

# Gender and Lexical Representation in Enid Blyton's *The Famous Five: A Corpus-based Study*

---

Kriangkrai Vathanalaoha  
Srinakharinwirot University

Dhanakorn Jeeradhanawin  
Thammasat University

## Abstract

Corpus stylistics is deemed promising in literary school as it fosters cautious optimism in using empirical evidence to support individualistic interpretations. Also, it is established that we are living in a digital era where texts are produced more in form of electronic text. This study aims to use corpus-based analyses to offer empirical data that reveals about language behavior of what literary critics have been studied in *The Famous Five*, one of the most popular series of children's literature in the world, with respect to literary themes and gender representation. The study uses two corpus analysis tools, Wmatrix and AntConc, to generate information such as key semantic fields (thematic categorizations), collocations and concordances. Using the corpus-based information, this paper concludes that *The Famous Five* comprises the essence of adventures and mysteries and that the story contains conventional stereotypes of male and female protagonists. Amongst the entire group, George, or Georgina, is the most intriguing character because she has been represented as an impatient and outspoken character, characterized by the lexical choices that the author has made.

**Keywords:** corpus stylistics; gender representation; lexical representation

## บทคัดย่อ

ปัจุบันสื่ออิเลคทรอนิกด่างๆ มีผู้ใช้อย่างแพร่หลายมากยิ่งขึ้น ส่งผลให้ระเบียบวิธีวิจัยวัจวนลีลากลางข้อมูลเป็นที่นิยมในหมู่นักวิจัยด้านวรรณกรรม ศึกษาอันเนื่องมาจากวิธีการดังกล่าวสามารถให้ข้อมูลเชิงประจักษ์เพื่อสนับสนุน การตีความของผู้วิจัยได้เป็นอย่างดี งานวิจัยขึ้นนี้มีจุดมุ่งหมายที่จะนำกลังข้อมูล นักเรียนชาวอังกฤษ ซึ่งเป็นวรรณกรรมสำหรับเด็กที่มีชื่อเสียง เพื่อหา หลักฐานเชิงประจักษ์สนับสนุนการตีความของนักวิจารณ์ในด้านการนำเสนอ เพศสภาพ และเนื้อหาหลักของเรื่องผู้วิจัยใช้เครื่องมือในการวิเคราะห์กลังข้อมูล สองประเภท ได้แก่ Wmatrix คิดค้นโดย ดร. Paul Rayson และ AntConc สร้างโดย ดร. Laurence Anthony เพื่อหาข้อมูล key semantic fields (thematic categorization) collocations และ concordances ผลการวิเคราะห์สรุปได้ว่า วรรณกรรมเรื่อง *The Famous Five* มีเนื้อหาหลักคือการจอยภัยผ่านเรื่องลึกลับ ของตัวละครหลักทั้งชายและหญิงที่มีความคิดแบบสมัยนิยม โดยที่ตัวละครที่ชื่อ George หรือชื่อจริงของเธอคือ Georgina นั้นนำศึกษามากที่สุดเนื่องจากผู้ประพันธ์ มีการใช้คำรูปแบบต่างๆ เพื่อสร้างให้เธอมีบุคลิกโงพางเจ้าอารมณ์ ซึ่งผิดแผล ไปจากความคาดหวังของผู้คนในสังคมยุคนี้ที่มีต่อสตรีเพศอย่างเช่น

**คำสำคัญ:** วัจวนลีลากลางข้อมูล การนำเสนอเพศสภาพ การนำเสนอการใช้คำ

## Introduction

*The Famous Five* was deemed one of the most popular series of children's literature, written by Enid Blyton, one of the household members of British novelists. The first episode, *Five on a Treasure Island*, was published in 1946 and it has become one of the most successful novels since then. *The Famous Five* contains 21 episodes in total and the last series was published in 1960. As the title suggested, five characters consisted of two girls (Georgina, Anne), two boys (Julian, Dick) and a dog (Timmy). It was established by literary critics that the story expressed "gender bias" because the female protagonist, Georgina, was liked to be called as George and she would be angry if someone called her by her real name. Her behavior was even more aggressive than that of other male characters and divergent from the impression of femininity. She was outspoken in particular scenes and recurrently addressed by the narrator as a tomboy. In terms of pedagogy, *The Famous Five* has been cited in children's literature classroom as a phenomenon in the 1950s and 1960s as Blyton's voice of contemporary feminism in the sense that female characters could possibly imitate what male characters did.

Regarding literary themes, its content has been investigated by literary critics, politicians and philosophers. Literary interpretations, by means of textual analysis and gender studies, were geared into inarguable consensus that there was a trace of gender bias throughout the series. For example, Macdonald (2013) addressed that Blyton vividly illustrated female characters as inferior as can be seen from Georgina who wanted to dress "like boys" or immensely discussed how Anne would end up with "washing-up" clothes like a submissive woman. Besides, Georgina's preference to be called as George was evidently an adaptation of "masculine demeanour". Julian, a male protagonist, has been deemed an intelligent and respectful character that would be against all criminal acts and wrongdoings. However, Georgina (George henceforth) was mostly the one who would lead the group out of predicaments through her wits and intuition.

