlett reassures that Melanie is not going to die when delivering her baby, besides, she utters the phrase when she does not want to share the watermelon with a guest and persuades Melanie and Carreen that they need to hide it. Example 64 reveals the situation when Rhett says the phrase to Scarlett when she confesses that she is afraid. Afterwards Scarlett adds that she fears to die and go to hell because she cheated Frank in order to marry him and made his life miserable after all.

65. *Don't be a goose*, Scarlett, you've never been afraid in your life."

Continuing with Scarlett's phrases, there is one more that appears to be frequent as well. Scarlett exclaims "How do you run on"especially when she reacts on Rhett's remarks. For instance,

66. Oh, Rhett, *how you do run on*, teasing a country girl like me!

The recurrent 5-gram is *her heart in her throat* leads to important experiences in Scarlett's life as she is very nervous when she is about to meet Ahsley and tell him about her love. Besides, she experiences the same feeling when she first sees the Yankee soldiers and reads casualty lists because she is anxious to find Ashley dead.

> 66. Quick, Melly, cried Scarlett, *her heart in her throat*, exasperation sweeping her as she saw that Melly's hands were shaking so that it was impossible for her to read.

The phrase *I'll think of it* is usually continued with the adverb *later*, again pointing to Scarlett's way of pushing back unpleasant thoughts. The same connotations carry the 5-grams *I can't stand it* and *I mustn't think about.* As to 5-gram *hundred and fifty thousand dollars* representing a sum of money, it proves to be symbolic in this novel as Scarlett needs this money in order to save her plantation Tara.

The noun phrase *the end of the world* is a symbol how characters perceive the war. Ashley sees the end of the war is also the end of the world. The life he has led before changes completely after the war is over. When Scarlett hears the noises and the lighted sky, she thinks of the end of the world. Moreover, the 5-gram comes up in the conversation between Scarlett and Rhett. Scarlett remembers what Ashley says to her:

> 67. Once at Tara he said something about the-a-dusk of the gods and about *the end of the world* and some such foolishness."

Scarlett considers Ashley's words to be *foolishness* that again reveals Scarlett's narrow-mindedness.

The next 5-gram *but she did not care* can be ascribed to Scarlett. The conjunction *but* stresses that Scarlett is aware of what is going on, but she does not bother about what others say about her behaviour. Examples 68-70 highlight Scarlett's egoistic nature.

> 68. She thought dully what Aunt Pitty would say, if she could see her sprawled here on the front porch with her skirts up and her drawers showing, *but she did not care*.
> 69. She knew Atlanta was talking *but she did not care*.
> 70. She hoped he was equally pleased *but she did not care* very much whether he was or not.

The independent clause describing Scarlett *She looked up at him* is also repeated in the novel. The pronoun *she* refers to Scarlett and Melanie in one instance. The pronoun *him* denotes Rhett.

71. *She looked up at him* imploringly, her face crimson with the shame of their last meeting, and met two of the blackest eyes she had ever seen, dancing in merciless merriment.

The contexts of the incomplete past tense clause *there were so many things* are related to Scarlett.

72. But *there were so many things* about him that she did not understand.
73. And then there were-Oh, *there were so many things* to do to bachelors and she knew them all, the nuance of the sidelong glance, the half-smile behind the fan, the swaying of the hips so that skirts swung like a bell, the tears, the laughter, the flattery, the sweet sympathy.
74. Oh, *there were so many things* she would preface with Do you remember?
75. Perhaps it was that there was nothing she would not do, and *there were so many things* these people would rather die than do. "
76. But of course, in all her conversations with Ashley *there were so many things* which could not be said, for honor's sake, that the sheer force of them inhibited other remarks.

Taking into account the above examples, it is striking that Scarlett cannot fully understand Ashley (Example 72), besides, she cannot be frank with him as he is honourable (Example 76), meanwhile, she can freely discuss daily wants with Rhett. Example 73 shows that Scarlett is a skillful flirt. Example 75, in turn, reveals Scarlett's strong personality as she is a fighter who is ready to do everything in order to prosper. One more 5-gram contains the noun *thing,* which is *thing in the world that.*

77. Land is the only *thing in the world that* amounts to anything,he shouted, his thick, short arms making wide gestures of indignation, for 'tis the only thing in this world that lasts, and don't you be forgetting it! 'tis the only thing worth working for, worth fighting for-worth dying for."
78. For 'tis the only *thing in the world that* lastsand to anyone with a drop of Irish blood in them the land they live on is like their mother'tis the only thing worth working for, fighting for, dying for."
79. If there's one *thing in the world* that gives me more amusement than anything else, he remarked, it's the sight of your mental struggles when a matter of principle is laid up against something practical like money.

80. There was only one *thing in the world that* was a certain bulwark against any calamity which fate could bring, and that was money.

81. With the sound of his feet dying away in the upper hall was dying the *last thing in the world that* mattered.

The contexts of the 5-gram indicate that the superlative *the only, the last* or the quantifier *one* preceede the noun phrase. The emphasis is put on the noun phrases the 5-gram modifies. The noun phrases used in Examples 77-81 are *land, the sight of your mental struggles when a matter of principle is laid up against something practical like money* and *dying.* Thus, the novel's themes become noticeable for the readers: these are land and money. Furthermore, the extended contexts contribute to the characterization of Scarlett. Land turns out to be the most precious thing for Scarlett though when she was sixteen she did not realize it. It is Scarlett's father who shows her the real value of Tara (Example 77). As underlined before, money has huge value as they give her the security she needs (Example 80). Besides, Rhett makes fun of her, saying that she is ready to put aside her moral principles if she can get money. However, Scarlett is not afraid of dying as nothing matters anymore (Example 81).

Almost all labels found among the 5-grams are associated with Scarlett. In addition, several of them are negatively coded when representing Scarlett's thoughts and words:*I won't think of, Don't be a fool,Don't be a goose,I can't stand it* and *I mustn't think about.* Apart from these 5-grams that contribute to the formation of Scarlett's character, there are also the observations of the narrator that help readers to comprehend what kind of person Scarlett is. The narrator's remarks are easily distinguishable from Scarlett's own thoughts and the speech clusters as the verbs forms are not contracted - *She did not want to*and *but she did not care.* The contexts of the labels do not only reveal Scarlett's character, but also pinpoint the novel's themes. The discussion of labels is followed by the analysis of speech clusters and time and place clusters.

3.4.2 Speech Clusters and Time and Place Clusters

Most of the speech 5-grams contain the personal pronoun *I* and reveal a person's unwillingness, intentions or inability*: I don' t want to, I don't know what, I'm going to have, I'm not going to, I don't care what, I don't know how, I don't see why, Do you think I'm.* Thus, the speech acts used in the novel are often negative. In addition, two 5-grams imply lack of faith: *I don't believe I* and *I don't believe it.* All in all, the verbs *want, know, think, see, believe* are

constantly repeated. Thus, Stubbs and Barth's claim that "frequent vocabulary from the lexical fields of saying, looking, thinking and wanting" is characteristic of the fictional texts holds true for the Mitchell's novel (2003: 79 cited in Section 2.4). The high number of speech signals signifies that the dialogues play an essential part in the creation of the characters' images and their relationships.

Time and place clusters are not that frequent as speech clusters and labels. The most frequent of them is a time adverbial with the preposition *for the first time in*, focusing on what happens for the first time. This time adverbial has appeared with the pronoun *she* and the noun phrases *her life* and *his life,* but it also appears in other contexts. For instance,

82. It was as if she were really seeing herself *for the first time in* a year.
83. Her green eyes were alert and sparkling again, and she laughed aloud *for the first time in* weeks when Rhett and Bonnie met her and Wade and Ella at the depot-laughed in annoyance and amusement.

In addition, the same time adverbial with the preposition *since* is also repeated among the 5-grams – *for the first time since.*

84. Here in the northwesternmost corner of the state, at Chickamauga, serious fighting had occurred on Georgia soil *for the first time since* the war began.

Several of the 5-grams turn out to be positional phrases that follow the pattern *PREP the N of the: at the end of the, at the head of the, at the top of the, in the back of the, at the bottom of the* and *at the foot of the.* This is the pattern discovered by Stubbs and Barth (2003) (Section 2.4).

The noun *house* appears to be dominant in two 5-grams – *in front of the house* and *out of the house and.* Another time 5-gram *Just a few more days* is used in the song "My Old Kentucky Home" that is mentioned above in connection with the 6-gram *for to tote the weary load* (Section 3.3).

The distribution of 5-grams types has showed that only three groups from Mahlberg's classification prove to be relevant in the analysis of the 5-grams – labels, speech clusters and time and place clusters. Labels slightly prevail over speech and time and place clusters. The second place is taken by speech 5-grams, but time and place clusters make the smallest functional group. Speech clusters assist in the creation of characters and the relationships between them, while time and place clusters are positional phrases or time adverbials that describe the novel's settings. Mahlberg's observation that long clusters are less frequent than

shorter ones is pertinent in the analysis of the recurrent N-grams in *Gone with the Wind.* 5-grams present in the novel are more recurrent and larger in number than 8-, 7- and 6-grams.

3.5 4-Grams Present in the Novel *Gone with the Wind*

The next step of the analysis is to identify 4-grams in the novel *Gone with the Wind*. As the search for 4-grams with the minimum frequency of 5 returned 724 results, I decided to set the frequency threshold for 4-grams to 10. The new query returnedone hundred and ten 4-grams, but only eighty-three of them are presented in Table 3.6 and Appendix 2 because twenty-seven 4-grams appeared as parts of 8-, 7-, 6- and 5-grams.

3.6 4-Grams in*Gone with the Wind*

Labels (L)	Speech (S)	Time and place (TP)	Other (d)
She did not know (32)	I'm going to (69)	for the first time (47)	and there was a (19)
did not want to (22)	I don't know (68)	the back of the (25)	the rest of the (18)
I've got to (22)	I don't want (52)	in front of the (22)	he looked at her (18)
but she did not (21)	And I don't (44)	at the top of (19)	and there was no (15)
For a moment she (20)	I don't care (35)	on the front porch (17)	did not seem to (13)
she would have to (19)	and I don't (32)	out of the house (14)	he was going to (13)
that she could not (17)	don't want to (31)	and for a moment (13)	it would be to (12)
she did not care (16)	I don't believe (20)	at the end of (12)	she would have to (12)
but she could not (15)	but I don't (17)	the top of the (12)	and some of the (11)
Don't be a (15)	don't know what (18)	the head of the (11)	was going to be (11)
if she could only (15)	I didn't know (18)	sat down on the (10	are going to be (10)
and she did not (15)	I don't see (18)	stood for a moment (10)	looked down at her (10)
she thought of the	don't know how (15)		

(15)			
came back to her (13)	I don't like (15)		
she had ever seen (13)	I won't be (14)		
that she did not (13)	I won't have (14)		
I'm not going (10)	if you don't (14)		
I'm tired of (10)	I wouldn't have (13)		
she could not understand (10)	's going to be (13)		
she had never seen (10)	I don't think (12)		
	I want you to (12)		
	I won't go (12)		
	'm going to have (12)		

As if Clusters	War (W)	Body Part (BP)
as if she had (13)	The Yankees are coming (13)	at the sight of (21)
as if he were (12)	the war was over (14)	And his voice was (11)
as she were (12)		in the face of (11)

3.7 Distribution of 4-gram types in*Gone with the Wind*

Labels (L)	Speech (S)	Time and place (TP)	*As if*(AI)	War (W)	Body Part (BP)	Other (d)
20	31	12	3	2	3	13

First of all, it is important to emphasize that quite many 5- and 4-grams reflect 8-, 7- and 6-grams. For that reason, twenty-five 4-grams that perform the label function have not been included in the analysis of 4-grams. The 4-grams signaling speech have also come up as parts of longer N-grams, but as 4-grams they have appeared in different contexts that were not encountered before. The same happens to all 4-grams that have been noticeable among longer N-grams, but if their extended contexts reveal other uses, then they are counted as 4-grams.

In contrast to the 5-gram types, the distribution of 4-gram types indicates that all Mahlberg's functional categories fully fit in the analysis of 4-grams present in*Gone with the Wind.* In addition, one functional group that I called *war* is added to the classification.

3.5.1 Labels

Almost half of the labels, nine of them to be precise, turn out to be negative constructions like*: she did not know, did not want to, but she did not, that she could not, she did not care, but she could not, Don't be a, that she did not* and *she could not understand.* Moreover, they clearly belong to the characterization of Scarlett: the 3rd person, feminine pronoun is used together with the predicates when the narrator refers to Scarlett. These 4-grams reveal Scarlett's nature: her stubbornness, egoism and unawareness of other people's feelings and actions. Several parallel constructions only with different conjunctions are used to refer to Scarlett's personality: *but she did not – that she did not; that she could not – but she could not.* The examples bellows show their occurrences.

85. To be sure, she still jumped at the sound of explosions *but she did not* run screaming to burrow her head under Melanie's pillow.
86. But Scarlett never lost the old feeling that he was watching her covertly, knew that if she turned her head suddenly she would surprise in his eyes that speculative, waiting look, that look of almost terrible patience *that she did not* understand.
87. She thought *that she could not* live any longer in the same house with the woman who was carrying Ashley's child, thought that she would go home to Tara, home, where she belonged.
88. She had immediately loathed nursing *but she could not* escape this duty because she was on both Mrs. Meade's and Mrs.Merriwether's committees.

The 4-grams *that she did not* and *that she could* not are also associated with Melanie. Example 89 reveals Melanie's kind nature, her ability to see the best in all people. Besides, three parallelisms that are deliberately used by the narrator emphasize that everyone, no matter servant, girl or man, stupid, ugly, worthless or boring, is precious.

89. There was no servant so stupid *that she did not* find some redeeming trait of loyalty and kind-heartedness, no girl so ugly and disagreeable *that she could not* discover grace of form or nobility of character in her, and no man so worthless or so boring *that she did not* view him in the light of his possibilities rather than his actualities.

Moreover, a pair of contrasts is found among the 4-grams:*she had ever seen* vs. *she had never seen.* Here are their uses:

90. She rose as he entered, thinking with proprietary pride that he was the handsomest soldier *she had ever seen*.

91. She thought it the most beautiful and most elegantly furnished house *she had ever seen*, but Rhett said it was a nightmare.

92. Scarlett thought *she had never seen* a man with such wide shoulders, so heavy with muscles, almost too heavy for gentility.

93. Scarlett thought *she had never seen* such elegant pants as he wore, fawn colored, shepherd's plaid, and checked.

As seen from Examples 90-91, the 4-gram *she had ever seen* is used together with the superlatives *the handsomest* and *the most beautiful and most elegantly furnished.* The contexts of both 4-grams signify that Scarlett is surprised by material things like houses, people's looks and clothes. It contributes to Scarlett's characterization: her interests are quite superficial as she is not amazed at people's qualities like kindness and generosity. In addition, Example 91 shows Scarlett's confidence, she believes that she has a good taste, but Rhett disagrees. Still he wants her to be happy and, therefore, allows her to do whatever pleases her.

Apart from the examples with the 5-gram *She was going to have* described in Section 3.4.1, there are other uses of its shorter version – the 4-gram *she was going to.*

94. Sugar always caught more flies than vinegar, as Mammy often said, and *she was going to* catch and subdue this fly, so he could never again have her at his mercy.

95. *She was going to* rush into life and wrest from it what she could.

Example 94 signifies Scarlett's intention to be kind to Rhett and then subjugate him as he knows about her secret love to Ashley. The 4-gram appears together with verbal metaphors *catch and subdue this fly* where *fly* implies Rhett. Thus, the stylistic devices provide a vivid picture for readers and create a humorous effect as Rhett is believed to be a strong character that is not associated with a fly. In addition, Example 95 reflects two other verbal metaphors *rush* and *wrest.* They demonstrate Scarlett's vitality, her willingness to get what she wants by all means.