As can be seen from literary studies, literary interpretations have been studied rather through subjective viewpoints of established scholars. The scholars criticize using excerpts or parts of textual clues to offer and crystallize individual insights. In terms of language, however, there has been no research study to date which mentions or utilize empirical data and explains how the readers arrive at particular meanings. A corpus-based approach will be used to generate empirical evidence for *The Famous Five* whether it would be either supportive or antithetical to the previous text-based studies. The approach is different from previous literary studies because it uses holistic data, as opposed to partial, to search for frequent patterns which could increase a degree of objectivity to literary criticism.

## Review of Literature

Application of corpus linguistics in literary work has been implemented to spell out the significance of empirical data for literary interpretation. Before the age of electronic text, a study of language in literature was known as “stylistics”. Stylistics is an interdisciplinary study focusing on using linguistic approach to explain how particular meanings are created through cognitive grammar, poetics and discourse analysis. Stylistics researchers stated that this is a kind of systematic means of literary studies. Stylistics studies have their own limitation because of the fact that they analyze only some extracts of the text and some other parts of the text might have been overlooked. As a result, corpus linguistics has been adopted to analyze literary texts through a holistic perspective, a.k.a. corpus stylistics. According to Biber et. al (1998), corpus analysis is “empirical, analyzing the actual patterns of language use in natural texts” (p. 7). Corpus analysis makes use of computer technology and utilizes a large collection of naturally-occurring texts such as spoken or written data as the basis for building corpora. With respect to corpus stylistics, literary meanings will be explained through interdisciplinary studies because it needs both

quantitative and qualitative analytical techniques. The nature of corpus stylistics, according Adolphs (2006), focuses on the distribution of words as a platform to identify textual features that are “especially characteristic of an author, particular text, or even a single character within a play or novel” (p. 66). Methodological approaches of corpus linguistics, which have been extensively used to scrutinize stylistic features of particular writings, are keyword analysis, extended lexical phrases and collocational analysis.

### *Keyword Analysis*

According to Stockwell (2002), foregrounding elements are essential in stylistics (p. 14), keyword analysis is compulsory as it can identify which word has been overused or underused throughout the text. For example, literary themes can be traced through investigation of “keyness” provided by a corpus tool called WordSmith. Mahlberg and Smith (2010) demonstrate crucial corpus methodologies which support stylistics analysis in Jane Austen’s *Pride and Prejudices* (PrPr henceforth). Such methodologies include keywords, concordances and intratextual comparisons which all provide evidence for substantiation of the civility theme in the story. They find “keywords” in PrPr by using a reference corpus containing 18 novels by nineteenth century authors (Austen’s excluded). Wijitsopon (2014) explores Jane Austen’s novels using “keywords” and found such key semantic domains and key part of speech that verify the author’s exaggerative writing style which comprises evaluative modalities and state verbs.

### *Extended Lexical Phrases*

Extended lexical phrases, or lexical bundles, which is the term used by Biber et. al (2004), demonstrate frequent use of formulaic patterns beyond lexical level. Knowing these lexical bundles and using them is especially important to the corpus-based analysis as it places the individual words within a larger context. When individual words

are embedded within different contextual environments they can take on different meanings. For example, Adolphs and Carter (2002) exemplifies the case of Virginia Woolf's *To the Lighthouse* so as to confirm how lexical bundles can help to comprehend Mrs. Ramsay's mental process. Compared with CANCODE corpus, Adolphs (2006) finds the semantic prosody with lexical bundles such as "happen to" and "up and down" related to negative expectation and negative shading.

### *Collocational Analysis*

Collocational analysis is a way in which particular words are associated with each other or their attraction to co-occur more frequently than would be expected by chance (Adolphs, 2006). Using collocational analysis through a corpus analysis tool will tell us how one particular word is related to a specific group of words. Grouping or categorizing such relationship among words can elaborate more on thematic or literary meanings. Techniques such as frequency lists of individual words, keyword analyses and type-token ratios are important. Barnbrook's (1996) investigation of frequency lists of words related to "monster" and "creature" in Frankenstein. A simple concordance can reveal a general idea of how a character is represented in the story by observing repetitive patterns in context. Adolphs (2006) illustrates more how concordance can achieve a comprehensive picture (p. 67) of Anna Karenina a passionate and spontaneous woman in the novel. In terms of language and gender, Kettemann (1995) also uses the same technique to analyze personal pronouns, i.e. "he" and "she", in a nearly emancipatory American short story and finds that "she" collocates mostly with cooking and baking while "he" collocates with lexical words related to confidence and control; however, Kettemann discovers that concordance lines change as the plot gradually develops.

### Research Questions

*The Famous Five* has been renowned for its gender bias reinforced through exciting adventures of the group. As regards the review of literature, this study aims to answer two main research questions as follows:

1. In what way does corpus-based analysis of Enid Blyton's *The Famous Five* provide key semantic fields as linguistic indicators of the series' thematic ideas
2. To what extent does corpus-based analysis of Enid Blyton's *The Famous Five* provide lexical information related to gender representation throughout the series?

### Research Methodology

This study used two corpus analysis tools: Wmatrix and AntConc. Wmatrix was developed by Rayson (2003) and can generate basic functions such as concordances, collocates and keywords. Wmatrix can offer three types of quantified data as follows: (a) lexical (b) grammatical (c) semantical ones. At the lexical level, Wmatrix can generate basic frequency lists of the text. On the grammatical level, Wmatrix applies CLAWS (the Constituent Likelihood Automatic Word-tagging System) to tag each word with POS (part of speech) to which they belong and the system will subsequently generate frequency lists based on grammatical groupings. At the semantic level, each word is tagged with UCREL Semantic Analysis System (USAS) and the system will subsequently generate frequency lists based on semantic groupings. USAS tag-set is divided into twenty one categories from A to Z (excluding these letters: D, J, R, U and V). Each letter has unique classification; for example, semantic field of A group will be in the categorization of "general & abstract terms" and that of K group will be in the categorization of "entertainment". As each semantic field is annotated into A to Z according to their thematic categorizations, it

can be used to provide empirical evidence of literary themes pervasive in *The Famous Five*.

Wmatrix can generate information of collocations; however, it is still under development. Therefore, another analysis tool, called AntConc (Version 3.43), will be used to generate information of collocates throughout the text. AntConc is a freeware corpus analysis toolkit, developed by Dr. Laurence Anthony (2014), and all basic functions (e.g. concordancing, collocating, and keyword search) of corpus-based analysis are also provided. AntConc will be used in this research to generate collocates.

Two corpora (specialized and general) were constructed and employed to give comparison. A specialized corpus was smaller in size and built through compilation of twenty-one books of *The Famous Five*. All twenty-one digital files were converted to .txt version and uploaded onto both corpus analysis programs. On the contrary, the general corpus was larger in size and constructed from the collection of imaginative texts available in the British National Corpus. Although the British National Corpus contains a variety of written texts (e.g. newspaper, magazines, historical documents), only imaginative texts were used in this study as a reference corpus because they consist of similar kind of writing. In fact, BNC has been widely used in *Longman Grammar of Spoken and Written English* (Biber et al. 1999). Both specialized and general corpora were converted and uploaded into Wmatrix and AntConc programs.

On the lexical level, both Wmatrix and AntConc can generate keywords which refer to lexical items that occur with a significantly higher frequency (positive keywords) or with a significantly lower frequency (negative keywords) in a collection of texts when compared with a larger reference corpus. Keywords are important to search for a particular relevance to literary themes and gender representation and help to explain what characterizes particular discourses (Adolphs, 2006).

Keywords will be taken into consideration in accordance with statistical analysis by the corpus analysis tools.

The application of Wmatrix and AntConc throughout the study is to generate empirical data to help answer both research questions as meticulous as possible. *The Famous Five* is a classic literature which is compulsory for literary scholars to discuss its eminent theme recounted by the author; therefore, Wmatrix serves this purpose as it can generate information about thematic categorizations. However, the Antconc program is also needed to generate information about collocations that will spell out gender representation of male and female characters. Information gained from both programs will be investigated and analyzed based on both quantitative and qualitative methods. Log-likelihood statistics is chosen for this study because of the fact that the data collection is relatively small when compared with the reference corpus and that it places more emphasis on lexical words, rather than grammatical words, (Baker, 2006) which relate to contents and thematic features in narrative fictions. The cut-off value of this research will be set at 100.00 and 5.00 for Wmatrix and AntConc, respectively, and these values are considered low so as to narrow the focus down onto the semantic concepts with stronger statistical significance (Balossi, 2014, p.109). As with gender representation, this research paper adopts Kettemann's (1995) methodology, which uses collocation and concordances to analyze personal pronouns, i.e. "he" and "she", in the American short story previously discussed. In a similar fashion, the concordances and collocates between "he" and "she" will be investigated through AntConc to generate lexical features related to both pronouns.

## Results and Discussion

The results can be divided into two parts in response to research questions one and two respectively. The first part will be the results retrieved from Wmatrix program which will be used to explicate thematic categorizations. The second part will be the results retrieved from the

AntConc program, generating such information related to gender representation of male and female characters.

### Part 1: Wmatrix

Regarding the thematic categorizations provided by USAS Tagging in Figure 1, it shows key semantic fields along with their Log-likelihood value with cut-off value at 100.00.