Other 4-grams are also important for the creation of the characters' images. These are *I've got to, For a moment she, if she could only, came back to her* and *I'm tired of.* There are many duties that lie on Scarlett's shoulders as she needs to provide not only for herself, but also for her family. For instance, she says several times what she needs to do in order to save Tara:

96. But *I've got* to have a new dress! Don't you see *I've got to* go to Atlanta and get money for the taxes? *I've got to* get some money. I've got to do it!

44

The context where the 4-gram *For a moment she* is used, describes Scarlett's reaction and it usually opens the sentence:

97. *For a moment she* was indignant that he should say other women were prettier, more clever and kind than she, but that momentary flare was wiped out in her pleasure that he had remembered her and her charm.

98. For a moment she stood irresolute and through the closed door came the sound of her boy's treble laughter.

99. *For a moment she* was on the verge of an outburst of childish wild tears.

The label *if she could only* is used with a number of verbs that reveal Scarlett's deepest wishes about coming back to Tara and her mother Ellen, getting to Ashley and lastly, regretting her bad attitude towards Melanie. The part of the conditional opens the sentences as well. Here are the examples from the novel:

100. *If she could only* be transferred by magic to Tara and to safety. If she could only be with Ellen, just to see her, to hold onto her skirt, to cry and pour out the whole story in her lap.

101. *If she could only feel* that sense of oneness with him for which she had yearned since that day, so long ago, when he had come home from Europe and stood on the steps of Tara and smiled up at her.

102. Oh, *if she could only* live those years over again! She would never even let her eyes meet those of Ashley.

Regarding the 4-gram *came back to her*, it is ascribed to Scarlett's memories of Rhett and Ashley's words:

103. Then the laughing words of Ashley *came back to her*: Only one person, Rhett Butler… who has money."

104. The bitter words Rhett had spoken in the early days of the war *came back to her*, and she remembered him saying he would never fight for a society that had made him an outcast.

105. Fragments of words *came back to her* and she quoted parrot-like: A glamor to it – a perfection, a symmetry like Grecian art."

In the end of the novel Rhett has basically the same view of life as Ashley: both of them yearn for the old days with their *slow charm* and *glamour*. Scarlett, on the contrary, is used to looking ahead, being attracted to new things and fighting for them, neglecting the traditions. This is the major difference between her and these two men: she belongs to a new generation, whereas Ashley and Rhett are from old guard.

The 4-gram *I'm tired of* depicts Scarlett and her reluctance to be like other women of her time so that men feel than they are dominant, more clever and stronger than her. In fact, she really proves to stand out from other women as she is self-willed, strong and always gets whatever she wants:

> 106. *I'm tired of* everlastingly being unnatural and never doing anything I want to do. *I'm tired of* acting like I don't eat more than a bird, and walking when I want to run and saying I feel faint after a waltz, when I could dance for two days and never get tired. I'm tired of saying, How wonderful you are!' to fool men who haven't got one-half the sense I've got, and *I'm tired of* pretending I don't know anything, so men can tell me things and feel important while they're doing it I can't eat another bite."

Another 4-gram *I'm not going to* can be mainly ascribed to Rhett as he clearly states that he is not going to send their daughter Bonnie to the North to study because she would not be accepted here in the South. Moreover, he does not want her to marry a Yankee "because no decent Southern family will have her-because her mother was a fool and her father a blackguard". At this point he realizes that he needs to appreciate the traditions and values of the South.

3.5.2 Speech Clusters, Time and Place Clusters, and *Other* Clusters

Turning to the second most frequent functional group *speech clusters,* it basically reflects a number of interactions such as *and I don't, Do you think I, I don't believe, I don't see, I don't like*. Just like labels, speech clusters are also negative. In addition, as already mentioned, some 4-grams prove to be formulaic like *I don't know, I don't care, I beg your pardon, What do you mean? What's the matter?* Just like with the 5-grams, the verbs *think, see, want* are prominent among the 4-grams. *I don't know* is a common colloquial phrase that tends to be frequent in general. BengtAltenberg finds the independent clause*I don t know* to be used 47 times in the London-Lund Corpus of Spoken English where it functions as an "epistemic tag", i.e. it appears

as "a modal comment clause that is either inserted parenthetically in, or added to the end of, an utterance" (Altenberg 1998: 104-105). Besides, he points out that this expression has "a clearly recognizable clause structure" (ibid.).

Time *and* place clustersare also encountered among the 4-grams. Some of them like *in front of the* and*the back of the* were seen as parts of 5-grams - *in front of the house, the back of the wagon.* Nevertheless, the positional phrase*in front of the* is used with other nouns such as*the mirror, avenue, and wooden awnings.* In addition, the 4-gram is also used together with animate objects like *others, hostile neighbors, men,* then they are in the focus, not the location. *The back of the* comes together with such nouns as *buggy* and *house.* Other frequent time and place clusters just like the 5-grams follow Stubbs and Barth's pattern (2003) (Section 2.4)*PREP/ the N of the: the end of the,on the front porch, out of the house,at the end of, the head of the* and *the top of the*.

According to the distribution, the functional group *other* contains one 4-gram more than time and place clusters. These are mostly 4-grams which contribute to the novel's narration, highlighting the progression of the story, events, actions and feelings of the characters: *and there was a*, *and there was no, he was going to, was going to be, it would be to, looked down at her, etc.* Two other 4-grams, in turn, imply quantity or presence of something: *the rest of the* "County, Charleston folks, world, etc.", *and some of the*"soldiers, vitality, real anguish, etc."

All in all, the functional groups labels, speech clusters, time and place clusters, and *other* are the largest categories. However, there are three other groups that also contribute to the novel's style.

3.5.3 Body Part Clusters, As If Clusters and War Clusters

As noted above, there are three small groups – body part clusters, *as if* clusters and *war*clusters that are relevant to the discussion of the novel. Body part clusters are represented by the 4-grams: *at the sight of, and his voice was* and *in the face of.*Though the 4-gram *at the sight of* does not comprise the body part, it implies "seeing".

107. Her spirits rose, as always *at the sight of* her white skin and slanting green eyes, and she smiled to bring out her dimples.

108. But he could not see her family evicted, and his disappointment soon faded *at the sight of* her radiant happiness, disappeared entirely at the loving way she took onover his generosity.

The 4-gram *and his voice was* mostly depicts Ashley and Rhett. When characterising Ashley, the narrator uses strong epithets like *deadly quiet, very quiet, toneless.* Thus, the readers understand that Ashley is almost dead himself as he cannot find the place in the new world.

109. Even if it is a lie,said Ashley *and his voice was* deadly quiet, it is not something which can be discussed."

As to Rhett's voice, it is described with the adjectives *swift, rough, gentle, a little thick.* Rhett is embodiment of manhood and his voice signals it.

110. You little fool,*and his voice was* swift and rough.

In the face of carrying its literal meaning is found among the 4-grams.

111. Nor was there any more interest in it than *in the face of* a man watching the last act of a none-tooamusing comedy.

With the help of comparison, the narrator shows that Rhett has lost interest in Scarlett.

It is worth noting that among 4-grams, we found *as if* clusters, which was not the case with 5-grams.*As if* clusters proved to be frequent among the 5-grams in the Dickens' Corpus, but they turned out to be recurrent only among the novel's 4-grams. These 4-grams signifying comparisons are *as if she had* and parallel constructions*as if he were* and *as if she were*. As Mahlberg points out,*as if* clusters often appear with the function words and that is also the case with *as if* 4-grams from *Gone with the Wind* (Section 2.2.3).Besides, Mahlberg underlines that the narrator"draws the reader's attention to features of the fictional characters"(2012:128 cited in Section 2.2.3).Consequently, *as if* 4-grams found in the novel refer to male and female characters, their experience or how others perceive them:

112.She had left not only that graceful dwelling but also the entire civilization that was behind the building of it, and she found herself in a world that was as strange and different *as if she had* crossed a continent. (describing Ellen's feelings when she moved from Savannah to Tara)

113. At first, Scarlett spoke the word *as if she had* never heard it before and had no comprehension of its meaning and then: The Klan!

114. But they treated her deferentially, *as if she were* old and finished, and their constant chatter of dances and beaux made her both envious of their pleasures and resentful that her widowhood barred her from such activities. (attitude of unmarried girls to Scarlett)

114.His dress was as debonaire*as if he were* going to a ball, well-tailored white linen coat and trousers, embroidered gray watered-silk waistcoat and a hint of ruffle on his shirt bosom.(in reference to Rhett's dress)

115. Moreover, he looked *as if he were* ashamed to be working convicts and he had little to say to her these days. (speaking of Ashley)

In addition to the groups presented by Mahlberg, there is one group that can be introduced for a more detailed classification of 4-grams found in the novel *Gone with the Wind* that is *war* clusters. Two 4-grams belong to the group: *The Yankees are coming* and *the war was over*. I decided to add this group because it reflects one of the main themes in the novel – the Civil War. As these 4-grams are recurrent, they create the novel's fictional world. The meanings of 4-grams *The Yankees are coming* and *the war was over* are fairly transparent. The Yankees were the enemies of the Southerners, the ones who initiated the war, therefore, Southerners were always worried where the Yankees' troops were located and the phrase *The Yankees are coming* is the admonition stating that the troops are really close to the town.

The 4-gram *the war was over* is related to the expectations, relief, but also anger. At first Southerners thought that they would win this war. The 4-gram is encountered as a part of the 6-gram *One more victory and the war* (Section 3.3). Besides, they hoped that everything would eventually change for the better when the war was over (Example 116):

116. When *the war was over*, everything would be settled, somehow. Somehow it would come out all right when *the war was over*.

Afterwards, when the war is finished, Scarlett realizes that Ashley will finally come home. Nevertheless, Scarlett realizes that the Yankees have a lot of power; therefore they can harm her.

117. *The war was over*, peace had been declared, but the Yankees could still rob her, they could still starve her, they could still drive her from her house.

The analysis of 4-grams has shown that all the functional groups proposed by Mahlberg in the discussion of 5-grams in the Dickens's corpus are applicable to 4-grams in the novel *Gone*

with the Wind. Speech clusters dominate among the 4-grams, while labels take second place. In contrast to 5-grams, where time and place clusters are more frequent than other clusters, in the case of 4-grams other clusters slightly prevail over time and place clusters. In contrast to the novel's 5-grams, *as if* clusters and body part clusters are recurrent among the 4-grams. In addition to Mahlberg's functional categories, the group called *war* clusters is represented as a separate type of the 4-grams. The study of 4-grams in the novels is followed by an analysis of the 3-grams.

3.6. 3-GramsPresent in the Novel *Gone with the Wind*

3-grams found in *Gone with the Wind* prove to be far more numerous than 4-, 5-, 6-, 7- and 8-grams; therefore, the minimum frequency threshold is set to 25. *AntConc* identified 331 3-grams, but only those having a lexical word in them will be discussed in this section due to the limited scope of thesis. Still, it is worth noting that the 3-gram *out of the* appears on the fifth position with 115 occurrences. This 3-gram is the most frequent three-word cluster (1,210 occurrences) in the Dickens Corpus compiled by Mahlberg as stated in Section 2.1. 38 3-grams are shown in Table with descending frequency. All of them are grouped according to the functional groups they belong to. It is worth noting that some 3-grams like *the first time, for a moment* are deliberately omitted as they have been analysed in the previous sections when dealing with longer N-grams.

Table 3.8 3-Grams in*Gone with the Wind*

Labels (L)	Body Part (BP)	Time and place (TP)	Other (d)
in the world (72)	the sight of (57)	of the house (47)	the sound of (57)
she thought of (57)	in his eyes (45)	in the house (41)	the thought of (47)
in love with (33)	in her eyes (31)	on the front (35)	looked at her (37)
the old days (34)	in the face (28)	in the County (34)	take care of (34)
in her heart (29)	in his voice (27)	the dining room (31)	the fact that (33)
War (W)		up the stairs (30)	came to her (32)

the Yankees were (34)		in the hall (29)	to tell you (28)
Of the Confederacy (29)		on the floor (27)	to her feet (28)
of the war (29)		of the town (25)	to do with (27)
			in the dark (26)
			of her own (26)
			and tried to (26)

3.6.1 Labels

Labels is the first functional group to be discussed in this section. The 3-gram *in the world* does not carry much meaning itself, but its contexts reveal that it is preceded by the superlatives with *land* being the head noun as seen in Examples 118-119.

118. It was a savagely red land, blood-colored after rains, brick dust in droughts, the best cotton land *in the world.*
119. I offer you the most beautiful land *in the world*-saving County Meath in the Old Country-and what do you do? You sniff!"
120. The honeysuckle which draped the gullied red sides of the road in tangled greenery was piercingly fragrant as always after rain, the sweetest perfume *in the world*.
121. Nowhere else *in the world* was there land like this.

As already underlined in Section 3.4.1, *land* proves to be one of the most important themes in the novel. Apart from her Tara, there is also a material thing that is of great importance for Scarlett that is money. Quantifiers *not enough* and *everything* are used together with the 3-gram (Examples 123-124).

122. I've found out that money is the most important thing *in the world* and, as God is my witness, I don't ever intend to be without it again.
123. When I wake up from that dream, it seems like there's not enough money *in the world* to keep me from being afraid of being hungry again.

51

124. She had thought, half an hour ago, that she had lost everything *in the world*, except money, everything that made life desirable, Ellen, Gerald, Bonnie, Mammy, Melanie and Ashley. She had to lose them all to realize that she loved Rhett-loved him because he was strong and unscrupulous, passionate and earthy, like herself.

Example 123 tells about Scarlett's nightmare where she is back to Tara, her mother has died, her father, sisters and the servants are hungry and she is hungry herself so she feels dreadful and runs somewhere. Afterwards when she wakes up, she fears she will be hungry again. Therefore Scarlett promises herself that she will never starve again and money can give her everything she wants. As seen from Example 124, money is equally important for her as her parents, daughter and friends. Still, due to the loss of loved ones, it dawns upon her that she loves Rhett, but this realization comes too late. There is another evidence of Scarlett's unawareness of what others feel:

125. She wondered forlornly if she had ever really understood anyone *in the world*.

Speaking of other characters in the novel, Ashley's despair is reflected through the question:

126. But where do I fit *in the world* anymore?

He does not know where he belongs, the war has taken everything he loved and devastated him; thus, there is no place for him in the new world. The 3-gram is also connected to Melanie: examples 127 and 128 show that her son Beau and her husband Ashley are the most important people for her and she considers this to be every woman's deepest dream:

127. There's nothing *in the world* more precious to me than Beau except Ash-except Mr. Wilkes.
128. I have everything *in the world* any woman could want."

Another 3-gram that proves to be a label is *she thought of.* It is mainly a narration of Scarlett's thoughts. According to the right collocate, Scarlett thinks of *Ashley*, also of *all the things she had intended to say to him* and *all the unsaid things that lay between them.* Besides, Scarlett thinks of Ellen, Gerald and Tara. Still Rhett is the one she thinks of most. He gives her presents:

129. When she thought of hats, she thought of Rhett Butler.

Furthermore, when Scarlett needs help, she approaches him:

130.Suddenly she thought of Rhett Butler and calm dispelled her fears. Why hadn't *she thought of* him this morning when she had been tearing about like a chicken with its head off? She hated him, but he was strong and smart and he wasn't afraid of the Yankees.

Another frequent 3-gram is *the old days.* As discussed before, Ashley belongs to the old guard and he longs for old times. The examples 131-132, 134 show how the life in the South before the war was.

131. But the South was aflame with war, events roared along as swiftly as if carried by a mighty wind and the slow tempo of *the old days* was gone.

132.The lazy luxury of *the old days* was gone, never to return.

133.I won't be a big-mouthed fool, she thought grimly. Let others break their hearts over *the old days* and the men who'll never come back.