USAS Tagging	LL-Value
Personal names (George, Julian, etc.)	948.54
Objects Generally (stones, rocks, things)	278.20
Light (shining, sunshine, gleaming)	241.05
Sailing, swimming (swim, sail, boat)	236.45
Speech: Communicative (said, talking)	211.06
Time: Future (will, shall, won't)	210.87
Location and Direction (here, there, away, back)	184.03
Degree: Boosters (very, enormously)	182.59
Degree (as)	170.74
Unexpected (surprisingly, unexpectedly)	164.03
Geographical Terms (beach, sea, cliffs)	160.40
Darkness (dark, darkness)	130.57
Moving, coming and going (go, come, off)	103.41
Degree: Compromisers (quite, rather)	101.37

Figure 1: Key Semantic Fields of *The Famous Five*

As discussed, the series recounts the adventurous lives of the five children; therefore, USAS tagging can reveal the gist of textual content through statistics methods. The highest value belongs to “personal names” categorization and this can tell us that the story has been narrated through repetition of main characters (George, Julian, Dick, Anne and Timmy) all over the series. It is established as one of the means to acquaint main characters to children and to make the readers feel close to individual characters. The second highest value demonstrates

the categorization of “objects generally” which is mostly related to ordinary things that readers can find in natural settings. Besides, the categorizations such as “Sailing, swimming”, “Geographical Terms” connotes the scenarios as the landscape for adventures. This can be affirmed by the first episode of the series: “Five on a Treasure Island” as most of the events occur on an isolated island where the group needs to protect against multiple villains through various types of transportation, e.g. sailing and boating. The series also consists of “Unexpected” categorization, which means plenty of surprising elements is included and needed to decipher in due course.

The highlight of USAS tagging belongs to that of the third highest value related to “light” where literal and figurative language are used extensively. The “light” categorization has been reinforced by the twelfth highest value, “darkness”, which foregrounds a contradiction between good and evil theme of the series. Fourteen concordances (C1–C14) are shown below to represent what makes up the thematic categorization of “light”.

C1 !It 's Tuesday ! And the sun 's	shining.	“Anne woke up with a jump and
C2 eaths of air, and feel the warm	sunshine	on his shoulders. He climbed int
C3 fine !” said Julian, his eyes	shining	with delight. George looked at h
C4 lly !” said Julian, his eye 's	shining.	“That would be marvellous .Wi
C5 eorge looked at him and her eyes	shone too,	as bright as the sea itself
C6 ne . It was so magnificent. The	lightning	tore the sky in half almost every
C7 .”As he spoke, a ray of pale	sunshine	came wavering out between a gap I
C8 dder. The others followed . The	light	from the torch showed a very quee
C9 uted back George, her blue eyes	shining	dangerously. “And you can wait
C10 eally ?” said George, her eyes	shining.	“Well, I 'll go then. I alwa
C11rs to the two artists, her eyes	gleamed	angrily. ‘The
thief ! They must		
C12 “ Golly ! “ said Dick, his eyes	gleaming.	“Souptins of
meattins of fruit		
C13e, yawned by his feet. Julian	shone	his torch in
front of him, and a		
C14 n to the study. She saw a torch	shining	on the floor,
dropped by someone		

C1–C14 demonstrates that the categorization of “light” can be extended from literal to figurative meaning. On the one hand, literal meaning can be covered from “shine” signifying sunlight in the morning to the “light” of torch used for guidance in the darkness. For example, C1 denotes the significance of sunlight as a time indicator and C8 shows the usefulness of light torch while being in the darkness. It could also reflect the author’s philosophy that draws on the sun’s necessity to society. On the other hand, figurative language of “light” can illustrate either positive or negative emotions through characters’ eye as of C3 “his eyes shining with delight” and of C9 “her blue eyes shining dangerously”. Manifestation of “light” reinforces two literary themes, respectively: “the light at the end of the tunnel” and “eyes are the window to the soul”. The first literary theme suggests that there would always be a solution or exit to every problem. The theme construes the genre of the series as detective fictions, where all characters brainstorm for ideas to seek the way out of predicaments. Secondly, the series was written based on the theme “eyes are the window to the soul” because whenever the characters encounter displeased or delightful situations, the narrator conveys such emotions through the eyes.

It is notable that the concordances and thematic categorization could serve as a means to seek for literary themes. In addition, Wmatrix can provide keywords which would provide how the series’ lexical items or phrases are unique when compared with the BNC (imaginative text). Figure 2 demonstrates positive (overused) and negative (underused) keywords. Positive keywords are such words or phrases that are frequently uncommon, whereas negative keywords are those of infrequently uncommon. Personal names are overused mostly throughout the series, especially those of the protagonists (e.g. George, Julian) and the antagonist (Mr. Roland). Of the most interesting ones are “Timothy” and “Timmy”, both referring to George’s intelligent dog that helps the group out of many perilous situations. Moreover, the word “said”

(680.61) represented the narrative style of the series as to use extensive structure of direct speech. While there is a tendency of dominant features of personal names in positive keywords, the negative keywords reveal that the series contain less use of relative pronouns (e.g. who, which), less use of kinship words such as (brother, son, daughter) and less use of womanhood words (woman, lady). This could substantiate that the narrative consists of simple, rather than complex, sentences and it serves the purpose as a reading selection for children. Kinship is not highlighted because the story is focused on the friendship assured among the group and this is the story of children who come from different background. Quantitatively, when compared with “man” (-29.09), the lack of womanhood words such as “woman” (-58.08) and “lady” (-32.73) leads to gender bias, where the female are numerically dominated throughout the narrative.