134. She was thinking with a leaden heart that in burying Gerald she was burying one of last links that joined her to *the old days* of happiness and irresponsibility.

135.Mammy, the last link with *the old days*.

Thus, slow tempo, luxury, happiness and irresponsibility belong to the old days for Southerners. Example 133 alludes to Scarlett's attitude to old days, for her they are gone, but a new life awaits and she gets excited about it. However, in the end of the novel she changes her mind, she herself wants to return to the old days; therefore, she decides to go to Tara, to her childhood home, where Mammy can take care of her as she is "the last link with the old days" (Example 135).

Three 3-grams *the Yankees were, of the Confederacy* and *of the war* fit into the *war* category. The 3-gram *the Yankees were* goes together with description how the Yankees appear to the Southerners and also it serves as a part of predicative implying their battle action. Firstly, Southerners believe that they are *cowards,* but afterwards they realize that they prove to be *good fighters*. *Of the Confederacy* are used together with the nouns denoting the military: *armies, stars, VicePresident, the leaders, a soldier.* Nevertheless, it is also personified through the nouns like *the fate, the Spirit, the heart, the very life, the fortunes, the fading hopes.* The Confederacy stands for the Southern states, thus, representing people – the Southerners – their fate, hopes, spirit, life and fortunes. The context of the next 3-gram *of the war* mostly provides the time frame for the war: *At the onset of the war, the first year of the war, The end of the war,the last*

days of the war, in the early days of the war, the last days of the war, etc. In addition, the 3-gram reveals how Scarlett perceives the war: *the full horror of the war, the terrors of the war, the cynical coolness of the war days.*

The 3-gram that is also associated with Scarlett is *in her heart.* Scarlett experiences different emotions: *the pain, a tingling, desolation, a swell of pride, a savage envy, fear, the empty spaces, The old irritation and antagonism, hatred of Rhett, the ache,The word lost-lost-lost dinged frighteningly in her heart, the triumph, the fury of love*. The nouns that carry negative connotations prevail over the ones that imply positive feelings, that makes eighteen vs. six. Judging from the negative nouns, Scarlett proves to suffer a lot and only when she realizes that she loves Rhett, she feels happy:

136.Suddenly she felt strong and happy. She was not afraid of the darkness or the fog and she knew with a singing in *her heart* that she would never fear them again.

3.6.2 Body Part Clusters

The second group that is presented among the functional groups is body part clusters, which were not prominent in the case of longer N-grams. The most frequent body part 3-gram with 57 occurrences is *the sight of.* Though this 3-gram does not have a body part noun, which is a prerequisite in Mahlberg's classification, it embodies Scarlett's vision of things, other characters and her emotions about them. Here come the examples of body language depicted with the help of *the sight of* and their interpretations.

137. It was a friendly and comfortable room and, ordinarily, Scarlett loved the quiet hours which the family spent there after supper; but tonight she hated *the sight of* it and, if she had not feared her father's loudly bawled questions, she would have slipped away, down the dark hall to Ellen's little office and cried out her sorrow on the old sofa.

Example 137 points to Scarlett's emotional state: she hates the room where they usually gather with her parents and sisters. The hatred is caused by the knowledge that Ashley is going to announce his engagement to Melanie the following day.

138. It was *the sight of* her mother's serene face upturned to the throne of God and His saints and angels, praying for blessings on those whom she loved.

Example 138 shows Scarlett's trust and belief in her mother Ellen. Taking into account the context where this sentence appears, it becomes clear that Ellen's prayers sooth Scarlett's anxiety about losing her Ashley. Besides, Scarlett believes that Heaven hears when exactly Ellen is praying.

> 139.Her spirits rose, as always at *the sight of* her white skin and slanting green eyes, and she smiled to bring out her dimples.

Here Scarlett's self-admiration is obvious. She loves herself, being aware how her appearance helps her impress men.

> 140. At *the sight of* them, three times a day, her memory would rush back to the old days, the meals of the old days, the candle-lit table and the food perfuming the air.

This example reflects Scarlett's hunger, her longing of abundant food at Tara. Though the war is going on, it seems distant, whereas a food shortage depresses Scarlett and only then she remembers the old days.

> 141. Her heart contracted a little with shame at *the sight of* that closed door, remembering the many nights of this last summer when Rhett had sat there alone, drinking until he was sodden and Pork came to urge him to bed.

Again Scarlett's emotions are revealed in this example: she realizes that she is guilty as in her blind love for Ashley she neglected Rhett.

The second most frequent body part 3-gram is *in his eyes* used 45 times. The 3-gram usually goes together with the noun *look* and then it characterizes Ashley's emotional state after the war, for example,

> "and that look of sadness and despair would be *in his eyes*; a look *in his eyes* which eluded her-not fear, not apology, but the bracing to a strain which was inevitable and overwhelming; *In his eyes*, there was a fagged, haunted look; a far-off look in his eye; a lost-dog look *in his eyes*; a queer dead look *in his eyes*."

The 3-gram also contributes to the description of Rhett, especially his facial expression when he is dealing with Scarlett:

142. There was contempt in his manner as *in his eyes*, contempt overlaid with an air of courtesy that somehow burlesqued their own manners.

Several occurrences of the body part 3-gram reveal Rhett's emotions:

"a warm dancing malice in *his eyes*; a diabolic gleam *in his eyes* as if something amused him greatly; anger flickering *in his eyes*; the glint *in his eyes*; a cold glitter *in his eyes*; that old, puzzling, watchful glint in *his eyes*-keen, eager as though he hung on her next words; a malicious light danced *in his eyes*; a savage bright glitter in his eyes."

Nevertheless, even Aunt Pitty and Scarlett notice that Rhett behaves differently towards Melanie than to other women, he highly respects her as seen from Example 143:

143. There was never that cool look of appraisal, never mockery *in his eyes*, when he looked at Melanie; and there was an especial note in his voice when he spoke to her, courteous, respectful, anxious to be of service.

There are instances where other characters are reflected through the 3-gram *in his eyes:* Gerald – "his Irishman's passion for horses shining *in his eyes*" andCharles – "his whole clean simple heart *in his eyes*".

The third most frequent body part cluster is *in her eyes.* Eyes reflect not only different emotions like *disgust,* sadness (*tears), exaltation, understanding, complacency, a fighting light, venom, the hate* and *impatience,* but also appearance:

144. She would have been a strikingly beautiful woman had there been any glow *in her eyes*, any responsive warmth in her smile or any spontaneity in her voice that fell with gentle melody on the ears of her family and her servants.

As seen in Example 144, Scarlett is not a person with a warm personality. Example 145, in turn, shows that Melanie's face becomes beautiful, when she proudly speaks of soldiers:

145.There was a deep, almost fanatic glow *in her eyes* that for a moment lit up her plain little face and made it beautiful.

Turning back to the emotions, most of them prove to be negative. Melanie has *tears in her eyes,* but she is never ascribed having disgust, hate, venom, whereas Scarlett can feel this:

146. There was venom *in her eyes* as she answered, venom that was too unmistakable to be missed, and the smile went from his face.

147. He took a quick step toward her as though to put a hand on her arm but she twisted away from him, and before the hate *in her eyes* his face hardened.

The poverty is again visible through the 3-gram:

148. The boys followed her to the wagon and assisted her in, courtly for all their rags, gay with the volatile Fontaine gaiety, but with the picture of their destitution *in her eyes*, she shivered as she drove away from Mimosa.

Scarlett is weary of poverty as all she wants is a prosperous life.

As to the 3-gram *in the face,* it functions as a body part cluster when co-occurring with the verb of "seeing":

149. When she looked God *in the face* on the Day of Judgment and read her sentence in His eyes, it would not be as bad as this.

Scarlett is conscience-striken as she is afraid that Melanie got to know about her feelings for Ashley. It is a turning point as Scarlett realizes that she cares what Melanie thinks of her and that Melanie truly loves her and proves to be her best friend.

Example 150 shows a peculiarity of Scarlett's character, her ability to go forward despite everything:

150. With the spirit of her people who would not know defeat, even when it stared them *in the face*, she raised her chin.

Speaking of Rhett, Examples 151-152 reflect his nature.

151. They both see the same unpleasant truth, but Rhett likes to look it in the face and enrage people by talking about it-and Ashley can hardly bear to face it.

In addition, he is rebellious enough to patronize bad women and have intercourse with Belle Watling:

152. And to think that Rhett would have the gall to look a decent woman *in the face* after being with that creature!

For instance, Melanie is scared that Uncle Peter can tell Aunt Pitty that she talked to Belle Watling and took money from her:

153. It will kill Auntie if she knows I ever even looked that woman *in the face*.

Scarlett also notices that her father Gerald has completely changed after Ellen's death:

154. *In the face* which she could only see dimly, there was none of the virility, the restless vitality of Gerald, and the eyes that looked into hers had almost the same fear-stunned look that lay in little Wade's eyes.

Regarding the 3-gram *in his voice,* it is associated with the main characters Ashley (Examples 155-158) and Rhett (Example 159-160, 162).

155. He went on speaking and there was a quality *in his voice*, a sadness, a resignation, that increased her fear until every vestige of anger and disappointment was blotted out.

156. The edge *in his voice* spoke of a raw, unhealed wound that ached within him and his words brought shame to her eyes.

157. Oh, Ashley! she cried despairingly, tears in her eyes at the jeering note *in his voice*.

158. There was music *in his voice*, the music of fiddles and banjos to which they had danced in the white house that was no more.

Ashley is crushed after having participated in military operations (Example 155). He asks Scarlett to take care of Melanie in case he dies. He resigns as he knows that the Southerners are going to lose the war. Example 156 brings out Ashley's guilt of letting Scarlett help him and his family instead of doing it himself. In addition, Scarlett's action makes Ashley think of his life. He does not consider himself worthy therefore he wants to go to New York, but Scarlett prevents him from doing this and makes him work for her in the mills. By contrast, Ashley livens up when he imagines himself in old days (Example 158).

159. Something *in his voice* made her heart beat pleasantly faster and she felt her face flush.

Scarlett anticipates that Rhett is in love with her and she also feels that she likes him, but tries to persuade herself that she only loves Ashley (Example 159). Furthermore, Rhett himself hides his feelings for Scarlett as he is afraid that Scarlett would hurt him. Rhett's voice is also characterized having *a malicious tenderness* when he decides to leave Scarlett and Melanie and join the troops; *a suave, almost teasing note, suave brutality, a vibrant note almost like anger, barely suppressed savagery* when he realizes that Scarlett wants money from him not only for herself, but also for Ashley.

160. At the faint echo of passion *in his voice*, pleasure and excitement crept back into her. Rhett confesses his love for Scarlett and explains the reason why kept silence so long:

161. I loved you but I couldn't let you know it. You're so brutal to those who love you, Scarlett. You take their love and hold it over their heads like a whip."

All in all, Rhett's love ceases and he wants to return to his roots, but without Scarlett:

162. No, he said, and there was finality *in his voice*.

3.6.3 Time and place clusters, and *Other*

The third group that is relevant in the discussion of the 3-grams is *time and place clusters.* They refer to the house – *to the front of the house, the back of the house, in the house, on the front steps, in the dining room, up the stairs, in the hall* and *on the floor.*

As to the functional group named *Other,* it comprises several 3-grams that contribute to the narrative. For instance, *looked at her, in the County, came to her, leaped to her feet, in the dark, and tried to.* Moreover, among *Other* 3-grams are encountered the 3-grams *the sound of, the thought of* that clearly correspond to the pattern *the N of,* introduced by Stubbs and Barth as described in Subchapter 1.2, as well as of-phrases signifying belonging – *of the town, of her own.* Fixed expressions are also present here – *take care of, the fact that, to tell you, to do with.*

To sum up, the 3-grams are too short to highlight stylistic features, but their contexts provide deeper insights into the novels: additional information about the characters and novels' themes. The next subchapter, by contrast, is devoted to the analysis of the keywords, consisting of just one word, that appear to be particularly frequent in the novel.

3.7Keywords Present in *Gone with the Wind*

The keywords facility in *WordSmithTools* returned 500 keywords in *Gone with the Wind* in comparison with the Brown corpus. As stated in the theory chapter, Mahlberg and McIntyre (2011) conclude that keywords represent the names of characters, the words that signify "aboutness" of the text and grammatical words relating to the author's style, thus, revealing the fictional world and thematic signals (Section 2.3). This proves to be the case for the keyword analysis of *Gone with the Wind.* The most frequent keywords are the pronouns *she* and *her* with the character names *Scarlett, Melanie, Ashley* and *Rhett* included in the top 10 most frequent keywords. "Aboutness" of the text is seen through the keywords *Tara, Yankees, money, darkies, black, negroes, Carpetbaggers, klan, Scallawag, blockade, etc.* They belong to both the fictional world and thematic signals as they refer to the events of the novel: the American Civil War. Table provides a comprehensive overview of fictional world and thematic signal keywords.

Table 3.9 Keywordsin *Gone with the Wind*

Categories:	Keywords:
Fictional world: *Characters:* Names	Scarlett, Melanie, Ashley, Rhett, Mammy, Gerald, Pitty, Ellen, Frank, Wade, Suellen, Meade, Prissy, Merriwether, Pork, Bonnie, Archie, Carreen, Belle, India, Beau, Fontaine, Uncle, Dilcey, Hugh, Brent, Cathleen, Honey, Tony, Charles, Ella, Stuart, Fanny, Maybelle, Robillard, Cade,Wilkerson, Rene, Emmie, Gallegher, Munroe
Body part nouns	eyes, face, heart, voice, mind, hands, arm, head, lips, shoulders, feet, ears, mouth, hair, lap, cheeks, bosom
Clothes and accessories	skirts, bonnet, dress, velvet, lace, handkerchief, boots, hat
Language	ain, de, dat, dey, lak, doan, Mist, wuz, wid, whut, gwine, fer, dem, bout, yas, din, mah, git, jes, en, dar, dis, nuthin, kin, chile, ze, fo, kain, ef, tis, den, Gawd, Miz, sho, tek, ole
Society	Yankees, ladies, old, folks, darkies, negroes, gentlemen, Carpetbaggers, niggers, convicts, trash, white, Scallawag
Emotions	frightened, tears, love, hurt, smile, afraid, crying, rage, pride, hate, dreadful, embarrassed, laughter, happy, fright, excitement, contempt, bewildered, indignation, annoyed, pleasure,
Emotions	upset, damned, sobbing, jealous, mad

Settings and props Places Vehicles and buildings	 Tara, Atlanta, Jonesboro, town, Oaks, Charleston, Georgia, Peachtree, Macon, Savannah, Fayetteville, horse, carriage, house, mill, buggy, wagon, yard, mules, reins
Thematic signals: American Civil War	 Yankees, darkies, Confederacy, Carpetbaggers, niggers, war, darky, nigger, Klan, Sherman, army, burned, Scallawag, Bullock, fighting, blockade Tara, money, gone, love, tomorrow, hungry, lost

As outlined in Table 3.9, there are several groups that can be distinguished. This classification was proposed by Mahlberg and McIntyre (2011) in the analysis of Fleming's novel *Casino Royale*(Section 2.3)*.* First of all, several keywords convey information about the characters: their names, body parts, clothes and accessories, language, the society they are involved in, and the emotions they experience. The names, body parts and clothes are predictable to be prominent among the keywords. By contrast, the fictional world keywords implying language and society are of special interest in the thesis. The mentioned keywords reflect a vernacular language spoken by the slaves or *darkies* the way they are referred to in *Gone with the Wind.* There is a clear stylistic distinction between the main characters, the ones who come from rich families and own cotton plantations and the slaves who work on these plantations. The main characters speak standard modern English, whereas the slaves use the language of uneducated people, for instance, they say *de, gwine, mah, kain, tek* instead of respective *the, going,my, cannot, take.* These forms were not frequent among longer N-grams. The *American Heritage Dictionary of the English Language* confirms that the etymology of such words is African American Vernacular English and they are chiefly spoken in the Southern and South Midland U.S. One of the main characters who speaks vernacular English is Scarlett's nanny Mammy. It is important to stress that African American words come up as keywords not necessarily because they are highly frequent, but due to their absence in the reference corpus.

As mentioned above, society, the layers of it in particular become visible through the keywords. The derogatory nouns *darkies, negroes*, *niggers,* that became taboos in present-day

English, designate the slaves. Thus, a disparaging attitude of the rich towards the slaves is apparent: they are merely seen as the workforce. In addition, there are people called *white trash.*

> 163. There was much about the South-and Southerners-that he would never comprehend: but, with the wholeheartedness that was his nature, he adopted its ideas and customs, as he understood them, for his own-poker and horse racing, red-hot politics and the code duello, States' Rights and damnation to all Yankees, slavery and King Cotton, contempt for *white trash* and exaggerated courtesy to women

The example above emphasizes the characteristics of the Southerners: they disliked Yankees, used slaves, grew cotton, disdained white trash and were too courteous to women. For that reason, the repetition of such words as *Yankees, cotton, white trash, gentlemen* and *ladies* is expected. Apart from *white trash* there are also *Carpetbaggers, Scallawags*and *convicts*. The *Merriam-Webster's Online Dictionary* gives a comprehensive definition for the term *Carpetbagger* – "a Northerner in the South after the American Civil War usually seeking private gain under the reconstruction governments." *Scallawag*, in turn, is "a White Southerner who supported the Republican Party and its policy of Black emancipation" (The *Collins English Dictionary*). Rhett acknowledges that he is a Scallawag:

> 164. I'm a renegade, a turncoat, a *Scallawag*.

Besides, as Scarlett married Rhett, she is also considered to be a Scallawag:

> 165. And, if the marriage went through, Scarlett also would be turning *Scallawag*.

Speaking of *convicts,* they were given freedom to fight against the Southerners. Scarlett, for example, leases convicts to work for her mills when the war is over.

Regarding keywords expressing emotions, these are adjectives and nouns like *afraid, happy, pride, hate* and *bewildered.* Most of them have negative connotations: nineteen are negatively charged words, whereas only five keywords are positive. Due to this, the mood created in the novel appears to be gloomy. All in all, the keywords contribute to the fictional world of *Gone with the Wind,* but there are also the keywords that highlight thematic signals.

The thematic signals that prevail among the keywords can be grouped into two main categories: the American Civil War and other novel's themes. The events in the novel take place during the American Civil War when the Southerners fight against the Yankees; thus, such keywords as *Confederacy, Klan, Sherman, Bullock* signal this. In addition, there are common nouns that belong to the semantic field of *war: war, army, fighting, blockade.* As to other thematic signals,

they have already come up as the parts of longer N-grams discussed in the previous sections and the fact that they are prominent as the keywords only underlines their significance in the novel. These are striking metaphors that assist in portraying the main characters:

> 166. I'll think of it all *tomorrow*, at *Tara*. I can stand it then. *Tomorrow*, I'll think of some way to get him back. After all, *tomorrow* is another day."

Though the keyword analysis returns a lot of predictable information about the novel, it points out essential features that are not dominant through the N-grams. In the case of *Gone with the Wind* the forms of a vernacular language spoken by the slaves are conspicuous among the keywords; besides, society described in the novel also becomes more noticeable. These keywords together with the keywords reflecting characters' names, their clothes and emotions form the fictional world of the novel. Additionally, several of them function as thematic signals as they convey the novel's themes. The reason why some keywords are not identified as parts of the N-grams can be as follows: one word tends to occur more often that a combination of two, three, four or five words. Besides, a keyword need not be extremely frequent in the corpus as long as it is used significantly more than in the reference corpus.

The next chapter is devoted to the analysis of N-grams found in the novel *A Farewell to Arms*. Just like with *Gone with the Wind* 8-grams will be first analysed in *A Farewell to Arms* and further shorter ones in the descending order will be examined, concluding with the keywords.

4. Corpus Stylistic Analysis of Ernest Hemingway's Novel *A Farewell to Arms*

Ernest Hemingway's novel *A Farewell to Arms* tells about the American lieutenant Frederic Henry who participates in the Italian campaign during the First World War. Through his friend the lieutenant Rinaldi he gets acquainted with a British nurse, Catherine Barkley, who later becomes his lover. After some time, there is to be an attack up at a place up the river, therefore Henry with the mechanics takes four cars there and they wait in the shelter. A mortar shell reaches their dugout and one of the mechanics, Manila, is killed, but Henry is wounded in the knee. He is sent to Milan to undergo surgery and medical treatment there. Catherine also comes to the hospital and takes care of Henry till he is completely recovered from his injury. They consider themselves married and Henry wants Catherine to be his wife, but she objects as then she would need to leave him. Before Henry goes to the front, he gets to know that Catherine is three months pregnant. Soon the Italian army retreats as Austrian troops break through. The car Henry is driving in is stuck in a muddy road and asks two sergeants to help to push the car, but they refuse and try to escape, Henry shoots and kills one of them. Without the car Henry and mechanics continue the retreat on foot. The battle police stop them as they suspect them to be Germans. Henry is to be executed, but he manages to save himself. After all, he is fed up with the war therefore he becomes a deserter. Then he reunites with Catherine and they stay in the hotel. The barman warns Henry that he is going to be arrested the next morning. For that reason, Catherine and Henry decide to go to Switzerland. After a while Catherine goes into labour: she gets a Caesarian section and delivers a dead boy, later she dies of a hemorrhage. Henry goes back to the hotel.

4.1 8-Grams Present in the Novel *A Farewell to Arms*

The corpus stylistic analysis of the novel *A Farewell to Arms* begins with the 8-grams. In contrast to the novel *Gone with the Wind,* there were three 8-grams with a minimum frequency of four as outlined in Table 4.1.

Table 4.18-Grams in *A Farewell to Arms*

Rank	Frequency	8-gram
1	4	I went to the window and looked out
2	4	what if she should die She can t
3	4	what if she should die She won t

Judging from the contexts of these 8-grams, it becomes clear that they point to the main character – Lieutenant Frederic Henry and his narration, signifying that it is a first-person narrator who tells the story. The first 8-gram describes Frederic's actions and it is represented by a complete clause. Only once it occurs with a co-ordinated clause:

1. *I went to the window and looked out* but could not see across the road.

The remaining 8-grams display Frederic's concerns about Catherine's condition. These are parallel constructions with the first part being implemented by the conditional*what if she should die* in addition to the reply represented by the third-person pronoun she referring to Catherine and the modals *can't, won't.* Frederic is afraid of Catherine's death and when the question *what if she should die* pops up in his mind, he is strongly reluctant to believe it. For Frederic Catherine simply cannot die and leave him alone – the continuations of the 8-grams run as follows – *She can't die, She won't die.* The second reply *She won't die* signals that Frederic tries to reassure himself that Catherine is not going to die.

2. But *what if she should die? She can't die*.
3. Yes, but *what if she should die? She won't die*.

All in all, the recurrence of the conditionals proves to be significant in the novel as it reveals Frederic's attitude to Catherine and also pinpoints the novel's climax.

4.2 7-grams Present in the Novel *A Farewell to Arms*

Turning to 7-grams found in the novel, they prove to be the reflections of the mentioned 8-grams, only the cut-off line for them to be 7-grams appears in different positions:

Table 4.2 7-Grams in *A Farewell to Arms*

Rank	Frequency	8-gram
1	4	But what if she should die She

2	4	I went to the window and looked
3	4	if she should die She can t
4	4	if she should die She won t
5	4	went to the window and looked out
6	4	what if she should die She can
7	4	what if she should die She won

As their contexts and meanings have been analysed above, there is no need to discuss them further. The next step is to look at the 6-grams.

4.3 6-Grams Present in the Novel *A Farewell to Arms*

The search for 6-grams with minimum frequency of 5 returned just four results; therefore, 6-grams with minimum frequency of 4 were also taken into account. Thus, *AntConc* identified fourteen 6-grams. Only four of them did not appear as parts of the above 7- and 8-grams; they are:

Table 4.3 6-Grams in *A Farewell to Arms*

Rank	Frequency	N-gram
1	6	on the other side of the
2	6	What's the matter with you
3	4	I don t want you to
4	4	to go back to the front

The most frequent 6-gram *on the other side of the* is a prepositional phrase that signifies location: *room, tracks, lake, mountain* and *table.* This phrase does not reveal much about the novel *A Farewell to Arms* as it used often used to describe the position of something.

4. As I lay on the bed I could see the big mirror *on the other side of the* room but could not see what it reflected.

5. The mountains were sharp and steep *on the other side of th*e lake and down at the end of the lake was the plain of the Rhone Valley flat between the two ranges of mountains; and up the valley where the mountains cut it off was the Dent du Midi.

6. The nurse was on one side and the doctor stood *on the other side of the* table beside some cylinders.

The 6-gram *What's the matter with you* is "a routine formulae." Sabine Fiedler explains that

Routine formulae (also called 'pragmatic idioms' [Roos 2001: 70] and 'functional idioms' [Cowie et al. 1983: xvii]) are conventionalized utterances that the language puts at our disposal as ready-made units for recurrent situations (Fiedler 2007: 50).

The phrase is used to inquire whether there is something wrong with the person. It is mainly uttered by the main character Frederic, but also the major and Rinaldi.

7. *What's the matter with you*?This war is killing me,Rinaldi said, I am very depressed by it.

Thus, the wrecking impact of the war becomes striking through Rinaldi's answer.

The 6-gram *I don't want you to* is associated with Catherine. She does not want Frederic to have a higher rank (Example 8); besides, she does not want Frederic to go away as her love for him is so strong that she wants to see him every minute as he is the meaning of her life (Example 9). Frederic does not want Catherine to speak nonsense. However, her words that precede Frederic's objection prove to be fatal. Catherine tells Frederic that she sees him and herself dead in the rain; therefore she is afraid of rain. Further when she claims that she can keep him from danger, he asks her to stop (Example 10).

8. *I don't want you to* have any more rank.
9. *I don't want you* to go away.
10. *I don't want you to* get Scotch and crazy tonight.

The infinitive*to go back to the front* characterizes Lieutenant Henry's and a nurse's awareness that he would sooner or later come back to the front despite his unwillingness. The 6-gram highlights one of the novel's themes.

11. Unless you find something else I'm afraid you will have *to go back to the front* when you are through with your jaundice.

These are the words of the nurse Miss Van Campen that provoke Henry. He asks her whether she had jaundice herself or seen patients enjoying it. Then he enquires again, using the blunt, vulgar language:

> Miss Van Campen, I said, did you ever know a man who tried to disable himself by kicking himself in the scrotum?

After a while, she answers that that she saw a lot of men who wounded themselves in order to escape the front. Her statement alludes to the fact that many men fear the war and Henry is one of them.

The analysis of 8-, 7- and 6-grams has shown that few long N-grams are encountered in the novel and their frequency is low. Besides, the 7-grams identified by *AntConc* prove to be the reflections of the 8-grams. The 8-grams are exclusive for the novel as they characterize Frederic and Catherine. The 6-grams are also associated with the novel when their contexts are explored. However, one 6-gram is a positional phrase that describes the novel's settings. The 5-grams are to be examined next. Mahlberg's functional groups will be applied to the extent that they define the types of 5-grams.

4.4 5-Grams Present in the Novel *A Farewell to Arms*

In contrast to 6-grams, 5-grams used at least five times were relatively common in the novel *A Farewell to Arms.* Seventeen 5-grams were identified with several of them previously discussed when dealing with 8-, 7-, and 6-grams. The 5-grams will be presented under the respective functional categories they belong to in Table 4.4. The distribution of 5-gram types is shown in Table 4.5.

Table 4.4 5-Grams in *A Farewell to Arms*

Labels (L)	Speech (S)	Time and Place (TP)	Other (d)
What if she should die (11)	I don't want to (15)	on the other side of (7)	looked out of the window (5)
don't let her die (8)	I don't know I (10)	on the side of the (7)	
I went back to the (7)	What's the matter with (9)	the side of the road (6)	

I did not say anything (6)	I don't think so (6)	on the floor of the (5)	
I did not want to (6)	I don't want you (5)		
to go back to the (5)	if you don't want (5)		

Table 4.5 Distribution of 5-gram types in *A Farewell to Arms*

Labels (L)	Speech (S)	Time and place (TP)	Other (d)
6	6	4	1

As seen from Table 4.4 only four groups presented by Mahlberg in the classification of the 5-grams from Dickens' Corpus can be used to sort the 5-grams from Hemingway's novel. The number of labels and speech clusters is equal, namely they appear six times. Time and place clusters make the third biggest functional group. Just one 5-gram *looked out of the window* does not fit into any of the mentioned functional categories.

4.4.1 Labels

Labels point to the protagonist – Frederic. As already discussed, he worries about Catherine's health; therefore, 5-grams pinpointing this are frequently used. Other 5-grams, in turn, show Henry's actions. He keeps silent when he does not know what to say. The 5-gram *I did not say anything* is a short, complete sentence. When Catherine tells about her former beau, Frederic does not comment it (Example 12)

12. *I did not say anything*.

The 5-gram *I went back to the* is encountered with several nouns – *drivers, major's dugout, cars, kitchen, stern, papers, hospital.* The nouns can be placed under the semantic category *war.* As to the 5-gram *I did not want to,* Examples illustrate its occurrences in different contexts:

13. *I did not want to* go there anyway.
14. *I did not want to* see the bank.
15. *I did not want to* look conspicuous.

16. I had the paper but I did not read it because *I did not want to* read about the war.
17. If we were that close to the border *I did not want to* be hailed by a sentry along the road.

The 5-gram *to go back to the* has been already studied with regards to the 6-gram *to go back to the front.* Nevertheless, one occurrence has not come up before.

18. It was too soon *to go back to the* hospital.

4.4.2 Speech Clusters and Time and Place Clusters

Speech clustersrepresent negative utterances, signaling that characters do not want to perform something or they are unsure, or they do not believe. Again the verbs *want, know* and *think* are frequently repeated,which again confirms the propriety of Stubbs and Barth's (2003) claim (Section 2.4) that "vocabulary from the lexical fields of saying, looking, thinking and wanting" is characteristic of this novel as well.

19. *I don't want to* do anything more to you.
20. *I don't know I*'m going to get it.
21. *I don't believe it.*

The speech 5-grams depict atmosphere in the novel: events in the novel take place at World War I when people are sceptical and frustrated, and lose their hopes.

The 5-gram *What's the matter with* has been discussed as the formulae *What's the matter with you* (Section 4.3). Apart from the pronoun *you,* it also appears together with the noun phrase *this machine,* the pronoun *the others* and the noun *baby.*

22. *What's the matter* with the baby?I asked.

Example 15 alludes to the significant point in the novel. Frederic asks the doctor what is wrong with the baby and he gets to know that it never was alive.

The functional group time and place clustersreveals mostly 5-grams meaning location: *on the other side of, on the side of the, the side of the road* and *on the floor of the.* They follow Stubbs and Barth's pattern (2003) (Section 2.4) *PREP/ the N of the. On the other side of* has occurred as a part of the 6-gram *on the other side ofthe,* but it appears with the geographical name in one instance (Example 22). Here are the uses of the time and place 5-grams in the text:

23. Then *on the other side of*Isola Madre go with the wind.

70

24. She sat up *on the side of the* bed.

25. They had me at *the side of the road* now.

26. Lying *on the floor of the* flat-car with the guns beside me under the canvas I was wet, cold and very hungry.

To sum up, the 5-grams are more frequent and diverse than the 8-, 7- and 6-grams. Still only four of Mahlberg's groups can be used in the distribution of the 5-grams. Labels and speech clusters contain the same number of the 5-grams, followed by time and place clusters and *other* being the least prominent group. As seen above, short sentences carrying clear and concise information are characteristic of the novel's style. The corpus stylistic analysis of *A Farewell to Arms* is continued by the examination of the 4-grams.