Figure 2: Positive and Negative Keywords (Log-likelihood Value)

Keywords can be ranged from a word to phrases and Wmatrix can generate multi-word expressions results as well. These multi-word expressions are deemed promising as to emphasize literary themes discussed.

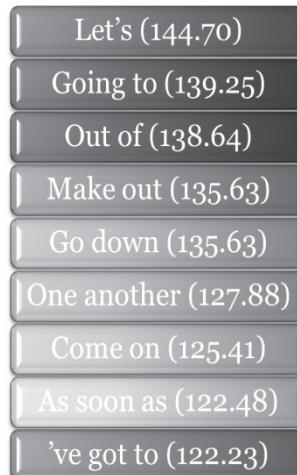


Figure 3: Multi-word Expression (Log-likelihood value)

Figure 3 affirms the theme of friendship through phrases such as “let’s”, “one another”, “come on” and a phrase of commitment “ve got to”; the theme of adventure through “going to”, “out of”, “go down” which are phrases related to directions; the theme of mysteries through the use of “make out” and “as soon as” when a particular character understands the situation and then make a quick decision to take the group away from dangers.

C1 I ‘m glad you ‘ve come after all.  
 C2 ant to lose our boat, do we? “ ”  
 C3 ,crossly, when his aunt had gone  
 C4 !” said George, and she stepped  
 C5 After a little while they could  
 C6 ld rag. We ‘ll try our best to  
 C7 soon as we yawn . I did want to  
 C8 soon came back , and he wanted to  
 C9 d, too. The children stared at

Let ‘stake a boat out this afternoon and  
 Let ‘sexplore the island , let ‘s explore  
 out of the room. “You know quite well wh  
 out of the door. “I ‘m going straight ho  
 make out the outlines of a dark hulk,  
 make out what ‘s on it and tell you. Hur  
 go down on the beach for a while. “I ‘m  
 go down into the dungeons again. “No, you  
 one another, feeling sorry and dismayed.

C10 y again. The children looked at  
 C11 aloud, “ We ‘ve found the gold.  
 C12 here to be seen, although she had  
 C13 meanwhile the men were not idle !  
 C14 g to the children ‘s breathing.  
 C15 quite happy on my own. Now I  
 C16 way to the entrance because we  
 one another in the greatest joy and excitemen  
 Come on down at once and see it. Georgina  
 come on the station platform with Anne. ‘W  
 As soon as they heard the door slam they spu  
 As soon as he knew they were asleep he got u  
 ‘ve got to put up with a silly girl who li  
 ‘ve got to follow my chalk-marks. “Right

## Part 2: AntConc

### 2.1) Frequency and Voice

There are four characters which are the focus of gender representation. Two male characters (Julian and Dick) and two female characters (Georgina and Anne) are at the center of attention. Julian and Georgina are two protagonists that propel the plot throughout the series; therefore these two characters are intriguing to scrutinize how the author particularly uses her word choice. AntConc is particularly used to generate concordances and collocates among male and female characters. AntConc can also calculate the tokens of the search words. As “said” is one of the positive keywords (uncommonly frequent), it is important how the word is distributed to both male and female protagonists.

Julian said (2,884)	George said (1,899)
he said (1,208)	she said (825)
“he” (14,187)	“she” (6,666)
Julian (6,620)	George (6,681) or Georgina (112)
Dick (4,195)	Anne (3,828)
Timothy (251) or Timmy (4,264)	

Table 1: A comparison of male and female characters

An overview of tokens between male and female characters as shown in Table 1 reveals that the male characters' voice governs the narrative. The use of pronoun "he" outnumbers that of "she" and represents that male characters have been frequently addressed. When compared with Julian, George speaks less than almost fifty percent. It is also emphasized by the greater number of "he said", when compared to "she said". Unlike Julian, the male protagonist, George has been equally addressed. Interestingly, Timothy, the male dog, is addressed more often than Dick and Anne. This empirical evidence proves that the animal can be seen as more important than both Dick and Anne because its name was repeated a higher number of times.

## *2.2) Language and Gender*

Counting tokens of pronouns and proper names can suggest that male characters have more space to speak. AntConc can generate data of concordance and collocation analyses which are considered crucial for a comparison between male and female characters. As pronouns "he" and "she" are of most interest, it is imperative to investigate how each pronoun forms collocates.

List of collocates between male and female characters (In order of Log-likelihood value)	
“He”	“Hhe”
climbs, watchfully, unzip, unexplained, twiddled, suggests, spectators, snarlings, service, rickety, respected, raves, pesters, offence, mistakes, knowledgeable, intelligently, heartbroken, growly, flops, disapprove, despised, competitor, changeable, boney, astride	weedy, vowed, unhappiness, tantrum, Spanish, shied, prettiest, persevered, overmantel, muddling, impish, grieving, dreamily, dependable, crawls, canes, praises, dozing

Table 2: Collocates of “he” and “she”

The collocation analysis from Table 2 reveals that the personal pronoun “he” collocates with a particular set of words, which can be divided into positive and negative traits of male characteristics. Positive traits of the male characters are described as being active, logical, dependable, and respectful, while their negative traits are associated with being disagreeable, unsuccessful and furious. Foregrounded with the positive and negative traits, readers may have a particular impression of male characters through a variety of lexical items employed; however, collocations provide empirical evidence to confirm how that impression can be formed through a given set of words.