4.5 4-Grams Present in the Novel *A Farewell to Arms*

The frequency threshold for 4-grams in *Gone with the Wind* has been set to ten; therefore, the same frequency threshold will be applied in the analysis of 4-grams found in *A Farewell to Arms* in spite of the shorter length of this novel. The query returned twenty-seven 4-grams which are found in the table below.

Table 4.64-Grams Present in the Novel *A Farewell to Arms*

Labels (L)	Speech (S)	Time and place (TP)
I could see the (16)	I don t know (52)	the side of the (17)
I'm going to (15)	I don t want (28)	on the other side (16)
I went to the (14)	I don t think (21)	in front of the (15)
I went back to (12)	What's the matter (21)	for a long time (11)
I looked at the (11)	don't want to (19)	the edge of the (10)
If she should die (11)	I don't believe (17)	
and went back to (10)	All right I said (11)	
I did not know (10)	don t have to (11)	
went back to the (10)	don t believe it (10)	

t let her die (10)	I don t care (10)	
	I have to go (10)	
	It's all right (10)	

Table 4.7 Distribution of 4-gram types in *A Farewell to Arms*

Labels (L)	Speech (S)	Time and place (TP)
10	12	5

Regarding the functional groups that describe 4-grams, these are *labels, speech clusters* and *time and place clusters*. Neither body part clusters clusters nor *other* clustersare relevant in the classification of the 4-grams.

4.5.1 Labels

Labels take the second position among the functional groups and they portray the main hero – Lieutenant Frederic Henry. The 4-grams that have not been touched upon are *I could see the, I'm going to, I went to the,and went back to,went back to the,I looked at the* and*I did not know*. The 4-gram *I could see the* reflects Henry's perception of the things and people around him.

27. *I could see the* snow falling.
28. *I could see*the priest was disappointed but he said, That's all right.
29. *I could see* the yellow on their hats and the yellow marks on their cape collars.

Other 4-grams like *I went to the, and went back to,went back to the,* and *I looked at the* mainly describe Henry's actions and progression in narrative. The 4-gram*I'm going to* can be ascribed to several characters as it highlights their intentions:

30. *I'm going to* be a captain.
31. *I'm going to* the front, I said.
32. *I'm going to* sleep on the major's bed, Bonello said.

72

However, this 4-gram also occurs with the verb *die* and it is used by Catherine. She is aware that she is going to die and her utterances anticipate the novel's denouement. The last 4-gram to be discussed here is *I did not know* that again describes Frederic's unawareness of the situation, e.g.

33. It was not a pleasant trip to the dressing room and *I did not know* until later that beds could be made with men in them.
34. Crossing the field, *I did not know* but that someone would fire on us from the trees near the farmhouse or from the farmhouse itself.
35. *I did not know* where we were and I wanted to get into the Swiss part of the lake.

Though the labels refer to Frederic, they do not focus on his personality; on the contrary, they reveal bare facts stating what he does.

4.5.2 Speech clusters and Time and Place Clusters

Speech clusters are predominant among other types of 4-grams. The five most frequent of them are *I don't know, I don't want,I don't think* and *don't want to*. The rest represent such phrases as *I don't believe,All right I said,I do' t care*and *I have to go*, thus, they comprise the personal pronoun *I*together with the contracted, negative auxiliaries and verbs, which is predictable as the novel is told by a first-person narrator and it contains a number of dialogues. Apart from the uses of the 4-gram *What's the matter* as a part of 6- and 5-gram, it goes often with address, for instance *father/darling/Catherine/Emilio*. The 4-gram is associated with Frederic and Catherine. When speaking to Frederic Catherine uses the noun *darling.*

With regards to *time and place clusters,* there are just five of them among the recurrent 4-grams in the novel. Four of them belong to place clusters and they point to the particular location like the side, front or edge. Frederic gives a detailed account of where he is and what happens around him. For instance, *the side of the house, road, timber, bed, mountain; on the other side of the room, of the tracks, of the lake; in front of the cathedral, window, wheels; the edge of the slope, bathtub, woods, etc.* The time cluster,*for a long time,*proves to be a set expression.

It is striking that only three functional groups can be applied in the description of 4-grams found in *A Farewell to Arms.* Thus, there is little variety among the 4-grams: they mostly describe the actions of the main hero based on his objective narration, the speech signals and positional phrases.

4.6 Analysis of 3-grams present in *A Farewell to Arms*

In contrast to 331 3-grams encountered 25 times in *Gone with the Wind, AntConc* found only 110 3-grams with 15 occurrences in *A Farewell to Arms.* As in case of 3-grams from *Gone with the Wind,* the 3-grams present in *A Farewell to Arms* will also be discussed if they contain a lexical word. The 3-gram *out of the* is ranked in the ninth place with 36 uses compared to 115 uses in *Gone with the Wind.* Turning back to the 3-grams that carry lexical meanings, there are twenty seven of them included in Table 4.8.

Table 4.8 3-Grams Present in the Novel *A Farewell to Arms*

Labels	Time and Place Clusters	Other
I saw the (30)	in the morning (29)	in the rain (35)
I went out (20)	all the time (24)	back to the (31)
I want to (18)	on the bed (24)	in the dark (24)
	in the room (21)	looked at me (22)
	to the hospital (20)	to the front (19)
	down the hall (20)	the major said (18)
	a little while (18)	I'm afraid (16)
	after a while (18)	shook his head (15)
	in the night (22)	
	on the floor (18)	
	the road and (18)	
	a long way (16)	
	on the road (16)	
	the door and (16)	
	in the mountains (15)	
	the window and (14)	

The functional groups that can be applied to the description of the 3-grams are labels, time and place clusters, and *other*. The 4-grams *to the front, the major said* and *shook his head*appear in the category *other*; however, they could be placed under the functional categories

war, body part clusters and speech clusters, but I decided not to establish categories only for one 3-gram.

4.6.1 Labels

Labels comprise the first-person pronoun and describe the main character Henry: *I saw the, I went out* and *I want to.* The surroundings and people are described through the 3-gram *I saw the,* e.g. *town with the hill and the old castle above it, sun coming through the window, river, doors, carabiniere start for me, motorman and conductor coming out of the station, etc.* Henry sees his newborn son, but when the nurse asks whether he is proud of him, he answers that he almost killed his mother. Henry seems to be shocked by the consequences of the delivery; therefore, he does not feel happy about the son's birth and he does not have any bond of affection for him. Catherine proves to be the only one who preoccupies him.

Next 3-gram *I went out* is used as a separate clause or with the precision of a place like *the door, the hall, the kitchen.* Generally the 3-gram *I want to* is associated with Catherine. She wants to see *Niagara falls*and *San Francisco.* Apart from that, she wants to cut her hair short to be like Henry:

36. Oh, darling, I want you so much *I want to* be you too."

Catherine is desperately in love with Henry, thus, her speech is expressive and emotionally coloured. The second example is evidence to this as well:

37. *I want to* ruin you.

4.6.2 Time and Place Clusters, and *Other*

Time and place clusters is the largest functional category among the 3-grams. On the other hand, only one body part cluster is found. This leads to conclusion that a story is more concerned with external events and narration of time and place. Regarding the time and place clusters, they make the story accurate, providing the specification of time and place something happens. Here are examples of some 3-grams:

38. We got into Milan early *in the morning* and they unloaded us in the freight yard.
39. I went over and sat *on the bed* beside her and kissed her.
40. It was bright sunlight *in the room* when I woke.

41. I heard it buzz *down the hall* and then someone coming on rubber soles along the hall.

42. *After a while* the stream from the stretcher above lessened and started to drip again and I heard and felt the canvas above move as the man on the stretcher settled more comfortably.

The third functional group *other* consists of the phraseological units like a routine formulae *I'm afraid,* a phrasal verb *looked at me* and a fixed expression *back to the.* Moreover, the 3-grams *in the rain*, *in the dark, to the front*, which resemble place clusters are significant in the novel's interpretation. The dark and the rain make the perception of events even bleaker: nature appears to be against the war as seen from Examples 43-47:

43. There was much shelling and many rockets *in the rain* and machine-gun and rifle fire all along the line.

44. We were very close to Germans twice *in the rain* but they did not see us.

45. As we moved out through the town it was empty *in the rain* and the dark except for columns of troops and guns that were going through the main street.

46. Sometimes *in the dark* we heard the troops marching under the window and guns going past pulled by motor-tractors.

47. *In the dark* and the rain, making our way along the side of the road I could see that many of the troops still had their rifles.

The rain has a symbolic meaning: Catherine confesses that she is afraid of it, while Henry likes it. Still afterwards Catherine promises:

48. I'll love you *in the rain* and in the snow and in the hail and what else is there?"

In the end of the novel when Catherine dies, it is again raining outside:

49. After a while I went out and left the hospital and walked back to the hotel *in the rain*.

In the rain appears again when Henry reflects on the war: he does not consider soldiers' sacrifices to be something worthy as it just an empty word for him.

50. We had heard them, sometimes standing *in the rain* almost out of earshot, so that only the shouted words came through, and had read them, on proclamations that were slapped up by billposters over other proclamations, now for a long time, and I had seen nothing sacred, and the things that were glorious had no glory and the sacrifices were like the stockyards at Chicago if nothing was done with the meat except to bury it.

Speaking of the 3-gram *to the front,* it is expected to be found in the novel as the setting is World War I. As noted above, it would fit into the category *war* proposed in the analysis of 4- and 3-grams from *Gone with the Wind*(Sections 3.5.3.and 3.6.1).It is repeatedly employed with the verbs *go back, get back, return* or the adjective *close.*The use of the 3-gram as a part of the 6-gram *to go back to the front* have been already examined in Section 4.3.

51. This is close *to the front*, isn't it?
52. I was to have three weeks' convalescent leave and then return *to the front*.
53. The night I was to return *to the front* I sent the porter down to hold a seat for me on the train when it came from Turin."

All in all, the analysis of 3-grams indicates that the 3-grams are useful in the novel's interpretation: they characterize mood in the novel, show characters' perception of the events and describe the language they use. Moreover, the corpus stylistic analysis of *A Farewell to Arms* will be complemented by the keywords' analysis that is found in the following subchapter.

4.7 Keywords Present in *A Farewell to Arms*

The comparative analysis of keywords found in the Brown Corpus and the novel *A Farewell to Arms* has shown that the groups presented by Mahlberg and McIntyre (2011) are relevant to the discussion of the novel (Section 2.3). The categories that are also associated with the keywords from *A Farewell to Arms* are the ones that describe the fictional world – Character names, Body part nouns and Clothes and accessories. Still some categories like Italian words, Physical and emotional states can be added to the depiction of the characters. As to the Settings and props, Places, Nature, Vehicles and buildings, and Food and drinks are also pertinent to the creation of the fictional world. Thematic signals, in turn, are essential to the novel's interpretation as they reflect the novel's themes. The above mentioned categories with respective keywords are included in Table.

Table 4.9 Keywords in*A Farewell to Arms*

Categories:	Keywords:
Fictional world:	
Characters: Names	Catherine, Rinaldi, Tenente, Piani, Bonello, priest, Aymo, doctor, nurse, Ferguson, Ettore, Passini, Meyers, barman,

	Gage, Campen, Carabinieri, Manera, Fergy, captain, Gordini, Crowell, Greffi, Lieutenant, Gino, waiter, Valentini, Simmons, Guttingen, Gavuzzi, Bartolomeo,
Body part nouns	knee, legs, head, hand, arm
Clothes and accessories	umbrella, dressing, bags, boots, cap, rucksack, tunic, cape, puttee
Italian words	Tenente, Ciaou
Physical and emotional states	tired, wounded, afraid, smiled, love, glad, drunk, crazy, happy, hurt, hungry, sleepy, cheerful, crying, sick, cold, awake, asleep
Settings and props	
Places	Milan, Montreux, bridge, Udine, café, villa, embankment,Bainsizza, station, town, hotel, Locarno, Gorizia, Capri, Abruzzi, Gabriele, Stresa, Pallanza, Plava, Switzerland, Mestre, Tagliamento, Pordenone, galleria, farmhouse, Cova, Carso, Scala,
Nature	rain, mountains, trees, river, raining, wind, plain, muddy, bank, woods, hedge, plateau
Vehicles and buildings	hospital, carriage, cars, ambulance, train, carts, van, toboggan
Food and drinks	wine, glass, brandy, cheese, bottle, cognac, vermouth, grappa, whiskey, beer, soda,choucroute, spaghetti, coffee
Thematic signals:	
World War I	Austrians, stretcher, wounded, Italians, war, front, retreat, girls, officers, mess, Germans, soldier, guns, troops, sergeants, helmets, canvas, bearers, attack, Bersaglieri, pistol, gunner, shell, wounds, carabiniere, shelling, armoire, bombardment, dugout, fighting,
	die, war, love, gone, rotten, killed

Despite the fact that there are thirty-one keywords stating the names of the characters, there are just five body part nouns and eight nouns describing clothes and accessories. Descriptions of people's appearance in the novel are scarce, with the nouns being used mostly to refer to the body parts where one gets injured, or fatigues. As to the communication language used by the main hero and other characters, it proves to be Italian apart from Henry's conversations with Catherine. Still the novel is written in English with two words in Italian, for

example, Henry's rank in the military – *Tenente (Lieutenant)* and a greeting phrase *Ciaou (Hi).*Just like with the African American keywords from *Gone with the Wind,* the Italian words are counted as keywords because they are non-existent in the Brown Corpus.Moreover, there are adjectives and nouns that display physical and emotional states: five of them such as *love, glad, happy* reflect positive emotions, ten others like *tired, crazy, hungry* imply negative emotions with two adjectives *awake* and *asleep* being emotionally neutral.

Among the group "Settings and props" the keywords alluding to Nature in addition to Food and drinks stand out. For instance, the 3-gram *in the rain* was discussed in the previous section and again the noun *rain* and *raining* are prominent as keywords. Besides, the narrator describes where the battles take place; therefore, an accurate description of surroundings is prevalent in the novel. Here is an example of a long sentence, highlighting a detailed description which provides a vivid picture of the settings (all the highlighted words are in the keywords list above):

> 54. The mountain that was beyond the valley and the hillside where the chestnut forest grew was captured and there were victories beyond the *plain* on the *plateau* to the south and we crossed the river in August and lived in a house in *Gorizia* that had a fountain and many thick shady trees in a walled garden and a wistaria vine purple on the side of the house.

As emphasized above, Food and drinks are prominent in the keyword analysis. Seven of them are the words that denote various types of alcoholic beverages – *wine,cognac, vermouth, grappa, whiskey*and only three nouns refer to foodstuff – *cheese, choucroute,*and *spaghetti*. The repetition of many alcoholic drinks can be explicable: the main hero Henry is male and he narrates about himself and other men who are involved in the war. In order toforget about the war and entertain themselves, they drink. Consequently, the adjective *drunk* is also frequent among the keywords.

Regarding the thematic signal keywords, a lot of them refer to World War I. Warring sides, ammunition, and participants are found in the semantic field of war: *Austrians, Italians, Germans, guns, helmet, pistol, officers, soldier* and *sergeants.* Furthermore, the keyword *girls* fits in here in spite of not being semantically related to the war terms; it characterizes the lives of Henry and his comrades. Apart from drinking, the men spend their spare time with girls. Other thematic signals are the keywords: *die, war, love, gone, rotten* and *killed.* The first three have been already studied as parts of longer N-grams. *Gone, rotten* and *killed* are closely associated

with Henry and his vision of war. The war is rotten for him: he is emotionally devastated and he loses his Catherine. In the end everything in the world is about killing:

> 55. They threw you in and told you the rules and the first time they caught you off base they *killed* you. Or they *killed* you gratuitously like Aymo. Or gave you the syphilis like Rinaldi. But they *killed* you in the end.

In this example there are two sentences beginning with *Or*. Sentences are not usually started with the conjunction, but here it is done on purpose: to stress the action and show that several ways of killing are possible.

The keyword analysis of *A Farewell to Arms* completes the discussion of the whole novel. It reveals a lot of keywords that are significant in the characterization of the fictional world and in the pinpointing of the novel's themes. Firstly, the names of the characters, their body parts, clothes and accessories, physical and emotional states portray the novel's characters. Secondly, places, nature, vehicles and buildings, food and drinks create the settings of the novel. Thirdly, the keywords implying World War I and other matters point out the novel's themes. The next chapter is dedicated to the comparison of the N-grams found in *Gone with the Wind* and *A Farewell to Arms* to the N-grams present in the Brown Corpus. In addition, the similarities and differences between two novels will be considered.

5. The Comparative Analysis

The following chapter presents comparison of 5-grams, 4-grams and 3-grams found in the Brown Corpus to 5-grams, 4-grams, and 3-grams found in the novels *Gone with the Wind* and *A Farewell to Arms.* Further the novels are compared to each other with regard to N-grams as well as keywords. The Brown Corpus consists of 500 samples of 2000 words each representing informative and imaginative prose.[4] Though these are incomplete texts, they are written by the American authors and, therefore, might reveal general patterns characteristic of American prose which are similar or different from the patterns occurring in the novels.

5.1 5-Grams Present in the Brown Corpus

As the recurrent 5-grams have been already identified in *Gone with the Wind* and *A Farewell to Arms* (Sections 3.4 and 4.4)*,* the first step of the present analysis is to find the 5-grams in the Brown Corpus. The minimum frequency is set to 5, the same frequency that was applied in the analyses of 5-grams in *Gone with the Wind* and *A Farewell to Arms. AntConc* returns sixty-one 5-grams. The first 30 of them are shown in Table in the descending order, and the rest are found in Appendix 3.

Table 5.1 5-Grams in the Brown Corpus

Labels (L)	Speech (S)	Time and place (TP)	Other (d)
the Secretary of the Treasury (8)	I don't know what (11)	at the end of the (33)	on the part of the (12)
Governorof the state of (5)	I don't want to (6)	the United States of America (24)	in the form of a (10)
		Government of the United States (16)	in the basic wage rate (8)
		in the middle of the (12)	as a result of the (10)
		the other side of the (11)	is one of the most (8)

		at the time of the (9)	turned out to be a (8)
		at the beginning of the (9)	as a matter of fact (12)
		for the first time in (9)	there are a number of (9)
		and at the same time (8)	on the basis of the (8)
		at the foot of the (8)	in the case of (7)
		for the rest of the (7)	of the State to be (7)
		in the center of the (7)	one or more of the (6)
		the Department of Economic Affairs (7)	the purposes of this Act (5)
		the far end of the (7)	This is one of the (5)
		the one hundred and eighty (7)	
		the turn of the century (6)	
		the other end of the (5)	
		the result obtained in item (5)	

Table 5.2 Distribution of 5-gram types in the Brown Corpus

Labels (L)	Speech (S)	Body Part Clusters (BP)	Time and place (TP)	Other (d)
3	2	3	33	20

5.1.1 Time and place Clusters

Taking into account the distribution of the functional groups among the 5-grams, it is striking that time and place clusters significantly outweigh other groups. The prose texts fromthe

Brown Corpus seem to focus on when and where something happened, or is situated. The pattern *PREP the N of the* turns out to be dominant among the 5-grams: *at the end of the*, *in the middle of the, at the time of the, at the beginning of the, at the foot of the* and *in the center of the, etc*. Examples 1-5 show their uses:

1. When we'd finished our regular rounds, Pete pointed me toward the small ward *at the end of the* floor. (Romance and Love story)

2. He would wake up *in the middle of the* night and fret about it. (Romance and Love story)

3. Our old one blew down in a storm *at the time of the*<pokeneu> festival fifteen moonsago. (Adventure and Western fiction)

4. These entries should normally appear *at the beginning of the* program or immediately following each ~LITORIGIN statement. (Learned)

5. Roberts careened backward,his back arched, fought for balance and, failing, stumbled against the newel post *at the foot of the* stairs. (Mystery and Detective Fiction)

Apart from the *of*-phrases beginning with the preposition, there are also 5-grams that start with the definite article and fall into the pattern *theADJ N of the*, for instance, *the other side of the* and *the other end of the:*

6. He would not be like the "rich Americans" who lived in white-columnedhouses *on the other side of* the park. (Romance and Love story)

7. Phil Rossoff, seated nextto Deegan, got up and moved to *the other end of the* bench.(Romance and Love story)

Since the samples from the Brown Corpus represent American English and to a great extent informative prose such as newspaper language, the word combinations *the United States of America* and *Government of the United States* are prominent.

5.1.2 *Other* Clusters and Body Part Clusters

The second largest functional group includes the 5-grams that cannot be attached to the groups proposed by Mahlberg. It is striking that many *other* 5-gramsconsist of general vocabulary and are relatively fixed phrases. Since the Brown Corpus does not constitute a single

text, but is composed of many short extracts, 5-grams are not specific to any particular topic. For example, several 5-grams comprise the quantifier *one* that serves as a pronoun and is used together with the superlative:*is one of the most* and *This is one of the.* Besides, there are set expressions that mention the result, conditions or means: *asa result of the, as a matter of fact, in the case* of and*on the basis of*. In addition to these phrases, there are found more phraseological units that are not typical of the novels like*on the part of the,in the form of a,in the basic wage rate*and *there are a number of.* There is also a part of conditional clause among the 5-grams. It is found in the Catholic Bible and it designates people:

8. *_IF WE ARE BORN OF* GOD WE HAVE FAITH IN CHRIST AS THE ONLY SAVIOUR._

During the analysis of 5-grams, three body part clusters have been spotted – *set my hand and caused*, *in the hands of the*and *in the minds of the.* The context of the first mentioned 5-gram reveals that the 5-gram occurs in Rhode Island Governor's Proclamations:

9. IN TESTIMONY WHEREOF, I have hereunto *set my hand and caused* the seal of the State to be affixed this 17th day of May, in the year of Our Lord, one thousand nine hundredand sixty-one, and of Independence, the one hundred and eighty-sixth.

Other two body part 5-grams*in the hands of the*and *in the minds of the* are used in different genres: Editorials, Religion, Popular Lore, Learned, General Fiction.

10. The news of battle on Breed's Hill had already seeped through, and New York itself was now left *in the hands of the* local Provincial Congress. (General fiction)

11. The teachers of Mr& Uno's school gave me a small gift to thank me for coming. Hiroshima is a better city than it was before – in the minds of the people I met was a strong determination for peace and understanding. (Popular lore)

5.1.3 Labels and Speech Clusters

Only two labels are found among the 5-grams in the Brown Corpus - *the Secretary of the Treasury*and*Governor of the state of*. The label *the Secretary of the Treasury*is repeated in the U.S. Legislation on Foreign Relations:

12. The Secretary of State shall certify to *the Secretary of the Treasury* the total cost of adjudication, not borne by the claimants, attributable to the Yugoslav Claims Agreement of 1948.

The next 5-gram that performs the function of label is*Governor of the state of.* It is encountered only in one text, the one mentioned above, Rhode Island Governor's Proclamations:

9. […] AND NOW, THEREFORE, DO I, JOHN A&NOTTE, JR&, *GOVERNOR OF THE STATE OF* RHODE ISLAND AND PROVIDENCE PLANTATIONS,PROCLAIM MONDAY, MAY 22nd, 1961, AS NATIONAL MARITIME DAY […]

Speech clusters are least frequent together with the labels. Their contexts have shown that *I don't know what* occurs in imaginative prose: general fiction, mystery and detective fiction, adventure and western fiction, romance and love story. Example 11 shows how the 5-gram is used in the genre Romance and Love story:

10. "*I don't know what* you heard that would make youthink so, but I assure you I don't even know myself, so how can you be so sure?

Generally, it appears together with the noun clauses, e.g. *I don't know whatI'm going to do with you/goes on around here/you're up to, etc.* Another 5-gram that signals speech is *I don't want to*. It is followed by the verbs such as *talk, see, leave (2 instances), go* and *be:*

11. *I don't want to* leave it. (Adventure and Western fiction)

As the texts from the Brown Corpus are both varied and incomplete, it is explicable why recurrent labels and speech clusters found there are too few. Still there are a lot of time and place clusters, and the 5-grams which cannot be categorized according to the meanings they carry.

5.2 A Comparative Analysis of 5-grams from the Brown Corpus and the Novels *Gone with the Wind* and *A Farewell to Arms*

Without doubt, the major difference between the Brown Corpus and *Gone with the Wind* is distribution of 5-gram types. Time and place 5-grams greatly outweigh other functional groups in the Brown Corpus, while labels dominate among the 5-grams in*Gone with the Wind.* In addition, speech 5-grams, time and place 5-grams and "Other" follow after in the descending order, while in the Brown Corpus "Other" 5-grams are the second largest group with labels and

speech 5-grams being least frequent. Despite the uneven correlation of the groups, certain 5-grams are characteristic of both the novel and the corpus. These are, for example, positional phrases *at the end of the*, *the other side of the, at the foot of the, at the bottom of the.* Thus, the patterns *PREP the N of the* and *the ADJ N of the* are common for the corpus and the novels. Besides these patterns, there are such frequent words as *edge* and *top.* They are used in the positional phrases like *at the edge of the* and *at the top of the* in Mitchell's novel and *on the edge of the, to the edge of the,* and *to the top of the* in the Brown Corpus. It is worth noting that speech 5-grams *I don t want to* and *I don t know what* are present in *Gone with the Wind* and the Brown Corpus and they express the characters' unwillingness to do something or their uncertainty.

In contrast to *Gone with the Wind,* the most prominent group in *A Farewell to Arms* is *speech clusters.* Fewer 5-grams stand for time and place clusters, and labels, with *other* 5-gram being just a single instance. Nevertheless, it is striking that though the time and place 5-grams make the second biggest group in the novel, the only common 5-gram for the Brown Corpus and *A Farewell to Arms* is *on the other side of*. The same speech cluster, in turn, is found in the corpus and the novel, that is *I don't want to.*

As seen from the comparison, the novel *Gone with the Wind* shares more 5-grams with the Brown Corpus than *A Farewell to Arms* does. The reason for this may lie in the number of the 5-grams because *Gone with the Wind* simply has more 5-grams than *A Farewell to Arms.* The 5-grams that appear in both novels and the Brown Corpus are the ones that refer to time or location, or mark the speech. The Brown Corpus is compilation of many incomplete texts of different genres; therefore, the labels that are presumably present in the texts are not identified by *AntConc* due to their infrequency.

5.3 4-Grams Present in the Brown Corpus

The 4-grams present in the Brown Corpus are greater in number than the 5-grams: one hundred twenty-six 4-grams as opposed to sixty-one 5-grams. Thirty-six of the 4-grams are given in Table 5.2, while the others are included in Appendix 4. Many of the 4-grams are part of the 5-grams presented in Table 5.1.

Table 5.2 4-Grams in the Brown Corpus

Speech (S)	Time and place (TP)	Other (d)	Body Part (BP)

I don't know (45)	of the United States (91)	on the basis of (47)	in the minds of (10)
don't want to (14)	at the same time (73)	the rest of the (46)	
I don't think (13)	the end of the (70)	one of the most (43)	
I'm going to (13)	at the end of (55)	on the other hand (41)	
didn't want to (12)	in the United States (53)	is one of the (36)	
I've got to (12)	for the first time (36)	as well as the (34)	
I would like to (11)	in front of the (34)	per cent of the (29)	
	the center of the (28)	a member of the (29)	
	the top of the (22)	the fact that the (29)	
	the United States and (21)	as a result of (28)	
	from time to time (21)	in the form of (23)	
	the middle of the (20)	in the case of (23)	
	at the time of (20)	The Government of the (23)	
	the side of the (19)	was one of the (21)	
	on the other side (19)	to be able to (21)	
	the edge of the (18)	In addition to the (21)	
	in the middle of (18)	on the part of (20)	
	in and out of (17)	was going to be (20)	
	the beginning of the (15)		
	the bottom of the (15)		

Table 5.4 Distribution of 4-gram types in *A Farewell to Arms*

Speech (S)	Time and place (TP)	Other (d)	Body Part Clusters (BP)
7	34	84	1

The distribution of 4-gram types demonstrates that the 4-grams that cannot be attached to specific groups and fall into the category *other* are highly prevalent over time and place, speech and body part 4-grams. As underlined before, their dominance is due to the informative prose taking the biggest part of the corpus. For that reason, time and place 4-grams are expected to be widely occurring as well. During the analysis there also found more speech clusters than with the 5-grams, whereas there is only one body part cluster.

5.3.1 *Other* Clusters

As eighty-four clusters are placed to the functional group *other,* it is clear that Mahlberg's classification does not quite fit to in the discussion of incomplete texts. *Other* 4-grams represent different prepositional, noun and adverbial phrases, as well as parts of predicates. For instance, the pronoun *one* is frequent in the pronoun phrase used together with the superlative *one of the most* and as a part of the predicates – *is one of the* and *was one of the.* Here are some examples of their contexts in the Brown Corpus:

16. Mary Jane might not be the most intelligent woman, but she was *one of the most*determined. (Adventure and Western Fiction)

17. It is not merely a new thing; it *is one of the* very few new things in <Utopia>; most of the rest is medieval or humanist or part of an old tradition of social criticism. (Learned)

18. This *was one of the* Navy'scrossroads- you find them all around the world. (Romance and Love story)

Prepositional phrases tend to be repeated too. The most frequent *other* 4-gram is an *of*-phrase signifying belonging – *of the United States* –then come such phrases as *on the basis of, in the form of, in the case of, on the part of, of the fact that, with respect to, etc.* Besides, noun and adverbial phrases prove to be phraseologicalunits – *the rest of the, on the other hand, a member of the, as a result of, in addition to the, etc.* Their occurrences are shown in Examples 19-23:

19. It serves merely as an excess cushion for *the rest of the* carcass. (General Fiction)

20. Shivering, people talked andargued; all this government spending would have to be paid for somehow, but *on the other hand* desperate circumstances called for desperate remedies and something had to be done. (Romance and Love story)

21. Not that he had supposed, considering the evidence, that he wassharing this refrigerator with *a member of the* Beach Patrol. (Romance and Love story)

22. *As a result of* these changed conditions, the impactof the organizational effort on agricultural labor-management relation has been much greater than in the past. (Learned)

23. Present at the scene- *in addition to the* deadman, who was indeed Louis Thor- had been Thor's partner BillBlake, and Antony Rose, an advertising agency executive who handledthe ZING account. (Adventure and Western fiction)

Furthermore, there are phrases implying quantity - *a large number of, a good deal of the part of the, of some of the:*

24. Although *a large number of* coatings systems,particularly at low thicknesses fail cohesively by the cutting mechanism,frequently a second type of cohesive failure may also take place. (Learned)

25. They hire*a good deal of* local labor, including two members of our Trustee Board. (General fiction)

26. He made a point of frowning,of acting out *the part of the* fond father-confessor. (Mystery and Detective fiction)

27. She convinced him that he ought to be a member *of some of the* small tea-drinking parties she held at her roomsand in the end he complied with her wishes, although it was only rarelythat he added anything to the random conversations. (Romance and Love story)

5.3.2 Time and Place Clusters, Speech Clusters and a Body Part Cluster

As stated above, time and place 4-grams make up the second largest functional group in the Brown Corpus. The nouns *end, top, side,* and *edge* again prominent – *the end of the, at the end*

of, the top of the, the side of the, the other side of, the edge of the and *on the edge of.* What is more, there are prepositional phrases that point to the concrete parts of location – *front, middle, center* and *bottom:*

28. Even at a car's length I could sense that something was wrong, and so I followed her up to the turnaround *in front of the* house.