As opposed to the male characters, the collocation analysis demonstrates that the personal pronoun “she” collocates largely with a particular set of negative emotions. Although some words are positive(e.g. “dependable” and “praises”) and related to outlook (“prettiest”), the number of negative words statistically outweighs. The female characters are represented as weak, unhappy, inactive, submissive and illogical.

C1 not a scrap of sun-tan! And she looks weedy, doesn't she? I'm  
 C2 one except for Tim. She had always vowed that she never, never would  
 C3 in the kennel? She would die of unhappiness! She turned to Julian  
 C4 when he saw her flying into a tantrum. She smiled unwillingly.  
 C5 disliked her, or whom she disliked, she shied She was always je  
 C6 Anne thought she was the prettiest away like a frightened  
 C7 hands were still so tightly tied. But she persevered. horse – and kicked like  
 C8 came into her sharp little mind. With an impish doll she had ever seen  
 C9 came to the conclusion that she was grieving. She grew so tired that  
 she had to  
 grin she turned back.  
 She pulled the door  
 about Timmy. Still she  
 didn't look very s

While collocation analyses show only a broad overview of gender differences, it may not be absolutely accurate. It is worth noting that the pronoun "he" also includes "Timmy" or "Timothy," the dog, but in the case of "she," it is solely used to refer to all human beings. Yet, both personal pronouns appear sporadically and the readers may focus on repetition of the personal names, which is a common style of most children's literature. Therefore, more thorough analyses of characters' personal names are obligatory to see how male and female protagonists are actually represented. As George and Julian are two main protagonists of the series, collocates of both characters will be scrutinized and compared. Besides, collocations of George and Anne will be used to compare how George's tomboy traits are different from those of Anne, who is reckoned representative of idealistic womanhood of that time.

Collocation Analysis: George (LL-Value)
commanded (5.75022)
desperately (5.54377)
crossly (5.42829)
indignantly (5.38150)
sulky (5.32138)
patting (5.29223)
obstinately (5.28073)
scowled (5.22184)
wildly (5.20273)
demanded (5.02637)

Table 3: Collocation Analysis [5R–5L<sup>1</sup>]: George

According to Table 3, it becomes apparent from the collocation analysis that George is highlighted by a group of speech utterances and that her behavior revolves around being emotional, dependable, and controllable. The highest value of all is “commanded,” which shows that George likes to control the other members. Apparently, the words such as “crossly, indignantly, obstinately, and wildly” affirm that George is moody and disobedient in nature. George expresses her emotions and desires directly through words, such as “desperately” and “demanded”. Despite of being bossy, she encourages her friends and gives support by “patting” their shoulders as can be seen in multiple scenes. The words such as “crossly” and “indignantly” substantiate what literary criticism discussed that George is an impatient character. According to Ward (2004), George is “headstrong and courageous by nature and, like her father, scientist Quentin Kirrin, has a hot and fiery

<sup>1</sup> Any words that co-occur in the word search (node) within the range of 5 words on its left and right positions. This span, we discover, comprises more variety of lexis (verbs, adjectives and adverbs) than any other span positions.

temper". The following concordances (C1–C5) show that George gives the readers an impression of being "emotional" and "disagreeable" when something is not right.

C1	bit rushed to its burrow. 'Timmy!' called George, crossly, and poor Timmy	put his tail down;
		looking
C2	won't! Don't be silly,' said George, rather	crossly. And anyway it
		wouldn't smell
		like him. He
C3	said Anne, shaking her. George shrugged away	crossly, not believing her.
		It was just a
		trick to
C4	We'll leave him outside then," said George,	crossly "What a silly rule!
		Timmy, stay here.
		We
C5	George was just about to answer	crossly when Anne gave
		an exclamation.

The concordances, when overtly repeated, shows that "crossly" is used widely when George demands Timmy to accompany her in C1 or when she disagrees with Mr. Roland, who expels Timmy out of the house in C4. In addition, the collocation This list is followed by "indignantly" which is also a negative adverb. Because of the fact that the word has been used extensively as key in the story, George is perceived as a headstrong and impatient character.

Collocation Analysis: Julian (LL-Value)
thoughtfully (6.47900)
nudged (6.19855)
cheerfully (6.18503)
clapped (6.14196)
deftly (6.00446)
winked (5.89404)
politely (5.82937)
clutched (5.82004)
considered (5.78833)
lazily (5.76345)
soberly (5.72693)
suggested (5.45391)

Table 4: Collocation Analysis [5R-5L]: Julian

While George is generally perceived as the emotional character in the group, the results of Julian through collocation analysis (Table 4) illustrate what is deemed positive. According to (2004), Julian has been seen as “strong and intelligent as well as caring, responsible, kind and very protective towards Anne”. Having the highest value of 6.47900, the word “thoughtfully” represents Julian as a kind-hearted character. Concordances (C1-C6) apparently show that “thoughtfully” usually co-occurs with “said”, “stared”, “said” and “asked” and this shows that Julian takes appropriateness into consideration regarding verbal and non-verbal acts. This can be recapped by words such as “cheerfully”, “politely”. Moreover, the other words, such as “deftly”, “soberly” and “suggested” can refer to his aptitude or capacity that is required for being the leader of the group.