29. Then when Miss Langford was on *the end of the* line of girls,Jack, *in the middle of the line*, gave an extra hard pull and the young teacher sprawled backwards, sitting down hard, her dress flying over her head. (Adventure and Western fiction)

30. Little boys crowded together on long wooden benches, and*in the center of* the room sat the teacher.

31. I was slowly swimming down to *the bottom of the* sea. She mademe welcome**.**

The third functional group that characterizes the 4-grams is speech clusters. Similar to the 5-grams, the 4-grams returned by *AntConc* contain the verbs *know, want* that are used together with the contracted auxiliary verbs and imply unawareness or unwillingness: *I don't know, don't want to, I don't think, didn't want to.* The clauses*I don't know* and *I don't think* and the parts of predicates *don't want to* and *didn't want to* are encountered in several genres:

32. "*I don't know*", Phil said. (Romance and Love story)

33. The girl took a couple of steps toward theman in shorts when Benson, in that barefoot courtliness Ramey couldnever decide was real, said, "You *don't want to* go around there,Ma'am". (Adventure and Western Fiction)

34. And thenagain perhaps the reason why he couldn't find time to do any of thethings he had planned to do after retirement: reading, roaming, gardening,lying on his back and watching the clouds go by, was becausehe *didn't want to* do them. (Humour)

35. "*I don't think*you've been quite honest, Cady. (Romance and Love Story)

The first-person pronouns appear also in other 4-grams that refer to intentions or what one has to do.

36. "*I'm going to* tell everybody how big a snake I killed". (Romance and Love Story)

37. I do seem to snapat everybody these days, but *I would like to* think of a way to makea little extra money". (Romance and Love Story)

38. "*I've got to* run an errand on theway. (Romance and Love Story)

To finish the analysis of 4-grams present in the Brown Corpus, there is found one body part cluster which is *in the minds of.* It occurs in the genres like Editorial, Religion, Skills and Hobbies, Popular Lore, Belles Lettres and Learned. Thus, it is characteristic of the informative prose collected in the Brown Corpus:

39. Thisleads one to conclude, as you have, that there is inevitably more prestige in a management position *in the minds of* our people" (Skills and Hobbies)

In contrast to the 5-grams, where time and place clusters is the most prominent functional group, *other* clustersdominate significantly over time and place clusters among 4-grams. As described above, there are a number of prepositional, noun and adverbial phrases that are basically common to the informative prose. Further comparison of 4-grams present in the Brown Corpus and the novels*Gone with the Wind* and*A Farewell to Arms*will reveal the similarities and differences between the corpus texts and the novels.

5.4 A Comparative Analysis of 4-grams from the Brown Corpus and the Novels*Gone with the Wind*and *A Farewell to Arms*

The comparative analysis of 4-grams from the Brown Corpus and*Gone with the Wind*shows that several speech 4-grams like *I'm going to, I don't know, don't want to, I've got to, didn't want* and *I don't think to* prove to be common for the corpus and the novel. Time and place clusters such as *the end of the, in front of the, the top of the, the back of the* are often repeated in Mitchell's novel and the Brown Corpus. The 4-gram *for the first time* is also common, however: in the Brown Corpus it is a time expression, while in *Gone with the Wind*it is a label associated with Scarlett. The *Other* 4-gram *the rest of the* is also present among the 4-grams in the Brown Corpus and the novel. All in all, the distributions of the functional groups in the Brown Corpus and in *Gone with the Wind*differ markedly. Though *other* 4-grams are seen in *Gone with the Wind,* they appear to be too few in comparison to eighty-four 4-grams from the Brown Corpus. Labels is the biggest group in *Gone with the Wind,* but it is not relevant in the classification of the 4-grams from the Brown Corpus. Another significant difference between the corpus and the novel is that *as if* 4-grams just like the *war* 4-grams are not found in the Brown Corpus. For that reason, the recurrent 4-grams from*Gone with the Wind*are more diverse than the 4-grams from the Brown Corpus.

The comparison of the 4-gram types found in the Brown Corpus opposed to the 4-grams from *A Farewell to Arms* indicates that there are four functional groups among the 4-grams from the corpus, while there are only three in the novel. Labels as recurrent 4-grams are missing in the corpus texts, but they are encountered in the novel. On the contrary, *other* 4-grams greatly outweigh the functional groups present in the Brown Corpus, but they are not pertinent to the classification of the novel's 4-grams. Still three speech 4-grams prove to be common for the Brown Corpus and Hemingway's novel: *I don't know, don't want to* and *I don't think.* The positional phrases and a time expression are typical of both the novel and the corpus: *the side of the, on the other side, in front of the, the edge of the,* and *for a long time.*

Despite the fact that there are fewer recurrent 4-grams in *A Farewell to Arms* than in *Gone with the Wind*(due to different text lengths)*,* the number of identical 4-grams for the Brown Corpus and the novels does not differ radically: eleven instances in *Gone with the Wind*and eight instances in *A Farewell to Arms.* The findings of Stubbs and Barth (2003), whoanalysed the chains from Fiction, Belles Lettres, and Learned, have already been presented in Section 2.2. They underline that the verbs *said, know,* and *want* are frequently used in the novels. This holds completely true for the 5-grams and 4-grams present in the Brown Corpus and Mitchell's and Hemingway's novels. For that reason, they can be regarded as text-type discriminators. The same happens to the positional phrases that are common for the novels and the corpus, they are characteristic of fiction.As the 4-grams are rather short, it is predictable that they appear more frequently in the prose than the longer N-grams. The 4-grams might be the same for the novels and the texts from the corpus, but their contexts can reveal differences between them.

Below, the present analysis is supplemented with the comparison of 8-, 7-, 6-, 5-, 4-, 3-grams and keywords in both novels in order to comprehend how these two texts are different or similar to each other.

5.5 A Comparison of 8-, 7- and 6-grams found in *Gone with the Wind* and *A Farewell to Arms*

Though the novels *Gone with the Wind* and*A Farewell to Arms*vary considerably in size**,** the number of 8-grams is almost the same: two and three respectively.The minimum frequency for the 8-grams is set to four. The found 8-grams prove to be exclusive for each novel: the 8-grams describe Scarlett and Frederic. If one compares the structures of these 8-grams, they contain personal pronouns like *she* in *Gone with the Wind* and *I* and *she* (in reference to

Catherine)in *A Farewell to Arms.* They reflect the narration of the both novels: in case of*Gone with the Wind* it is a third-person narrative, but in case of *A Farewell to Arms*it is a first-person narrative. One of the 8-grams from*Gone with the Wind* is the adverbial of time – *For the first time in her life she,* while another is a colloquial phrase with a narrator's word *I won't think of it now she*, conveying one of the main messages in the novel. As to *A Farewell to Arms,* the first 8-gramis a main complete clause *I went to the window and looked out* functioning as a sentence in two instances and once being a part of a complex sentence. The other two 8-grams are similar conditional clauses starting with *what if* – *what if she should die She can't* and *what if she should die She won't.* Thus, the 8-grams prove to point to peculiar features of the novels' styles.

The comparison of 7-grams indicates that the 7-grams from*A Farewell to Arms*are reflections of the 8-grams only with slightly different cut-off points like *But what if she should die She, I went to the window and looked.* By contrast, new 7-grams are encountered in *Gone with the Wind,* e.g.*she was going to have a baby, Do you think I'm going to, It did not occur to her that, like a duck on a June bug*. These were discussed in greater detail in Section 3.2. Judging from the 7-grams, they also prove to be characteristic of the novel as the use of the pronouns *she*, *her* and *I* points to Scarlett, still *Do you think I'm going to* is associated with Rhett. However, the 7-gram *like a duck on a June bug* is a phraseological unit that apparently does not belong to the description of one particular character as it mainly refers to the Yankees or how the Southerners perceive the Yankees' attacks. All in all, the 7-grams represent the novel and again Mahlberg's finding about the longer clusters being character-specific holds true for both novels.

Regarding the 6-grams, they appear to be more recurrent than 8- and 7-grams. Nevertheless, of the fourteen 6-grams from *A Farewell to Arms,* only four have not appeared before as parts of 8- and 7-grams, whereas in *Gone with the Wind* twenty-two out of forty-one 6-grams have not been encountered before. This means that *Gone with the Wind* has more diverse 6-grams than*A Farewell to Arms.* Another difference that is found between two novels lies in the use of the personal pronouns. In spite of the fact that *A Farewell to Arms*is told in the first-person, the 6-grams with the first-person pronoun *I* are not as frequent as one would expect. For instance, there are only three occurrences of *I* among fourteen 6-grams, but eleven occurrences of the pronoun *she* in reference to Catherine. Still the 6-grams concentrate on Frederic's feelings, thus, creating the image of the main hero. In turn, the calculation for*Gone with the Wind* shows that there are sixteen uses of the third-person pronoun *she* and the possessive pronouns *her, his* in contrast to ten uses of the first-person pronoun *I.* Taking this into consideration, one can

conclude that both the narrator's comments and the characters' utterances, especially those of Scarlett and Rhett, play an important role in the novel *Gone with the Wind.*

Apart from the time adverbials and clauses, there are also other patterns that can be distinguished. Speech acts denoting future plans or unwillingness are frequent in *Gone with the Wind.* For example, *I'm going to have, I'm coming home to you, I don't want to hear.* Besides, there is an adverbial clause of reason – *if it hadn't been for,* the pattern *PREP the N of Possesive N – at the top of her voice*, a postmodified noun, labelling Belle Watling – *the Girl of the Period Saloon,* a phraseological unit – *for to tote the weary load,* prepositional phrases defining place directly and indirectly – *in the back of the wagon,* and *into the back of her mind.A Farewell to Arms* contains speech acts as well, one of them functions as routine formulae - *What's the matter with you* and another one expresses unwillingness *I don't want you to.* Furthermore, there is a positional phrase – *on the other side of the* and the predicative – *to go back to the front.* None of the 6-grams are identical for both novels.

To sum up, 8-, 7- and 6-grams from *Gone with the Wind* focus a lot on what happens to Scarlett for the first time and how she deals with difficulties; therefore, time expressions are used excessively with the help of prepositional phrases and adverbs – *for the first time, now,* and *later.* Without doubt, more N-grams are expected to be found in*Gone with the Wind* than in *A Farewell to Arms* due to text length*,* but in case of *Gone with the Wind* 8-, 7- and 6-grams turn out to be more diverse, that is they reveal additional features of the novel despite several of them being parts of 8- and 7-grams. By contrast, 7-grams from *A Farewell to Arms*are various reflections of the 8-grams and so are most of the 6-grams.

5.6 A Comparison of 5-grams found in *Gone with the Wind* and *A Farewell to Arms*

The search for the 5-grams with a minimum frequency of five returns completely different results: sixty-five 5-grams in *Gone with the Wind* vs. nineteen 5-grams in *A Farewell to Arms*(Sections 3.4and 4.4)*.* As the frequency threshold is set to five, the 5-grams returned are less in quantity than they could be if the frequency of four was preserved. Nevertheless, the functional groups proposed by Mahlberg can be discerned for both novels. A general overview of 5-gram types present in the novel is as follows: the most frequent functional group in *A Farewell to Arms*is speech clusters, whereas labels outweigh other groups in *Gone with the Wind.* Labels in *A Farewell to Arms*and speech clusters in case of *Gone with the Wind* take the second position. Time and place clusters, and other clusters are the third and fourth functional

groups in both novels. All in all, the same groups are pertinent in the analysis of the novels with body part clusters and *as if* clusters being irrelevant.

To start with labels, they are associated with the main characters in the two novels – Scarlett and Frederic. As labels prevail in *Gone with the Wind*, the depiction of the novel's characters is more specified than in*A Farewell to Arms*. Through the contexts of the 5-grams, readers get to know Scarlett's thoughts and feelings that help to comprehend her nature. By contrast, the 5-grams from*A Farewell to Arms*do not say much of Frederic as a person despite revealing his inner thoughts and feelings. In addition, the 5-grams in*A Farewell to Arms* characterize only Frederic, while in*Gone with the Wind* they highlight other characters than Scarlett, too – Rhett, Melanie and Ashley.

Further analysis indicates that speech clusters in the novels are negative constructions that signal unwillingness, inability or disbelief, e.g. *I don't want to, I don't know what, I don't know I, I don't see why, I don't believe it.* The 5-gram *I don't want to* is common for both novels and it is the most frequently used one – twenty-one occurrences in *Gone with the Wind* and fifteen occurrences in*A Farewell to Arms.* Some ofthe right collocates for *I don't want to*coincide for both novels: these are *hear*and *do.* Besides, the 5-gram *I don't believe it* is seen in both novels: used 9 times in*A Farewell to Arms*and 5 times in *Gone with the Wind.* In contrast to*Gone with the Wind,* 5-grams representing speech acts are in the foreground in*A Farewell to Arms*; therefore, the colloquial language of the characters assists in understanding the relationships between the characters and themselves as well. The dialogues between the characters are also essential for the creation of the fictional world in*Gone with the Wind,* but labels make a novel's imagery and its characters more clear than speech clusters.

Time and place clusters appear to be prominent as there are more of them among the 5-grams than among longer N-grams. Several 5-grams found in *Gone with the Wind* follow a certain pattern, namely *at the Nof the*: *at the end of the, at the head of the, at the top of the, at the bottom of the, at the foot of the.* In addition, prepositional phrases with the head noun *house* are frequently used – *in front of the house, out of the house and.* Similar patterns just with the preposition *on* are visible among the 5-grams from *AFarewell to Arms*: *on the N of the – on the side of the, on the floor of the* and *on the ADJ N of – on the other side of.* A prepositional phrase with the head noun *road* is a recurrent 5-gram in *AFarewell to Arms.* Except the nouns *house* and *road,* there are no other concrete locations, only the triggers that imply the wider context of time or place.

Turning to the 5-grams that do not fit into the previously discussed functional groups, there are too few of them in the novels to form any pattern – one in *A Farewell to Arms* and one in *Gone with the Wind.* The first one describes a person's action – *looked out of the window,* the second one describes how a person sobs or screams – *at the top of her (voice, lungs).*

The comparison of the 5-grams has shown that there are not only differences, but also similarities between the novels. First and foremost, there are two 5-grams that are present in both novels and one of them is the most frequent one – *I don't want to.* The speech 5-grams present in the novels carry negative connotations – unwillingness or inability. Time and place clusters follow similar patterns: in case of *Gone with the Wind* these are constructions with the preposition *at*, whereas in case of *A Farewell to Arms* these are positional phrases beginning with the preposition *on.*

As to the differences, the depiction of the characters is more detailed in *Gone with the Wind* and the 5-grams encountered are immediately associative with particular characters. Speech 5-grams prominent in *A Farewell to Arms* highlight peculiarities about the characters and their relationship, but the 5-grams themselves do not provide enough information; therefore, extended contexts are always necessary.

5.7 A Comparison of 4-grams found in *Gone with the Wind* and *A Farewell to Arms*

To continue the comparison with 4-grams, there are striking differences between the novels: *Gone with the Wind* contains more functional groups, namely two additional groups – *as if* 5-grams and 5-grams referring to war. In *A Farewell to Arms*, however, 5-grams representing speech outnumber other types of 5-grams, but with difference of only two 5-grams and labels taking the second position. For *Gone with the Wind other* 5-grams is the third most frequent group, while for *A Farewell to Arms* it is time and place clusters which also is the least frequent group as other 5-gram types are not pertinent in the classification. Time and place clusters take the fourth place followed by *as if* and *war* 5-grams.

In contrast to labels designated by 5-grams, there are more labels present among 4-grams which comprise the first-person pronoun *I*, thus, the narration of what Frederic is preoccupied with becomes more noticeable. Labels from *Gone with the Wind* depict Scarlett's concerns; therefore, the narrator of the story excessively uses the pronoun *she.* Speech clusters found among the 4-grams express reluctance to do something, disbelief or unawareness. It is not a

surprising observation as most of them showed up as part of 5-grams. Several of them – *I don't know, I don't want, I don't think, I don't care, don't want to* and *What's the matter* – are found in both novels. With regards to time and place clusters, there is a pattern that was characteristic of 5-grams in *A Farewell to Arms*, that is, *on the ADJ N of*, only without *of* – *on the front porch* is present in *Gone with the Wind* and *on the other side* is present in *A Farewell to Arms*. In addition, the positional phrase *in front of the* is typical of both novels. All in all, there are more 4-grams that are common for both novels than it was with 5-grams.

5.8 A Comparison of 3-grams found in *Gone with the Wind* and *A Farewell to Arms*

As only lexical 3-grams have been chosen for the analysis in the thesis, the functional group speech clusters is not applicable in the classification of 3-grams (sections 3.6 and 4.6). Nonetheless, the functional group *Body part clusters* proves to be significant in the discussion of the 3-grams from *Gone with the Wind,* which is not actually the case with *A Farewell to Arms* as only one such 3-gram is found there. Time and place clusters greatly outnumber other categories in *A Farewell to Arms*, whereas *other* 3-grams prevail over other types of 3-grams in *Gone with the Wind.* The second largest group for *A Farewell to Arms* is *other* 3-grams with labels being in a minority. As to *Gone with the Wind,* the second prominent functional group is labels, the third one – time and place clusters and the last category is body part clusters in addition to three war clusters.

A major difference between labels in the two novels is that labels in *A Farewell to Arms* contain the first-person pronoun *I,* while in *Gone with the Wind* apart from the expected third-person pronouns *she* are found phrases like *in the world, in love with* and *the old days* that portray Scarlett, her attitude towards material things and other characters and also her vision of the old times. Another difference is seen in time and place clusters. The time and place 5-grams in *Gone with the Wind* illustrate the house and its parts as the main scene, but in *A Farewell to Arms* it is mostly the bed, room, hospital, road or mountains that are depicted as the main scenes. Just like with the 5-grams the nouns *house* and *road* are frequently used again. Still the 3-gram *on the floor* is present in both novels. In addition, the 3-gram *in the dark,* from the category *other,* has almost the same frequency in two novels – 26 instances in *Gone with the Wind* and 24 instances in *A Farewell to Arms.* Corresponding 3-grams pointing to the main characters are also frequent in both novels: *looked at me* in *A Farewell to Arms* and *looked at her* in *Gone with the Wind.*

Due to the fact that only lexical 3-grams have been analysed, there are fewer 3-grams shared by both novels than was the case with 4-grams. Consequently, the 3-grams that carry the messages of the novels become prominent, especially in *Gone with the Wind.* Further comparison of keywords of two novels emphasizes considerable differences and similarities.

5.9 A Comparison of Keywordsfound in *Gone with the Wind* and *A Farewell to Arms*

In the description of characters and settings certain groups of keywords can be distinguished and they generally suit one of the novels, but not the other. For instance, keywords denoting nature, food and drinks are not dominant in*Gone with the Wind*, whereas they are frequent in*AFarewell to Arms.* Keywords representing society are not relevant to the discussion of*AFarewell to Arms,* but they are significant in the interpretation of the novel*Gone with the Wind.* However, the groups proposed by Mahlberg and McIntyre (2011) prove to fit perfectly in the keyword analyses of the novels, but it is striking that *Gone with the Wind* comprises more body part nouns than*AFarewell to Arms* and the ones occurring in*AFarewell to Arms*are associated with injuries. Thus, more attention is devoted to the characters' appearance in*Gone with the Wind* than in *AFarewell to Arms.* This refers also to clothes: women's clothing is demonstrated through the keywords in Mitchell's novel, while fatigues are prevalent among the keywords from Hemingway's novel.

Taking into account the thematic signals, the keywords signal clearly that the stories narrate about the American Civil War and World War I. For that reason, the keywords from the novels differ as they represent different participant in the war, still the keyword *war* is prominent in both novels. The semantic field of war keywords from *AFarewell to Arms*is directed to the troops, their ammunition and what happens in the front. By contrast, the keywords from *Gone with the Wind* are focused on the most important events and people in the war. Other keywords that are common for both novels are *love* and *gone.Gone with the Wind* reflects Scarlett's love for Ashley and Rhett's love for Scarlett, and in the end Scarlett's love for Rhett. *AFarewell to Arms*depicts love of Frederic and Catherine. The keyword *gone* has different symbolic meanings in the novels: the generation of prosperous Southerners owing slaves vanishes; therefore, the adjective *lost* is also found in the keywords' analysis. In the case of *AFarewell to Arms,* a merry life of Lieutenant Frederic has gone as the war proves to be a rotten thing that is merely about killing and dying.

5.10 Summary

The comparative analysis of 5- and 4-grams present in the Brown Corpus and the novels has shown that there are several time, place and speech clusters that are common for both novels and the corpus. Thus, certain patterns of the positional phrases have become striking. The distribution of the 5- and 4-grams in the Brown Corpus has been uneven as time and place clusters dominate over other functional groups among the 5-grams, whereas *other* clusters outweigh vastly the functional categories among the 4-grams. Due to the fact that the Brown Corpus contains more texts of the informative prose than the imaginative prose, speech clusters and especially labels have not been that frequent as opposed to labels and speech clusters in the analysed novels.

Regarding the comparison of *Gone with the Wind* to *A Farewell to Arms*, 8- and 7-grams prove to be specific to the novels, consequently, revealing information about the main characters. The 6-grams in *Gone with the Wind* prove to be more character specific than the 6-grams from *A Farewell to Arms.* By contrast, 5-, 4- and 3-grams are more frequent and higher in number than 8-,7- and 6-grams, but the most part of them reveals not only the main characters, but also other characters' speech, the novel's settings and other matters important in the narration. The discussion of the novels is concluded by the comparison of the keywords. Although some of them appear to be identical for both novels, the connotations they carry appear to be exclusive for each novel.

6. Conclusions

The thesis has shown that Mahlberg's corpus stylistic approach can be applied to the analysis of the novels *Gone with the Wind* and *A Farewell to Arms*. Longer N-grams like 8- and 7-grams prove to be character specific and exclusive for the novels' styles. Shorter ones as 6-, 5-, 4- and 3-grams are partly character specific, otherwise they can be grouped according to the functions they perform: labels, speech clusters, time and place clusters, body part clusters, *as if* clusters and other. Mahlberg's observation about the higher frequency of shorter clusters holds true for the N-grams from *Gone with the Wind* and *A Farewell to Arms*. 5-, 4- and 3-grams are more frequent and larger in number than 8-, 7- and 6-grams.The novels display certain tendency: speech clusters dominate over other functional groups among 5- and 4-grams.

There were also clear differences between the novels within functional groups. Most oflabels from *A Farewell to Arms*present the eventsthat happen with Frederic, while labels *found in*Gone with the Wind*describe inner thoughtsof Scarlett. Besides, the variety of functional categories is more distinctive in*Gone with the Wind*than in *A Farewell to Arms*. As if* clusters have not been frequent in Hemingway's novel. The reason for this may be found in the novel's style as quite short sentences prevail in the novels, while theuse of the conjunction would introduce simile and, therefore, longer sentences.*A Farewell to Arms*contains very few body part clusters, thus, the appearance of the characters is not revealed.

The contrastive analysis of 5- and 4-grams present in the Brown Corpus and the novels has shown that several of them functioning as speech and time and place clusters are common for the texts of informative and imaginative prose included in the Brown Corpus and the novels*Gone with the Wind* and *A Farewell to Arms*. It is striking that the most prominent functional groups identified among the 5- and 4-grams are time and place clusters, and *other* respectively. This may happen due to the fact that the texts collected in the Brown Corpus represent various genres and are incomplete. As mentioned above, both novels contain 5- and 4-grams that are also recurrent in the American texts from 1961, but they have the 5-, 4-, and 3-grams that are characteristic only of them as well. In addition, some keywords are present in both novels, but they have different connotations. The more detailed comparison of the N-grams used in the Brown Corpus and the N-grams encountered in the novels is found in Chapter 5.

Mahlberg's functional categories were quite sufficient for the analysis of the novels though not all groups were prominent among the 5-, 4- and 3-grams. Still I needed to add the category of war-related vocabulary because it proved to be characteristic of Mitchell's novel,

pinpointing one of the major themes of the novel. Functional groups did not work extremely well in the analysis of the Brown Corpus because most N-grams ended up in the *Other* category, which indicates that a corpus-stylistic analysis functions best on complete, long texts such as novels.

Limitations

After having written this master's thesis, we understand that the different lengths of the novels may be considered a limitation of the current study. In contrast to the N-grams from*Gone with the Wind,* much fewer N-grams have come up from *A Farewell to Arms.* If the novel was of equal size with *Gone with the Wind*, the functional groups may have been distributed more evenly. Another limitation can be ascribed to the Brown Corpus because it comprises 374 samples of the informative prose and only 126 samples of the imaginative prose. For that reason, the frequently recurring 5- and 4-grams that appear in the Corpus mainly highlight the peculiarities of non-fiction, but the aim of the thesis was to found out whether the N-grams present in the novels reveal the novels' styles or general tendencies among other fictional texts. Further research could be undertaken to analyse the novels equal in size in order to pinpoint the similarities and differences between the styles of American English author and British English author, or the styles from different time periods, such as the Renaissance or Postmodernism.

References

Altenberg, Bengt. 1998. On the Phraseology of Spoken English: The Evidence of Recurrent Word-combinations. In A.P. Cowie (ed.). *Phraseology.Theory, Analysis, and Applications*.Oxford University Press.101-122.

Anthony, Laurence. 2011. *AntConc*3.2.4w (Corpus Tools) Available at: http://www.antlab.sci.waseda.ac.jp/antconc_index.html[Accessed 10 May 2013]

Bondi, Marina and Mike Scott (ed.). 2010. *Keyness in Texts.* Amsterdam : John Benjamins

Fiedler, Sabine. 2007. *English Phraseology*. *A Coursebook*.Tübingen: Gunter Narr Verlag.

Hemingway, Ernest. 1929. *A Farewell to Arms.* An electronic version. Available at:http://webreading.ru/prose_/prose_classic/ernest-hemingway-a-farewell-to-arms.html[Accessed 5 May 2013]

Kucera, Henry and W. Nelson Francis. 1979. The Brown Corpus. Available at the University of Oslo. [Accessed 10 May 2013]

Leech, Geoffrey N. and Michael H. Short. 1981. *Style in fiction: a linguistic introduction to English fictional prose.* London : Longman

Lindebjerg, Anne. 1997. *Brown Corpus Manual*. Available at: http://icame.uib.no/brown/bcm.html[Accessed 10 May 2013]

Mahlberg, Michaela. 2007. Corpus stylistics: bridging the gap between linguistic and literary studies. In M. Hoey, M. Mahlberg, M. Stubbs, and W. Teubert (eds.). *Text, Discourse and Corpora.Theory and Analysis*. London: Continuum, 219-246.

Mahlberg, M. 2010. Corpus linguistics and the study of nineteenth century fiction.*Journal of Victorian Culture*. 15(2), 292-298. Available at: http://www.tandfonline.com/doi/abs/10.1080/13555502.2010.491667#.UZPT1aLIYQM[Accessed 10 May 2013]

Mahlberg, Michaela and Dan McIntyre. 2011. A case for corpus stylistics: Ian Fleming's *Casino Royale*. *English Text Construction*. 4(2), 204-227.Available at the University of Oslo. [Accessed 10 May 2013]

Mahlberg, Michaela. 2012. *Corpus Stylistics and Dickens's Fiction*. London: Routledge.

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

Mitchell, Margaret. 1936. *Gone with the Wind*. An electronic version. Available at: http://webreading.ru/prose_/prose_classic/margaret-mitchell-gone-with-the-wind.html [Accessed 5 May 2013]

Semino, Elena and Mick Short. 2004. *Corpus stylistics: speech, writing and thought presentation in a corpus of English writing.* London :Routledge

Scott, Mike. 2008. *WordSmith Tools.* Version 5.0.Available at the University of Oslo.[Accessed 5 May 2013]

Stubbs, Michael and Isabel Barth. 2003. Using Recurrent Phrases as Text-type Discriminators. *Functions of Language* 10(1): 65-108.

The American Heritage Dictionary of the English Language, Fourth Edition.Houghton Mifflin Company, 2004. Available at:http://education.yahoo.com/reference/dictionary/ [Accessed 10 May 2013]

The Collins English Dictionary Online.2013. HarperCollins Publishers Limited. Available at: http://www.collinsdictionary.com/dictionary/english/scallawag [Accessed 10 May 2013]

The Merriam-Webster's Online Dictionary, 11th Edition. 2008. Available at: http://www.merriam-webster.com/dictionary/carpetbagger

Appendix 1.

Speech (S)
I don't know why (5)
if it wasn't for (5)
I don't want a (5)
I don't want any (5)
's going to be a (5)

Appendix 2.

Speech (S)
What's the matter (12)
Don't you think (11)
I'll have to (11)
I'll tell you (11)
you think I'm (11)
I beg your pardon (10)
What do you mean (10)

Appendix 3

Body Part Clusters (BP)	Time and place (TP)	Other (d)
set my hand and caused (5)	at the bottom of the (6)	per pound of body weight (6)
in the hands of the (5)	at the far end of (6)	as a member of the (5)
in the minds of the (5)	in and out of the (6)	in the history of the world (5)
	in the United States (6)	in the same manner as (5)
	on the edge of the (6)	was going to be a (5)
	United States of America in (5)	Of Rhode Islandand Providence (5)
	at the center of a (5)	
	in the early years of (5)	
	in the year of Our (5)	
	is to be found in (5)	
	to the edge of the (5)	
	to the Government of India (5)	
	to the top of the (5)	

Appendix 4

Time and place (TP)	Other (d)
in the midst of (15)	in the face of (20)
to be found in (14)	In the first place (20)
at the beginning of (14)	of the fact that (19)
at the University of (14)	the nature of the (18)
in the center of (14)	The members of the (18)
the end of a (13)	the use of the (17)
in one of the (13)	the way in which (17)
on the side of (13)	a large number of (17)
for a long time (12)	a good deal of (16)
the back of the (12)	the part of the (16)
on the edge of (11)	of some of the (15)
to the United States (11)	that the United States (15)
in front of him (11)	with respect to the (15)
the time of the (11)	in the name of (14)
over a period of (10)	a part of the (14)
	for the most part (14)
	for the benefit of (13)
	in the history of (13)
	is likely to be (13)
	of the United Nations (13)

	Other (d)
	as well as a (12)
	at the age of (13)
	but he did not (13)
	would have to be (13)
	The fact that he (13)
	in the light of (13)
	the way to the (13)
	can be used to (12)
	in terms of the (12)
	the form of a (12)
	in an effort to (12)
	the order of the (12)
	the value of the (12)
	a great deal of (12)
	the size of the (12)
	the spirit of the (12)
	of the New York (11)
	the basic wage rate (11)
	will be able to (11)
	at the foot of (11)
	in the development of (11)
	in the way of (11)

	Other (d)
	of a number of (11)
	one of the best (11)
	that it would be (11)
	that most of the (11)
	the basis of the (11)
	the development of the (11)
	the life of the (11)
	to be one of (11)
	to go to the (11)
	in accordance with the (11)
	of the Department of (11)
	the plane of the (11)
	at the expense of (10)
	by the fact that (10)
	from the fact that (10)
	index word or electronic (10)
	is a matter of (10)
	may or may not (10)
	of one of the (10)
	the power of the (10)
	the state of the (10)
	to the fact that (10)

	in the basic wage (9)