C1 the rest of us.” “That’s true,” *said* Julian, thoughtfully. “Well- we’ll suppose that Dick, or w

C2 me –a Big Adventure!” Julian *stared* at his cousin thoughtfully. Could there possibly be anything in

C3 eyes gleaming. “Yes\97it might be,” *said* Julian, thoughtfully, trying to undo the straps of the tru

C4 97we must warn the police of that,” *said* Julian, thoughtfully. “Better not spread the news abroad t

C5at we saw the light flashing from?” *asked* Julian, thoughtfully. ‘The seaward side – or the landward?

C6 e been signalling to the smugglers,’ *said* Julian, thoughtfully. ‘Let me see – probably they came over

However, for the negative word “lazily” in the collocation analysis is rather interesting to investigate to see whether it will show possible negative traits of Julian. The word may relate to inactiveness or idleness that the character conveys throughout the series. The concordances (C1–C4) reveal that “lazily” co-occurs with “asked”, “said”, “thought” and “pointed”. The word does not relate much to negative attribute, but rather on the way that the character uses his verbal and non-verbal acts with lack of enthusiasm or with contentment during the conditions given.

C1 to come and peep into this cave?" asked Julian, lazily, finishing up a cup of cocoa. "Lions? Tiger  
 C2 were being rude to my Grandad," said Julian, lazily "Now ' it's settled, is it, that  
 C3 it was only April. How marvellous! Julian *thought* lazily that they were really lucky to have such we  
 C4 the cool breeze and get dry." Julian *pointed* lazily to a big notice not far off. It said

Collocation Analysis: Anne (LL-Value)
giggled (7.25413)
indignant (6.93220)
giggle (6.76228)
admiringly (6.70981)
fearfully (6.65210)
mummy (6.61028)
shiver (6.51717)
hugged (6.42970)
sleepily (6.39164)
comforted (6.34724)
shivering (6.29477)
squealed (6.26924)
dolls (6.12485)

Table 5: Collocation Analysis [5R ' 5L]: Anne

Throughout the story, Anne is portrayed as a conventional girl in the group as she takes care of domestic duties during the summer holidays. She is the youngest person who needs protection and she becomes easily frightened of and anxious about unexpected determination and adventures. Regarding the collocation analysis in Table 5, it is clearly shown that the list of collocates is different from that of George and

Julian. The innocence of Anne is highlighted by the first and the third collocates of the analysis (“giggled” and “giggle”) and they are the word which is frequently used with female characters. In terms of fear and anxiety, words such as “fearfully” and “shivering” show that Anne is frightened and weak. She needs protection from her friends as can be seen by words such “hugged”, “comforted”. Despite being young, domestic words such as “mummy” and “dolls” represent Anne as a mother figure for the group. Although all collocates substantiate what literary critics discussed, the word “indignant”, the second-highest value, could add that she is occasionally emotional, if not entirely passive.

C1 like Tim, when Mr. Roland is in the room,’	giggled Anne. ‘Laugh all you like,’ said George, i
C2 book at him. It crashed to the floor. Anne	giggled. ‘Oh stop it, you two. We’ll never get
C3 know every stream and path and er–volcano!’ Anne	giggled. The others stared at MrLuffy, wondering
C4 who thought she was sitting on a volcano.’ Anne	giggled. ‘You’re not to tease me. No, honestly, Mr.
C5 a ninny and an idjit,’ said Jock, and Anne	giggled. ‘Look – there’s Cecil. You might tell him
C6 , putting on a face exactly like the twins. Anne	giggled. They all followed the twins, who were dre
C1 in a temper.’ ‘Where shall we hide?’ said Anne,	fearfully. ‘In the secret room!’ said Julian. The
C2 be drowned, will he?’ said Anne,	fearfully. ‘Oh, I expect he can look after him
C3 or hear anything of the boys’ said Anne,	fearfully. ‘George, let’s go to the quarry
C4 They all stared out of the window, said Anne half	fearfully. ‘It’s gone!’ cried Dick, in
C5 that shouting like that?’ said Anne,	fearfully. ‘It can’t be the robber, can

Not only do her emotions, but also the illustrative examples from Anne's concordance lines (C1–C6) that undoubtedly reveal her age as the youngest person. She laughs when something crashes to the floor or when she hears funny words like "a ninny and an idjit" referring to "idiotic person". In addition, concordance lines (C1–C5) of "fearfully" clearly demonstrate how Anne expresses her apprehension verbally ("said"). Her characteristics reflect her worldview as an innocent child, filled with glee and outward expressions.

A comparison of collocations among the three major characters can reveal intriguing information about their differences in terms of language and gender. Collocations of George reflect her negative emotions and indignant expression. Collocations of Julian reveal how he is represented as a dependable and cherished figure among the group as he becomes thoughtful when it comes to a final decision. Collocations of Anne apparently demonstrate how readers would think of her as a naïve girl that connotes acquiescence and delicateness of womanhood.

## Conclusions and Recommendations

As with language, gender and themes in *The Famous Five*, this study aims to apply corpus-based approaches to substantiate what literary critics have discussed. Corpus-based analyses can give empirical evidence to support two literary themes: adventures and mysteries and gender representation. Adventures and mysteries can be seen from thematic categorization by USAS tagging of Wmatrix program, providing semantic fields such as "sailing", "light" "geographical terms" and "darkness". Gender representation has been substantiated through collocations of both "he" and "she" and concordances of individual major characters: Julian, George and Anne. AntConc generates different sets of collocations among the three characters and shows that Julian, a male character, is inclined to be dependable, treasured and thoughtful and that female characters comprise two polarities. George, or Georgina, as a tomboy, is rather emotional and outspoken, whereas Anne is passive, innocent and delighted.

Corpus stylistics is proven to be useful to assist literary interpretations; however, its limitation is that the researcher needs to have sufficient background of a particular text under investigation, such as its themes, motif, and characterization prior to implementing a corpus analysis tool. A corpus analysis tool only provides information without explanation and interpretation; therefore, it is the researcher's role to figure out how to use the information with cautious optimism.

As opposed to conventional methodology of stylistics teaching, corpus stylistics can lead to pedagogical implications in terms of lexical bundles and phraseology through the lens of holistic and objective viewpoint. If the story is used as a tool to expand students' creativity and critical thinking through analyzing characters, a corpus tool can give general information of words or phrases that have been repeated and foregrounded throughout the story. Teachers may apply those words and phrases to their lesson plans and curriculum as a platform for students to figure out how and why particular characters are highlighted with words through discovery of concordances, collocations and keywords.

## Biodata

Kriangkrai Vathanalaoha is a lecturer at Language Centre, International College for Sustainability Studies, Srinakharinwirot University, Bangkok. He is now a PhD student in English Language Teaching, Language Institute, Thammasat University. He has published academic papers on language and literature. His research interests include cognitive poetics, corpus linguistics, discourse analysis and pedagogical stylistics. He can be reached at [kstylistics@gmail.com](mailto:kstylistics@gmail.com).

Dhanakorn Jeeradhanawin is a lecturer at the Department of English, Faculty of Liberal Arts, Thammasat University, Bangkok. His research interests include critical discourse analysis, computer-assisted language learning, and corpus linguistics. He can be reached at [dhanakorn@live.com](mailto:dhanakorn@live.com).

## References

Adolphs, S. & Carter, R.A. (2002). Corpus stylistics: point of view and semantic prosodies in “To The Lighthouse.” *Poetica*, 58: 7–20.

Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies*. Oxon: Routledge.

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from .

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Balossi, G. (2014). *A corpus linguistic approach to literary language and characterization: Virginia Woolf's 'The Waves'*. Philadelphia. John Benjamins

Barnbrook, G. (1996). *Language and Computers: A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.

Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Biber, D. Conrad, S. & Cortes, V. (2004). If you look at ...: lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3): 371–405.

Kettemann, B. (1995). Concordancing in stylistics teaching. In W. Grosser, J. Hogg and K. Hubmayer (Eds.), *Style: Literary and Non-Literary: Contemporary Trends in Cultural Stylistics* (pp. 307–18). New York: The Edwin Mellen Press.

Macdonald, S. (2013). *The gender bias in children's books*. Retrieved from <http://www.dailylife.com.au/life-and-love/parenting-and-families/the-gender-bias-in-childrens-books-20131118-2xqzz.html>.

Mahlberg, M. & Smith, C. (2010). Corpus approaches to prose fiction: civility and body language in *Pride and Prejudice*. In B. Busse & D. McIntyre (Eds.), *Language and Style* (pp.449–467). Basingstoke: Palgrave.

Rayson, P. (2003). 'Matrix: A statistical method and software tool for linguistic analysis through corpus comparison', Ph.D. thesis, Lancaster University.

Rayson, P. (2005). Wmatrix: a web-based corpus processing environment. Computing Department, Lancaster University. Retrieved from <http://www.comp.lancs.ac.uk/ucrel/wmatrix/>.

Stockwell, P. (2002). *Cognitive Poetics: An introduction*. London: Routledge

Ward, D. (2004). Enid Blyton 'the grown-ups' favourite. *The Guardian*. Retrieved from <http://www.theguardian.com/uk/2004/aug/23/books.books for children and teenagers>.

Wijitsopon, R. (2014). *Where the computer meets language and literature in English studies: A case of corpus-based approach to analysis of nineteenth-century British novels*. Published Project Report in Thailand Research Fund. Retrieved from [http://elibrary.trf.or.th/project\\_content.asp?PJID=MRG5380179](http://elibrary.trf.or.th/project_content.asp?PJID=MRG5380179).